Journal of Privacy and Confidentiality
Vol. 10 (1) 2020

TPDP 2018

Submitted Mar 2019
Published Jan 2020

DIFFERENTIALLY PRIVATE INFERENCE FOR BINOMIAL DATA

JORDAN AWAN AND ALEKSANDRA SLAVKOVIĆ

Department of Statistics, The Pennsylvania State University, University Park, PA 16802,
e-mail address: awan@psu.edu

Department of Statistics, The Pennsylvania State University, University Park, PA 16802,
e-mail address: sesa@psu.edu

Abstract. We derive uniformly most powerful (UMP) tests for simple and one-sided hypotheses for a population proportion within the framework of differential privacy (DP), optimizing finite sample performance. We show that in general, DP hypothesis tests can be written in terms of linear constraints, and for exchangeable data can always be expressed as a function of the empirical distribution. Using this structure, we prove a 'Neyman-Pearson Lemma' for binomial data under DP, where the DP-UMP only depends on the sample sum. Our tests can also be stated as a post-processing of a DP summary statistic, whose distribution we coin "Truncated-Uniform-Laplace" (Tulap), a generalization of the Staircase and discrete Laplace distributions.

We show that by post-processing the Tulap statistic, we are able to obtain exact p-values corresponding to the DP-UMP, uniformly most accurate (UMA) one-sided confidence intervals, optimal confidence distributions, uniformly most powerful unbiased (UMPU) two-sided tests, and uniformly most accurate unbiased (UMAU) two-sided confidence intervals. As each of these quantities are a post-processing of the same summary statistic, there is no increased cost to privacy by including these additional results, allowing for a complete statistical analysis at a fixed privacy cost. We also show that our results can be applied to distribution-free hypothesis tests for continuous data. Our simulation results demonstrate that all our tests have exact type I error, and are more powerful than current techniques.

1. Introduction

Differential privacy (DP), introduced by Dwork et al. (2006), offers a rigorous measure of disclosure risk and more broadly, a formal privacy framework, such that privacy guarantees hold regardless of the assumed knowledge of the malicious user. To satisfy DP, a procedure cannot be a deterministic function of the sensitive data, but must incorporate additional randomness, beyond sampling. Subject to the DP constraint, it is natural to search for a procedure which maximizes the utility of the output. Many works address the goal of minimizing the distance between the outputs of the randomized DP procedure and standard

Key words and phrases: Bernoulli, Hypothesis Test, Confidence Interval, Frequentist, Statistical Disclosure Control, Neyman-Pearson, Confidence Distribution.

This paper is an expansion of an earlier version, Awan and Slavković (2018*a*).

www.journalprivacyconfidentiality.org
DOI:10.29012/jpc.725

non-private algorithms, but fewer attempt to infer properties about the underlying population (for notable exceptions, see related work), which is typically the goal in statistics and scientific research. In this paper, we focus on the setting where each individual contributes a sensitive binary value, and we wish to infer the population proportion via hypothesis tests and confidence intervals, subject to DP. While a simple setting, there are many important problems that fit this format, where privacy is a concern (such as determining the proportion of binge drinkers in universities, the proportion of illegal immigrants in the US, or the proportion of citizens participating in an illegal activity).

In particular, our main results are focused on deriving *uniformly most powerful* (UMP) and *uniformly most powerful unbiased* (UMPU) tests, related p-values, and confidence intervals, which optimize finite sample performance. Crucially, all of these statistical tools can be expressed as a post-processing of a DP summary statistic, which has the distribution we coin *Truncated-Uniform-Laplace* (*Tulap*). We show that by combining an understanding of the Tulap distribution and classical statistical methods, we are able to compute several private statistical results at a fixed privacy cost. Furthermore, while these tests are designed for binary data, we also show that they can be used to construct certain hypothesis tests for continuous data as well.

UMP tests are fundamental to classical statistics, being closely linked to sufficiency, likelihood inference, and confidence sets. However, finding UMP tests can be hard and in many cases they do not even exist (see Schervish, 1996, Section 4.4). Our results are the first to achieve UMP tests under $(\epsilon, \delta)-$DP, and are among the first steps towards a general theory of optimal inference under DP.

Related work Vu and Slavković (2009) were the first to perform classical hypothesis tests under DP. They develop private tests for population proportions as well as for independence in 2×2 contingency tables. In both settings, they fix the noise adding distribution, and use approximate sampling distributions to perform these DP tests. A similar approach is used by Solea (2014) to develop tests for normally distributed data. The work of Vu and Slavković (2009) was extended by Wang, Lee and Kifer (2015) and Gaboardi et al. (2016), developing additional tests for multinomial data. To implement their tests, Wang, Lee and Kifer (2015) develop asymptotic sampling distributions, verifying via simulations that the type I errors are reliable. On the other hand, Gaboardi et al. (2016) use simulations to compute an empirical type I error. Uhler, Slavković and Fienberg (2013) develop DP chi-squared tests and p-values for GWAS data, and derive the exact sampling distribution of their noisy statistic. Working under "local differential privacy," a stronger notion of privacy than DP, Gaboardi and Rogers (2018) develop multinomial tests based on asymptotic distributions. Given a DP output, Sheffet (2017) and Barrientos et al. (2019) develop significance tests for regression coefficients. Following a common strategy in the field of Statistics, Wang et al. (2018) develop approximating distributions for DP statistics, which can be used to construct hypothesis tests and confidence intervals. In a recent work, Canonne et al. (2019) show that for simple hypothesis tests, a DP test based on a clamped likelihood ratio test achieves optimal sample complexity.

While not directly related to the testing problems we consider, Wasserman and Zhou (2010) showed that the constraint of differential privacy can be interpreted in terms of hypothesis tests on the database, and Kairouz, Oh and Viswanath (2017) leverage the connection between DP and hypothesis tests to derive tight privacy bounds for the composition of several mechanisms.

Outside the hypothesis testing setting, there is some additional work on optimal population inference under DP. Duchi, Jordan and Wainwright (2018) give general techniques to derive minimax rates under local DP, and in particular give minimax optimal point estimates for the mean, median, generalized linear models, and nonparametric density estimation. Karwa and Vadhan (2017) develop nearly optimal confidence intervals for normally distributed data with finite sample guarantees, which could potentially be inverted to give approximately UMP unbiased tests.

Related work on developing optimal DP mechanisms for general loss functions such as Geng and Viswanath (2016*a*) and Ghosh, Roughgarden and Sundararajan (2009), give mechanisms that optimize symmetric convex loss functions, centered at a real-valued statistic. Similarly, Awan and Slavković (2018*b*) derive optimal mechanisms among the class of K-Norm Mechanisms for a fixed statistic and sample size.

Our contributions The previous literature on DP hypothesis testing has a few characteristics in common: 1) nearly all of the proposed methods first add noise to the data, and perform their test as a post-processing procedure, 2) all of the hypothesis tests use either asymptotic distributions or simulations to derive approximate decision rules, and 3) while each procedure is derived intuitively based on classical theory, none show that they are optimal among all possible DP algorithms.

In contrast, in this paper we search over all DP hypothesis tests at level α, deriving the *uniformly most powerful* (UMP) test for a population proportion. We find that our DP-UMP test can be stated as a post-processing of a noisy statistic, which allows us to efficiently compute exact p-values, confidence intervals, and confidence distributions as post-processing.

Sections 2.1-2.5 appeared in an earlier version of this work (see, Awan and Slavković (2018*a*)), and focus on developing DP-UMP simple and one-sided tests for binomial data. In Section 2.2, we show that arbitrary DP hypothesis tests, which report 'Reject' or 'Fail to Reject', can be written in terms of linear inequalities. In Theorem 2.2, we show that for exchangeable data, DP tests need only depend on the empirical distribution. We use this structure to find closed-form DP-UMP tests for simple hypotheses in Lemmas 2.7 and 2.9, and extend these results to obtain one-sided DP-UMP tests in Theorem 2.10. These tests are closely tied to our proposed *Truncated-Uniform-Laplace* (*Tulap*) distribution, which extends both the discrete Laplace distribution (studied in Ghosh, Roughgarden and Sundararajan (2009)), and the Staircase distribution of Geng and Viswanath (2016*a*) to the setting of (ϵ, δ)-DP. We prove that the Tulap distribution satisfies (ϵ, δ)-DP in Theorem 2.11. While the tests developed in the previous sections only result in the output 'Reject' or 'Fail to Reject', in Section 2.5, we show that our DP-UMP tests can be stated as a post-processing of DP summary statistic, distributed as Tulap. From this formulation, we obtain exact p-values via Theorem 2.12 and Algorithm 1 which agree with our DP-UMP tests. In fact, since releasing the Tulap summary statistic satisfies (ϵ, δ)-DP, we can also produce two-sided p-values, confidence intervals, and confidence distributions all in terms of the private summary statistic, thus offering a comprehensive DP statistical analysis for binomial data at a fixed privacy cost.

To go beyond the simple tests and one-sided hypothesis results of Awan and Slavković (2018*a*), we use a Bonferroni correction for multiple comparisons and the one-sided DP-UMP tests to construct two-sided tests, which we detail in Section 2.6. In Section 2.7, we study unbiased tests for two-sided hypotheses and derive a two-sided DP-UMPU test, using similar techniques as in Sections 2.3 and 2.4. While these unbiased tests often do not have a convenient form, we propose a close approximation, which can be used to efficiently compute

p-values in Section 2.8. In Section 3 we develop methods to construct DP confidence intervals for binomial data. We show in Section 3.2 that our one-sided DP-UMP tests give uniformly most accurate one-sided confidence intervals. In Section 3.3, we show that the DP-UMPU test, Bonferroni test and the approximately unbiased two-sided tests can all be used to construct two-sided DP confidence intervals. Furthermore, we show that the DP-UMPU test leads to uniformly most accurate unbiased confidence intervals. In Section 4.1, we derive stochastically optimal confidence distributions in terms of the one-sided DP-UMP tests. In Section 4.2, we apply our results to develop private distribution-free hypothesis tests of continuous data.

In Section 5, we study each of our proposed tests and confidence intervals through simulations. In Section 5.1, we compare our DP-UMP to the Normal approximation test of Vu and Slavković (2009) as well as the non-private UMP test. In Section 5.2, we compare the power of our different proposed two-sided tests, and in Section 5.3 we study the average width of our two-sided confidence intervals. We conclude in Section 6 with discussion. Detailed proofs and technical lemmas are postponed to Appendix A.

2. Hypothesis testing

2.1. **Background and notation.** We use capital letters to denote random variables and lowercase letters for particular values. For a random variable X, we denote F_X as its cumulative distribution function (cdf), f_X as either its probability density function (pdf) or probability mass function (pmf), depending on the context.

For any set $\mathscr{X}$, the n-fold cartesian product of $\mathscr{X}$ is $\mathscr{X}^n = \{(x_1, x_2, \ldots, x_n) \mid x_i \in \mathscr{X}\}$. We denote elements of $\mathscr{X}^n$ with an underscore to emphasize that they are vectors. The *Hamming* metric on $\mathscr{X}^n$ is $H : \mathscr{X}^n \times \mathscr{X}^n \to \mathbb{Z}^{\geq 0}$, defined by $H(\underline{x}, \underline{x}') = \#\{i \mid x_i \neq x_i'\}$.

Differential privacy, introduced by Dwork et al. (2006), provides a formal measure of disclosure risk. The notion of DP that we give in Definition 2.1 more closely resembles the formulation in Wasserman and Zhou (2010), which uses the language of distributions rather than random mechanisms. It is important to emphasize that the notion of differential privacy in Definition 2.1 does not involve any distribution model on $\mathscr{X}^n$.

Definition 2.1 (Differential Privacy: Dwork et al. (2006); Wasserman and Zhou (2010))**.** Let $\epsilon > 0$, $\delta \geq 0$, and $n \in \{1, 2, \ldots\}$ be given. Let $\mathscr{X}$ be any set, and $(\mathscr{Y}, \mathscr{F})$ be a measurable space. Let $\mathscr{P} = \{P_{\underline{x}} \mid \underline{x} \in \mathscr{X}^n\}$ be a set of probability measures on $(\mathscr{Y}, \mathscr{F})$. We say that $\mathscr{P}$ satisfies (ϵ, δ)-*differential privacy* ((ϵ, δ) - DP) if for all $B \in \mathscr{F}$ and all $\underline{x}, \underline{x}' \in \mathscr{X}^n$ such that $H(\underline{x}, \underline{x}') = 1$, we have $P_{\underline{x}}(B) \leq e^{\epsilon} P_{\underline{x}'}(B) + \delta$.

In Definition 2.1, we interpret $\underline{x} \in \mathscr{X}^n$ as the database we collect, where $\mathscr{X}$ is the set of possible values that one individual can contribute, and $Y \sim P_{\underline{x}}$ as the statistical result we report to the public. With this interpretation, if a set of distributions satisfies (ϵ, δ)-DP for small values of ϵ and δ, then if one person's data is changed in the database, the change in the distribution of Y is small. Ideally ϵ is a value less than 1, and $\delta \ll \frac{1}{n}$ allows us to disregard events which have small probability. A special case is when $\delta = 0$, and $(\epsilon, 0)$-DP is referred to as pure DP.

One of our main goals in this paper is to find uniformly most powerful (UMP) hypothesis tests, subject to DP. As the output of a DP method is necessarily a random variable, we work with randomized hypothesis tests, which we review in Definition 2.2. Our notation follows that of Schervish (1996, Chapter 4).

Definition 2.2 (Hypothesis Test)**.** Let $(X_1, \ldots, X_n) \in \mathscr{X}^n$ be distributed $X_i \overset{\text{iid}}{\sim} f_\theta$, where $\theta \in \Theta$. Let Θ_0, Θ_1 be disjoint subsets of Θ. We call Θ_0 the *null* and Θ_1 the *alternative*. A *(randomized) test* of $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$ is a measurable function $\phi : \mathscr{X}^n \to [0, 1]$. The *power* of ϕ at θ is denoted $\beta_\phi(\theta) = \mathbb{E}_{f_\theta}\phi$. We say a test ϕ is at *level* α if $\sup_{\theta\in\Theta_0} \beta_\phi(\theta) \leq \alpha$, and at *size* α if $\sup_{\theta\in\Theta_0} \beta_\phi(\theta) = \alpha$.

Let Φ be a set of tests. We say that $\phi^* \in \Phi$ is the *uniformly most powerful level* α (UMP-α) test among Φ for $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$ if 1) $\sup_{\theta\in\Theta_0} \beta_{\phi^*}(\theta) \leq \alpha$ and 2) for any $\phi \in \Phi$ such that $\sup_{\theta\in\Theta_0} \beta_\phi(\theta) \leq \alpha$ we have $\beta_{\phi^*}(\theta) \geq \beta_\phi(\theta)$, for all $\theta \in \Theta_1$.

In Definition 2.2, $\phi(\underline{x})$ is the probability of rejecting the null hypothesis, given that we observe $\underline{x} \in \mathscr{X}^n$. That is, the output of a test is either 'Reject', or 'Fail to Reject' with respective probabilities $\phi(\underline{x})$, and $1 - \phi(\underline{x})$. While the condition of (ϵ, δ)-DP does not involve the randomness of X, for hypothesis testing, the level/size, and power of a test depend on the model for X. In Section 2.2, we study the set of hypothesis tests which satisfy (ϵ, δ)-DP.

2.2. **Problem setup and exchangeability condition.** We begin this section by considering arbitrary hypothesis testing problems under DP. Let $\phi : \mathscr{X}^n \to [0, 1]$ be any test. Since the only possible outputs of the mechanism are 'Reject' or 'Fail to Reject' with probabilities $\phi(\underline{x})$ and $1 - \phi(\underline{x})$, the test ϕ satisfies (ϵ, δ)-DP if and only if for all $\underline{x}, \underline{x}' \in \mathscr{X}^n$ such that $H(\underline{x}, \underline{x}') = 1$,

$$\phi(\underline{x}) \leq e^\epsilon \phi(\underline{x}') + \delta \quad \text{and} \quad (1 - \phi(\underline{x})) \leq e^\epsilon (1 - \phi(\underline{x}')) + \delta. \tag{2.1}$$

Remark 2.1. For any simple hypothesis test, where Θ_0 and Θ_1 are both singleton sets, the DP-UMP test ϕ^* is the solution to a linear program. If $\mathscr{X}$ is finite, this observation allows one to explore the structure of DP-UMP tests through numerical linear program solvers.

Given the random vector $\underline{X} \in \mathscr{X}^n$, initially it may seem that we need to consider all ϕ, which are arbitrary functions of $\underline{X}$. However, assuming that $\underline{X}$ is exchangeable, Theorem 2.2 below says that for any DP hypothesis tests, we need only consider tests which are functions of the empirical distribution of $\underline{X}$. In other words, ϕ need not consider the order of the entries in $\underline{X}$. This result is reminiscent of De Finetti's Theorem (see Schervish, 1996, Theorem 1.48) in classical statistics.

Theorem 2.2. *Let Θ be a set and $\{\mu_\theta\}_{\theta\in\Theta}$ be a set of exchangeable distributions on $\mathscr{X}^n$. Let $\phi : \mathscr{X}^n \to [0, 1]$ be a test satisfying (2.1). Then there exists $\phi' : \mathscr{X}^n \to [0, 1]$ satisfying (2.1) which only depends on the empirical distribution of X, such that $\int \phi'(\underline{x})\, d\mu_\theta = \int \phi(\underline{x})\, d\mu_\theta$, for all $\theta \in \Theta$.*

Proof Sketch. Define ϕ' by $\phi'(\underline{x}) = \frac{1}{n!} \sum_{\pi\in\sigma(n)} \phi(\pi(\underline{x}))$, where $\sigma(n)$ is the symmetric group on n letters. For any $\pi \in \sigma(n)$, $\phi(\pi(\cdot))$ satisfies (ϵ, δ)-DP. By exchangeability, $\int \phi(\pi(\underline{x}))\, d\mu_\theta = \int \phi(\underline{x})\, d\mu_\theta$. Since condition 2.1 is closed under convex combinations, and integrals are linear operators, the result follows. □

We now state the particular problem which is the primary focus of the remainder of Section 2. Each individual contributes a sensitive binary value to the database, and the database can be thought of as a random vector $\underline{X} \in \{0, 1\}^n$, where X_i represents the sensitive data of individual i. We model $\underline{X}$ as $X_i \overset{\text{iid}}{\sim} \text{Bern}(\theta)$, where θ is unknown. Then the statistic $X = \sum_{i=1}^n X_i \sim \text{Binom}(n, \theta)$ encodes the empirical distribution of $\underline{X}$. By

Theorem 2.2, we can restrict our attention to tests which are functions of X. Such tests $\phi : \{0, 1, \ldots, n\} \to [0, 1]$ satisfy (ϵ, δ) -DP if and only if for all $x \in \{1, 2, \ldots, n\}$,

$$\phi(x) \leq e^{\epsilon}\phi(x-1) + \delta \tag{2.2}$$
$$\phi(x-1) \leq e^{\epsilon}\phi(x) + \delta \tag{2.3}$$
$$(1 - \phi(x)) \leq e^{\epsilon}(1 - \phi(x-1)) + \delta \tag{2.4}$$
$$(1 - \phi(x-1)) \leq e^{\epsilon}(1 - \phi(x)) + \delta. \tag{2.5}$$

We denote the set of all tests which satisfy (2.2)-(2.5) as $\mathscr{D}^n_{\epsilon,\delta} = \big\{\phi : \phi \text{ satisfies (2.2)-(2.5)}\big\}$.

