A. General Techniques in Differential Privacy

A standard mechanism in the privacy literature, the *Laplace mechanism*, perturbs the output of an algorithm by adding Laplace noise to make the output private. Assume the algorithm computes a function $f : [n]^s \to \mathbb{R}$. The amount of noise required depends on the privacy parameter, ξ, and how much f varies over two neighboring datasets. More precisely, this variation of f is called *sensitivity* of the function and it is defined as:

$$\Delta f = \max_{\text{neighboring } x,y} |f(x) - f(y)| .$$

The noise is drawn from a Laplace distribution with parameter $b = \Delta f/\xi$. We denote the noise by $\mathbf{Lap}(b)$. More precisely,

$$\mathbf{Pr}[\mathbf{Lap}(b) = x] = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right) .$$

The following is well-known:

Lemma A.1 (The Laplace mechanism (Theorem 3.6 in (Dwork & Roth, 2014)))**.** *Assume there is an algorithm $\mathcal{A}$ that on input x, outputs $f(x) + \mathbf{Lap}(\Delta f/\xi)$. Then $\mathcal{A}$ is ξ-private.*

Note that the expected value of $\mathbf{Lap}(b)$ is zero. Therefore, the expected value of the output remains $\mathbf{E}[f(x)]$. Since we draw the noise independently from x, the variance of the output is increased by $\mathbf{Var}[\mathbf{Lap}(b)] = 2b^2$.

Moreover, the following lemmas help us understand how the privacy guarantee changes if we process the output of one or more private algorithm.

Lemma A.2 (Post-processing (Proposition 2.1 in (Dwork & Roth, 2014)))**.** *Assume $\mathcal{A}$ is a ξ-private algorithm. Any algorithm that on input x outputs a function $f(\mathcal{A}(x))$ is also ξ-private.*

Lemma A.3 (Composition Theorem (Theorem 3.16 in (Dwork & Roth, 2014)))**.** *Let $\mathcal{A}_i : [n]^s \to \mathbb{R}$ be a ξ_i-private algorithm for $i = 1, \ldots, k$. Any algorithm that on input x outputs a function $f(\mathcal{A}_1(x), \mathcal{A}_2(x), \ldots, \mathcal{A}_k(x))$ is $\left(\sum_{i=1}^k \xi_i\right)$-private.*

B. Generic Differentially Private Tester

In this section, we describe a simple generic method to convert a non-private tester into a private tester with a multiplicative overhead in the sample complexity. While this method is known in the differential privacy community, it is useful to contrast its sample complexity with the (substantially smaller) sample complexity of our testers in Sections 3, 4, and 5.

Algorithm 4 Reduction to a non-private tester

Input: Sample access to p, explicit access to q, n, ϵ
$m \leftarrow \lceil \frac{6}{\xi} \rceil$
$s' \leftarrow m \cdot s(n, \epsilon)$
$x_1, x_2, \ldots, x_{s'} \leftarrow s'$ samples from p.
$r \leftarrow$ Pick a random number from $[m]$.
$O \leftarrow \mathcal{A}\left(\{x_{(r-1)s+1}, x_{(r-1)s+2}, \ldots, x_{rs}\}\right)$.
With probability 1/6, $O \leftarrow \{\mathsf{accept}, \mathsf{reject}\} \setminus O$. ⟨⟨*flip the answer with probability 1/6.*⟩⟩
Output O.

Assume $\mathcal{A}$ is a tester that draws $s(n, \epsilon)$ samples. The idea is to draw $m \cdot s(n, \epsilon)$ samples for a sufficiently large m, and from this sample, to pick a random subset of size $s(n, \epsilon)$ samples. Then, the new tester runs $\mathcal{A}$ on the randomly chosen subset and outputs $\mathcal{A}$'s output. Given two sample sets that differ in one sample, the new private tester will give the same output whenever a chunk that does not contain the differing sample is chosen, which happens with probability at most $1/m$. This reduction to a non-private tester is described in Algorithm 4. We formally show its correctness in Theorem B.1.

Theorem B.1. *Let $\mathcal{A}$ be an ϵ-tester for property $\mathcal{P}$ that uses $s(n, \epsilon)$ samples from distribution p over $[n]$. Algorithm 4 is an (ϵ, ξ)-private property tester for property $\mathcal{P}$ using $O\left(s(n, \epsilon)/\xi\right)$ samples.*

Proof: Suppose $\mathcal{A}$ is an ϵ-tester for property $\mathcal{P}$ that uses $s(n, \epsilon)$ samples. Without loss of generality, assume the tester $\mathcal{A}$ errs with probability at most $1/6$[2]. Since the output of $\mathcal{A}$ is then flipped with probability $1/6$, by the union bound, the probability that Algorithm 4 errs is at most $1/3$, and it is thus an ϵ-tester for uniformity.

To prove the privacy guarantee, let m be $\lceil 6/\xi \rceil$, and let $X = \{x_1, x_2, \ldots, x_{s'}\}$ and $Y = \{y_1, y_2, \ldots, y_{s'}\}$ be two sample sets of size $s' := m \cdot s(n, \epsilon)$ that differ in exactly one sample. Without loss of generality, we assume they differ in the first sample: $x_i = y_i$ for $i > 1$ and $x_1 \neq y_1$. Algorithm 4 picks a random number, r, in $[m]$ and feeds $\mathcal{A}$ with the r-th chunk of size $s(n, \epsilon)$ from the input sample set. If $r \neq 1$, the distribution of the output is identical X and Y. Let $T(X)$ indicate the output of Algorithm 4 on input X. More precisely, we have

[2] This can be achieved by the standard amplification method (i.e., running the tester $O(1)$ times and taking the majority answer). The new sample complexity grows by at most a constant multiplicative factor.

$$
\begin{aligned}
&\mathbf{Pr}[T(X) = \mathsf{reject}] \\
&= \sum_{i=1}^{m} \mathbf{Pr}[T(X) = \mathsf{reject} | r = i] \cdot \mathbf{Pr}[r = i] \\
&= \tfrac{1}{m} \sum_{i=1}^{m} \mathbf{Pr}[T(X) = \mathsf{reject} | r = i] \\
&= \tfrac{1}{m} \sum_{i=2}^{m} \mathbf{Pr}[T(Y) = \mathsf{reject} | r = i] \\
&+ \tfrac{1}{m} \mathbf{Pr}[T(X) = \mathsf{reject} | r = 1] \\
&\leq \tfrac{1}{m} \sum_{i=1}^{m} \mathbf{Pr}[T(Y) = \mathsf{reject} | r = i] + \tfrac{1}{m} \\
&\leq \mathbf{Pr}[T(Y) = \mathsf{reject}] + \tfrac{1}{m} \ .
\end{aligned}
$$

