Journal of Machine Learning Research 26 (2025) 1-79 Submitted 11/24; Revised 9/25; Published 11/25

Minimax Optimal Two-Sample Testing under Local Differential Privacy

Jongmin Mun jongmin.mun@marshall.usc.edu
Data Sciences and Operations Department, Marshall School of Business
University of Southern California
Los Angeles, CA 90089, USA

Seungwoo Kwak swkwak@skhu.ac.kr
Division of Future Convergence
Sungkonghoe University
Guro-gu, Seoul, 08359, Republic of Korea

Ilmun Kim ilmunk@kaist.ac.kr
Department of Mathematical Sciences
Korea Advanced Institute of Science & Technology
Yuseong-gu, Daejeon, 34141, Republic of Korea

Editor: Mahdi Soltanolkotabi

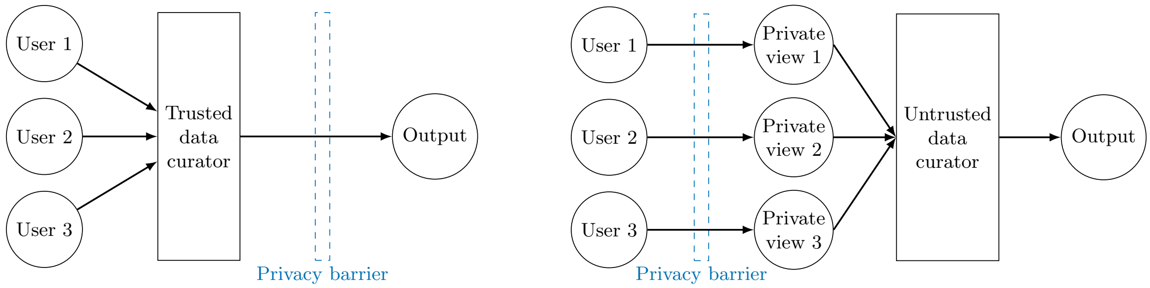
Abstract

We explore the trade-off between privacy and statistical utility in private two-sample testing under local differential privacy (LDP) for both multinomial and continuous data. We begin with the multinomial case, where we introduce private permutation tests using practical privacy mechanisms such as Laplace, discrete Laplace, and Google's `RAPPOR`. We then extend this approach to continuous data via binning and study its uniform separation under LDP over Hölder and Besov smoothness classes. The proposed tests for both discrete and continuous cases rigorously control type I error for any finite sample size, strictly adhere to LDP constraints, and achieve minimax optimality under LDP. The attained minimax rates reveal inherent privacy-utility trade-offs that are unavoidable in private testing. To address scenarios with unknown smoothness parameters in density testing, we propose a Bonferroni-type adaptive test that ensures robust performance without prior knowledge of the smoothness parameters. We validate our theoretical findings with extensive numerical experiments and demonstrate the practical relevance and effectiveness of our proposed methods.

Keywords: local differential privacy, minimax optimality, permutation tests, U-statistics, two-sample testing

1. Introduction

Large-scale internet services such as Netflix and Amazon collect sensitive data from massive user bases, enabling cost-effective randomized experiments by assigning users to two different interfaces or marketing campaigns. By testing whether the resulting two independent sets of samples originate from the same distribution (a procedure known as A/B testing or two-sample testing), companies can statistically assess the impact of new designs on user behavior. However, the sensitivity of detailed personal data raises substantial privacy con-



(a) Central differential privacy **(b)** Local differential privacy

Figure 1: Graphical illustration of central differential privacy and local differential privacy.

cerns. Since privacy protection inherently conceals information and thereby compromises statistical utility, it is crucial to characterize and balance the trade-off between statistical utility and data privacy. Differential privacy (DP; Dwork et al., 2006) provides a rigorous framework for this trade-off, defining data privacy as a mathematical concept that supports such balancing.

We briefly review two notions of DP: central DP and local DP (LDP; Kasiviswanathan et al., 2011). The central DP constraint, illustrated in Figure 1(a), assumes that a trusted data curator (or distributor) has access to the entire original data set and computes a noisy statistical result. This centralized approach requires that the probability of any event based on the noisy result remains essentially the same when a single data entry is arbitrarily perturbed. Under this constraint, one cannot reliably extract any individual-level information from the noisy statistical result. In contrast, under the local DP constraint, illustrated in Figure 1(b), data owners do not trust the curator. Instead, each data owner independently reports a noisy version of their data (see Definition 1 for the formal definition). This stricter local constraint inevitably reduces statistical utility compared with the central constraint. However, it provides a higher level of privacy and separates the data curator from the responsibility for disclosure risk.