Remark 2.3. For arbitrary DP hypothesis testing problems, the number of constraints generated by (2.1) could be very large, even infinite, but for our problem we only have $4n$ constraints.

2.3. **Simple DP-UMP tests when $\delta = 0$.** In this section, we derive the DP-UMP test when $\delta = 0$ for simple hypotheses. In particular, given $n, \epsilon > 0, \alpha > 0, \theta_0 < \theta_1$, and $X \sim \text{Binom}(n, \theta)$, we find the UMP test at level α among $\mathscr{D}^n_{\epsilon,0}$ for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$.

Before developing these tests, we introduce the *Truncated-Uniform-Laplace* (Tulap) distribution, defined in Definition 2.3, which is central to all of our main results. To motivate this distribution, recall that Geng and Viswanath (2016*a*) show for general loss functions that adding discrete Laplace noise $L \sim \text{DLap}(e^{-\epsilon})$ to X is optimal under $(\epsilon, 0)$-DP. For this reason, it is natural to consider a test which post-processes $X + L$. However, we know by classical UMP theory that since $X + L$ is discrete, a randomized test is required. Instead of using a randomized test, by adding uniform noise $U \sim \text{Unif}(-1/2, 1/2)$ to $X + L$, we obtain a continuous sampling distribution, from which a deterministic test is available. We call the distribution of $(X + L + U) \mid X$ as $\text{Tulap}(X, b, 0)$. In the setting of (ϵ, δ)-DP, we require an additional parameter q, which is linked to the value δ. The distribution $\text{Tulap}(X, b, q)$ is obtained by truncating between the $(q/2)^{th}$ and $(1 - q/2)^{th}$ quantiles of $\text{Tulap}(X, b, 0)$.

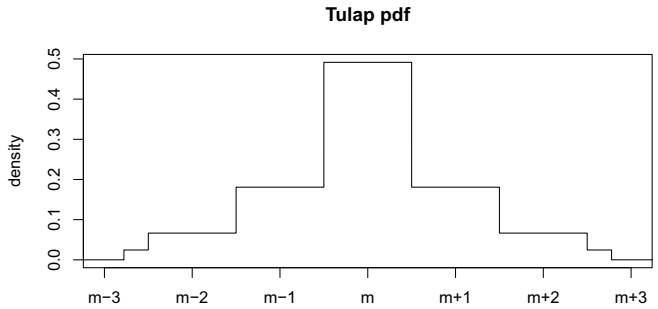

FIGURE 1. Plot of the Tulap density function for arbitrary m, $b = e^{-1}$, and $q = .06$.

In Definition 2.3, we use the *nearest integer function* $[\cdot] : \mathbb{R} \to \mathbb{Z}$. For any real number $t \in \mathbb{R}$, $[t]$ is defined to be the integer nearest to t. If there are two distinct integers which are nearest to t, we take $[t]$ to be the even one. Note that, $[-t] = -[t]$ for all $t \in \mathbb{R}$.

Definition 2.3 (Truncated-Uniform-Laplace (Tulap))**.** Let N and N_0 be real-valued random variables. Let $m \in \mathbb{R}$, $b \in (0,1)$ and $q \in [0,1)$. We say that $N_0 \sim \text{Tulap}(m,b,0)$ and $N \sim \text{Tulap}(m,b,q)$ if N_0 and N have the following cdfs:

$$F_{N_0}(x) = \begin{cases} \frac{b^{-[x-m]}}{1+b}\left(b + (x - m - [x-m] + \frac{1}{2})(1-b)\right) & \text{if } x \leq [m] \\ 1 - \frac{b^{[x-m]}}{1+b}\left(b + ([x-m] - (x-m) + \frac{1}{2})(1-b)\right) & \text{if } x > [m], \end{cases}$$

$$F_N(x) = \begin{cases} 0 & \text{if } F_{N_0} < q/2 \\ \frac{F_{N_0}(x) - \frac{q}{2}}{1-q} & \text{if } \frac{q}{2} \leq F_{N_0}(x) \leq 1 - \frac{q}{2} \\ 1 & \text{if } F_{N_0} > 1 - \frac{q}{2}. \end{cases}$$

Note that a Tulap random variable $\text{Tulap}(m,b,q)$ is continuous and symmetric about m. An illustration of the Tulap pdf is in Figure 1. While the name Tulap reflects the ability for the parameter q to restrict the support of the distribution, in the case where $q = 0$ the support is in fact unbounded. We will see in this section that for $(\epsilon, 0)$-DP, the untruncated distribution arises in the development of the DP-UMP test. In Section 2.4, we see that when $\delta > 0$ the parameter q is dependent on δ and ϵ to limit the support.

Remark 2.4. The Tulap distribution extends the staircase and discrete Laplace distributions as follows: $\text{Tulap}(0,b,0) \overset{d}{=} \text{Staircase}(b, 1/2)$ and $[\text{Tulap}(0,b,0)] \overset{d}{=} \text{DLap}(b)$, where $\text{Staircase}(b,\gamma)$ is the distribution in Geng and Viswanath (2016*a*). Geng and Viswanath (2016*a*) show that for a real valued statistic T and convex symmetric loss functions centered at T, the optimal noise distribution for $(\epsilon, 0)$-DP is $\text{Staircase}(b,\gamma)$ for $b = e^{-\epsilon}$ and some $\gamma \in (0,1)$. If the statistic is a count, then Ghosh, Roughgarden and Sundararajan (2009) show that $\text{DLap}(b)$ is optimal. Our results agree with these works when $\delta = 0$, and extend them to the case of arbitrary δ.

Now that we have defined the Tulap distribution, we are ready to develop the UMP test among $\mathscr{D}^n_{\epsilon,0}$ for the simple hypotheses $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$. In classical statistics, the UMP for this test is given by the *Neyman-Pearson lemma*, however in the DP framework, our test must satisfy (2.2)-(2.5). Within these constraints, we follow the logic behind the Neyman-Pearson lemma as follows. Let $\phi \in \mathscr{D}^n_{\epsilon,0}$. Thinking of $\phi(x)$ defined recursively, equations (2.2)-(2.5) give upper and lower bounds for $\phi(x)$ in terms of $\phi(x-1)$. Since $\theta_1 > \theta_0$, and binomial distributions have a monotone likelihood ratio (MLR) in x, larger values of x give more evidence for θ_1 over θ_0. Thus, $\phi(x)$ should be increasing in x as much as possible, subject to (2.2)-(2.5). Lemma 2.5 shows that taking $\phi(x)$ to be such a function is equivalent to having $\phi(x)$ be the cdf of a Tulap random variable.

Lemma 2.5. *Let $\epsilon > 0$ be given. Let $\phi : \{0,1,2,\ldots,n\} \to (0,1)$. The following are equivalent:*

(1) *There exists $m \in (0,1)$ such that for $x = 0,\ldots,n$,*

$$\phi(x) = \begin{cases} m & \text{if } x = 0 \\ \min\{e^{\epsilon}\phi(x-1), 1 - e^{-\epsilon}(1 - \phi(x-1))\} & \text{if } x > 0. \end{cases}$$

(2) *There exists* $m \in (0,1)$ *such that for* $x = 0, \dots, n$,

$$\phi(x) = \begin{cases} m & \text{if } x = 0 \\ e^{\epsilon}\phi(x-1) & \text{if } x > 0 \text{ and } \phi(x-1) \leq \frac{1}{1+e^{\epsilon}} \\ 1 - e^{-\epsilon}(1-\phi(x-1)) & \text{if } x > 0 \text{ and } \phi(x-1) > \frac{1}{1+e^{\epsilon}}. \end{cases}$$

(3) *There exists* $m \in \mathbb{R}$ *such that* $\phi(x) = F_{N_0}(x-m)$ *for* $x = 0, 1, 2, \dots, n$, *where* $N_0 \sim \text{Tulap}(0, b = e^{-\epsilon}, 0)$.

Proof Sketch. First show that (1) and (2) are equivalent by checking which constraint is active. We then verify that $F_{N_0}(x-m)$ satisfies the recurrence of (2). This can be done using the properties of the Tulap cdf, stated in Lemma A.2, found in Section A. □

While the form of (1) in Lemma 2.5 is intuitive, the connection to the Tulap cdf in (3) allows for a usable closed-form of the test. This connection with the Tulap distribution is crucial for the development in Section 2.5, which shows that the test in Lemma 2.5 can be achieved by post-processing $X + N$, where N is distributed as Tulap.

It remains to show that the tests in Lemma 2.5 are in fact UMP among $\mathscr{D}^n_{\epsilon,0}$. The main tool used to prove this is Lemma 2.6, which can be viewed as an abstraction of the standard Neyman-Pearson Lemma. The proof of Lemma 2.6 uses the same trick as the proof of the Neyman-Pearson Lemma given in Lehmann and Romano (2008, Theorem 3.2.1(ii)).

Lemma 2.6. *Let* $(\mathscr{X}, \mathscr{F}, \mu)$ *be a measure space and let* f *and* g *be two densities on* $\mathscr{X}$ *with respect to* μ. *Suppose that* $\phi_1, \phi_2 : \mathscr{X} \to [0,1]$ *are such that* $\int \phi_1 f \, d\mu \geq \int \phi_2 f \, d\mu$, *and there exists* $k \geq 0$ *such that* $\phi_1 \geq \phi_2$ *when* $g \geq kf$ *and* $\phi_1 \leq \phi_2$ *when* $g < kf$. *Then* $\int \phi_1 g \, d\mu \geq \int \phi_2 g \, d\mu$.

Proof. Note that $(\phi_1 - \phi_2)(g - kf) \geq 0$ for almost all $x \in \mathscr{X}$ (with respect to μ). This implies that $\int (\phi_1 - \phi_2)(g - kf) \, d\mu \geq 0$. Hence, $\int \phi_1 g \, d\mu - \int \phi_2 g \, d\mu \geq k \left(\int \phi_1 f \, d\mu - \int \phi_2 f \, d\mu \right) \geq 0$. □

Next we present one of our key results, Lemma 2.7, which can be viewed as a 'Neyman-Pearson lemma' for binomial data under $(\epsilon, 0)$-DP. Lemma 2.7 is the simplest case of our general DP-UMP result in Theorem 2.10, in which we have a simple hypothesis and $\delta = 0$. Since $\delta = 0$, the Tulap random variable is actually untruncated in this setting. While Lemma 2.7 is the simplest case of Theorem 2.10, it contains most of the big ideas required to prove the general result.

Lemma 2.7. *Let* $\epsilon > 0$, $\alpha \in (0,1)$, $0 \leq \theta_0 < \theta_1 \leq 1$, *and* $n \geq 1$ *be given. Observe* $X \sim \text{Binom}(n, \theta)$, *where* θ *is unknown. Set the decision rule* $\phi^* : \mathbb{Z} \to [0,1]$ *by* $\phi^*(x) = F_{N_0}(x-m)$, *where* $N_0 \sim \text{Tulap}(0, b = e^{-\epsilon}, 0)$ *and* m *is chosen such that* $E_{\theta_0}\phi^*(x) = \alpha$. *Then* ϕ^* *is UMP-*α *test of* $H_0 : \theta = \theta_0$ *versus* $H_1 : \theta = \theta_1$ *among* $\mathscr{D}^n_{\epsilon,0}$.

Proof Sketch. Let ϕ be any other test which satisfies (2.2)-(2.5) at level α. Then, since ϕ^* can be written in the form of (1) in Lemma 2.5, there exists $y \in \mathbb{Z}$ such that $\phi^*(x) \geq \phi(x)$ when $x \geq y$ and $\phi^*(x) \leq \phi(x)$ when $x < y$. By the MLR property of the binomial distribution and applying Lemma 2.6, we have $\beta_{\phi^*}(\theta_1) \geq \beta_{\phi}(\theta_1)$. □

While the classical Neyman-Pearson lemma results in an acceptance and rejection region, the DP-UMP always has some probability of rejecting the null, due to the constraints (2.2)-(2.5). As $\epsilon \uparrow \infty$, the DP-UMP converges to the non-private UMP, explored in Remark 2.16.

2.4. **Simple and one-sided DP-UMP tests when $\delta \geq 0$.** In this section, we extend the results of Section 2.3 to allow for $\delta \geq 0$. We begin by proposing the form of the DP-UMP test for simple hypotheses. As in Section 2.3, the DP-UMP test is increasing in x as much as (2.2)-(2.5) allow. Lemma 2.8 states that such a test can be written as the cdf of a Tulap random variable, where the parameter q depends on ϵ and δ. We omit the proof of Lemma 2.9, which mimics the proof of Lemma 2.7.

Lemma 2.8. *Let $\epsilon > 0$ and $\delta \geq 0$ be given and set $b = e^{-\epsilon}$ and $q = \frac{2\delta b}{1-b+2\delta b}$. Let $\phi : \{0, 1, 2, \ldots, n\} \to [0, 1]$. The following are equivalent:*

(1) *There exists $y \in \{0, 1, 2, \ldots, n\}$ and $m \in (0, 1)$ such that for $x = 0, \ldots, n$,*

$$\phi(x) = \begin{cases} 0 & \text{if } x < y \\ m & \text{if } x = y \\ \min\{e^{\epsilon}\phi(x-1) + \delta, \quad 1 - e^{-\epsilon}(1 - \phi(x-1)) + e^{-\epsilon}\delta, \quad 1\} & \text{if } x > y. \end{cases}$$

(2) *There exists $y \in \{0, 1, 2, \ldots, n\}$ and $m \in (0, 1)$ such that for $x = 0, \ldots, n$,*

$$\phi(x) = \begin{cases} 0 & \text{if } x < y \\ m & \text{if } x = y \\ e^{\epsilon}\phi(x-1) + \delta & \text{if } x > y \text{ and } \phi(x-1) \leq \frac{1-\delta}{1+e^{\epsilon}} \\ 1 - e^{-\epsilon}(1 - \phi(x-1)) + e^{-\epsilon}\delta & \text{if } x > y \text{ and } \frac{1-\delta}{1+e^{\epsilon}} \leq \phi(x-1) \leq 1 - \delta \\ 1 & \text{if } x > y \text{ and } \phi(x-1) > 1 - \delta. \end{cases}$$

(3) *There exists $m \in \mathbb{R}$ such that $\phi(x) = F_N(x - m)$ where $N \sim \text{Tulap}(0, b, q)$.*

Proof Sketch. The equivalence of (1) and (2) only requires determining which constraints are active. To show the equivalence of (2) and (3), we verify that $F_N(x - m)$ satisfies the recurrence of (2), using the expression of $F_N(x)$ in terms of $F_{N_0}(x)$ given in Definition 2.3, and the results of Lemma 2.5. □

Lemma 2.9. *Let $\epsilon > 0$, $\delta \geq 0$, $\alpha \in (0, 1)$, $0 \leq \theta_0 < \theta_1 \leq 1$, and $n \geq 1$ be given. Observe $X \sim \text{Binom}(n, \theta)$, where θ is unknown. Set $b = e^{-\epsilon}$ and $q = \frac{2\delta b}{1-b+2\delta b}$. Define $\phi^* : \mathbb{Z} \to [0, 1]$ by $\phi^*(x) = F_N(x - m)$ where $N \sim \text{Tulap}(0, b, q)$ and m is chosen such that $E_{\theta_0}\phi^*(x) = \alpha$. Then ϕ^* is UMP-α test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ among $\mathscr{D}^n_{\epsilon,\delta}$.*

So far we have focused on simple hypothesis tests, but since our test only depends on θ_0, and not on θ_1, our test is in fact the DP-UMP for one-sided tests, as stated in Theorem 2.10. Theorem 2.10 also shows that we can use our tests to build DP-UMP tests for $H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$ as well. Hence, Theorem 2.10 is our most general result so far, containing Lemmas 2.7 and 2.9 as special cases.

Theorem 2.10. *Let $X \sim \text{Binom}(n, \theta)$. Set $\phi^*(x) = F_N(x - m_1)$ and $\psi^*(x) = 1 - F_N(x - m_2)$, where $N \sim \text{Tulap}\left(0, b = e^{-\epsilon}, q = \frac{2\delta b}{1-b+2\delta b}\right)$ and m_1, m_2 are chosen such that $E_{\theta_0}\phi^*(x) = \alpha$ and $E_{\theta_0}\psi^*(x) = \alpha$. Then $\phi^*(x)$ is UMP-α among $\mathscr{D}^n_{\epsilon,\delta}$ for testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, and $\psi^*(x)$ is UMP-α among $\mathscr{D}^n_{\epsilon,\delta}$ for testing $H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$.*

2.5. **Optimal one-sided private p-values.** For the DP-UMP tests developed in Sections 2.3 and 2.4, the output is simply to 'Reject' or 'Fail to Reject' H_0. In scientific research, however, p-values are often used to weigh the evidence in favor of the alternative hypothesis over the null. Intuitively, a p-value is the smallest level α, for which a test outputs 'Reject'. Definition 2.4 gives a formal definition of a p-value.

Definition 2.4 (p-Value: Casella and Berger (2002))**.** For a random vector $X_i \overset{\text{iid}}{\sim} f_\theta$, a *p-value* for $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$ is a statistic $p(\underline{X})$ taking values in $[0, 1]$, such that for every $\alpha \in [0, 1]$,

$$\sup_{\theta \in \Theta_0} P_\theta(p(X) \leq \alpha) \leq \alpha.$$

The smaller the value of $p(X)$, the greater evidence we have for H_1 over H_0.

In this section, we show that our proposed DP-UMP tests can be implemented as a simple threshold test, where the test statistic is a function of a Tulap random variable. We show that the test statistic itself satisfies (ϵ, δ)-DP. This allows for differentially private p-values to be computed as a post-processing of the test statistic. We prove in Theorem 2.12 that these private p-values agree with the DP-UMP tests in Sections 2.3 and 2.4. While we state our p-values for one-sided tests, they also apply to simple tests as a special case.

Since our DP-UMP test from Lemma 2.9 rejects with probability $\phi^*(x) = F_N(x - m)$, given $N \sim F_N$, $\phi^*(x)$ rejects the null if and only if $X + N \geq m$. So, our DP-UMP tests can be stated as a post-processing of $X + N$. We prove in Theorem 2.11 that releasing $X + N$ satisfies (ϵ, δ)-DP. By the post-processing property of DP (see Dwork and Roth, 2014, Proposition 2.1), once we release $X + N$, any function of $X + N$ also satisfies (ϵ, δ)-DP. Thus, we can compute our private UMP-α tests as a function of $X + N$ for any α. The smallest α for which we reject the null is the p-value for that test. In fact Algorithm 1 and Theorem 2.12 give a more elegant method of computing this p-value.