Since we change the output of $\mathcal{A}$ with probability $1/6$, it is not hard to see that $\mathbf{Pr}[T(Y) = \mathsf{reject}]$ is at least $1/6$ for any input y. Thus,

$$
\frac{\mathbf{Pr}[T(X) = \mathsf{reject}]}{\mathbf{Pr}[T(Y) = \mathsf{reject}]} \leq 1 + \frac{6}{m} \leq 1 + \xi < e^{\xi} \ .
$$

Similarly, we can show the above inequality when the output is accept. Thus, the algorithm is ξ-private. □

C. Amplification of Confidence Parameter in the Private Setting

For convenience, throughout this paper we work with testing algorithms that have failure probability at most $1/3$. Here we point out that this is without loss of generality, since a standard amplification method also succeeds in the differentially private setting.

Algorithm 5 Amplified confidence parameter

```
m ← 18⌈ln 1/δ⌉ + 1
s ← s(n, ε, ξ)
c ← 0
for i = 1, ..., m do
    X^(i) ← a set of s samples from p
    Run A using samples in X^(i).
    if A accepts then
        c ← c + 1
    end if
end for
if c ≥ m/2 then
    Output accept.
else
    Output reject.
end if
```

Theorem C.1. *Given $\mathcal{A}$, an (ϵ, ξ)-private tester for property $\mathcal{P}$, such that $\mathcal{A}$ uses $s(n, \epsilon, \xi)$ samples for any input distribution p over $[n]$. Algorithm 5 is an (ϵ, ξ)-private tester for property $\mathcal{P}$, using $O\left(\log 1/\delta \cdot s(n, \epsilon, \xi)\right)$ samples from p, that outputs the correct answer with probability $1 - \delta$.*

Proof: First, we show that algorithm 5 is ξ-private: Let X and Y be two sample sets of size $m \cdot s$ (where m and s are as defined in Algorithm 5) that differ only in one sample. Without loss of generality, assume they differ in the first sample. Therefore, $X^{(1)}$ and $Y^{(1)}$ differ in only one sample, and for $i > 1$, $X^{(i)}$ and $Y^{(i)}$ are identical. Hence, the distribution of the output of $\mathcal{A}$ in all of the iterations except the first one is identical for both X and Y. For the first iteration, the distribution over the output of $\mathcal{A}$ cannot change drastically, because $\mathcal{A}$ is a ξ-private algorithm. More formally, we have the following:

$$
\mathbf{Pr}[\mathcal{A}(X^{(i)}) = \mathsf{accept}] = \mathbf{Pr}[\mathcal{A}(Y^{(i)}) = \mathsf{accept}] \quad \text{for } i > 1,
$$

and

$$
\mathbf{Pr}[\mathcal{A}(X^{(1)}) = \mathsf{accept}] \leq e^{\xi} \cdot \mathbf{Pr}[\mathcal{A}(Y^{(1)}) = \mathsf{accept}].
$$

An analogous argument holds when the output is reject. Let $T(X)$ indicate the output of Algorithm 4 on input X. Let $\sigma(X^{(i)})$ be an indicator variable that is one if $\mathcal{A}$ outputs accept on input $X^{(i)}$ and zero otherwise. Since iterations of the algorithm are independent, we have:

$$
\begin{aligned}
&\mathbf{Pr}[T(X) = \mathsf{accept}] = \mathbf{Pr}\left[\sum_{i=1}^{m} \sigma(X^{(i)}) \geq 9\lceil \ln \tfrac{1}{\delta} \rceil + 1\right] \\
&= \mathbf{Pr}\left[\sigma(X^{(1)}) = 1\right] \cdot \mathbf{Pr}\left[\sum_{i=2}^{m} \sigma(X^{(i)}) = 9\lceil \ln \tfrac{1}{\delta} \rceil\right] \\
&+ \mathbf{Pr}\left[\sum_{i=2}^{m} \sigma(X^{(i)}) \geq 9\lceil \ln \tfrac{1}{\delta} \rceil + 1\right] \\
&\leq e^{\xi} \cdot \mathbf{Pr}\left[\sigma(Y^{(1)}) = 1\right] \cdot \mathbf{Pr}\left[\sum_{i=2}^{m} \sigma(Y^{(i)}) = 9\lceil \ln \tfrac{1}{\delta} \rceil\right] \\
&+ \mathbf{Pr}\left[\sum_{i=2}^{m} \sigma(Y^{(i)}) \geq 9\lceil \ln \tfrac{1}{\delta} \rceil + 1\right] \\
&\leq e^{\xi} \cdot \mathbf{Pr}\left[\sum_{i=1}^{m} \sigma(Y^{(i)}) \geq 9\lceil \ln \tfrac{1}{\delta} \rceil + 1\right] \\
&\leq e^{\xi} \cdot \mathbf{Pr}[T(Y) = \mathsf{accept}] \ .
\end{aligned}
$$

An analogous inequality holds for the case where the output is reject. Therefore, Algorithm 5 is ξ-private. Moreover, the output of the algorithm is wrong only if the majority of the invocations of $\mathcal{A}$ return the wrong answer (i.e. more than $9\lceil \ln 1/\delta \rceil$ times). However, $\mathcal{A}$ errs with probability at most $1/3$ by definition. By the Hoeffding bound, the probability of outputting the wrong answer is

$$\mathbf{Pr}\left[T(X) \text{ is wrong}\right] \leq e^{-2m/36} \leq \delta \ .$$

Thus, the total error probability is at most δ. Therefore, Algorithm 5 is an (ϵ, ξ)-private tester that outputs the correct answer with probability $1 - \delta$. □

D. Proof of Theorem 3.1

Theorem 3.1. *Given an (ϵ, ξ)-private uniformity tester using $s(n, \epsilon, \xi)$ samples, there exists an (ϵ, ξ)-private tester for identity using $s = s(6n, \epsilon/3, \xi)$ samples.*

Proof: Given s samples from p, we map them to s samples from p' using the following mapping:

1. Given sample i from p, the process $F_1(i)$ flips a fair coin. If the coin is Heads, $F_1(i)$ outputs i, otherwise, $F_1(i)$ outputs j drawn uniformly from $[n]$. Let p_1 denote the output distribution of $F_1(i)$'s. It is clear that $p_1(i) = (1/2)p(i) + 1/(2n)$. We define $q_1(i)$ similarly.