These benefits have driven widespread adoption of the LDP framework in Internet-scale data analysis. Examples include Apple's iOS and MacOS diagnostics (Apple Differential Privacy Team, 2017), Google's `RAPPOR` on their Chrome browser (Erlingsson et al., 2014), and Microsoft's telemetry data in Windows 10 Fall Creators Update (Ding et al., 2017). With massive user bases, these companies require stringent privacy protections, while also having the capacity to obtain large samples that enable statistically meaningful analyses under strong privacy constraints. Industrial implementations of LDP have naturally spurred a substantial body of theoretical research (Acharya et al., 2021d; Duchi et al., 2018; Lam-Weil et al., 2022; Lalanne et al., 2023; Cai et al., 2021), examining the intrinsic trade-off between privacy and statistical utility across different problems under LDP. Our paper contributes to this literature by exploring the trade-off in the context of the two-sample testing problem.

The two-sample testing problem, which originates from the classical two-sample t-test (Student, 1908), has gained renewed interest in recent years due to the emergence of high-dimensional and complex data. Notably, several novel methodological approaches have been developed, including kernel-based tests (Gretton et al., 2009, 2012), distance-based tests (Székely and Rizzo, 2004, 2005) and regression/classification-based tests (Kim et al., 2019, 2021), which demonstrate promising capabilities for modern data sets. On the theoretical front, researchers have explored the fundamental limits of this problem through minimax analysis both in the statistics (for example, Arias-Castro et al., 2018; Kim et al., 2022a; Schrab et al., 2023) and computer science (for example, Batu et al., 2000; Chan et al., 2014; Diakonikolas and Kane, 2016; Goldreich and Ron, 2011) literature. In addition to methodological and theoretical advances, two-sample testing has found contemporary applications in areas such as education research (Rabin et al., 2019), network traffic analysis (Kohout and Pevný, 2018), and audio segmentation (Harchaoui et al., 2009). Despite its long history and fundamental roles in practice, most existing work on two-sample testing has focused on non-private settings, with a few exceptions. Exceptions include private versions of multinomial tests (Acharya et al., 2018; Aliakbarpour et al., 2019, 2018), traditional nonparametric tests (for example, Mann-Whitney and Wilcoxon signed-rank tests; Couch et al., 2019; Task and Clifton, 2016), partitioning-based test of univariate continuous distributions (Sheffet and Omer, 2024) and kernel tests (Raj et al., 2020; Kim and Schrab, 2023) under central DP. The literature on two-sample testing under LDP is even more scarce. Notable contributions in this area include Ding et al. (2018) and Waudby-Smith et al. (2023), both primarily focused on detecting differences in location. In contrast, our primary goal is to develop two-sample tests for general alternatives under LDP, focusing on both multinomial and multivariate continuous data. Additionally, we shed new light on the fundamental limits of the two-sample problem under LDP.

In the following subsection, we first review related work and then present our techniques and contributions.

1.1 Related Prior Work

Private hypothesis testing has been extensively studied in the statistics and computer science literature. Among these studies, we briefly review those most closely related to our work. Initially motivated by privacy attacks on genome-wide association studies (GWAS) (Homer et al., 2008), the early work on private testing mainly focused on private versions of chi-square tests and explored their asymptotic properties (Gaboardi et al., 2016; Gaboardi and Rogers, 2018; Johnson and Shmatikov, 2013; Rogers and Kifer, 2017; Uhler et al., 2013; Vu and Slavkovic, 2009; Wang et al., 2015; Yu et al., 2014). In contrast, a more recent line of work in computer science community studied non-asymptotic properties of private tests designed for multinomial data sets, and investigated optimal sample complexities of testing problems from a minimax perspective. This line of work was initiated by Cai et al. (2017) for central DP and Sheffet (2018) for LDP, and continued by Acharya et al. (2018), Aliakbarpour et al. (2019, 2018), and Sheffet and Omer (2024) for central DP and Acharya et al. (2020, 2021b) for LDP, respectively. Optimal sample complexity is typically established through systematic analyses of both upper and lower bounds. For the upper bounds, previous works extended non-private multinomial tests (Acharya et al., 2015; Chan et al., 2014; Diakonikolas et al., 2018; Diakonikolas and Kane, 2016; Goldreich and Ron, 2011; Valiant and Valiant, 2014) to private counterparts by incorporating randomization mechanisms. Lower bound analysis relied on information-theoretic techniques, such as Le Cam's method, treating the DP requirement as information constraints (see, for example, Acharya et al., 2020; Duchi et al., 2018, for detailed discussions).