Theorem 2.11. *Let $\mathscr{X}$ be any set, and $T : \mathscr{X}^n \to \mathbb{Z}$, with $\sup |T(\underline{x}) - T(\underline{x}')| \leq 1$, where the supremum is over the set $\{(\underline{x}, \underline{x}') \in \mathscr{X}^n \times \mathscr{X}^n \mid H(\underline{x}, \underline{x}') = 1\}$. Then the set of distributions $\left\{\text{Tulap}\left(T(\underline{x}), b = e^{-\epsilon}, q = \frac{2\delta b}{1-b+2\delta b}\right) \middle| \underline{x} \in \mathscr{X}^n\right\}$ satisfies (ϵ, δ)-DP.*

Proof Sketch. Since Tulap random variables are continuous and have a MLR in $T(\underline{x})$, by Lemma A.3 in Section A, it suffices to show that for all $t \in \mathbb{R}$, the cdf of a Tulap random variable $F_N(t - T(\underline{x}))$ satisfies (2.1), with $\phi(\underline{x})$ replaced with $F_N(t - T(\underline{x}))$. This is already established in Lemma 2.8, by the equivalence of (1) and (3). □

Theorem 2.12. *Let $\epsilon > 0$, $\delta \geq 0$, $X \sim \text{Binom}(n, \theta)$ where θ is unknown, and $Z|X \sim \text{Tulap}(X, b = e^{-\epsilon}, q = \frac{2\delta b}{1-b+2\delta b})$. Then*

1. *$p(\theta_0, Z) := P(X + N \geq Z \mid Z)$ is a p-value for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, where the probability is over $X \sim \text{Binom}(n, \theta_0)$ and $N \sim \text{Tulap}(0, b, q)$.*
2. *Let $0 < \alpha < 1$ be given. The test $\phi^*(x) = P_{Z \sim \text{Tulap}(x,b,q)}(p(\theta_0, Z) \leq \alpha \mid X)$ is UMP-α for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ among $\mathscr{D}^n_{\epsilon,\delta}$.*
3. *For all $\theta_1 > \theta_0$, $p(\theta_0, Z)$ is the stochastically smallest (ϵ, δ)-DP p-value for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta \geq \theta_0$.*
4. *The output of Algorithm 1 is equal to $p(\theta_0, Z)$.*

In the following corollary, we see that $1 - p(\theta_0, Z) = P(X + N \leq Z \mid Z)$ is the corresponding p-value for $H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$, with all the analogous properties.

Corollary 2.13. *In the same setup as Theorem 2.12,* $1 - p(\theta_0, Z) = P(X + N \leq Z \mid Z)$ *is the stochastically smallest* (ϵ, δ)*-DP p-value for* $H_0 : \theta \geq \theta_0$ *versus* $H_1 : \theta < \theta_0$*, and* $\psi^*(x) = P_{Z \sim \text{Tulap}(x,b,q)}(1 - p(\theta_0, Z) \leq \alpha \mid X)$ *agrees with the UMP-*α *test in Theorem 2.10.*

Algorithm 1 UMP one-sided p-value for binomial data under (ϵ, δ)-DP

INPUT: $n \in \mathbb{N}$, $\theta_0 \in (0,1)$, $\epsilon > 0$, $\delta \geq 0$, $Z \sim \text{Tulap}\left(X, b = e^{-\epsilon}, q = \frac{2\delta b}{1-b+2\delta b}\right)$

1: Set F_N as the cdf of $N \sim \text{Tulap}(0, b, q)$
2: Set $\underline{F} = (F_N(0 - Z), F_N(1 - Z), \ldots, F_N(n - Z))^\top$
3: Set $\underline{B} = (\binom{n}{0}\theta_0^0(1-\theta_0)^{n-0}, \binom{n}{1}\theta_0^1(1-\theta_0)^{n-1}, \ldots, \binom{n}{n}\theta_0^n(1-\theta_0)^{n-n})^\top$

OUTPUT: $\underline{F}^\top \underline{B}$

To implement Algorithm 1, we must be able to sample a Tulap random variable, which Algorithm 2 provides. The algorithm is based on the expression of $\text{Tulap}(m, b, 0)$ in terms of geometric and uniform variables, and uses rejection sampling when $q > 0$ (see Bishop, 2006, Chapter 11 for an introduction to rejection sampling). A proof that the output of this algorithm follows the correct distribution can be found in Lemma A.1 in Section A.

Algorithm 2 Sample from Tulap distribution: $N \sim \text{Tulap}(m, b, q)$

INPUT: $m \in \mathbb{R}$, $b \in (0,1)$, $q \in [0,1)$

1: Draw $G_1, G_2 \overset{\text{iid}}{\sim} \text{Geom}(1-b)$ and $U \sim \text{Unif}(-1/2, 1/2)$
2: Set $N = G_1 - G_2 + U$
3: If $F_{N_0}(N) < q/2$ or $F_{N_0}(N) > 1 - q/2$, where $N_0 \sim \text{Tulap}(0, b, 0)$, go to 1:

OUTPUT: $N + m$

Remark 2.14. It has been noted in the DP literature, such as in Haeberlen, Pierce and Narayan (2011), that the running time of an algorithm can potentially leak information about the value of the noise used or entries of the database itself. While Algorithm 2 has variable runtime, in fact the runtime is independent of the output value and the sensitive statistic m. To see this, observe that the value N generated at each iteration, is independent of all previous values. So, the number of iterations and the final output are independent. Since m is only added to N at the end, the runtime does not depend on m either.

Furthermore, since $F_{N_0}(N) \sim U(0,1)$, we observe that the runtime follows the distribution $\text{Geom}(1-q)$, as the probability of "success" is $(1-q)$ and we want to know when the first "success" is. In particular, the expected running time is $1/(1-q)$ iterations.

Remark 2.15. Since we know that releasing $Z = X + N$, where N is a Tulap random variable, satisfies (ϵ, δ)-DP, one could release Z and compute all of the desired inference quantities as a post-processing of Z, at no additional cost to privacy. In the remainder of the paper, we show that private two-sided p-values, confidence intervals, and confidence distributions can all be expressed as a post-processing of the summary statistic Z, leading to a more complete DP statistical analysis of binomial data at a fixed privacy cost.

Remark 2.16 (Asymptotic Relative Efficiency)**.** One may wonder about the asymptotic properties of the DP-UMP test compared to the non-private UMP test. It is not hard to show that for any fixed $\epsilon > 0$, δ, and $\theta_0 \in (0,1)$, our proposed DP-UMP test has asymptotic relative efficiency (ARE) of 1, relative to the non-private UMP test (see Van der Vaart, 2000, Section 14.3 for an introduction to ARE). Let $X \sim \text{Binom}(n, \theta_0)$. Define the two test

statistics as $T_1 = X$ and $T_2 = X + N$, where $N \sim \text{Tulap}(0, b, q)$. The ARE of the DP-UMP relative to the non-private UMP test is $\lim_{n\to\infty}(C_2(n)/C_1(n))^2$, where

$$C_i(n) = \left(\frac{d}{d\theta}\mathbb{E}_\theta T_i\Big|_{\theta=\theta_0}\right) \Big/ \sqrt{n \operatorname{Var}_{\theta_0}(T_i)}, \text{ for } i = 1, 2.$$

We compute $\mathbb{E}_\theta T_i = n\theta$, $\operatorname{Var}_{\theta_0}(T_1) = n\theta_0(1-\theta_0)$, and $\operatorname{Var}_{\theta_0}(T_2) = n\theta_0(1-\theta_0) + \operatorname{Var}(N)$. Since $\operatorname{Var}(N)$ is a constant, we have that $C_1 = (\theta_0(1-\theta_0))^{-1/2}$ and $C_2 \to (\theta_0(1-\theta_0))^{-1/2}$, and so the ARE is 1.

For $\delta = 0$, we have a simple closed formula for the variance of N: $\operatorname{Var}(N) = \frac{2\exp(-\epsilon)}{(1-\exp(-\epsilon))^2}$. In this setting, we can also consider the rate at which $(C_2(n)/C_1(n))^2$ converges to 1, as a function of n and ϵ. We have that

$$\left(\frac{C_2}{C_1}\right)^2 = \frac{\operatorname{Var}_{\theta_0}(T_1)}{\operatorname{Var}_{\theta_0}(T_2)} = \frac{1}{1 + \frac{2\exp(-\epsilon)}{(1-\exp(-\epsilon))^2}\frac{1}{n\theta_0(1-\theta_0)}}.$$

Applying the Taylor expansion of $1/(1+x)$ about $x = 0$, we arrive at two formulas. If $\epsilon \to 0$ such that $\frac{1}{\epsilon} = o(\sqrt{n})$ we have that $(C_2/C_1)^2 = 1 - O\left(\frac{1}{n\epsilon^2}\right)$ and if $\epsilon \to \infty$, then $(C_2/C_1)^2 = 1 - O\left(\frac{1}{n\exp(\epsilon)}\right)$. We see that as long as $1/\epsilon = o(\sqrt{n})$, the DP-UMP has the same ARE as the non-private UMP. On the other hand, if $\epsilon \to \infty$, the DP-UMP converges to the non-private UMP at an exponential rate.

2.6. **Bonferroni two-sided tests.** In this section as well as in Sections 2.7 and 2.8, we develop "two-sided" tests for hypotheses of the form $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. One way of viewing this problem is as a *multiple testing problem*, where we test both $H_0 : \theta \leq \theta_0$ and $H_0 : \theta \geq \theta_0$. It is well known that if p_1 is a p-value for $H_0 : \theta \leq \theta_0$ and p_2 is a p-value for $H_0 : \theta \geq \theta_0$, then $p = 2\min(p_1, p_2)$ is a p-value for $H_0 : \theta = \theta_0$.

More generally, if we are interested in testing $H_0 : \theta \in \cap_{i=1}^k \Theta_k$, and p_i is a p-value for testing $H_0 : \theta \in \Theta_k$, then $p = k\min\{p_1, \ldots, p_k\}$ is a p-value for $H_0 : \theta \in \cap_{i=1}^k \Theta_k$. We call this setting a *multiple testing problem* or an *intersection-union test* (See Casella and Berger, 2002, Section 8.2.3). The factor k in the computation of p is called the *Bonferroni correction*. Other concepts related to multiple testing are false discovery rate (Benjamini and Hochberg, 1995), and adaptive data analysis which each offer methods of adjusting global error rates when running multiple statistical analyses on the same database. Recently, there has been an interesting line of work, beginning with Dwork et al. (2015), which shows that the tools of DP can be used to address the problem of adaptive data analysis.

Using the Bonferroni correction, we can combine the DP-UMP p-values from Theorem 2.12 and Corollary 2.13 to derive a DP two-sided test given in Proposition 2.17. The proof of Proposition 2.17 is fairly mechanical, and is left to Section A.

Proposition 2.17. *Let $\epsilon > 0$, $\delta \geq 0$, $X \sim \text{Binom}(n, \theta)$ where θ is unknown, and $Z \mid X \sim \text{Tulap}\left(X, b = e^{-\epsilon}, q = \frac{2\delta b}{1-b+2\delta b}\right)$. Then*

1. *$p'(\theta_0, Z) = 2\min(p(\theta_0, Z), 1 - p(\theta_0, Z))$ is a (ϵ, δ)-DP p-value for $H_0 : \theta = \theta_0$ versus $H_0 : \theta \neq \theta_0$ for any $\theta_0 \in (0, 1)$, where p is the one-sided p-value from Theorem 2.12.*
2. *The test $\phi'(X) = P_Z(p'(\theta_0, Z) \leq \alpha \mid X)$ is in $\mathscr{D}^n_{\epsilon,\delta}$ and can be written in the form $\phi'(x) = \phi^*(x) + \psi^*(x)$, where ϕ^*, ψ^* are as defined in Theorem 2.10 at size $\alpha/2$.*

(3) *The test $\phi'(X)$ is uniformly more powerful than any level $\alpha/2$ test in $\mathscr{D}^n_{\epsilon,\delta}$ for $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.*

The major benefit of the tests in Proposition 2.17 is in their simplicity, as they generally do not optimize any particular criteria. However, since they are more powerful than any DP test of level $\alpha/2$, they are not unreasonable tests, and perform relatively well compared to other candidate tests.

2.7. **DP-UMP unbiased two-sided tests.** In Section 2.6, we developed two-sided DP tests using a Bonferroni correction. While we were able to show that they are preferred over any level $\alpha/2$ test, they are not optimal when compared to other size α DP tests.

In this section, we continue our exploration of DP tests for $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. While one may hope to develop UMP tests in this setting, it is well known that among all tests, there is no UMP test, even without privacy. Indeed, the left-side DP-UMP and the right-side DP-UMP have higher power in different regions. Instead, we must restrict to a smaller class of tests. In classical statistics, it is common to restrict to *unbiased* tests. We show that there exists a DP-UMP unbiased test (DP-UMPU) for $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, and write the test in terms of the Tulap distribution.

Definition 2.5 (Unbiased Test)**.** A test $\phi : \mathscr{X}^n \to [0, 1]$ is unbiased for $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$ if for all $\theta_0 \in \Theta_0$ and all $\theta_1 \in \Theta_1$ we have that $\mathbb{E}_{\theta_0}\phi \leq \mathbb{E}_{\theta_1}\phi$.

Intuitively, unbiased means that the marginal probability of 'Reject' is always higher in the alternative than in the null.

While in the one-sided case, the DP-UMP test increases as much as possible in terms of either x or $-x$, now that we restrict to unbiased tests, the DP-UMP needs to increase as fast as possible in both directions. It turns out that there exists a center k, where the DP-UMPU test is symmetric about k, and increases as much as possible in both directions, subject to (2.2)-(2.5). This gives the form in Theorem 2.18.

Theorem 2.18. *Let $X \sim \text{Binom}(n, \theta)$, $0 < \theta_0 < 1$ and $0 < \alpha < 1$. There exists a UMPU size α test among $\mathscr{D}^n_{\epsilon,\delta}$ for $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, which is of the form*

$$\phi^*(x) = \begin{cases} F_N(x - k - m) & \text{if } x \geq k \\ F_N(k - x - m) & \text{if } x < k, \end{cases}$$

where k and m are chosen such that $\mathbb{E}_{X\sim\theta_0}(X - n\theta_0)\phi(X) = 0$ and $\mathbb{E}_{X\sim\theta_0}\phi(X) = \alpha$.

Proof Sketch. We must show that there exists k and m which solve the two equations, and then argue that ϕ^* is UMP among all level α unbiased tests in $\mathscr{D}^n_{\epsilon,\delta}$. The proof is inspired by the Generalized Neyman Pearson Lemma (Lehmann and Romano, 2008, Theorem 3.6.1), and has a similar strategy as Theorem 2.10. □

In Figure 2, we illustrate the DP-UMPU test, demonstrating that the value k depends on the size α. In particular k generally does not equal $n\theta_0$, an intuitive guess for the point of symmetry of the test.

In Theorem 2.10, we were able to derive DP p-values that agree with these tests. However, in Theorem 2.18, the quantity k depends on n, α, and θ_0, and there is no clear functional form of k in terms of these quantities. Thus it does not seem that there is a simple formula for the p-values of the test in Theorem 2.18. The following corollary shows that in the case when $\theta_0 = \frac{1}{2}$, the value of k is $\frac{n}{2}$ and a convenient form for the p-value exists.

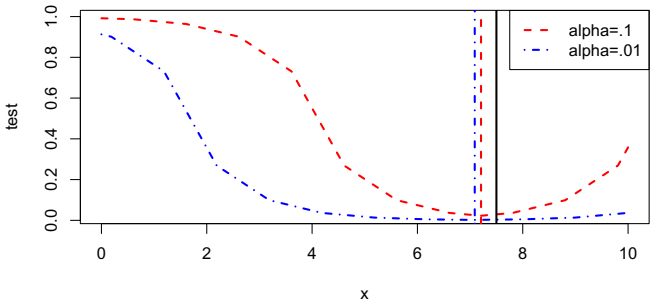

FIGURE 2. Plot of the DP-UMPU test $\phi(x)$ for $H_0 : \theta = .75$ versus $H_1 : \theta \neq .75$ at size $\alpha = .1$ and $\alpha = .01$, with $n = 10$, $\epsilon = 1$, $\delta = 0$. The solid vertical line indicates the value $n\theta_0 = 7.5$, and the dashed vertical lines indicate the value of k in Theorem 2.18, which are approximately 7.2 and 7.1 for $\alpha = .1$ and $\alpha = .01$, respectively.

Corollary 2.19. *In the setup of Theorem 2.18, if* $\theta_0 = \frac{1}{2}$ *then* $k = \frac{n}{2}$. *Let* $Z \mid X \sim \text{Tulap}\left(X, b = e^{-\epsilon}, \frac{2\delta b}{1-b+2\delta b}\right)$, *then the corresponding p-value is*

$$p(Z) = P_{X\sim\theta_0,N}\left(|X + N - n/2| \geq |Z - n/2| \middle| Z\right),$$

which can be computed via Algorithm 3, setting $\theta_0 = \frac{1}{2}$.

Proof Sketch. It suffices to check that when $k = \frac{n}{2}$, the test is unbiased. This is done using the symmetry of both $\phi(x)$ and $f_X(x)$ about $x = \frac{n}{2}$. □

Remark 2.20. While Corollary 2.19 only applies in the case that $\theta_0 = \frac{1}{2}$, this is in fact a common setting. This arises when we are interested in testing whether two mutually exclusive (and collectively exhaustive) events are equally likely, such as whether the probability of being born male versus female is $\frac{1}{2}$. In Section 4.2, we see that for the sign and median test, when testing whether the medians of two random variables are equal or not, this can be expressed as testing $H_0 : \theta = \frac{1}{2}$ versus $H_1 : \theta \neq \frac{1}{2}$.

2.8. **Asymptotically unbiased two sided tests.** In the previous section, we developed the DP-UMPU two-sided test, and showed that in the case where $\theta_0 = 1/2$, we can easily compute p-values. When $\theta_0 \neq 1/2$, k depends on α and there is no natural test statistics, making the problem more challenging. Nevertheless, based on Corollary 2.19 we conjecture that $|X + N - n\theta_0|$ is a test statistic which provides a close approximation to the DP-UMPU test. In Figure 2, we see that k is approximately equal to $n\theta_0$, even for n as small as 10. We show in Proposition 2.22 that this test statistic in fact leads to an asymptotically unbiased test. In Section 5.2, we see that this asymptotically unbiased test performs very similarly to the UMPU test from Section 2.7 for finite samples.

Algorithm 3 Asymptotically unbiased DP p-value

INPUT: $n \in \mathbb{N}$, $\theta_0 \in (0,1)$, $\epsilon > 0$, $\delta \geq 0$, $Z \sim \text{Tulap}\left(X, b = e^{-\epsilon}, q = \frac{2\delta b}{1-b+2\delta b}\right)$

1: Set $T = |Z - n\theta_0|$
2: Call $p(\theta, Z)$ the p-value computed by Algorithm 1
3: Set $p = p(\theta_0, T + n\theta_0) + 1 - p(\theta_0, n\theta_0 - T)$

OUTPUT: p

Proposition 2.21. *The output of Algorithm 3 is*

$$p(\theta_0, Z) = P_{X \sim \theta_0, N}\left(|X + N - n\theta_0| > |Z - n\theta_0| \middle| Z\right),$$

which is a p-value for $H_0 : \theta = \theta_0$ versus $H_0 : \theta \neq \theta_0$ and satisfies (ϵ, δ)-DP. The corresponding test $\phi(x) = P_N(p(\theta_0, Z) \leq \alpha \mid X)$ is of the form of Theorem 2.18, with $k = n\theta_0$.