2. Let $m_i = \lfloor 3n(q(i) + 1/n) \rfloor$. Given j and the output of process $F_1(i)$ where i is drawn from p, process $F_2(i)$ outputs j with probability $m_i / \left(3n(q(i) + 1/n)\right)$ and $n + 1$ otherwise. Let p_2 denote the output distribution of the $F_2(i)$'s. It is not hard to see that

$$\begin{aligned} p_2(j) &= p_1(j) \cdot \frac{m_j}{3n(q(j) + 1/n)} \\ &= \frac{1}{2} \cdot \left(p(j) + \frac{1}{n}\right) \cdot \frac{m_j}{3n(q(j) + 1/n)} \end{aligned}$$

   for all $i \in [n]$, and $p_2(n+1) = 1 - \sum_{\ell=1}^{n} p_2(\ell)$. We define $q_2(i)$ similarly.

3. Given k, the output of process $F_2(i)$ where i is drawn from p, we output $F_3(i) = (k, a)$ such that a is uniformly chosen from $[6nq_2(k)]$. Note that for $k \in [n]$, $6nq_2(k)$ is equal to m_k and it is an integer, so the set $[6nq_2(k)]$ is well-defined. We denote the distribution of $F_3(k)$'s as p'. It is not hard to see that if $p = q$, then

$$\begin{aligned} p'\left((k, a)\right) &= \frac{1}{2} \cdot \left(q(j) + \frac{1}{n}\right) \cdot \frac{m_j}{3n(q(j) + 1/n)} \cdot \frac{1}{m_j} \\ &= \frac{1}{(6n)} \end{aligned}$$

   for $j \in [n]$. For $k = n + 1$, we have:

$$6n\, q_2(n+1) = 6n - 6n \sum_{\ell=1}^{n} \frac{m_\ell}{6n} = 6n - \sum_{\ell=1}^{n} m_\ell.$$

   is also an integer. Therefore, $p'((n+1, a))$ is also $q_2(n+1)/(6n\, q_2(n+1)) = 1/6n$.

Thus, if $p = q$, then p' will be a uniform distribution. Similarly, if $\|p - q\|_1 \geq \epsilon$ then $\|p' - U\|_1 \geq \epsilon/3$. For a detailed proof, see (Goldreich, 2016).

Then, we run the private uniformity tester using the samples from p', and output the answer of the tester. As shown in (Goldreich, 2016), if p is ϵ-far from q, then p' is $\epsilon/3$-far from uniform; and if p is identical to q, then p' is uniform. Therefore, the algorithm is an ϵ-tester for identity. It suffices to show that the algorithm preserves differential privacy.

Assume X is the set of samples drawn from p, and denote by π the bits of randomness that the mapping used to build X'_π, the set of samples from p'. Assume Y is a sample set from p that differs from X in exactly one location. Then X'_π also differs from X'_π in at most one location, because each sample from p is used in generating exactly one sample from p'. Let $\mathcal{A}$ be the (ϵ, ξ)-private uniformity tester and denote by $\mathcal{A}(X'_\pi)$ the output of the tester on input X'_π. Since the algorithm is ξ-private, we have:

$$\mathbf{Pr}[\mathcal{A}(S'_\pi) = \mathsf{accept}] \leq e^{\xi} \cdot \mathbf{Pr}[\mathcal{A}(Y) = \mathsf{accept}] \ .$$

Let $T(X)$ denote the output of our algorithm. By construction, we have

$$\begin{aligned} \frac{\mathbf{Pr}[T(X) = \mathsf{accept}]}{\mathbf{Pr}[T(Y) = \mathsf{accept}]} &= \frac{\sum_\pi \mathbf{Pr}[\mathcal{A}(X'_\pi) = \mathsf{accept}] \cdot \mathbf{Pr}[\pi]}{\sum_\pi \mathbf{Pr}[\mathcal{A}(Y'_\pi) = \mathsf{accept}] \cdot \mathbf{Pr}[\pi]} \\ &\leq \frac{\sum_\pi e^{\xi} \cdot \mathbf{Pr}[\mathcal{A}(Y'_\pi) = \mathsf{accept}] \cdot \mathbf{Pr}[\pi]}{\sum_\pi \mathbf{Pr}[\mathcal{A}(Y'_\pi) = \mathsf{accept}] \cdot \mathbf{Pr}[\pi]} \\ &\leq e^{\xi} \ . \end{aligned}$$

By the same argument, we can show the above inequality holds when the output is reject. Therefore, our algorithm is an (ϵ, ξ)-private tester. □

E. Proof of Theorem 4.1

Theorem 4.1. *Given $s = O(\sqrt{n}/(\epsilon\sqrt{\xi}) + \sqrt{n}/\epsilon^2)$ samples from a distribution p over $[n]$, Algorithm 1 is an (ϵ, ξ)-private uniformity tester, if s is sufficiently smaller than the domain size n.*

Proof: Algorithm 1 draws s samples from the underlying distribution p. We use the Laplace mechanism to make the algorithm private: Let K be the number of unique elements in the sample set. Since changing one sample in the sample set can change the number of unique elements by no more than two, adding Laplace noise with parameter $2/\xi$ to K makes it ξ-private. Using the composition Theorem A.3, the algorithm is ξ-private.

To show the algorithm is an ϵ-tester, we prove the statistic K' concentrates well around its expected value in both the

soundness and completeness cases. Using Lemmas 1 and 2 in (Paninski, 2008), we have the following inequalities for the number of unique elements:

$$\mathbf{E}_{\mathcal{U}}[K] - \mathbf{E}_{\mathcal{P}}[K] \geq \frac{s^2\|p - U_n\|_1^2}{n} \tag{7}$$

and

$$\mathbf{Var}[K] \leq \mathbf{E}_{\mathcal{U}}[K] - \mathbf{E}_{\mathcal{P}}[K] + \frac{s^2}{n}. \tag{8}$$

First, we show the algorithm is an ϵ-tester for uniformity. Then, we prove that it is ξ-private.