A more recent line of work in statistics community explored univariate goodness-of-fit testing under LDP. Specifically, Dubois et al. (2023) proposed minimax optimal goodness-of-fit tests for Hölder densities under LDP. Lam-Weil et al. (2022) also considered goodness-of-fit testing under LDP, and developed minimax optimal tests for multinomials and for continuous densities over Besov balls. Our work builds on their framework and extends the focus from goodness-of-fit testing to two-sample testing for both (i) multinomials and (ii) multivariate Hölder and Besov densities. It is worth highlighting that previous works in this line (Lam-Weil et al., 2022; Dubois et al., 2023) relied solely on the Laplace mechanism (Dwork and Roth, 2013) to establish their theoretical results. In contrast, we explore various LDP mechanisms that achieve similar optimality properties and empirically demonstrate that the Laplace mechanism can underperform in practical scenarios. Specifically, we delve into the Google's `RAPPOR` (Erlingsson et al., 2014), generalized randomized response (Gaboardi and Rogers, 2018), and (discrete) Laplace mechanisms (Ghosh et al., 2009), illustrating their theoretical and empirical performance.

1.2 Techniques and Results

Previous work on hypothesis testing under local differential privacy primarily focused on goodness-of-fit testing (Dubois et al., 2023; Lam-Weil et al., 2022). We instead target a broader and arguably more challenging setting: two-sample testing of multinomials and multivariate densities. In particular, we provide testing methods that are both practically reliable and theoretically optimal. Our practical reliability stems from both the privatization mechanism and the testing procedure. For privatization, one of our methods leverages Google's `RAPPOR`, a widely adopted open-source privacy mechanism which has demonstrated effectiveness through years of large-scale deployment in the Chrome browser. Although prior works (Duchi et al., 2013; Acharya et al., 2019, 2021c) analyzed `RAPPOR`'s statistical performance under the minimax framework, their focus is limited to ℓ_1 separation and multinomial data. Our work expands upon them by establishing minimax optimality for both multivariate continuous and multinomial data under ℓ_2 separation, marking the first result of its kind. For the testing procedure, our methods rigorously control type I error in all scenarios. At the heart of achieving the blend of practicality and theory lies the permutation test. Practical reliability of a test partly depends on the accurate calibration of the critical value within the non-asymptotic regime, taking into account the randomization effects introduced by local differential privacy. Under a composite null hypothesis of two-sample testing, critical values cannot be determined through Monte-Carlo-approximated population quantile of test statistics, unlike in goodness-of-fit testing. Critical values obtained via concentration inequalities, on the other hand, typically depend on unspecified constants and thus are not reliable in practice. By employing a permutation procedure, our tests guarantee type I error control at any sample size and with a sufficiently large number of permutations (which does not depend on the sample size of data dimension). The permutation procedure also facilitates theoretical analysis of power. In particular, it allows us to leverage the technique of Kim et al. (2022a), namely the two moments method therein. This technique allows us to avoid directly analyzing the permutation distribution under LDP, and provides a sufficient condition for type II error control based solely on the first two moments of the test statistic. Equipped with this tool, we analyze our test statistics, which are U-statistics derived from

	Non-private rate	Private rate under LDP
Testing for multinomials in ℓ_2 separation	$n_1^{-1/2}$ (Chan et al., 2014; Kim et al., 2022a)	$\frac{k^{1/4}}{(n_1\alpha^2)^{1/2}} \vee n_1^{-1/2}$ (Theorem 3)
Testing for Hölder and Besov densities in $\mathbb{L}_2$ separation	$n_1^{\frac{-2s}{4s+d}}$ (Arias-Castro et al., 2018)	$(n_1\alpha^2)^{\frac{-2s}{4s+3d}} \vee n_1^{\frac{-2s}{4s+d}}$ (Theorem 11)

Table 1: Non-private and private minimax testing separations for two-sample multinomial and density testing in ℓ_2 and $\mathbb{L}_2$ separation where n_1 denotes the minimum sample size and α denotes the privacy level. For multinomial testing, k stands for the number of categories, and for density testing, s represents the smoothness parameter. The rates exhibit phase transitions at specific levels of privacy (elbow effects). See Sections 3 and 4 for details.

perturbed data with a carefully selected perturbation level. A substantial portion of our technical effort is dedicated to bounding the moments related to the U-statistic under the data perturbation.