Proof Sketch. It is easy to verify that when $\theta = \theta_0$, $p(\theta_0, Z)$ is marginally distributed as $U(0,1)$. Since $p(\theta_0, Z) \sim U(0,1)$, we have that $P_{\theta_0}(p(\theta_0, Z) \geq \alpha) = \alpha$. The p-value satisfies (ϵ, δ)-DP since it is a post-processing of Z. □

Proposition 2.22. *In the setting of Theorem 2.18, holding ϵ, δ, α, and θ_0 all fixed, the test in Proposition 2.21 is asymptotically unbiased.*

Proof Sketch. In the proof of Theorem 2.18, we saw that if ϕ is of the form in Theorem 2.18 and $\mathbb{E}_{\theta_0}(X - n\theta_0)\phi(X) = 0$, then ϕ is unbiased. Let ϕ be the test in Proposition 2.21. Then it suffices to show that $\lim_{n\to\infty} \mathbb{E}_{\theta_0} \frac{X - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}} \frac{\phi(X)}{\sqrt{n}} = 0$. Recall that if $X \sim \text{Binom}(n, \theta_0)$, then $\frac{X - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}} \xrightarrow{d} N(0,1)$. Using the fact that $\phi(x)$ is symmetric about $k = n\theta_0$, we see that the expectation is the integration of the product of two even functions and one odd function. Hence the expectation is zero. □

Remark 2.23. Since Proposition 2.22 shows that the test ϕ' in Proposition 2.21 is asymptotically unbiased, and since it is of the form of the UMPU test ϕ^* of Theorem 2.18, as the sample size increases, the power of the test ϕ' is very similar to that of ϕ^*. In Section 5.2, we see that even at $n = 30$, the performance is very close between ϕ' and ϕ^*.

3. Confidence intervals

3.1. **Background and notation.** A confidence set is a popular method of expressing uncertainty about a population quantity. Since all estimates have some error in them, a confidence set communicates the set of values in which we expect the population quantity to lie. While confidence sets can be of arbitrary forms, typically we prefer confidence sets which are intervals, since this simpler form improves interpretability.

Definition 3.1 (Confidence Interval)**.** Let $X_i \overset{\text{iid}}{\sim} f_\theta$, where $\theta \in \Theta \subset \mathbb{R}$. A *(random) confidence interval (CI)* is a set of random variables $\mathscr{C} = \{C(\underline{x}) \mid \underline{x} \in \mathscr{X}^n\}$, each of which takes values in $\{[a,b] \in \mathbb{R}^2 \mid a \leq b\}$. We say that $\mathscr{C}$ has coverage γ if for all $\theta \in \Theta$,

$$P_{\underline{X} \sim \theta}(\theta \in C(\underline{X})) \geq \gamma.$$

If one of a or b in a confidence interval $\mathscr{C}$ is constant, then we call $\mathscr{C}$ a *one-sided confidence interval*, otherwise we call $\mathscr{C}$ a *two-sided confidence interval*.

For convenience, we will often suppress the dependence of a confidence interval on x, and simply write C rather than $C(x)$. In classical statistics, there is a well known connection between hypothesis tests and confidence sets (see Casella and Berger, 2002, Chapter 9). For a deterministic test of level α, with rejection region R, the set $\Theta \setminus R$ is a confidence set with coverage $1-\alpha$. For randomized tests, it is more convenient to work with p-values. The following Proposition shows how one can use a p-value to build a confidence set. See Geyer and Meeden (2005) for a deeper understanding of randomized tests, p-values, and confidence sets in terms of fuzzy set theory.

Proposition 3.1. *If $p(\theta \mid X)$ is a p-value, then $C(X) = \{\theta \mid p(\theta \mid X) \geq \alpha\}$ is a confidence set with coverage $1-\alpha$.*

In order to decide whether one confidence interval is to be preferred over another, we require some criteria. In classical statistics, one considers *uniformly most accurate* (UMA) confidence intervals, which have properties related to UMP tests. UMA confidence intervals are defined in terms of *false coverage*, which is the analogue of the power of the corresponding test. The UMA property is important in theory and practice, because it results in smaller confidence intervals, and more accurately communicates the uncertainty of the parameter in question. Our definitions of false coverage and UMA follow that of Casella and Berger (2002, Section 9.3.2).

Definition 3.2 (False Coverage and Uniformly Most Accurate)**.** Let C be a confidence interval for $\theta \in \Theta$. The *false coverage probability* of θ' under θ is defined as $P_\theta(\theta' \in C)$. However, the allowable values of θ and θ' vary depending on the structure of C:

$$\begin{array}{ll} \text{if } C(X) = [L(X), U(X)], & \text{then only consider } \theta' \neq \theta \\ \text{if } C(X) = [L(X), \max\{\Theta\}], & \text{then only consider } \theta' < \theta \\ \text{if } C(X) = [\min\{\Theta\}, U(X)], & \text{then only consider } \theta' > \theta. \end{array}$$

The *Uniformly Most Accurate (UMA)* confidence interval among a set of confidence intervals, minimizes the false coverage for all valid pairs θ and θ'.

To understand the reason for limiting the values of θ and θ' in false coverage, consider the case where $C(X) = [L(X), \max\{\Theta\}]$. In this setting, we are only concerned when we cover a false value θ' which is lesser than the true θ, as this represents an unnecessarily wide interval.

3.2. **One-sided confidence intervals.** In this section, we show how we can use our DP-UMP tests to produce private confidence intervals for Bernoulli data. In the following Theorem, we show that the one-sided confidence interval based on our DP-UMP one-sided test is UMA. Furthermore, this interval is still a function of our private test statistic $Z = X + N$, and so after releasing Z, there is no additional cost to privacy when providing this confidence interval.

Theorem 3.2. *Let $Z = X + N$, where $X \sim \text{Binom}(n, \theta)$ and $N \sim \text{Tulap}(0, b = e^{-\epsilon}, q = \frac{2\delta b}{1-b+2\delta b})$, and let $p(\theta_0, Z)$ be the one-sided private p-value for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta \geq \theta_0$, defined in Theorem 2.12. Let $\alpha \in (0,1)$ be given. The confidence interval $C_\alpha^* = \{\theta_0 \mid p(\theta_0, Z) \geq \alpha\}$ is the UMA (ϵ, δ)-DP confidence interval of the form $[L, 1]$ with coverage $1-\alpha$.*

Proof Sketch. Releasing C_α^* satisfies (ϵ, δ)-DP by the post-processing property of DP. The fact that C_α^* is UMA is shown by observing that the false coverage probability can be interpreted as the power of a corresponding test, which is the DP-UMP. □

Corollary 3.3. *Using the same setup as Theorem 3.2, the interval* $C_\alpha = \{\theta_0 \mid (1-p(\theta_0, Z)) \geq \alpha\}$ *is the UMA* (ϵ, δ)*-DP confidence interval of the form* $[0, U]$*, with coverage* $1-\alpha$*.*

Remark 3.4. The value L^* in the interval $C^*_\alpha = [L^*, 1]$ of Theorem 3.2 can be easily computed by minimizing $(p(\theta_0, Z) - \alpha)^2$ over the interval $\theta_0 \in [0, 1]$. This can be done using standard optimization software.

3.3. **Two-sided confidence intervals.** As we saw in Theorem 3.2, when $p(\theta_0, Z)$ is a one-sided p-value, the set $\{\theta_0 \mid p(\theta_0, Z) \geq \alpha\}$ forms a one-sided confidence interval. Similarly, if $p(\theta_0, Z)$ is a two-sided p-value, then $\{\theta_0 \mid p(\theta_0, Z) \geq \alpha\}$ is of the form $[L, U]$. In this section, we will consider the DP confidence intervals produced by each of our proposed two-sided tests from Sections 2.6-2.8.

First, we introduce our optimality criterion for two-sided confidence intervals. Just as there is generally no UMP two-sided test, there does not exist a UMA two-sided confidence interval. In Section 2.7, we saw that for two-sided tests, we imposed the condition of unbiasedness in order to obtain a UMP test. Similarly, we will consider an analogous notion of *unbiasedness* for confidence intervals. Intuitively, a confidence interval is unbiased if the probability of false coverage is always smaller than the true coverage. Unbiased confidence intervals and unbiased hypothesis tests are in one-to-one correspondence via the connection in Proposition 3.1.

Definition 3.3 (Unbiased Confidence Interval)**.** Let $C(\underline{X})$ be a confidence interval for θ. We call C *unbiased* if for all $\theta \neq \theta'$, $P_\theta(\theta' \in C) \leq P_\theta(\theta \in C)$.

The next result shows that our DP-UMPU test from Theorem 2.18 leads to a DP-UMA unbiased (DP-UMAU) confidence interval.

Theorem 3.5. *Let* $\epsilon > 0$*,* $\delta \geq 0$*,* $X \sim \text{Binom}(n, \theta)$ *where* θ *is unknown. Let* $Z \mid X \sim \text{Tulap}\left(X, b = e^{-\epsilon}, q = \frac{2\delta b}{1-b+2\delta b}\right)$*. We construct the corresponding randomized p-value as*

$$p^*(Z) = \min\left\{\alpha \middle| |Z - k(\alpha)| \geq m(\alpha)\right\},$$

where $k(\cdot)$ *and* $m(\cdot)$ *satisfy the requirements of Theorem 2.18. The set* $C^* = \{\theta \mid p^*(x, U) \geq \alpha\}$ *is an unbiased, DP confidence interval with coverage* $(1-\alpha)$*, and* C^* *is the DP-UMAU confidence interval with coverage* $(1-\alpha)$*.*

The proof of Theorem 3.5 is similar to the proof of Theorem 3.2, and is postponed to Section A.

While Theorem 3.5 gives the DP-UMAU confidence interval, it is not easy to implement, since k and m do not have simple closed forms, as discussed in Section 2.7. Instead, we can use Proposition 3.6 to produce computationally convenient confidence intervals based on the p-values from Proposition 2.17 and Algorithm 3.

Proposition 3.6. *Let* $\epsilon > 0$*,* $\delta \geq 0$*,* $X \sim \text{Binom}(n, \theta)$ *where* θ *is unknown, and* $Z \mid X \sim \text{Tulap}\left(X, b = e^{-\epsilon}, q = \frac{2\delta b}{1-b+2\delta b}\right)$*. Consider the two quantities*

(1) $C^1_\alpha = \{\theta_0 \mid p'(\theta_0, Z) \geq \alpha\} = (C^*_{\alpha/2}) \setminus (C^*_{1-\alpha/2})$*, where* $p'(\theta, Z)$ *is the Bonferroni p-value from Proposition 2.17 and* C^*_α *is the one-sided confidence interval from Theorem 3.2.*
(2) $C^2_\alpha = \{\theta_0 \mid p(\theta_0, Z) \geq \alpha\}$*, where* $p(\theta_0, Z)$ *is the p-value from Proposition 2.21.*

Both C^1_α *and* C^2_α *are* (ϵ, δ)*-DP confidence intervals of the form* $[L, U]$ *with coverage* $(1-\alpha)$*.*

Remark 3.7. Since the confidence interval C_α^2 from Proposition 3.6 is based on the approximation to the DP-UMPU test, it serves as an approximation to the DP-UMAU confidence interval from Proposition 3.5.

Similar to Proposition 2.17, which stated that the Bonferroni two-sided test is uniformly more powerful than any level $\alpha/2$ DP test, the following Corollary shows that C_α^1 in Proposition 3.6 is uniformly more accurate than any DP confidence interval with coverage $1-\alpha/2$. The proof is found in Section A.

Corollary 3.8. *In the setting of Proposition 3.6, C_α^1 is uniformly more accurate than any (ϵ, δ)-DP confidence interval with coverage $1-\alpha/2$.*

4. Confidence distributions and distribution-free inference

4.1. **Confidence distributions.** A confidence distribution is a frequentist estimator, which contains information to produce hypothesis tests, confidence intervals, p-values, point estimates, etc (see Xie and Singh, 2013 for an introduction to Confidence Distributions). Much like in Bayesian statistics, where the posterior distribution is used to do inference, a confidence distribution contains the relevant information for frequentist statistics. Intuitively, a confidence distribution μ is a probability measure on Θ such that for $S \subset \Theta$, $\mu(S)$ is the coverage of S. Confidence distributions also have the property that the cdf of μ evaluated at θ_0 is a p-value for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.

The goal of this section is to release a confidence distribution, which satisfies DP. In particular, we show that using our one-sided DP-UMP tests we can produce optimal DP confidence distributions.

Definition 4.1 (Confidence Distribution: Xie and Singh (2013))**.** Let $X_i \overset{\text{iid}}{\sim} f_\theta$ for $\theta \in \Theta$ and $X_i \in \mathscr{X}$. A *confidence distribution* is a family of random variables $\{H_n(\underline{x}, \theta) \mid \underline{x} \in \mathscr{X}^n, \theta \in \Theta\}$ (we will suppress the dependence on $\underline{x}$ and write $H_n(\theta)$), each of which takes values in $[0,1]$ such that

(1) for each $\underline{x} \in \mathscr{X}^n$, $H_n(\cdot)$ is a cdf on Θ, and
(2) at the true value $\theta = \theta_0$, $H_n(\theta_0) = H_n(X, \theta_0) \sim U[0,1]$ (over the randomness of H_n and X).

Supposing that we have two methods of constructing confidence distributions, what criteria should we use to choose between them? In the following definition, we say that one confidence distribution is superior to another if the mass is more closely distributed near the true value θ_0.

Definition 4.2 (Xie and Singh, 2013)**.** For real-valued random variables X, Y, $X \overset{sto}{\leq} Y$ means that $P(X \leq t) \geq P(Y \leq t)$ for all $t \in \mathbb{R}$. Let H_1 and H_2 be two confidence distributions. We say that H_1 is *superior* to H_2 at $\theta = \theta_0$ if for all $\epsilon > 0$, $H_1(\theta_0 - \epsilon) \overset{sto}{\leq} H_2(\theta_0 - \epsilon)$ and $1 - H_1(\theta_0 + \epsilon) \overset{sto}{\leq} 1 - H_2(\theta_0 + \epsilon)$.

In Section 5 of Xie and Singh (2013), they discuss how using a UMP one-sided test results in the optimal confidence distribution. Theorem 4.1 below similarly shows that our DP-UMP one-sided test results in the optimal DP confidence distribution.

Theorem 4.1. *Let $Z = X + N$, where $X \sim \text{Binom}(n, \theta)$ and $N \sim \text{Tulap}(0, b = e^{-\epsilon}, q = \frac{2\delta b}{1-b+2\delta b})$, and let $p(\theta_0, Z)$ be the one-sided private p-value for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta \geq \theta_0$. Define $H_n^*(\theta_0) = p(\theta_0, Z)$. Then H_n^* is a confidence distribution which satisfies (ϵ, δ)-DP, and is superior to any other (ϵ, δ)-DP confidence distribution.*

Proof. That H_n^* satisfies (ϵ, δ)-DP follows by the post-processing property of DP. The fact that H_n^* is a confidence distribution follows from the fact that $p(\theta_0, Z)$ is monotonic in θ_0, and $p(\theta_0, Z) \in [0, 1]$. If H_n^* were not superior, then this contradicts that $p(\theta_0, Z)$ corresponds to the UMP test among $\mathscr{D}_{\epsilon,\delta}^n$ for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. □

4.2. **Application to distribution-free inference.** In this section, we show how our DP-UMP tests for count data can be used to test certain hypotheses for continuous data. In particular, we give a DP version of the sign and median test allowing one to test the median of either paired or independent samples. For an introduction to the sign and median tests, see Sections 5.4 and 6.4 of Gibbons and Chakraborti (2014). Let $\epsilon > 0$ and $\delta \in [0, 1)$ be given, and let $N \sim \text{Tulap}(0, b, q)$ for $b = e^{-\epsilon}$ and $q = \frac{2\delta b}{1-b-2\delta b}$.

Sign test: We observe n iid pairs (X_i, Y_i) for $i = 1, \ldots, n$. Then for all $i = 1, \ldots, n$, $X_i \overset{d}{=} X$ and $Y_i \overset{d}{=} Y$ for some ordinal random variables X and Y. The variables X and Y need not be numeric, but we assume that for any pair (X_i, Y_i) we can determine if $X_i > Y_i$ or not. For simplicity, we also assume that there are no pairs with $X_i = Y_i$. Denote the unknown probability $\theta = P(X > Y)$. We want to test a hypothesis such as $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. The sign test uses the test statistic $T = \#\{X_i > Y_i\}$. Since the sensitivity of T is 1, by Theorem 2.11, $T + N$ satisfies (ϵ, δ)-DP. Note that the test statistic is distributed as $T \sim \text{Binom}(n, \theta)$. Using Algorithm 1, we obtain a private p-value for the sign test as a post-processing of $T + N$.

To test whether $\text{median}(X) = \text{median}(Y)$, we consider the hypothesis $H_0 : \theta = \frac{1}{2}$ versus $H_1 : \theta \neq \frac{1}{2}$. Using the same test statistic $Z = T + N$, we obtain a p-value for the sign test via Algorithm 3.

Median test: We observe two independent sets of iid ordinal data $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$, where all X_i and Y_i are distinct values, and we have a total ordering on these values. We assume that there exists random variables X and Y such that $X_i \overset{d}{=} X$ and $Y_i \overset{d}{=} Y$ for all i. We want to test $H_0 : \text{median}(X) \leq \text{median}(Y)$ versus $H_1 : \text{median}(X) > \text{median}(Y)$. The median test uses the test statistic $T = \#\{i \mid \text{rank}(X_i) > n\}$, where $\text{rank}(X_i) = \#\{X_j \leq X_i\} + \#\{Y_j \leq X_i\}$. Since the sensitivity of T is 1, by Theorem 2.11, $T + N$ satisfies (ϵ, δ)-DP. When $\text{median}(X) = \text{median}(Y)$, $T \sim \text{HyperGeom}(n = n, m = n, k = n)$. Using Algorithm 1, with $\underline{B}$ replaced with the pmf of $\text{HyperGeom}(n = n, m = n, k = n)$, we obtain a private p-value for the median test as a post-processing of $T + N$.

To test whether $\text{median}(X) = \text{median}(Y)$, we consider the hypothesis $H_0 : \theta = \frac{1}{2}$ versus $H_1 : \theta \neq \frac{1}{2}$. Using the same test statistic $Z = T + N$, we obtain a p-value for the sign test via Algorithm 3, with $\underline{B}$ replaced with the pmf of $\text{HyperGeom}(n = n, m = n, k = n)$.

5. Simulations

5.1. **One-sided hypothesis testing simulations.** In this section, we study the empirical performance of our DP-UMP test compared to the Normal approximation proposed by Vu

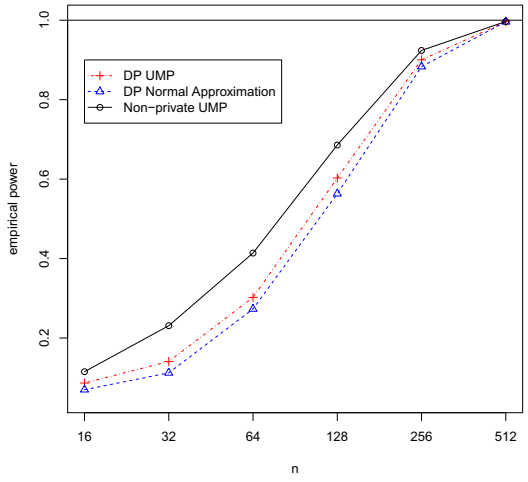

FIGURE 3. Empirical power for UMP and Normal approximation tests for $H_0 : \theta \leq .9$ versus $H_1 : \theta \geq .9$. The true value is $\theta = .95$. $\epsilon = 1$ and $\delta = 0$. n varies along the x-axis.

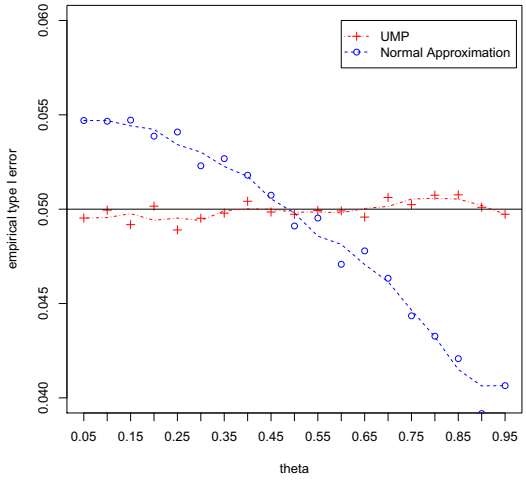

FIGURE 4. Empirical type I error α for UMP and Normal approximation tests for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta \geq \theta_0$. x-axis is θ_0. $n = 30$, $\epsilon = 1$, and $\delta = 0$. Target is $\alpha = .05$.

and Slavković (2009) and the non-private UMP test. Our goal is to understand the difference in performance between these tests.