Assume that the underlying distribution is the uniform distribution. Note that $\mathbf{E}[K] = \mathbf{E}[K']$. Then, by the Chebyshev inequality and Equation 8 we have that:

$$\begin{aligned}
&\mathbf{Pr}[|K' - \mathbf{E}_{\mathcal{U}}[K]| \geq \frac{C^2}{2\epsilon^2}] \\
&= \mathbf{Pr}[|K' - \mathbf{E}_{\mathcal{U}}[K']| \geq \frac{C^2}{2\epsilon^2}] \\
&\leq \frac{4\epsilon^4}{C^4}\,\mathbf{Var}[K'] \\
&\leq \frac{4\epsilon^4}{C^4}\,(\mathbf{Var}[K] + \mathbf{Var}[\mathbf{Lap}(2/\xi)]) \\
&\leq \frac{4\epsilon^4}{C^4}\,(\frac{s^2}{n} + \frac{8}{\xi^2}) \\
&\leq \frac{4}{C^2} + \frac{32\epsilon^4}{C^4\xi^2} \\
&\leq \frac{1}{3}\,,
\end{aligned}$$

where the last inequality comes from the fact that $C \geq \max(3.73\ \epsilon/\sqrt{\xi}, 4.9)$. Thus, the probability of rejecting $\mathcal{P}$ is less than $1/3$.

Now suppose $\mathcal{P}$ is a distribution which is ϵ-far from uniform. Again by the Chebyshev inequality and Equation (8) we have that:

$$\begin{aligned}
&\mathbf{Pr}[|K' - \mathbf{E}_{\mathcal{P}}[K]| \geq (\mathbf{E}_{\mathcal{U}}[K] - \mathbf{E}_{\mathcal{P}}[K])/2] \\
&= \mathbf{Pr}[|K' - \mathbf{E}_{\mathcal{P}}[K']| \geq (\mathbf{E}_{\mathcal{U}}[K] - \mathbf{E}_{\mathcal{P}}[K])/2] \\
&\leq \frac{4\mathbf{Var}[K']}{(\mathbf{E}_{\mathcal{U}}[K] - \mathbf{E}_{\mathcal{P}}[K])^2} \\
&= \frac{4(\mathbf{Var}[K] + \mathbf{Var}[\mathbf{Lap}(2/\xi)])}{(\mathbf{E}_{\mathcal{U}}[K] - \mathbf{E}_{\mathcal{P}}[K])^2} \\
&= \frac{4(\mathbf{Var}[K] + 8\xi^{-2})}{(\mathbf{E}_{\mathcal{U}}[K] - \mathbf{E}_{\mathcal{P}}[K])^2} \\
&\leq \frac{4(\mathbf{E}_{\mathcal{U}}[K] - \mathbf{E}_{\mathcal{P}}[K] + s^2/n + 8\xi^{-2})}{(\mathbf{E}_{\mathcal{U}}[K] - \mathbf{E}_{\mathcal{P}}[K])^2} \\
&\leq \frac{4}{\mathbf{E}_{\mathcal{U}}[K] - \mathbf{E}_{\mathcal{P}}[K]} + \frac{4s^2/n + 32/\xi^2}{(\mathbf{E}_{\mathcal{U}}[K] - \mathbf{E}_{\mathcal{P}}[K])^2}\,.
\end{aligned}$$

On the other hand by Equation 7, $\mathbf{E}_{\mathcal{U}}[K] - \mathbf{E}_{\mathcal{P}}[K]$ is at least C^2/ϵ^2. Thus,

$$\begin{aligned}
&\mathbf{Pr}\left[|K' - \mathbf{E}_{\mathcal{P}}[K]| \geq (\mathbf{E}_{\mathcal{U}}[K] - \mathbf{E}_{\mathcal{P}}[K])/2\right] \\
&\leq \frac{4\epsilon^2}{C^2} + \frac{4s^2\epsilon^4}{C^4 n} + \frac{32\epsilon^4}{C^4\xi^2} \\
&\leq \frac{1}{3}\,,
\end{aligned}$$

where the last inequality is true when $C \geq \max\left(6\epsilon, 6, 4.12\epsilon/\sqrt{\xi}\right)$. Thus, the probability of accepting is less than $1/3$. □

F. Proof of Theorem 4.2

Theorem 4.2. *Algorithm 2 is an (ϵ, ξ)-private tester for uniformity.*

Proof: Let $X = \{x_1, \ldots, x_s\}$ be a set of s samples from p. Let $f(X)$ be the number of collisions in X. All variables are as defined in Algorithm 2. First, we show that $\hat{f}(X)$ and $n_{\max}(X)$ concentrate well around their expected values.

Lemma F.1. *If s is $\Theta\left(\frac{\sqrt{n}}{\epsilon^2} + \frac{\sqrt{n\log n}}{\epsilon\,\xi^{1/2}} + \frac{\sqrt{n\max(1,\log 1/\xi)}}{\epsilon\,\xi} + \frac{1}{\epsilon^2\,\xi}\right)$, the following holds with probability at least $11/12$:*

- *If p is the uniform distribution, then $\hat{f}(X)$ is less than $\frac{1+\epsilon^2/6}{n}\binom{s}{2}$.*
- *If p is ϵ-far from uniform, then $\hat{f}(X)$ is greater than $\frac{1+\epsilon^2/6}{n}\binom{s}{2}$.*

Proof: First, we compute the expected value of $\hat{f}(X)$. Since the expected value of the noise is zero, $\mathbf{E}[\hat{f}(X)]$ is equal to $\mathbf{E}[f(X)]$. So, if p is uniform, then $\mathbf{E}[\hat{f}(X)]$ is $\binom{s}{2}/n$, and if p is ϵ-far from uniform $\mathbf{E}[\hat{f}(X)]$ is at least $(1+\epsilon^2)\binom{s}{2}/n$. Let α satisfy $\|p\|_2^2 = (1+\alpha)/n$ and σ be the standard deviation of $\hat{f}(X)$. We make an assumption that $|\epsilon^2/6 - \alpha|\binom{s}{2}/n$ is at least $\sqrt{12}\sigma$. Below, this assumption concludes the statement of the lemma. Later, we prove that the assumption holds for sufficiently large s.