We define the empirical power to be the proportion of times a test 'Rejects' when the alternative is true, and the empirical type I error as the proportion of times a test 'Rejects' when the null is true. For our simulations, we focus on small samples as the noise introduced by DP methods is most impactful in this setting.

In Figure 3, we plot the empirical power of our UMP test, the Normal approximation from Vu and Slavković (2009), and the non-private UMP. For each n, we generate 10,000 samples from Binom$(n, .95)$. We privatize each X by adding $N \sim \text{Tulap}(0, e^{-\epsilon}, 0)$ for the DP-UMP and $L \sim \text{Lap}(1/\epsilon)$ for the Normal approximation. We compute the UMP p-value via Algorithm 1 and the approximate p-value for $X + L$, using the cdf of $N\left(X, n/4 + 2/\epsilon^2\right)$. The empirical power is given by $(10000)^{-1}\#\{p\text{-value} < .05\}$. The DP-UMP test indeed gives higher power compared to the Normal approximation. The largest discrepancy in power between these two tests is at $n = 128$ with a difference of .039. In terms of a percentage, the greatest difference between the DP-UMP and the Normal approximation is at $n = 32$, with the approximate test achieving only 79.5% the power as the DP-UMP.

In Figure 4 we plot the empirical type I error of the DP-UMP and the Normal approximation tests. We fix $\epsilon = 1$ and $\delta = 0$, and vary θ_0. For each θ_0, we generate 100,000 samples from Binom$(30, \theta_0)$. For each sample, we compute the DP-UMP and Normal approximation tests at type I error $\alpha = .05$. We plot the proportion of times we reject the null as well as moving average curves. The DP-UMP, which is provably at type I error $\alpha = .05$ achieves

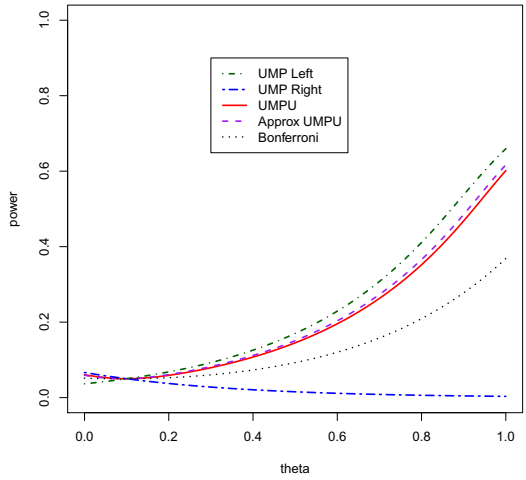

FIGURE 5. Power of DP tests as the true θ varies. $n = 30$, $\theta_0 = .1$, $\epsilon = .1$, $\delta = 0$, $\alpha = .05$. Testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta_0 \neq \theta_0$. x-axis is the truth.

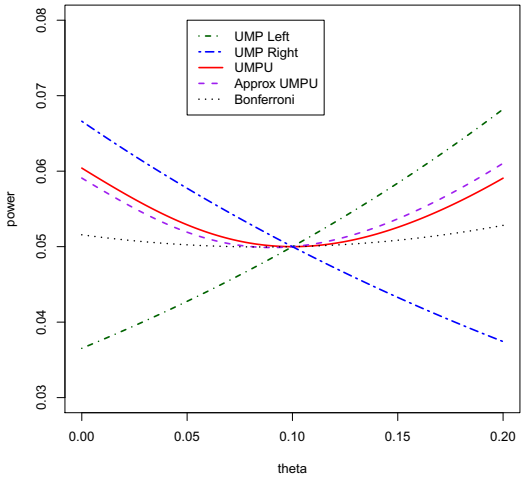

FIGURE 6. Power of DP tests as the true θ varies. Same parameters as Figure 5. Testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta_0 \neq \theta_0$. x-axis is the truth for $0 \leq \theta \leq .2$.

type I error very close to .05, but the Normal approximation has a higher type I error for small values of θ_0, and a lower type I error for large values of θ_0.

5.2. **Two-sided hypothesis testing simulations.** In this section, we compare the various tests we have developed for $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. As the DP-UMPU does not have a simple formula for p-values, we are particularly interested in understanding how closely the test of Algorithm 3 approximates the DP-UMPU. We also include the one-sided DP-UMPU tests, since these provide upper bounds for the power of the two-sided tests. For each of the simulations, we were able to compute the power exactly, since we have closed forms for the tests in terms of the Tulap distribution, and power is just an expected value.

In Figures 5-8, we plot the power of our proposed DP test for varying values of the true θ and sample size n. The label "UMP Left" corresponds to the DP-UMP test for $H_0 : \theta \leq \theta_0$, "UMP Right" corresponds to the DP-UMP test for $H_0 : \theta \geq \theta_0$, "UMPU" corresponds to the test from Theorem 2.18, "Approx UMPU" corresponds to the test from Section 2.8, and "Bonferroni" corresponds to the test from Proposition 2.17.

In Figures 5 and 6, we see how the tests perform when the null value is more extreme ($\theta_0 = .1$). As our theory showed, the DP-UMP test for $H_0 : \theta \leq \theta_0$ is the most powerful for true values $> \theta_0$, and the DP-UMP test for $H_0 : \theta \geq \theta_0$ is the most powerful for true values $< \theta_0$. We see that the DP-UMPU test and the approximately unbiased test perform well on both sides. However, the Bonferroni test suffers a loss in power, demonstrating that either the DP-UMPU, or the approximately unbiased test should be preferred.

In Figure 7 we study our tests when $n = 100$ and $\theta_0 = .5$. In this case, the approximate test is identical to the DP-UMPU test, which we know from Corollary 2.19. The Bonferroni

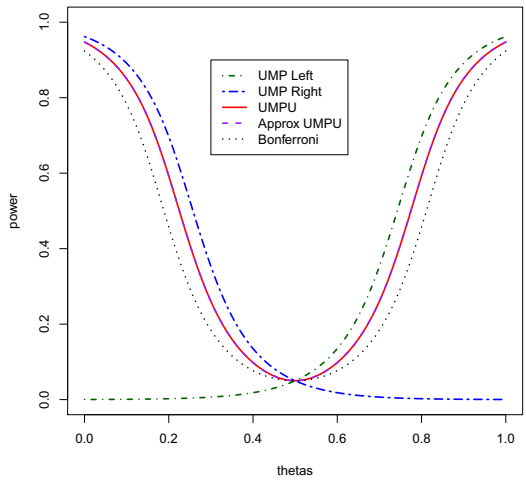

FIGURE 7. Power of DP tests as the true θ varies. $n = 100$, $\theta_0 = .5$, $\epsilon = .1$, $\delta = 0$, $\alpha = .05$. Testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta_0 \neq \theta_0$. x-axis is the truth.

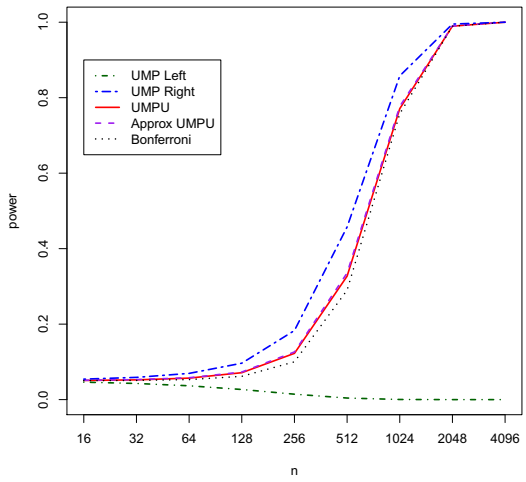

FIGURE 8. Power of DP tests as the sample size n increases. $\theta_0 = .8$, truth$= .75$, $\epsilon = .1$, $\delta = 0$, $\alpha = .05$. Testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta_0 \neq \theta_0$. x-axis is the sample size n.

test can also be shown to be unbiased in this setting, however it still suffers a loss in power since it is not UMP. As in Figure 5, the one-sided tests give upper bounds on the power.

In Figures 5, 6, and 7, we see that all of the proposed tests have power equal to $\alpha = .05$ when the true value of θ is equal to the null. This confirms that all of our tests have type I error exactly α, as claimed.

Figure 8 compares the power of the tests as the sample size increases. In this simulation, we are testing $H_0 : \theta_0 = .8$ versus $H_1 : \theta_0 \neq .8$ where the true value is $\theta = .75$. We use the values $\epsilon = .1$, $\delta = 0$, and $\alpha = .05$. In this plot, we see again that the DP-UMP test for $H_0 : \theta \leq \theta_0$ has more power than any of the other tests. The power of the UMPU and the approximate UMPU are indistinguishable, and the power of the Bonferroni test is slightly lower than either the UMPU or approximate UMPU tests. As we expect, the power of the DP-UMP test for $H_0 : \theta \geq \theta_0$ goes to zero as $n \to \infty$.

5.3. **Two-sided confidence interval simulations.** In this section, we study the performance of the private confidence intervals given in Proposition 3.6. The label "Approx UMPU" corresponds to the interval C_α^2 defined in Proposition 3.6, and "Bonferroni" corresponds to the interval C_α^1 defined in Proposition 3.6. As the DP-UMAU cannot be easily computed, our goal is to determine which of these two confidence intervals provides better performance.

In Figure 9, we compute the average width of the intervals depending on the true value of θ over 1000 replicates for each value of θ. For this simulation, $n = 30$, $\epsilon = 1$, $\delta = 0$, and $\alpha = .05$. In Figure 9, we see that the approximately unbiased confidence interval achieves smaller width than the Bonferroni confidence interval for moderate θs, at the expense of larger widths for more extreme θs. At $\theta = \frac{1}{2}$, the approximately unbiased confidence interval

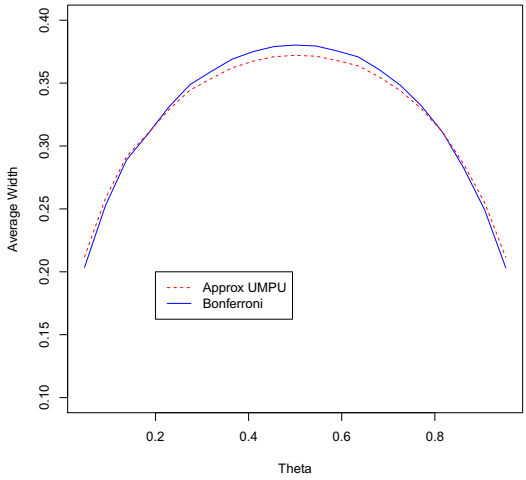

FIGURE 9. Average width of DP confidence intervals as the true θ varies. $n = 30$, $\epsilon = 1$, $\delta = 0$, $\alpha = .05$. Average width of DP confidence intervals, over 1000 replications. x-axis is the true value of θ.

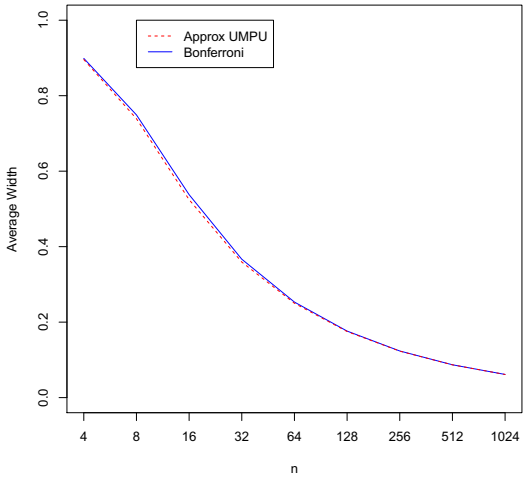

FIGURE 10. Average width of DP confidence intervals as the sample size n increases. $\theta = 1/2$, $\epsilon = 1$, $\delta = 0$, $\alpha = .05$. Average width of DP confidence intervals, over 1000 replications. x-axis is the sample size n.

is 97.8% the width of the Bonferroni confidence interval, but at θ close to 0 or 1, the approximately unbiased confidence interval is 4.1% wider than the Bonferroni confidence interval. The empirical coverage varied between .93 and .962 for both confidence intervals, with an average coverage of 0.9496. The Monte Carlo standard error of these estimates is $\sqrt{.95 * (1 - .95)/(1000)} = 0.0069$, suggesting that the confidence intervals have coverage $(1 - \alpha) = .95$ as claimed by Proposition 3.6.

In Figure 10, we explore the average width, but vary the sample size instead of θ. For this simulation, the true value of θ is $\frac{1}{2}$, $\epsilon = 1$, $\delta = 0$, and $\alpha = .05$. The average width is computed at each value of n, over 1000 replicates. We see that the approximately unbiased confidence interval consistently achieves a slightly smaller average width than the Bonferroni confidence interval. The largest discrepancy is at $n = 16$, where is approximately unbiased confidence interval has an average width 97.53% of the Bonferroni confidence interval. We see that as the sample size increases, the average width of these two confidence intervals becomes more similar.

6. Discussion and future directions

In this paper, we derived uniformly most powerful simple and one-sided tests for Bernoulli data among all DP α-level tests. Previously, while various hypothesis tests under DP have been proposed, none have satisfied such an optimality criterion. While our initial DP-UMP tests only output 'Reject' or 'Fail to Reject', we showed that they can be achieved by

post-processing a noisy sufficient statistic based on the Tulap distribution. This allows us to produce private p-values which agree with the DP-UMP tests. We also applied our techniques to produce two-sided tests, confidence intervals, and confidence distributions.

The ability to produce private p-values and confidence intervals, rather than simply an accept/reject decision, has practical importance as well, since both the statistics and scientific community have been strongly arguing for providing more complete information on basic statistical inference when determining statistical significance, as the latter cannot and should not be equated with scientific significance (Nuzzo, 2014; Wasserstein, Lazar et al., 2016). Gardner and Altman (1986) argues that instead of only reporting significance, scientists should report "sample estimates, confidence intervals, test statistics, and p-values" to provide full statistical information of the study. We have shown in this paper, that all of these results can be produced within the constraint of differential privacy, at a fixed privacy budget.

A simple, yet fundamental observation that underlies our results is that DP tests can be written in terms of linear constraints. This idea alone allows for a new perspective on DP hypothesis testing, which is particularly applicable to other discrete problems, such as multinomial models or difference of population proportions. Stating the problem in this form allows for the consideration of all possible DP tests, and allows the exploration of UMP tests through numerical linear program solvers.

We showed that for exchangeable data, DP tests need only depend on the empirical distribution. For binary data, the empirical distribution is equivalent to the sample sum, which is a complete sufficient statistic for the binomial model. However, in general it is not clear whether optimal DP tests are always a function of complete sufficient statistics as is the case for classical UMP tests. It would be worth investigating whether there is a notion of sufficiency which applies for DP tests.

When $\delta = 0$, our optimal noise adding mechanism, the proposed Tulap distribution, is related to the discrete Laplace distribution, which Ghosh, Roughgarden and Sundararajan (2009) and Geng and Viswanath (2016*a*) also found is optimal for a general class of loss functions. For $\delta > 0$, a truncated discrete Laplace distribution is optimal for our problem. Little previous work has looked into optimal noise adding mechanisms for (ϵ, δ)-DP. Geng and Viswanath (2016*b*) studied this problem to some extent, but did not consider truncated Laplace distributions. Steinke (2018) and Bun et al. (2018) propose that truncated Laplace can be viewed as the "canonical distribution" for (ϵ, δ)-DP in a way that Laplace is "canonical" for $(\epsilon, 0)$-DP. Further exploration in the use of truncated Laplace distributions in the (ϵ, δ)-DP setting may be fruitful.

It is interesting to consider whether there exist DP-UMP tests in the settings of either concentrated DP (Dwork and Rothblum, 2016; Bun and Steinke, 2016), or Renyi DP (Mironov, 2017). Unfortunately, our DP-UMP tests rely on the fact that the inequalities of (ϵ, δ)-DP produce linear constraints on the test ϕ. However, for concentrated DP and Renyi DP, the constraints are not linear. More sophisticated techniques may be required to understand these settings.

In our work, we found that there was a close connection between our UMP tests and the discrete Laplace distribution, which Ghosh, Roughgarden and Sundararajan (2009) showed is universal utility maximizing for binary data. However, Brenner and Nissim (2014) show that when the data are non-binary, there is no universal utility maximizing mechanism. As Canonne et al. (2019) discuss, this result seems to imply that in settings where the data is non-binary, it may not be possible to develop DP-UMP tests. Even if this is the case, finding

statistically sound techniques to account for the additional randomness due to privacy, is still an essential problem to consider, to help with the usability of DP.

Acknowledgments

We thank Vishesh Karwa and Matthew Reimherr for helpful discussions and feedback on previous drafts; Anton Xue for helping to develop some of the R code; and the reviewers for their helpful comments and suggestions, which improved the completeness and accessibility of this paper. This work is supported in part by NSF Award No. SES-1534433 to The Pennsylvania State University. We thank the Simons Institute for the Theory of Computing and the Center for Research on Computation and Society at Harvard Univerisity for their hospitality during part of this work.

References

Awan, Jordan, and Aleksandra Slavković. 2018*a*. "Differentially Private Uniformly Most Powerful Tests for Binomial Data." In *Advances in Neural Information Processing Systems 31*. , ed. S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett, 4208–4218. Curran Associates, Inc. http://papers.nips.cc/paper/7675-differentially-private-uniformly-most-powerful-tests-for-binomial-data.pdf.

Awan, Jordan, and Aleksandra Slavković. 2018*b*. "Structure and sensitivity in differential privacy: Comparing K-norm mechanisms." *arXiv:1801.09236*. http://arxiv.org/abs/1801.09236.

Barrientos, Andrés F, Jerome P Reiter, Ashwin Machanavajjhala, and Yan Chen. 2019. "Differentially private significance tests for regression coefficients." *Journal of Computational and Graphical Statistics*, 1–24. https://doi.org/10.1080/10618600.2018.1538881 .

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1): 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x .

Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA:Springer-Verlag New York, Inc.

Brenner, Hai, and Kobbi Nissim. 2014. "Impossibility of differentially private universally optimal mechanisms." *SIAM Journal on Computing*, 43(5): 1513–1540. https://doi.org/10.1137/110846671 .

Bun, Mark, and Thomas Steinke. 2016. "Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds." *CoRR*, abs/1605.02065. http://arxiv.org/abs/1605.02065.

Bun, Mark, Cynthia Dwork, Guy N Rothblum, and Thomas Steinke. 2018. "Composable and versatile privacy via truncated CDP." 74–86, ACM. https://doi.org/10.1109/FOCS.2007.41 .

Canonne, Clément L, Gautam Kamath, Audra McMillan, Adam Smith, and Jonathan Ullman. 2019. "The structure of optimal private tests for simple hypotheses." 310–321, ACM. https://doi.org/10.1145/3313276.3316336 .

Casella, G., and R.L. Berger. 2002. *Statistical Inference. Duxbury advanced series in statistics and decision sciences*, Thomson Learning. https://books.google.com/books?id=0x_vAAAAMAAJ.