The conditions of the lemma hold if $\hat{f}(X)$ is closer to its expected value than the distance of the threshold, $\frac{1+\epsilon^2/6}{n}\binom{s}{2}$, to its expected value. Using the Chebyshev inequality, the probability that the conditions do not hold is at most

$$\begin{aligned}
&\mathbf{Pr}\left[\left|\hat{f}(X) - \mathbf{E}[f(X)]\right| > \left|\frac{1+\epsilon^2/6}{n}\binom{s}{2} - \mathbf{E}[f(X)]\right|\right] \\
&= \mathbf{Pr}\left[\left|\hat{f}(X) - \mathbf{E}[f(X)]\right| > \frac{|\epsilon^2/6 - \alpha|}{n}\binom{s}{2}]\right] \\
&\leq \mathbf{Pr}\left[|\hat{f}(X) - \mathbf{E}[f(x)]| \geq \sqrt{12}\sigma\right] \leq \frac{1}{12}\,.
\end{aligned}$$

Thus, it is sufficient to show that

$$\frac{|\epsilon^2/6-\alpha|}{n}\binom{s}{2} \geq \sqrt{12}\sigma. \tag{9}$$

Recall that σ^2 is equal to $\mathbf{Var}[f(X)] + \mathbf{Var}[\mathbf{Lap}(2\eta_f/\xi)]$, so σ is at most:

$$\sqrt{2\max(\mathbf{Var}[f(x)], 8\eta_f^2/\xi^2)}\,.$$

Hence, we prove two stronger inequalities that yield to Equation (9):

$$s \geq \sqrt{\frac{20\,n\,\sqrt{\mathbf{Var}[f(X)]}}{|\epsilon^2/6-\alpha|}}, \tag{10}$$

and

$$s \geq \sqrt{\frac{28\,n\,\eta_f}{\xi|\epsilon^2/6-\alpha|}}\,. \tag{11}$$

Using a similar proof to the proof of Lemma 4 in (Diakonikolas et al., 2016), the inequality of Equation (10) holds for $s = c\sqrt{n}/\epsilon^2$ for sufficiently large constant c. Now, we focus on Equation (11). If p is a uniform distribution, α is zero, and if p is ϵ-far form being uniform, then α is at least ϵ^2. Therefore, the denominator is at least $\epsilon^2/6$. Solving Equation (11) for s, we have:

$$s \geq c' \cdot \left(\frac{1}{\epsilon^2\,\xi} + \frac{\sqrt{n\log n}}{\epsilon\,\xi^{1/2}} + \frac{\sqrt{n\max(1,\log 1/\xi)}}{\epsilon\,\xi}\right)\,.$$

Hence, for sufficiently large constant c', Equation (9) holds and the proof is complete. □

We have the following lemma:

Lemma F.2. *Let X be a sample set of size s from the uniform distribution over $[n]$. With probability $11/12$, we have*

$$\hat{n}_{\max} \leq \max\left(\frac{3}{2}\cdot\frac{s}{n}, 12\,e^2\ln 24n\right) + \frac{2\ln 12}{\xi}\,.$$

Proof: First, we show that $n_{\max}(X)$ is at most $\max\left(3s/(2n), 12\,e^2\ln 24n\right)$ with probability at least 23/24. It suffices to show that all of the $n_i(X)$'s are smaller than this bound. Consider the following cases: First, assume s is at most $12\,n\cdot\ln(24n)$. Let $k := 12\,e^2\cdot\ln(24n) \geq e^2 s/n$. If $s \leq k$, then $n_{\max}(X)$ is at most $\max\left(3s/(2n), 12e^2\ln 24n\right)$. Otherwise,

$$\mathbf{Pr}\left[n_i(X) > k\right] \leq \binom{s}{k}\cdot\frac{1}{n^k} \leq \left(\frac{s\cdot e}{k}\right)^k\cdot\frac{1}{n^k} \leq e^{-k} \leq \frac{1}{24\,n}.$$

Second, assume s is greater than $12\,n\cdot\ln(24n)$. By the Chernoff bound, we have

$$\mathbf{Pr}\left[n_i(X) > \frac{s}{n}\left(1+\frac{1}{2}\right)\right] \leq \exp(-\frac{s}{12\,n}) \leq \frac{1}{24\,n}\,.$$

Thus,

$$\mathbf{Pr}\left[n_i(X) > \max\left(3s/(2n), 12\,e^2\ln 24n\right)\right] \leq \frac{1}{24\,n}\,.$$

Using the union bound, with probability $23/24$ all the $n_i(X)$'s, and consequently $n_{\max}(X)$, are smaller than $\max\left(3s/(2n), 12\,e^2\ln 24n\right)$.

Moreover, based on the properties of the Laplace distribution, we have

$$\mathbf{Pr}\left[\mathbf{Lap}(2/\xi) \geq \frac{2\ln 12}{\xi}\right] \leq \frac{\exp(-\ln 12)}{2} \leq \frac{1}{24}\,.$$

By the union bound, $n_{\max}(X)$ and $\mathbf{Lap}(2/\xi)$ are not exceeding the aforementioned bounds with probability $11/12$. Therefore, we have

$$\mathbf{Pr}\left[\hat{n}_{\max} < \max\left(3s/(2n), 12\,e^2\ln 24n\right) + \frac{2\ln 12}{\xi}\right] \geq \frac{11}{12}\,.$$

Thus, the proof is complete. □

Given X, we define two probabilistic events, $E_1(X)$ and $E_2(X)$, to be

$$E_1(X)\colon \hat{n}_{\max} < T \qquad E_2(X)\colon \hat{f}(X) < \frac{6+\epsilon^2}{6n}\binom{s}{2}\,,$$

where the probability is taken over the randomness of the noise. Observe that $E_1(X)$ and $E_2(X)$ are independent. We use $\overline{E_1}(X)$ and $\overline{E_2}(X)$ to indicate the complementary events. Let $\mathcal{M}(X)$ denote the output of the algorithm when the input sample set is X. We set the output, O, to accept, if both $E_1(X)$ and $E_2(X)$ are true, and at the end of the algorithm we may flip the output with small probability. Here, we prove the probability of outputting the correct answer is at least $2/3$. Consider two following cases:

***(i)* p is uniform:** Using Lemma F.2, with probability at least $11/12$ we have that $\hat{n}_{\max}$ is less than T. By Lemma F.1, $\hat{f}$ is less than $\binom{s}{2}|1+\epsilon^2/6|/n$ with probability at least $11/12$. Therefore, $\mathbf{Pr}[\overline{E_1}(X)]$ and $\mathbf{Pr}[\overline{E_2}(X)]$ are at most $1/12$. At the end of the algorithm, we flip the output with probability at most $1/6$. Using the union bound, we have

$$\begin{aligned}&\mathbf{Pr}\left[\mathcal{M}(X) = \mathsf{accept}\right]\\ &\geq 1 - \mathbf{Pr}\left[\overline{E_1}(X)\right] - \mathbf{Pr}\left[\overline{E_2}(X)\right] - \frac{1}{6} \geq \frac{2}{3}\,.\end{aligned}$$

***(ii)* p is ϵ-far from uniform:** By Lemma F.1, $\hat{f}(X)$ is greater than $\binom{s}{2}|1+\epsilon^2/6|/n$ with probability at least $11/12$, so $\mathbf{Pr}[E_2(X)]$ is at most $1/12$. We flip the output of the algorithm with probability at most $1/6$. As a result, we have

$$\mathbf{Pr}\left[\mathcal{M}(X) = \mathsf{reject}\right] \geq 1 - \mathbf{Pr}\left[E_2(X)\right] - \frac{1}{6} \geq \frac{2}{3}\,.$$

Thus, with probability at least $2/3$ we output the correct answer.

In the rest of the proof, we focus on proving the privacy guarantee. It is not hard to see that $|n_{\max}(X) - n_{\max}(Y)|$ is at most one. By the properties of the Laplace mechanism in Lemma A.1, $\hat{n}_{\max}(X)$ is $\xi/2$-private. Assume $|f(X) - f(Y)|$ is at most η_f. Then, $\hat{f}(X)$ is $\xi/2$-private as well. Since privacy preserved after post-processing (Lemma A.2), both $E_1(X)$ and $E_2(X)$ are $\xi/2$-private. Using the composition Lemma A.3, the output is ξ-private (by Lemma A.3).

Now, assume $|f(X) - f(Y)|$ is greater than η_f. In this case, we show that $n_{\max}(X)$ has to be large. Therefore, the output is reject with high probability regardless of $\hat{f}(X)$. Although $\hat{f}(X)$ is not private, it cannot affect the output drastically and the output remains private. We prove this formally below. Without loss of generality, assume we replace a sample i in X with j to get Y. Thus, we have

$$\begin{aligned}
&|f(X) - f(Y)| \\
&= \left| \tbinom{n_i(X)}{2} + \tbinom{n_j(X)}{2} - \tbinom{n_i(Y)}{2} - \tbinom{n_j(Y)}{2} \right| \\
&= \left| \tbinom{n_i(X)}{2} + \tbinom{n_j(X)}{2} - \tbinom{n_i(X)-1}{2} - \tbinom{n_j(X)+1}{2} \right| \\
&= |n_i(X) - 1 - n_j(X)| \\
&\leq n_{\max}(X) \ ,
\end{aligned}$$

where the inequality comes from the assumption that there is at least one copy of i in X. Therefore, $n_{\max}(X)$ is greater than η_f as well. Since T is even smaller than η_f, it is very unlikely that $\hat{n}_{\max}$ be smaller than the threshold T. More formally, by the properties of the Laplace distribution, we have:

$$\begin{aligned}
\mathbf{Pr}[E_1(X)] &= \mathbf{Pr}[\hat{n}_{\max}(X) \leq T] \\
&=\leq \mathbf{Pr}[\hat{n}_{\max}(X) - n_{\max}(X) \leq T - \eta_f] \\
&\leq \mathbf{Pr}\left[\mathbf{Lap}(2/\xi) \leq -\tfrac{2\max(\ln 3, \ln 3/\xi)}{\xi}\right] \\
&\leq \exp\left(-\max(\ln 3, \ln 3/\xi)\right)/2 \\
&\leq \min(1/6, \xi/6) \ .
\end{aligned} \tag{12}$$

Now, consider the case that the algorithm output accept on input X. It is not hard to see that

$$\begin{aligned}
&\mathbf{Pr}\left[\mathcal{M}(X) = \mathsf{accept}\right] \\
&= (5/6) \cdot \mathbf{Pr}[E_1(X) \wedge E_2(X)] \\
&+(1/6) \cdot (1 - \mathbf{Pr}[E_1(X) \wedge E_2(X)]) \\
&= (2/3) \cdot \mathbf{Pr}[E_1(X) \wedge E_2(X)] + 1/6 \\
&= (2/3) \cdot \mathbf{Pr}[E_1(X)] \cdot \mathbf{Pr}[E_2(X)] + 1/6 \ .
\end{aligned} \tag{13}$$

Observe that since we flip the answer with probability $1/6$ at the end, $\mathbf{Pr}[\mathcal{M}(X) = \mathsf{accept}]$ and $\mathbf{Pr}[\mathcal{M}(Y) = \mathsf{accept}]$ are at least $1/6$. By this fact, Equation (12), and Equation (13), we have:

$$\frac{\mathbf{Pr}[\mathcal{M}(X) = \mathsf{accept}]}{\mathbf{Pr}[\mathcal{M}(Y) = \mathsf{accept}]} \leq \frac{\mathbf{Pr}[E_1(X)] + 1/6}{1/6} \leq \xi + 1 < e^{\xi} \ .$$

Now, consider the case where the output of the algorithm is reject on the input X. Similar to Equation (12), we can prove $\mathbf{Pr}[E_1(Y)]$ is at most $\min(1/6, \xi/6)$. Similar to Equation (13), it is not hard to see that

$$\mathbf{Pr}\left[\mathcal{M}(X) = \mathsf{reject}\right] = (2/3) \cdot (\mathbf{Pr}[\overline{E_1}(X) \vee \overline{E_2}(X)]) + 1/6 \ . \tag{14}$$

If $\mathbf{Pr}[\mathcal{M}(X) = \mathsf{reject}]$ is at most $\mathbf{Pr}[\mathcal{M}(Y) = \mathsf{reject}]$, then clearly, we have:

$$\frac{\mathbf{Pr}[\mathcal{M}(X) = \mathsf{reject}]}{\mathbf{Pr}[\mathcal{M}(Y) = \mathsf{reject}]} \leq 1 < e^{\xi} \ .$$