Duchi, John C., Michael I. Jordan, and Martin J. Wainwright. 2018. "Minimax Optimal Procedures for Locally Private Estimation." *Journal of the American Statistical Association*, 113(521): 182–201. https://doi.org/10.1080/01621459.2017.1389735 .

Dwork, Cynthia, and Aaron Roth. 2014. "The Algorithmic Foundations of Differential Privacy." *Foundations and Trends® in Theoretical Computer Science*, 9(3–4): 211–407. https://doi.org/10.1561/0400000042 .

Dwork, Cynthia, and Guy N. Rothblum. 2016. "Concentrated Differential Privacy." *arXiv:1603.01887*. http://dblp.uni-trier.de/rec/bib/journals/corr/DworkR16.

Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. "Calibrating Noise to Sensitivity in Private Data Analysis." *Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings*, 265–284. Berlin, Heidelberg:Springer Berlin Heidelberg. https://doi.org/10.1007/11681878_14 .

Dwork, Cynthia, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Leon Roth. 2015. "Preserving Statistical Validity in Adaptive Data Analysis." *STOC '15*, 117–126. New York, NY, USA:ACM. https://doi.org/10.1145/2746539.2746580 . http://doi.acm.org/10.1145/2746539.2746580.

Gaboardi, Marco, and Ryan Rogers. 2018. "Local Private Hypothesis Testing: Chi-Square Tests." Vol. 80 of *Proceedings of Machine Learning Research*, 1626–1635. Stockholmsmässan, Stockholm Sweden:PMLR. http://proceedings.mlr.press/v80/gaboardi18a.html.

Gaboardi, Marco, Hyun Lim, Ryan Rogers, and Salil Vadhan. 2016. "Differentially Private Chi-Squared Hypothesis Testing: Goodness of Fit and Independence Testing." Vol. 48 of *Proceedings of Machine Learning Research*, 2111–2120. New York, New York, USA:PMLR. http://dl.acm.org/citation.cfm?id=3045390.3045613.

Gardner, Martin J, and Douglas G Altman. 1986. "Confidence intervals rather than P values: estimation rather than hypothesis testing." *Br Med J (Clin Res Ed)*, 292(6522): 746–750. https://doi.org/10.1136/bmj.296.6631.1210 .

Geng, Quan, and Pramod Viswanath. 2016*a*. "The optimal noise-adding mechanism in differential privacy." *IEEE Transactions on Information Theory*, 62(2): 925–951. https://doi.org/10.1109/TIT.2015.2504967 .

Geng, Quan, and Pramod Viswanath. 2016*b*. "Optimal Noise Adding Mechanisms for Approximate Differential Privacy." *IEEE Trans. Information Theory*, 62(2): 952–969. https://doi.org/10.1109/TIT.2015.2504972 . https://doi.org/10.1109/TIT.2015.2504972.

Geyer, Charles J., and Glen D. Meeden. 2005. "Fuzzy and Randomized Confidence Intervals and P -Values." *Statist. Sci.*, 20(4): 358–366. https://doi.org/10.1214/088342305000000340 . https://doi.org/10.1214/088342305000000340.

Ghosh, Arpita, Tim Roughgarden, and Mukund Sundararajan. 2009. "Universally Utility-maximizing Privacy Mechanisms." *STOC '09*, 351–360. New York, NY, USA:ACM. https://doi.org/10.1145/1536414.1536464 . http://doi.acm.org/10.1145/1536414.1536464.

Gibbons, J.D., and S. Chakraborti. 2014. *Nonparametric Statistical Inference, Fourth Edition: Revised and Expanded.* Taylor & Francis. https://books.google.com/books?

id=kJbVO2G6VicC.

Haeberlen, Andreas, Benjamin C. Pierce, and Arjun Narayan. 2011. "Differential Privacy Under Fire." *SEC'11*, 33–33. Berkeley, CA, USA:USENIX Association. http://dl.acm.org/citation.cfm?id=2028067.2028100.

Inusah, Seidu, and Tomasz J. Kozubowski. 2006. "A discrete analogue of the Laplace distribution." *Journal of Statistical Planning and Inference*, 136(3): 1090 – 1102. https://doi.org/https://doi.org/10.1016/j.jspi.2004.08.014 . http://www.sciencedirect.com/science/article/pii/S0378375804003519.

Kairouz, Peter, Sewoong Oh, and Pramod Viswanath. 2017. "The Composition Theorem for Differential Privacy." *IEEE Trans. Information Theory*, 63(6): 4037–4049. https://doi.org/10.1109/TIT.2017.2685505 .

Karwa, Vishesh, and Salil P. Vadhan. 2017. "Finite Sample Differentially Private Confidence Intervals." *arXiv:1711.03908*. http://arxiv.org/abs/1711.03908.

Larson, Ron, and Bruce H. Edwards. 2010. *Calculus.* . 9 ed., Cengage Learning. https://www.cengage.com/c/calculus-11e-larson/9780547167022/.

Lehmann, E.L., and J.P. Romano. 2008. *Testing Statistical Hypotheses. Springer Texts in Statistics*, Springer New York. https://books.google.com/books?id=Y7vSVW3ebSwC.

Mironov, Ilya. 2017. "Rényi Differential Privacy." *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, 263–275. https://doi.org/10.1109/CSF.2017.11 .

Nuzzo, Regina. 2014. "Scientific method: statistical errors." *Nature News*, 506(7487): 150. https://doi.org/10.1038/506150a .

Royden, Halsey Lawrence, and Patrick Fitzpatrick. 1988. *Real analysis.* Vol. 32, Macmillan New York.

Schervish, M.J. 1996. *Theory of Statistics. Springer Series in Statistics*, Springer New York. https://books.google.com/books?id=F9A9af4It10C.

Sheffet, Or. 2017. "Differentially Private Ordinary Least Squares." Vol. 70 of *Proceedings of Machine Learning Research*, 3105–3114. International Convention Centre, Sydney, Australia:PMLR. http://proceedings.mlr.press/v70/sheffet17a.html.

Solea, Eftychia. 2014. "Differentially private hypothesis testing for normal random variables." Master's diss. The Pennsylvania State University. https://etda.libraries.psu.edu/catalog/21486.

Steinke, Thomas. 2018. *Private correspondence.*

Uhler, Caroline, Aleksandra Slavković, and Stephen Fienberg. 2013. "Privacy-Preserving Data Sharing for Genome-Wide Association Studies." *Journal of Privacy and Confidentiality*, 5. https://doi.org/10.29012/jpc.v5i1.629 .

Van der Vaart, A.W. 2000. *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press. https://books.google.com/books?id=Ocg2AAAAQBAJ.

Vu, Duy, and Aleksandra Slavković. 2009. "Differential Privacy for Clinical Trial Data: Preliminary Evaluations." *ICDMW '09*, 138–143. Washington, DC, USA:IEEE Computer Society. https://doi.org/10.1109/ICDMW.2009.52 .

Wang, Y., J. Lee, and D. Kifer. 2015. "Revisiting Differentially Private Hypothesis Tests for Categorical Data." *arXiv:1511.03376*. http://arXiv/abs/1511.03376.

Wang, Yue, Daniel Kifer, Jaewoo Lee, and Vishesh Karwa. 2018. "Statistical Approximating Distributions Under Differential Privacy." *Journal of Privacy and Confidentiality*, 8(1). https://doi.org/doi.org/10.29012/jpc.666 .

Wasserman, Larry, and Shuheng Zhou. 2010. "A Statistical Framework for Differential

Privacy." *Journal of the American Statistical Association*, 105:489: 375–389. https://doi.org/10.1198/jasa.2009.tm08651 .

Wasserstein, Ronald L, Nicole A Lazar, et al. 2016. "The ASA's statement on p-values: context, process, and purpose." *The American Statistician*, 70(2): 129–133. https://doi.org/10.1080/00031305.2016.1154108 .

Xie, Min Ge, and Kesar Singh. 2013. "Confidence Distribution, the Frequentist Distribution Estimator of a Parameter: A Review." *International Statistical Review*, 81(1): 3–39. https://doi.org/10.1111/insr.12000.

Appendix A. Detailed proofs and technical lemmas

Proof of Theorem 2.2. Define ϕ' by $\phi'(\underline{x}) = \frac{1}{n!}\sum_{\pi\in\sigma(n)}\phi(\pi(\underline{x}))$, where $\sigma(n)$ is the symmetric group on n letters. First note that $\phi(\pi(\underline{x}))$ satisfies (2.1) for all $\pi \in \sigma(n)$, and that $\int \phi(\pi(\underline{x}))\, d\mu_\theta = \int \phi(\underline{x})\, d\mu_\theta$. Then by exchangeability,

$$\begin{aligned}\int \phi'(\underline{x})\, d\mu_\theta = \int \frac{1}{n!}\sum_{\pi\in\sigma(n)}\phi(\pi(\underline{x}))\, d\mu_\theta &= \frac{1}{n!}\sum_{\pi\in\sigma(n)}\int \phi(\pi(\underline{x}))\, d\mu_\theta \\ &= \frac{1}{n!}\sum_{\pi\in\sigma(n)}\int \phi(\underline{x})\, d\mu_\theta = \int \phi(\underline{x})\, d\mu_\theta.\end{aligned}$$

To see that ϕ' satisfies (ϵ,δ)-DP, we check condition (2.1):

$$\begin{aligned}\phi'(\underline{x}) = \frac{1}{n!}\sum_{\pi\in\sigma(n)}\phi(\pi(\underline{x})) &\le \frac{1}{n!}\sum_{\pi\in\sigma(n)}(e^\epsilon\phi(\pi(\underline{x}')) + \delta) \\ &= \frac{1}{n!}\sum_{\pi\in\sigma(n)} e^\epsilon\phi(\pi(\underline{x}')) + \frac{1}{n!}\sum_{\pi\in\sigma(n)}\delta = e^\epsilon\phi'(\underline{x}') + \delta\end{aligned}$$

$$\begin{aligned}(1-\phi'(\underline{x})) = \left(1 - \frac{1}{n!}\sum_{\pi\in\sigma(n)}\phi(\pi(\underline{x}))\right) &= \frac{1}{n!}\sum_{\pi\in\sigma(n)}(1-\phi(\pi(\underline{x}))) \\ \le \frac{1}{n!}\sum_{\pi\in\sigma(n)}\left(e^\epsilon(1-\phi(\pi(\underline{x}'))) + \delta\right) &= e^\epsilon\left(1-\phi'(\pi(\underline{x}'))\right) + \delta.\end{aligned}$$

□

In Lemma A.1 which follows, we use the notation $L \sim \mathrm{DLap}(b)$ to denote that the random variable L follows a *discrete Laplace* distribution, which has pmf $f_L(x) = \frac{1-b}{1+b}b^{|x|}$ for $x \in \mathbb{Z}$ (Inusah and Kozubowski, 2006). Furthermore, we write $G \sim \mathrm{Geom}(p)$ to denote that G is a geometric random variable with pmf $f_G(x) = (1-p)^x p$ for $x \in \{0,1,2,\ldots\}$.

Lemma A.1.

(1) *Let* $L \sim \mathrm{DLap}(b)$, $U \sim \mathrm{Unif}(-1/2, 1/2)$, $G_1, G_2 \overset{iid}{\sim} \mathrm{Geom}(1-b)$, *and* $N_0 \sim \mathrm{Tulap}(m,b,0)$. *Then* $L + U + m \overset{d}{=} G_1 - G_2 + U + m \overset{d}{=} N_0$.
(2) *Let* N *be the output of Algorithm 2 with inputs* m, b, q. *Then* $N \sim \mathrm{Tulap}(m,b,q)$.
(3) *The random variable* $N \sim \mathrm{Tulap}(m,b,q)$ *is continuous and symmetric about* m.

Proof of Lemma A.1.

(1) We know that $L \overset{d}{=} G_1 - G_2$, as shown in Inusah and Kozubowski (2006). Let $f_U(\cdot)$ denote the pdf of U, and F_U denote the cdf of U. We will use the property that $f_U(x) = f_U(-x)$ and $F_U(-x) = 1 - F_U(x)$. Then the pdf of $L + U$ is

$$f_{L+U}(x) = f_U(x - [x])\left(\frac{1-b}{1+b}\right)b^{|[x]|} = \begin{cases} f_U(x-[x])\left(\frac{1-b}{1+b}\right)b^{-[x]} & [x] \le 0 \\ f_U(x-[x])\left(\frac{1-b}{1+b}\right)b^{[x]} & [x] > 0. \end{cases}$$

If $[x] \leq 0$, then we have

$$\begin{aligned}
F_{L+U}(x) &= \int_{-\infty}^{x} f_U(t-[t]) \left(\frac{1-b}{1+b}\right) b^{-[t]}\, dt \\
&= \int_{-\infty}^{[x]-1/2} f_U(t-[t]) \left(\frac{1-b}{1+b}\right) b^{-[t]}\, dt + \int_{[x]-1/2}^{x} f_U(t-[x]) \left(\frac{1-b}{1+b}\right) b^{-[x]}\, dt \\
&= \sum_{t=-\infty}^{[x]-1} \left(\frac{1-b}{1+b}\right) b^{-t} + \int_{[x]-1/2}^{x} f_U(t-[x]) \left(\frac{1-b}{1+b}\right) b^{-[x]}\, dt \\
&= \frac{b^{-[x]+1}}{1+b} + F_U(x-[x]) \left(\frac{1-b}{1+b}\right) b^{-[x]} \\
&= \frac{b^{-[x]}}{1+b}(b + F_U(x-[x])(1-b)).
\end{aligned}$$

Since, $L+U$ is symmetric about zero, as both L and U are symmetric about zero, for $[x] \geq 0$ we have $F_{L+U}(x) = 1 - F_{L+U}(-x)$. The rest follows by replacing x with $x-m$, and $F_U(x) = x + 1/2$.

(2) If $q = 0$, then by part (1), it is clear that the output of Algorithm 2 has the correct distribution. If $q > 0$, then by rejection sampling, we have that $N \sim \text{Tulap}(m, b, q)$. For an introduction to rejection sampling, see Bishop (2006, Chapter 11).

(3) This property follows immediately from (1), and that $\text{Tulap}(m, b, q)$ is truncated equally on both sides of m. □

Lemma A.2. *Let $N \sim \text{Tulap}(m, b, q)$ and let $t \in \mathbb{Z}$. Then $F_N(t) = \begin{cases} b^{-t}C(m) & t \leq [m] \\ 1 - b^t C(-m) & t > [m], \end{cases}$ where $C(m) = (1+b)^{-1} b^{[m]}(b + ([m] - m + 1/2)(1-b))$. $C(m)$ is positive, monotone decreasing, and continuous in m. Furthermore, $b^{-[m]}C(m) = 1 - b^{[m]}C(-m)$.*

Proof of Lemma A.2. The form of the cdf at integer values is easily verified from Lemma A.1. It is clear that $C(m)$ is positive. It is also clear that $C(m)$ is continuous and monotone decreasing for all $m \in \mathbb{R} \setminus \{z + 1/2 \mid z \in \mathbb{Z}\}$. So, we will check that C is continuous at $m = z + 1/2$ for $z \in \mathbb{Z}$:

$$\lim_{\epsilon \downarrow 0} (1+b) C(z + 1/2 + \epsilon) = \lim_{\epsilon \downarrow 0} b^{z+1}(b + (1-\epsilon)(1-b)) = b^{z+1}$$

$$\lim_{\epsilon \downarrow 0} (1+b) C(z + 1/2 - \epsilon) = \lim_{\epsilon \downarrow 0} b^{z}(b + \epsilon(1-b)) = b^{z+1}.$$

Since C is continuous on $\mathbb{R}$ and monotone decreasing almost everywhere, it follows that C is monotone decreasing on $\mathbb{R}$ as well.

Call $\alpha(m) = [m] - m + 1/2$, which lies in $[0, 1]$. Note that $\alpha(-m) = -[m] + m + 1/2 = 1 - \alpha(m)$. Then

$$\begin{aligned}
(1+b) b^{-[m]} C(m) &= b + \alpha(m)(1-b) = b + (1 - \alpha(-m))(1-b) = b + (1-b) - \alpha(-m)(1-b) \\
&= (1+b) - (b + \alpha(-m)(1-b)) = (1+b)(1 - b^{[m]} C(-m)).
\end{aligned}$$

□

Proof of Lemma 2.5. First we show that (1) and (2) are equivalent. Clearly the m is the same for both. We must show that for $p \in (0,1)$, $e^\epsilon p \leq 1-e^{-\epsilon}(1-p)$ whenever $p \leq \frac{1}{1+e^\epsilon}$, and $e^\epsilon p > 1-e^{-\epsilon}(1-p)$ when $p > \frac{1}{1+e^\epsilon}$. Setting equal $e^\epsilon p = 1-e^{-\epsilon}(1-p)$ we find that $p = \frac{1}{1+e^\epsilon}$. As $p \to 1$, we have that $e^\epsilon p > 1 - e^{-\epsilon}(1-p)$ and as $p \to 0$, we have $e^\epsilon p < 1 - e^{-\epsilon}(1-p)$. We conclude that (1) and (2) are equivalent.

Next we show that (2) and (3) are equivalent. First we show that $F_{N_0}(x-m)$ satisfies the recurrence relation in (2). Set $b = e^{-\epsilon}$. First we show that for $t \in \mathbb{Z}$ such that $t \leq [m]-1$, $F_{N_0}(t-m) \leq \frac{1}{1+e^\epsilon}$ and for $t \geq [m]$, $F_{N_0}(t-m) \geq \frac{1}{1+e^\epsilon}$. Since, $F_{N_0}(t-m)$ is increasing in t, it suffices to check $t = [m]-1$ and $t = [m]$:

$$F_{N_0}([m]-1-m) = b^{-[m]+1+[m]}\frac{(b+([m]-m+1/2)(1-b))}{1+b} \leq \frac{b}{1+b} = \frac{1}{1+e^\epsilon}$$

$$F_{N_0}([m]-m) = b^{-[m]+[m]}\frac{(b+([m]-m+1/2)(1-b))}{1+b} \geq \frac{b}{1+b} = \frac{1}{1+e^\epsilon},$$

where we use the fact that $0 \leq [m]-m+1/2 \leq 1$. Now, let $t \in \mathbb{Z}$ and check three cases:

- Let $t < [m]$, then $e^\epsilon F_{N_0}(t-m) = e^\epsilon b^{-t}C(m) = b^{-(t+1)}C(m) = F_{N_0}(t+1-m)$.
- Let $t = [m]$. Using Lemma A.2, $1-e^{-\epsilon}(1-F_{N_0}(t-m)) = 1-b(1-b^{-[m]}C(m)) = 1-b+b(1-b^{[m]}C(-m)) = 1-b+b-b^{[m+1]}C(-m) = F_{N_0}(t+1-m)$.
- Let $t > m$. Then $1-e^{-\epsilon}(1-F_{N_0}(t-m)) = 1-b(b^tC(-m)) = 1-b^{t+1}C(-m) = F_{N_0}(t+1-m)$.

Finally, for any value $c \in (0,1)$, we can find m such that $F_{N_0}(0-m) = c$, by the Intermediate Value Theorem (Larson and Edwards, 2010, Theorem 1.13). On the other hand, given m, set $\phi(0) = F_{N_0}(0-m)$. □

Proof of Lemma 2.7. First note that $\phi^* \in \mathscr{D}^n_{\epsilon,0}$, since by Lemma 2.5, $\phi^*(x) = \min\{e^\epsilon\phi^*(x-1), 1-e^{-\epsilon}(1-\phi^*(x-1))\}$. So, ϕ^* satisfies (2.2)-(2.5). Next, since by Lemma A.2, $F_{N_0}(x-m)$ is a continuous, decreasing function in m with $\lim_{m\uparrow\infty} F_{N_0}(x-m) = 0$ and $\lim_{m\downarrow-\infty} F_{N_0}(x-m) = 1$, we can find m such that $E_{\theta_0}\phi^*(x) = \alpha$ by the Intermediate Value Theorem (Larson and Edwards, 2010, Theorem 1.13).