Thus, assume $\mathbf{Pr}[\mathcal{M}(X) = \mathsf{reject}]$ is less than $\mathbf{Pr}[\mathcal{M}(Y) = \mathsf{reject}]$. Then, we have:

$$\begin{aligned}
&\frac{\mathbf{Pr}[\mathcal{M}(X) = \mathsf{reject}]}{\mathbf{Pr}[\mathcal{M}(Y) = \mathsf{reject}]} \\
&= \frac{(2/3) \cdot (\mathbf{Pr}[\overline{E_1}(X) \vee \overline{E_2}(X)]) + 1/6}{(2/3) \cdot (\mathbf{Pr}[\overline{E_1}(Y) \vee \overline{E_2}(Y)]) + 1/6} \\
&\leq \frac{\mathbf{Pr}[\overline{E_1}(X) \vee \overline{E_2}(X)]}{\mathbf{Pr}[\overline{E_1}(Y) \vee \overline{E_2}(Y)]} \leq \frac{1}{1 - \mathbf{Pr}[E_1(Y)]} \\
&\leq \frac{1}{1 - \min(1/6, \xi/6)} < 1 + \xi < e^{\xi} \ .
\end{aligned}$$

The second to last inequality is true since we showed previously that $\mathbf{Pr}[E_1(Y)]$ is at most $\min(1/6, \xi/6)$. Hence, the proof is complete. □

G. Proof of Theorem 5.1

Theorem 5.1. *Given sample access to two distributions p and q, Algorithm 3 is an (ϵ, ξ)-private tester for equivalence of p and q.*

Proof: Our proof has two main parts. First, we show that the algorithm outputs the correct answer with probability $2/3$. Second, we show that the algorithm is private.

Proof of Correctness: First, assume p and q are equal. In the algorithm, we compute Z and add Laplace noise, η, to it. Then we compare it to threshold $T := \epsilon^2 m^2/(8n + 4m)$. Based on Equation (5), we have

$$\mathbf{E}[Z'] = \mathbf{E}[Z] + \mathbf{E}[\eta] = \mathbf{E}[Z].$$

Using the Chebyshev inequality and Equation (6),

$$\begin{aligned}\mathbf{Pr}[\text{outputting reject}] &= \mathbf{Pr}[Z' > T] \\ &\leq \frac{\mathbf{Var}[Z']}{T^2} \leq \frac{\mathbf{Var}[Z] + \mathbf{Var}[\eta]}{T^2} \\ &\leq \frac{2\min\{m,n\} + 128/\xi^2}{T^2} \leq \frac{1}{3}\,,\end{aligned}$$

where the last inequality is true for a sufficiently large universal constant C.

Case 1: Consider the case $m \leq n$. Then,

$$\begin{aligned}\frac{2\min\{m,n\}}{T^2} &= \frac{2\,m\,(8n+4m)^2}{m^4\epsilon^4} \leq \frac{2\,m\,(12n)^2}{m^4\epsilon^4} \\ &\leq 288\left(\frac{n^{2/3}}{\epsilon^{4/3}}\cdot\frac{1}{m}\right)^3 \\ &\leq \frac{288}{C^3} \leq \frac{1}{6}\,,\end{aligned}$$

where the last inequality is true for C greater than 12. Moreover,

$$\begin{aligned}\frac{128\,\xi^{-2}}{T^2} &\leq \frac{128\,(8n+4m)^2}{\xi^2\,m^4\,\epsilon^4} \leq \frac{128\,(12\,n)^2}{\xi^2\,m^4\,\epsilon^4} \\ &\leq 18432\left(\frac{\sqrt{n}}{\sqrt{\xi}\epsilon}\cdot\frac{1}{m}\right)^4 \\ &\leq \frac{18432}{C^4} \leq \frac{1}{6}\,,\end{aligned}$$

where the last inequality is true for C greater than 19. Thus, for sufficiently large C, the probability of rejecting two identical distribution p and q is less than $1/3$.

Case 2: Consider the case $n < m$. Then,

$$\begin{aligned}\frac{2\min\{m,n\}}{T^2} &= \frac{2\,n\,(8n+4m)^2}{m^4\epsilon^4} \leq \frac{2\,n\,(12m)^2}{m^4\epsilon^4} \\ &\leq 288\left(\frac{\sqrt{n}}{\epsilon^2}\cdot\frac{1}{m}\right)^2 \\ &\leq \frac{288}{C^2} \leq \frac{1}{6}\,,\end{aligned}$$

where the last inequality is true for C greater than 42. Moreover,

$$\begin{aligned}\frac{128\,\xi^{-2}}{T^2} &\leq \frac{128\,(8n+4m)^2}{\xi^2\,m^4\,\epsilon^2} \leq \frac{128\,(12\,m)^2}{\xi^2\,m^4\,\epsilon^2} \\ &\leq 18432\left(\frac{1}{\xi\,\epsilon}\cdot\frac{1}{m}\right)^2 \leq \frac{18432}{C^2} \leq \frac{1}{6}\,,\end{aligned}$$

where the last inequality is true for C greater than 136. Thus, for sufficiently large C the probability of rejecting two identical distribution p and q is less than 1/3.

Now, suppose p and q are at least ϵ-far from each other in ℓ^1-distance. We show that in this case Z' is greater than T with high probability using Chebyshev's inequality. Based on Equation (6), we bound the variance of Z' in terms of the expected value of Z'. First, observe that, by Equation (5), we have that $\mathbf{E}[Z']$ is at least $C/6$ for any setting of parameters. Thus, for sufficiently large C, we can assume $\mathbf{E}[Z']$ is at least 360. Let I_1 be the set of all indices i such that $(1-(1-e^{-m(p_i+q_i)})/(m(p_i+q_i)))$ is greater $1/2$, and let I_2 be the set of remaining indices, i.e., $I_2 = [n] \setminus I_1$. By Equation (5), we have

$$\begin{aligned}\mathbf{E}[Z']^2 &= \left(\sum_i \frac{(p_i-q_i)^2}{p_i+q_i} m\left(1-\frac{1-e^{-m(p_i+q_i)}}{m(p_i+q_i)}\right)\right)^2 \\ &\geq 360\sum_i \frac{(p_i-q_i)^2}{p_i+q_i} m\left(1-\frac{1-e^{-m(p_i+q_i)}}{m(p_i+q_i)}\right) \\ &\geq 360\sum_{i\in I_1} \frac{(p(i)-q(i))^2}{p(i)+q(i)} m\left(1-\frac{1-e^{-m(p(i)+q(i))}}{m(p(i)+q(i))}\right) \\ &\geq 36\sum_{i\in I_1} 5m\frac{(p(i)-q(i))^2}{p(i)+q(i)}\,.\end{aligned}$$

On the other hand, for any i in I_2, we can conclude that $m(p(i)+q(i))$ is less than 2. Therefore, $m\frac{(p(i)-q(i))^2}{p(i)+q(i)}$ is at most 2. Thus, $\sum_{i\in I_2} 5m\frac{(p(i)-q(i))^2}{p(i)+q(i)}$ is at most $10\,n$. Since $\frac{(p(i)-q(i))^2}{p(i)+q(i)}$ is less than $p(i)+q(i)$, $\sum_{i\in I_2} 5m\frac{(p(i)-q(i))^2}{p(i)+q(i)}$ is also less than $10\,m$. Hence, we have

$$\begin{aligned}\mathbf{Var}[Z] &\leq 2\min\{m,n\} + \sum_i 5m\frac{(p(i)-q(i))^2}{p(i)+q(i)} \\ &\leq 2\min\{m,n\} + \sum_{i\in I_1} 5m\frac{(p(i)-q(i))^2}{p(i)+q(i)} \\ &\quad + \sum_{i\in I_2} 5m\frac{(p(i)-q(i))^2}{p(i)+q(i)} \\ &\leq 12\min\{m,n\} + \frac{\mathbf{E}[Z']^2}{36}\,.\end{aligned}$$

By Equation (5), the expected value of Z' is at least $2T$.

Using Chebyshev's inequality, we obtain

$$
\begin{aligned}
&\mathbf{Pr}[\text{outputting "Accept"}] \\
&= \mathbf{Pr}[Z' \le T] \le \mathbf{Pr}[\mathbf{E}[Z'] - Z' \ge \mathbf{E}[Z'] - T] \\
&\le \mathbf{Pr}\left[\mathbf{E}[Z'] - Z' \ge \frac{\mathbf{E}[Z']}{2}\right] \\
&\le \frac{4\mathbf{Var}[Z']}{\mathbf{E}[Z']^2} \le \frac{4(\mathbf{Var}[Z] + \mathbf{Var}[\eta])}{\mathbf{E}[Z']^2} \\
&\le \frac{48\min\{m,n\}}{\mathbf{E}[Z']^2} + \frac{10}{\mathcal{A}} + \frac{512}{\mathbf{E}[Z']^2\,\xi^2} \\
&\le \frac{48\min\{m,n\}(4n+2m)^2}{m^4\epsilon^4} + \frac{1}{9} + \frac{512(4n+2m)^2}{m^4\,\epsilon^4\,\xi^2} \\
&\le \frac{1}{3}\,,
\end{aligned}
$$

where the last inequality is true for sufficiently large C.

Proof of Privacy Guarantee: First, observe that the value of Z does not change drastically over two neighboring datasets. More formally, we have the following simple lemma:

Lemma G.1. *The sensitivity of the statistic Z is at most 8.*

Proof: Assume two neighboring dataset x and y. Let $Z^{(x)}$ and $Z^{(y)}$ be the statistic for x and y respectively. We define Z_i as follows:

$$
Z_i := \begin{cases} \dfrac{|X_i + Y_i| - X_i - Y_i}{X_i + Y_i} & \text{if } X_i + Y_i \neq 0 \\ 0 & \text{otherwise.} \end{cases}
$$

We use a superscript (x) or (y) for X_i, Y_i, Z_i to indicate the corresponding dataset we calculate them from. Since x and y are two neighboring datasets, there is a sample i in the x which has been replaced by j. Without loss of generality, assume i was a sample from p. This implies that $X_i^{(x)} - X_i^{(y)} = 1$ and $Y_i^{(x)} = Y_i^{(y)}$.

If $X_i^{(y)} + Y_i^{(y)}$ is zero, then $Z_i^{(x)}$ is one. Thus, the difference of $Z_i^{(x)}$ and $Z_i^{(y)}$ is one. Now, assume $X_i^{(y)} + Y_i^{(y)}$ is at least one. Then, we have

$$
\begin{aligned}
&\left|Z_i^{(x)} - Z_i^{(y)}\right| \\
&= \left|\frac{\left(X_i^{(x)} - Y_i^{(x)}\right)^2}{X_i^{(x)} + Y_i^{(x)}} - \frac{\left(X_i^{(y)} - Y_i^{(y)}\right)^2}{X_i^{(y)} + Y_i^{(y)}}\right| \\
&= \left|\frac{\left(X_i^{(y)} - Y_i^{(y)} + 1\right)^2}{X_i^{(y)} + Y_i^{(y)} + 1} - \frac{\left(X_i^{(y)} - Y_i^{(y)}\right)^2}{X_i^{(y)} + Y_i^{(y)}}\right| \\
&= \left|\frac{\left(X_i^{(y)} - Y_i^{(y)}\right)^2 + 2\left(X_i^{(y)} - Y_i^{(y)}\right) + 1}{X_i^{(y)} + Y_i^{(y)} + 1} - \frac{\left(X_i^{(y)} - Y_i^{(y)}\right)^2}{X_i^{(y)} + Y_i^{(y)}}\right| \\
&= \left|\frac{2\left(X_i^{(y)} - Y_i^{(y)}\right) + 1}{X_i^{(y)} + Y_i^{(y)} + 1} - \frac{(X_i^{(y)} - Y_i^{(y)})^2}{\left(X_i^{(y)} + Y_i^{(y)} + 1\right)\cdot\left(X_i^{(y)} + Y_i^{(y)}\right)}\right| \\
&\le 4\,.
\end{aligned}
$$

Similarly, we can show $|Z_j^{(x)} - Z_j^{(y)}|$ is at most four. Hence, we can conclude that $|Z^{(x)} - Z^{(y)}|$ is at most eight. □

Therefore, using the property of the Laplace mechanism (Lemma A.1), Z' is ξ-private. Using Lemma A.2 and the fact that the output of the algorithm is a function of Z', we conclude the algorithm is ξ-private. □