Now that we have argued that ϕ^* is a valid test, the rest of the result is an application of Lemma 2.6. It remains to show that the assumptions are satisfied for the lemma to apply.

Let $\phi \in \mathscr{D}^n_{\epsilon,0}$ such that $E_{\theta_0}\phi(x) \leq \alpha$. We will show that there exists $y \in \{0,1,2,\ldots,n\}$ such that for $x < y$, $\phi^*(x) < \phi(x)$ and for $x \geq y$, $\phi^*(x) \geq \phi(y)$. Suppose to the contrary that for all $x \in \{0,1,\ldots,n\}$, $\phi^*(x) < \phi(x)$. It follows that $\mathbb{E}_{\theta_0}\phi^*(x) < \mathbb{E}_{\theta_0}\phi(x) \leq \alpha$, contradicting the fact that $\mathbb{E}_{\theta_0}\phi^*(x) = \alpha$. We conclude that there exists y such that $\phi^*(y) \geq \phi(y)$.

Let y be the smallest point in $\{0,1,2,\ldots,n\}$ such that $\phi^*(y) \geq \phi(y)$. Our induction hypothesis is that for all $x \in \{y, y+1, \ldots, n\}$, $\phi^*(y) \geq \phi(y)$. For induction, suppose that for some $x \geq y$ we know that $\phi^*(x) \geq \phi(x)$. We will show that $\phi^*(x+1) \geq \phi(x+1)$. By Lemma 2.5, we know that $\phi^*(x+1) = \min\{e^\epsilon\phi^*(x), 1-e^{-\epsilon}(1-\phi^*(x))\}$, and by constraints (2.2)-(2.5), we know that $\phi(x+1) \leq \min\{e^\epsilon\phi(x), 1-e^{-\epsilon}(1-\phi(x))\}$.

- Case 1: If $\phi^*(x) \leq \frac{1}{1+e^\epsilon}$, then by Lemma 2.5, $\phi^*(x+1) = e^\epsilon\phi^*(x) \geq e^\epsilon\phi(x) \geq \phi(x+1)$.
- Case 2: If $\phi^*(x) > \frac{1}{1+e^\epsilon}$, then by Lemma 2.5, $\phi^*(x+1) = 1-e^{-\epsilon}(1-\phi^*(x)) \geq 1-e^{-\epsilon}(1-\phi(x)) \geq \phi(x+1)$.

We conclude that $\phi^*(x+1) \geq \phi(x+1)$. By induction, the inequality $\phi^*(x) \geq \phi(x)$ holds for all $x \in \{y, y+1, y+2, \ldots, n\}$. So, we have that $\phi^*(x) \geq \phi(x)$ for $x \in \{y, y+1, y+2, \ldots, n\}$ and $\phi^*(x) < \phi(x)$ for $x \in \{0,1,2,\ldots,y-1\}$. Since $\text{Binom}(n,\theta)$ has a monotone likelihood

ratio in θ, by Lemma 2.6 we have that $E_{\theta_1}\phi^*(x) \geq E_{\theta_1}\phi(x)$. We conclude that ϕ^* is UMP-α among $\mathscr{D}^n_{\epsilon,0}$ for the stated hypothesis test. □

Proof of Lemma 2.8. We will abbreviate $F(x) := F_{N_0}(x-m)$, where $N_0 \sim \text{Tulap}(0, b = e^{-\epsilon}, 0)$ to simplify notation. First we will show that (1) and (2) are equivalent. It is clear that y and m are the same in both. Next consider $1 - e^{-\epsilon}(1-p) + e^{-\epsilon}\delta = e^{\epsilon}p + \delta$, solving for p gives $p = \frac{1-\delta}{1+e^{\epsilon}}$. Considering as $p \to 0$ and $p \to 1$, we see that $1 - e^{-\epsilon}(1-p) + e^{-\epsilon}\delta \geq e^{\epsilon}p + \delta$ when $p \leq \frac{1-\delta}{1+e^{\epsilon}}$ and $1 - e^{-\epsilon}(1-p) + e^{-\epsilon}\delta \leq e^{\epsilon}p + \delta$ when $p \geq \frac{1-\delta}{1+e^{\epsilon}}$.

Next solving $1 - e^{-\epsilon}(1-p) + e^{-\epsilon}\delta = 1$ for p gives $p = 1 - \delta$. So, $1 - e^{-\epsilon}(1-p) + e^{-\epsilon}\delta \leq 1$ when $p \leq 1-\delta$ and $1 - e^{-\epsilon}(1-p) + e^{-\epsilon}\delta \geq 1$ when $p \geq 1-\delta$. Lastly, solving $e^{\epsilon}p + \delta = 1$ for p gives $p = \frac{1-\delta}{e^{\epsilon}} \geq \frac{1-\delta}{1+e^{\epsilon}}$. Combining all of these comparisons, we see that (1) is equivalent to (2).

Before we justify the equivalence of (2) and (3), we argue the following claim. Let $\phi(x)$ be defined as in (3). Then $\phi(x) \leq \frac{1-\delta}{1+e^{\epsilon}}$ if and only if $F(x) \leq \frac{1}{1+e^{\epsilon}}$. Suppose that $\phi(x) \leq \frac{1-\delta}{1+e^{\epsilon}}$. Then $\frac{F(x)-q/2}{1-q} \leq \frac{1-\delta}{1+e^{\epsilon}}$. Thus,

$$\begin{aligned} F(x) &\leq \frac{(1-q)(1-\delta)}{1+e^{\epsilon}} + \frac{q}{2} \\ &= \frac{1}{1+e^{\epsilon}}\left((1-q)(1-\delta) + \left(\frac{b+1}{b}\right)\frac{q}{2}\right) \\ &= \frac{1}{1+e^{\epsilon}}\left(\frac{(1-b)(1-\delta)}{1-b+2\delta b} + \left(\frac{b+1}{b}\right)\frac{\delta b}{1-b+2\delta b}\right) \\ &= \frac{1}{1+e^{\epsilon}}(1-b+2\delta b)^{-1}((1-b)(1-\delta) + (b+1)\delta) \\ &= \frac{1}{1+e^{\epsilon}}. \end{aligned}$$

We are now ready to show that $\phi(x)$ as described in (3) fits the form of (2).

- Suppose that $0 < \phi(x) < \frac{1-\delta}{1+e^{\epsilon}}$. By the above, we know that $F(x) \leq \frac{1}{1+e^{\epsilon}}$. By Lemma 2.5,

$$\begin{aligned} e^{\epsilon}\phi(x) + \delta = \frac{e^{\epsilon}F(x) - \frac{q}{2b}}{1-q} + \delta &= \frac{F(x+1) - \frac{q}{2}}{1-q} + \frac{\frac{q}{2} - \frac{q}{2b}}{1-q} + \delta \\ &= \phi(x+1) + \frac{\delta b}{1-b}\left(1 - \frac{1}{b}\right) + \delta = \phi(x+1). \end{aligned}$$

- Suppose that $\frac{1-\delta}{1+e^{\epsilon}} < \phi(x) \leq 1-\delta$. Then we have $F(x) > \frac{1}{1+e^{\epsilon}}$. Then

$$\begin{aligned} 1 - e^{-\epsilon}(1-\phi(x)) + e^{-\epsilon}\delta &= 1 - e^{-\epsilon}\left(1 - \frac{F(x) - q/2}{1-q}\right) + e^{-\epsilon}\delta \\ &= (1-q)^{-1}\left(1 - q - e^{-\epsilon}\left(1 - F(x) - q/2\right)\right) + e^{-\epsilon}\delta \\ &= (1-q)^{-1}(1 - e^{-\epsilon}(1-F(x)) + bq/2 - q) + b\delta \\ &= (1-q)^{-1}(F(x+1) - q/2) + \frac{(b-1)q/2}{1-q} + b\delta \\ &= \phi(x+1) + \frac{\delta b(b-1)}{1-b} + b\delta \\ &= \phi(x+1). \end{aligned}$$

- Finally, we must show that if $\phi(x) = 1$ then $\phi(x-1) \geq 1-\delta$. It suffices to show that $F(x) \geq 1 - q/2$ implies that $F(x-1) \geq (1-\delta)(1-q) + q/2 = 1 - (1/b)(q/2)$. We prove the contrapositive. Suppose that $F(x-1) < 1 - (1/b)(q/2)$. Then since F satisfies property (2.4), we know that

$$\begin{aligned} F(x) \leq 1 - e^{-\epsilon}(1 - F(x-1)) &< 1 - b(1 - (1 - (1/b)(q/2))) \\ &= 1 - b(1 - 1 + (1/b)(q/2)) = 1 - q/2. \end{aligned}$$

We have justified that $\phi(x)$ in (3) satisfies the recurrence relation in (2). Given ϕ' of the form in (2), with first non-zero entry at y, by Lemma A.2 and the Intermediate Value Theorem (Larson and Edwards, 2010, Theorem 1.13), we can find $m \in \mathbb{R}$ such that $\phi(y) = \phi'(y)$. We conclude that (1), (2), and (3) are all equivalent. □

Proof of Theorem 2.10. First we show that ϕ^* is UMP-α for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. Since $\phi^*(x)$ is increasing and $\text{Binom}(n, \theta)$ has a monotone likelihood ratio in θ, $E_\theta \phi^* \leq E_{\theta_0}\phi^* = \alpha$ for all $\theta \leq \theta_0$ (property of MLR). By Lemma 2.7, we know that $\phi^*(x)$ is most powerful for any alternative $\theta_1 > \theta_0$ versus the null θ_0. So, ϕ^* is UMP-α.

Next we show that ψ^* is UMP-α for $H_1 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$. First note that $\sup_{\theta \geq \theta_0} \mathbb{E}_\theta \psi^* = \alpha$. Let ψ be another test with $\sup_{\theta \geq \theta_0} \mathbb{E}_\theta \psi \leq \alpha$. Let $\theta_1 < \theta_0$, we will show that $\mathbb{E}_{\theta_1}\psi^* \geq \mathbb{E}_{\theta_1}\psi$. Define $\widetilde{\psi}^*(x) = \psi^*(n-x) = 1 - F_{N_0}(n-x-m_2) = F_{N_0}(x+m_2-n)$ and $\widetilde{\psi}(x) = \psi(n-x)$. Then using the map $(x, \theta) \mapsto (n-x, 1-\theta)$, we have that $\mathbb{E}_{X \sim (1-\theta_0)} \widetilde{\psi}^*(X) = \mathbb{E}_{X \sim (1-\theta_0)} \psi^*(n - X) = \mathbb{E}_{Y \sim \theta_0} \psi^*(Y) = \alpha$. By a similar argument for ψ, we have that both $\widetilde{\psi}^*$ and $\widetilde{\psi}$ are level α for $H_0 : \theta \leq 1 - \theta_0$ versus $H_1 : \theta > 1 - \theta_0$. Since $\mathbb{E}_{(1-\theta_0)} \widetilde{\psi}^* = \alpha$, and $\widetilde{\psi}^*(x) = F_{N_0}(x - m')$, we have that $\widetilde{\psi}^*$ is UMP-α for $H_0 : \theta \leq (1-\theta_0)$ versus $H_1 : \theta > (1-\theta_0)$. Then for $\theta_1 < \theta_0$,

$$\mathbb{E}_{X \sim \theta_1} \psi^*(X) = \mathbb{E}_{Y \sim (1-\theta_1)} \widetilde{\psi}^*(Y) \geq \mathbb{E}_{Y \sim (1-\theta_1)} \widetilde{\psi}(Y) = \mathbb{E}_{X \sim \theta_1} \psi(X).$$

We conclude that ψ^* is UMP-α for $H_1 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$. □

Lemma A.3. *Observe $\underline{x} \in \mathscr{X}^n$. Let $T : \mathscr{X}^n \to \mathbb{R}$, and let $\{\mu_{\underline{x}} \mid \underline{x} \in \mathscr{X}^n\}$ be a set of probability measures on $\mathbb{R}$, dominated by Lebesgue measure. Suppose that $\mu_{\underline{x}}$ is parameterized by $T(\underline{x})$ and $\mu_{\underline{x}}$ has MLR in $T(\underline{x})$. Then $\{\mu_{\underline{x}}\}$ satisfies (ϵ, δ)-DP if and only if for all $H(\underline{x}_1, \underline{x}_2) = 1$ and all $t \in \mathbb{R}$ the following two inequalities hold:*

$$\mu_{\underline{x}_1}((-\infty, t)) \leq e^\epsilon \mu_{\underline{x}_2}((-\infty, t)) + \delta \tag{A.1}$$

$$\mu_{\underline{x}_1}((t, \infty)) \leq e^\epsilon \mu_{\underline{x}_2}((t, \infty)) + \delta. \tag{A.2}$$

Proof of Lemma A.3. Let $\alpha \in [0, 1]$ be given. We will only consider $B \subset \mathbb{R}$ (Lebesgue measurable) such that $\mu_{\underline{x}_2}(B) = \alpha$. Then demonstrating (ϵ, δ)-DP requires $\sup_{\{B | \mu_{\underline{x}_2}(B) = \alpha\}} \mu_{\underline{x}_1}(B) \leq e^\epsilon \alpha + \delta$. We interpret this problem as testing the hypothesis $H_0 : \underline{x} = \underline{x}_2$ versus $H_1 : \underline{x} = \underline{x}_1$, using the rejection region B, where α is the type I error, and $\mu_{\underline{x}_1}(B)$ is the power. We know that $\sup_{\{B | \mu_{\underline{x}_2}(B) = \alpha\}} \mu_{\underline{x}_1}(B)$ is achieved by the Neyman-Pearson Lemma. Since $\mu_{\underline{x}}$ has an MLR in $T(\underline{x})$, $\arg\sup_{\{B | \mu_{\underline{x}_2}(B) = \alpha\}} \mu_{\underline{x}_1}(B)$ is either of the form $(-\infty, t)$ or (t, ∞), depending on whether $T(\underline{x}_1)$ is greater or lesser than $T(\underline{x}_2)$. Since $\mu_{\underline{x}_1}$ is dominated by Lebesgue measure for all $\underline{x}_1$, $\mu_{\underline{x}_2}((-\infty, t))$ is continuous in t, which allows us to achieve exactly α type I error. □

Proof of Theorem 2.11. Let $Z \sim \text{Tulap}\left(T(x), b = e^{-\epsilon}, \frac{2\delta b}{1-b+2\delta b}\right)$. We know that the distribution of Z is symmetric with location $T(x)$, and the pdf $f_Z(t)$ is increasing as a function of $|t - T(x)|$. It follows that $f_Z(t)$ has a MLR in $T(x)$. By Lemma 2.8, we know that $\phi(x) = F_Z(m)$ satisfies (2.2)-(2.5), so by Lemma A.3, we have the desired result. □

Proof of Theorem 2.12. We denote by $F_{Z\sim\theta_0}(\cdot)$ the cdf of the random variable Z, distributed as $Z \mid X \sim \text{Tulap}(X, b, q)$ and $X \sim \text{Binom}(n, \theta_0)$.

(1) First we show that $p(\theta_0, Z)$ is a p-value, according to Definition 2.4. To this end, consider

$$\sup_{\theta\le\theta_0} \underset{X\sim\text{Binom}(n,\theta)}{P_{Z|X\sim\text{Tulap}(X,b,q)}}(p(\theta, Z) \le \alpha) = \underset{X\sim\text{Binom}(n,\theta_0)}{P_{Z|X\sim\text{Tulap}(X,b,q)}}(p(\theta_0, Z) \le \alpha),$$

where we use the fact that X has a monotone likelihood ratio in θ. Note that $p(\theta_0, Z) = 1 - F_{Z\sim\theta_0}(Z)$. When $X \sim \text{Binom}(n, \theta_0)$, we have that $p(\theta_0, Z) = 1 - F_{Z\sim\theta_0}(Z) \sim \text{Unif}(0, 1)$. So,

$$\underset{X\sim\text{Binom}(n,\theta_0)}{P_{Z|X\sim\text{Tulap}(X,b,q)}}(p(\theta_0, Z) \le \alpha) = P_{U\sim\text{Unif}(0,1)}(U \le \alpha) = \alpha.$$

(2) Let $N \sim \text{Tulap}(0, b, q)$, and recall from Lemma 2.9 that the UMP-α test for $H_0 : \theta \le \theta_0$ versus $H_1 : \theta > \theta_0$ is $\phi^*(x) = F_N(x - m)$, where m satisfies $E_{\theta_0}\phi^*(x) = \alpha$. We can write ϕ^* as

$$\begin{aligned}\phi^*(x) = F_N(x - m) &= P_{N\sim\text{Tulap}(0,b,q)}(N \le X - m \mid X)\\ &= P_N(X + N \ge m \mid X) = P_{Z|X\sim\text{Tulap}(X,b,q)}(Z \ge m \mid X),\end{aligned}$$

where m is chosen such that

$$\begin{aligned}\alpha = E_{X\sim\theta_0}\phi^*(X) &= E_{X\sim\theta_0}P_{Z|X\sim\text{Tulap}(X,b,q)}(Z \ge m \mid X)\\ &= \underset{X\sim\text{Binom}(n,\theta_0)}{P_{Z|X\sim\text{Tulap}(X,b,q)}}(Z \ge m) = 1 - F_{Z\sim\theta_0}(m),\end{aligned}$$

where F is the cdf of the marginal distribution of Z, where $Z|X \sim \text{Tulap}(X, b, q)$ and $X \sim \text{Binom}(n, \theta_0)$. From this equation, we have that m is the $(1 - \alpha)$-quantile of the marginal distribution of Z.

Let $R|X \sim \text{Bern}(\phi^*(X))$ and $Z|X \sim \text{Tulap}(X, b, q)$. Then

$$\begin{aligned}R|X \overset{d}{=} I(Z \ge m) \mid X &\overset{d}{=} I(F_{Z\sim\theta_0}(Z) \ge F_{Z\sim\theta_0}(m)) \mid X\\ &\overset{d}{=} I\left(1 - \alpha \le F_{Z\sim\theta_0}(Z)\right)|X\\ &\overset{d}{=} I\left(p(\theta_0, Z) \le \alpha\right)|X.\end{aligned}$$

Taking the conditional expected value $\mathbb{E}(\cdot \mid X)$ of both sides gives

$$\phi^*(x) = E(R \mid X) = P_{Z|X\sim\text{Tulap}(X,b,q)}(p(\theta_0, Z) \le \alpha \mid X).$$

(3) Let $p'(X)$ be any other (ϵ, δ)-DP p-value, and let $\theta_1 > \theta_0$. We wish to show that

$$P_{X\sim\theta_1,N}(p(\theta_0, X + N) \le \alpha) \ge P_{X\sim\theta_1}(p'(X) \le \alpha).$$

However, the left side is just $\mathbb{E}_{X\sim\theta_1}\phi^*(X)$, the power of the DP-UMP test, and the right is the power of the corresponding DP test of p'. Since ϕ^* is uniformly most powerful among (ϵ, δ)-DP tests, the inequality is justified.

(4) We can express $p(\theta_0, Z)$ in the following way:

$$\begin{aligned} p(\theta_0, Z) &= P_{\substack{X\sim\text{Binom}(n,\theta_0)\\ N\sim\text{Tulap}(0,b,q)}}(X + N \geq Z) = P_{X,N}(-N \leq X - Z) \\ &= E_{X\sim\text{Binom}(n,\theta_0)} P_N(N \leq X - Z \mid X) = E_{X\sim\text{Binom}(n,\theta_0)} F_N(X - Z) \\ &= \sum_{x=0}^{n} F_N(x - Z)\binom{n}{x}\theta_0^x(1-\theta_0)^{n-x}, \end{aligned}$$

which is just the inner product of the vectors $\underline{F}$ and $\underline{B}$ in algorithm 1. □

Proof of Proposition 2.17. First $p'(\theta_0, Z)$ satisfies (ϵ, δ)-DP by post-processing. To verify (2), let $\alpha \in (0, 1)$, and consider

$$\begin{aligned} P_{N\sim\text{Tulap}}(p'(\theta_0, X + N) \leq \alpha \mid X) &= P(2\min\{p(\theta_0, X + N), 1 - p(\theta_0, X + N)\} \leq \alpha \mid X) \\ &= P(p(\theta_0, X + N) \leq \alpha/2 \mid X) \\ &\quad + P(p(\theta_0, X + N) \geq 1 - \alpha/2 \mid X) \\ &\quad - P(1 - \alpha/2 \leq p(\theta_0, X + N) \leq \alpha/2 \mid X) \\ &= \phi^*(X) + \psi^*(X) - 0, \end{aligned}$$

where we use the fact that $1 - \alpha/2 \geq \alpha/2$, implying that the last probability is zero. To see that $p'(\theta_0, Z)$ is a p-value, we compute

$$\begin{aligned} P_{X\sim\theta_0,N}(p'(\theta_0, X + N) \leq \alpha) &= \mathbb{E}_{X\sim\theta_0} P(p'(\theta_0, X + N) \leq \alpha \mid X) \\ &= \mathbb{E}_{X\sim\theta_0}[\phi^*(X) + \psi^*(X)] \\ &= \alpha/2 + \alpha/2. \end{aligned}$$

Finally, to see that ϕ' is more powerful than any level $\alpha/2$ test, notice that ϕ^* and ψ^* are the most powerful DP tests depending on whether $\theta > \theta_0$ or $\theta < \theta_0$, respectively. Since $\phi' = \phi^* + \psi^*$, it is more powerful than either of these tests. □

Proof of Theorem 2.18. We must show that there exists k and m which solve the two equations, and then argue that ϕ^* is UMP among all level α tests in $\mathscr{D}^n_{\epsilon,\delta}$. The proof is inspired by the Generalized Neyman Pearson Lemma Lehmann and Romano (2008, Theorem 3.6.1), and has a similar strategy as Theorem 2.10.

Let $\theta_1 \neq \theta_0$. We will show that $\phi*$ is most powerful among unbiased size α tests in $\mathscr{D}^n_{\epsilon,\delta}$ for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$. Set $f_1(x) = \binom{n}{x}\theta_0^x(1-\theta_0)^{n-x}$, $f_2(x) = (x - n\theta_0)\binom{n}{x}\theta_0^x(1-\theta_0)^{n-x}$, and $f_3(x) = \binom{n}{x}\theta_1^x(1-\theta_1)^{n-x}$. Let $\phi \in \mathscr{D}^n_{\epsilon,\delta}$ be any unbiased size α test, not identical to ϕ^*.

(1) There exists $k, m \in \mathbb{R}$ such that ϕ^* satisfies $\mathbb{E}_{X\sim\theta_0}(X - n\theta_0)\phi(X) = 0$ and $\mathbb{E}_{X\sim\theta_0}\phi(X) = \alpha$.

Proof. Set $g_1(k, m) = \mathbb{E}_{X\sim\theta_0}\phi^*(X) - \alpha$ and $g_2(k, m) = \mathbb{E}_{X\sim\theta_0}(X - n\theta_0)\phi^*(X)$. We need to show that there exists k and m such that $g_1(k, m) = g_2(k, m) = 0$.

Then, g_2 partitions $\mathbb{R}^2$ into two disjoint regions: the pairs (k, m) such that $g_2 \leq 0$ and those such that $g_2 > 0$. We claim that the solutions to $g_1 = 0$ form a curve. To see this, notice that for any $k \in \mathbb{R}$, there exists a unique $m \in \mathbb{R}$ such that $g_1(k, m) = 0$. In particular, when $k = -1$, and $m(-1)$ is the value such that $g_1(-1, m(-1)) = 0$, then $g_2(-1, m(-1)) > 0$, since g_2 measures the covariance between X and $\phi^*(X)$ and both are increasing functions of X. Similarly, for $k = n + 1$, $g_2(n + 1, m(n + 1)) < 0$.

Since, both g_1 and g_2 are continuous functions, by the Intermediate Value Theorem (Royden and Fitzpatrick, 1988, Proposition 11, p. 182), there exists $-1 \leq k \leq n+1$ and $m \in \mathbb{R}$ such that $g_1(k,m) = g_2(k,m) = 0$. □

(2) Since ϕ is unbiased, it's power must have a local minimum at θ_0 so, $\frac{d}{d\theta}\mathbb{E}_\theta\phi\Big|_{\theta=\theta_0} = 0$. This is equivalent to requiring that $\sum \phi f_2 = 0$.

Proof. We calculate the derivative of the power:

$$\begin{aligned}
\frac{d}{d\theta}\beta_\phi &= \frac{d}{d\theta}\sum_{x=0}^{n}\binom{n}{x}\theta^x(1-\theta)^{n-x}\phi(x) \\
&= \sum_{x=0}^{n}\binom{n}{x}\theta^x(1-\theta)^{n-x}\phi(x)\left(\frac{x}{\theta}+\frac{x-n}{1-\theta}\right) \\
&= \frac{1}{\theta(1-\theta)}\mathbb{E}_\theta(X-n\theta)\phi(X) \\
&= \frac{1}{\theta(1-\theta)}\sum \phi f_2.
\end{aligned}$$

□

(3) There exists $y_l \leq k \leq y_u$ (integers) such that $\phi^*(x) \geq \phi(x)$ when $x \geq y_u$ or $x \leq y_l$, and $\phi^*(x) \leq \phi(x)$ when $y_l \leq x \leq y_u$.

Proof. Since ϕ is not identical to ϕ^*, there exists $x \in \{0,\ldots,n\}$ such that $\phi^*(x) \neq \phi(x)$. If $\phi^*(x) > \phi(x)$ for all $0 \leq x \leq n$, then set $y_l = \text{floor}(k)$ and $y_u = \text{ceil}(k)$. If $\phi^*(x) \leq \phi(x)$ for all $0 \leq x \leq n$, then it cannot be that ϕ is size α. We conclude that there exists a value y such that $\phi^*(y) > \phi(y)$. If $y > k$, then for every $x \geq y$, $\phi^*(x) > \phi(x)$, since ϕ^* increases as much as possible. Alternatively, if $y < k$ then for all $x \leq y$, $\phi^*(x) < \phi(x)$. So, we have that either $y_l \leq k$ exists or $y_u \geq k$ exists. We need to show that both exist.

Suppose without loss of generality that $y_u \geq k$ exists, and suppose to the contrary that $y_l \leq k$ does not exist. Then it is the case that $\phi^*(x) > \phi(x)$ when $x \geq y_u$ and $\phi^*(x) \leq \phi(x)$ when $x < y_u$. Notice that $\frac{f_2(x)}{f_1(x)} = x - n\theta_0$. So, $\phi^*(x) \geq \phi(x)$ if and only if $f_2(x) \geq cf_1(x)$ for the constant $c = y_u - \frac{1}{2} - n\theta_0$. Then for all $x = 0,1,2,\ldots,n$,

$$(\phi^*(x) - \phi(x))(f_2(x) - cf_1(x)) \geq 0,$$

and $(\phi^*(y_u) - \phi(y_u))(f_2(y_u) - cf_1(y_u)) > 0$. Summing over x gives

$$\begin{aligned}
\sum_{x=0}^{n}(\phi^*(x) - \phi(x))(f_2(x) - cf_1(x)) &> 0 \\
\sum \phi^* f_2 - \sum \phi f_2 - c\sum \phi^* f_1 + c\sum \phi f_1 &> 0 \\
-\sum \phi f_2 &> 0 \\
\sum \phi f_2 &< 0,
\end{aligned}$$

We see that ϕ is not unbiased, contradicting our initial assumption. We conclude that both y_l and y_u exist.

□

(4) There exists $k_1, k_2 \in \mathbb{R}$ such that $f_3(x) \geq k_1 f_1(x) + k_2 f_2(x)$ when $x \notin (y_l, y_u)$ and $f_3(x) \leq k_1 f_1(x) + k_2 f_2(x)$ when $x \in (y_l, y_u)$.

Proof. We need to consider what forms the set $\{x \mid f_3(x) \geq k_1 f_1(x) + k_2 f_2(x)\}$ can take on. These are solutions to

$$1 \geq \frac{k_1 f_1(x) + k_2 f_2(x)}{f_3(x)}$$

$$1 \geq (k_1 + k_2(x - n\theta_0)) \frac{(1-\theta_0)^n}{(1-\theta_1)^n} \frac{\theta_0^x (1-\theta_0)^x}{\theta_1^x (1-\theta_1)^x}$$

$$\left(\frac{1-\theta_1}{1-\theta_0}\right)^n \left(\frac{\theta_1(1-\theta_1)}{\theta_0(1-\theta_0)}\right)^x \geq k_1 - k_2 n\theta_0 + k_2 x.$$

The left side is either convex or constant in x. The right side is linear in x. If the left is strictly convex (when $\theta_1 \neq 1 - \theta_0$), we can always choose k_1 and k_2 such that the set of solutions is of the form $(-\infty, y_l] \cup [y_u, \infty)$. If $\theta_1 = 1 - \theta_0$, set $k_1 = \left(\frac{1-\theta_1}{1-\theta_0}\right)^n$ and $k_2 = 0$. □

(5) The test ϕ^* is more powerful than ϕ at any θ_1.

Proof. We have established that there exists k_1 and k_2 such that $\phi^*(x) \geq \phi(x)$ whenever $f_3(x) \geq k_1 f_1(x) + k_2 f_2(x)$ and $\phi^*(x) \leq \phi(x)$ when $f_3(x) \leq k_1 f_1(x) + k_2 f_2(x)$. Then

$$(\phi^*(x) - \phi(x))(f_3(x) - k_1 f_1(x) - k_2 f_2(x)) \geq 0.$$

Then

$$\sum_{x=0}^{n} (\phi^*(x) - \phi(x))(f_3(x) - k_1 f_1(x) - k_2 f_2(x)) \geq 0$$

$$\sum_{x=0}^{n} \phi^*(x) f_3(x) - \sum_{x=0}^{n} \phi(x) f_3(x) \geq 0$$

$$\mathbb{E}_{X \sim \theta_1} \phi^*(X) \geq \mathbb{E}_{X \sim \theta_1} \phi(X).$$

Since our argument does not depend on the choice of θ_1, we conclude that ϕ^* is more powerful than any other size α unbiased test in $\mathscr{D}^n_{\epsilon,\delta}$. Finally, noting that by taking $\phi(x) := \alpha$, we see that ϕ^* is indeed unbiased. Hence, ϕ^* is the UMP-α among unbiased tests in $\mathscr{D}^n_{\epsilon,\delta}$. □

□

Proof of Corollary 2.19. We have to show that when using $k = \frac{n}{2}$, ϕ is unbiased. Call $A = \mathbb{E}_{X\sim\text{Binom}(n,1/2)}(X - \frac{n}{2})\phi(X)$. Then

$$\begin{aligned} A &= \left(\frac{1}{2}\right)^n \sum_{x=0}^{n} \binom{n}{x} \left(x - \frac{n}{2}\right) \phi(x) \\ &= \left(\frac{1}{2}\right)^n \sum_{y=0}^{n} \binom{n}{y} \left(\frac{n}{2} - y\right) \phi(n-y) \\ &= \left(\frac{1}{2}\right)^n \sum_{y=0}^{n} \binom{n}{y} \left(\frac{n}{2} - y\right) \phi(y) \\ &= -A, \end{aligned}$$

where we made the substitution $y = n - x$, and used the fact that both $\binom{n}{\cdot}$ and $\phi(\cdot)$ are symmetric about $k = \frac{n}{2}$. We see that $A = -A$, which implies that $A = 0$. □

Proof of Proposition 2.21. Call $p^*(\theta_0, Z)$ the output of Algorithm 3. First we will understand the distribution of $p^*(\theta_0, Z)$ when $\theta = \theta_0$:

$$\begin{aligned} p^*(\theta_0, Z) &= p(\theta_0, T + n\theta_0) + 1 - p(\theta_0, n\theta_0 - T) \\ &= P_{Z\sim\theta_0}(Z \geq T + n\theta_0) + P_{Z\sim\theta_0}(Z \leq n\theta_0 - T) \\ &= P_{Z\sim\theta_0}(Z - n\theta_0 \geq T \text{ or } Z - n\theta_0 \leq -T) \\ &= P_{Z\sim\theta_0}(|Z - n\theta_0| \geq T) \\ &= 1 - F_T(T) \\ &\sim U(0,1). \end{aligned}$$

Since $p^*(\theta_0, Z) \sim U(0,1)$, we have that $P_{\theta_0}(p^*(\theta_0, Z) \geq \alpha) = \alpha$. The p-value satisfies (ϵ, δ)-DP since it is a post-processing of Z. □

Proof of Proposition 2.22. In the proof of Theorem 2.18, we saw that if ϕ is of the form in Theorem 2.18 and $\mathbb{E}_{\theta_0}(X - n\theta_0)\phi(X) = 0$, then ϕ is unbiased. Let ϕ be the test in Proposition 2.21. Then it suffices to show that $\lim_{n\to\infty} \mathbb{E}_{\theta_0} \frac{X - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}} \frac{\phi(X)}{\sqrt{n}} = 0$. We begin by recalling that if $X \sim \text{Binom}(n, \theta_0)$ then by the Central Limit Theorem, we have that

$$\frac{X - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}} \xrightarrow{d} N(0,1).$$

Using this we have

$$\begin{aligned} \lim_{n\to\infty} \mathbb{E}_{\theta_0} \frac{X - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}} \frac{\phi(X)}{\sqrt{n}} &= \lim_{n\to\infty} \mathbb{E}_{Z\sim N(0,1)} Z \frac{\phi(Z\sqrt{n\theta_0(1-\theta_0)} + n\theta_0)}{\sqrt{n}} \\ &= \lim_{n\to\infty} \mathbb{E}_{Z\sim N(0,1)} Z\phi'(Z), \end{aligned}$$

where

$$\phi'(z) = \frac{1}{\sqrt{n}}\phi(z\sqrt{n\theta_0(1-\theta_0)} + n\theta_0) = \frac{1}{\sqrt{n}} \begin{cases} F_N(z\sqrt{n\theta_0(1-\theta_0)} - m) & \text{if } z \geq 0 \\ F_N(-z\sqrt{n\theta_0(1-\theta_0)} - m) & \text{if } z < 0, \end{cases}$$

(we assume that $k = n\theta_0$). Notice that ϕ' is symmetric about 0. So,

$$\begin{aligned}\mathbb{E}_{Z\sim N(0,1)} Z\phi'(Z) &= \mathbb{E}(Z\phi'(Z)I(Z \leq 0) + \mathbb{E}(Z\phi'(Z)I(Z \geq 0)) \\ &= -\mathbb{E}(Z\phi'(Z)I(\geq 0) + \mathbb{E}(Z\phi'(Z)I(Z \geq 0)) \\ &= 0.\end{aligned}$$

□

Proof of Theorem 3.5. It is easy to verify that C^* is unbiased, and has the appropriate coverage. Suppose to the contrary that C^* is not UMA among DP unbiased confidence intervals. Then there exists two values $\theta \neq \theta_0$ and another DP confidence interval C', which is unbiased with coverage $(1-\alpha)$ such that

$$P_\theta(\theta_0 \in C') < P_\theta(\theta_0 \in C^*),$$

or equivalently,

$$P_\theta(\theta_0 \notin C') > P_\theta(\theta_0 \notin C^*). \tag{A.3}$$

At this point, note that $\phi'(x) = P(\theta_0 \notin C' \mid X)$ is a size α, unbiased DP test for $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Furthermore, $\phi^*(x) = P(\theta_0 \notin C^* \mid X)$ is the size α DP-UMPU test from Theorem 2.18. But then equation (A.3) is equivalent to $\mathbb{E}_\theta \phi' > \mathbb{E}_\theta \phi^*$, which implies that ϕ^* is not the DP-UMPU. □

Proof of Theorem 3.2. Releasing C^*_α satisfies (ϵ, δ)-DP by the post-processing property of DP. The object C^*_α is of the form $[L, 1]$ by the monotonicity of $p(\theta_0, Z)$ in θ_0 (for a fixed Z). The coverage of C^*_α is $1-\alpha$, by the fact that $p(\theta_0, Z)$ is a p-value.

Next we check that C^*_α is in fact UMA. Suppose to the contrary that there exists another DP test C'_α with coverage $1-\alpha$, and there exists $\theta_0 < \theta_1$ such that $P_{\theta_1}(\theta_0 \in C'_\alpha) < P_{\theta_1}(\theta_0 \in C^*_\alpha)$, which is equivalent to

$$P_{\theta_1}(\theta_0 \notin C'_\alpha) > P_{\theta_1}(\theta_0 \notin C^*_\alpha). \tag{A.4}$$

Notice that $\phi^*(x) = P(\theta_0 \notin C^*_\alpha \mid x)$ is the UMP size α test from Lemma 2.9 for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, and $\phi'(x) = P(\theta_0 \notin C'_\alpha)$ is another test, which is also level α for the same test. Observe that ϕ' satisfies (ϵ, δ)-DP since it outputs 'Reject' if and only if $I(\theta_0 \notin C'_\alpha) = 1$, which is a post-processing of the DP confidence interval C'_α.

Now, note that (A.4) can be equivalently expressed as $\mathbb{E}_{\theta_1}\phi' > \mathbb{E}_{\theta_1}\phi^*$, which says that ϕ' has more power at θ_1 than ϕ^*, which contradicts that ϕ^* is UMP-α among $\mathscr{D}^n_{\epsilon,\delta}$. We conclude that C^*_α is UMA.

□

Proof of Corollary 3.8. Suppose to the contrary that there exist $\theta_0 \neq \theta_1$ and a (ϵ, δ)-DP confidence interval C' with coverage $1 - \alpha/2$ such that

$$P_{\theta_1}(\theta_0 \in C') < P_{\theta_1}(\theta_0 \in C^1_\alpha),$$

or equivalently,

$$P_{\theta_1}(\theta_0 \notin C') > P_{\theta_1}(\theta_0 \notin C^1_\alpha). \tag{A.5}$$

We can then construct two hypothesis tests $\phi^1(x) = P(\theta_0 \notin C^1_\alpha(x) \mid x)$, and $\phi'(x) = P(\theta_0 \notin C'(x) \mid x)$ for $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Note that $\phi^1(x)$ is the test from Proposition 2.17 at size α, and ϕ' has size $\alpha/2$. Now, (A.5) implies that $\mathbb{E}_{\theta_1}\phi' > \mathbb{E}_{\theta_1}\phi^1$, which contradicts Proposition 2.17, which states that ϕ^1 must be uniformly more powerful than ϕ'. □