By obtaining matching information-theoretic lower bounds, we establish the minimax optimality of our methods and gain insight into the fundamental trade-off between statistical power and privacy. For the lower bound analysis, we leverage a recently developed technique by Lam-Weil et al. (2022). This technique builds upon Ingster's method (Ingster, 1993), a classical approach for deriving minimax separation rates in testing problems, and adapts it to incorporate the LDP constraint. Central to Ingster's method is bounding the chi-square divergence between a simple null distribution and a mixture of alternative distributions. To achieve a tight lower bound under LDP, we construct the mixture distribution using the singular values and singular vectors of the privacy mechanism. Such construction naturally imposes extra restrictions caused by the LDP constraint, enabling a tight lower bound. Our technical effort lies in extending the univariate result of Lam-Weil et al. (2022) to more general settings, including the multivariate Hölder ball and Besov ball.

Summary of our contributions. We highlight our contributions and contrast them with prior work as follows. We also refer readers to Table 1, which summarizes the non-private and private minimax testing separation for two-sample testing, derived from both prior work and our findings.

- **Optimal multinomial testing under LDP (Theorem 3):** We start by developing a private two-sample test for multinomials and present minimax separation in terms of the ℓ_2 distance under LDP. The prior work (Acharya et al., 2018; Aliakbarpour et al., 2019, 2018) on private two-sample testing generally focused on central DP and imposed conditions such as equal sample sizes and Poisson sampling that may not be practically relevant. In contrast, our approach does not rely on such assumptions and achieves minimax optimality under more practical settings. We also highlight that our upper bound result is established using three distinct LDP mechanisms, namely

Laplace, discrete Laplace and `RAPPOR` mechanisms, thereby diversifying the practical toolkit. As mentioned earlier, this contrasts with prior work (Lam-Weil et al., 2022; Dubois et al., 2023), which primarily focused on the Laplace mechanism. Moreover, we show that the generalized randomized response mechanism (Gaboardi and Rogers, 2018) can lead to suboptimal power in Appendix H.

- **Optimal density testing under LDP (Theorem 11):** We next consider the two-sample problem for multivariate continuous data and derive $\mathbb{L}_2$ minimax separation under LDP, by leveraging the prior work (Lam-Weil et al., 2022; Kim et al., 2022a). In particular, we examine both Hölder and Besov smoothness classes, and show that the proposed private test is minimax optimal for both classes with the finite-sample validity. This approach differs from the prior work on a similar topic. For instance, unlike Sheffet and Omer (2024) that consider central DP with Poissonization, we focus on the more stringent setting of LDP and consider the standard sampling with fixed sample sizes. Moreover, in contrast to the prior work under LDP (Ding et al., 2018; Waudby-Smith et al., 2023), which primarily focused on location alternatives, our private test is sensitive against a broad range of nonparametric alternatives. Lastly, we highlight that our method controls the type I error for any finite sample size, and exactly satisfies the LDP condition, distinguishing it from the prior work of Raj et al. (2020).

- **Adaptive density testing under LDP (Theorem 12):** Similar to other nonparametric density testing methods, the optimality of our proposed density test relies on the knowledge of the underlying smoothness parameter, which is typically unknown. To tackle this issue, we introduce a Bonferroni-type approach that adapts to the unknown smoothness parameter at the expense of extra logarithmic factors in the separation. The proof of the adaptation result leverages the exponential inequality for the permuted U-statistic (Kim et al., 2022a). As a result, our adaptive test improves upon the adaptive goodness-of-fit test of Lam-Weil et al. (2022), which resorts to a simple upper bound for the variance of the U-statistic along with Chebyshev's inequality.

- **Numerical validation (Section 5):** Lastly, we assess the empirical performance of the proposed tests under various scenarios and showcase the trade-off between privacy and utility through numerical simulations. It is important to emphasize that previous research on private testing has primarily centered on theoretical optimality, often lacking empirical validation of their findings. We address this gap by complementing theoretical justifications with empirical evaluation, thereby enhancing practical relevance. Since no previous methods exist for two-sample testing for multinomials or densities under LDP, we create baseline methods by extending one-sample LDP χ^2-tests (Gaboardi and Rogers, 2018) to the two-sample problem (Appendix G.1), and compare their empirical performance with our main proposals. To facilitate adoption, we provide a Python package `privateAB` that implements all proposed and baseline methods, available at `https://pypi.org/project/privateAB/0.0.2/`.

1.3 Notation

Throughout this paper, real numbers are represented by lowercase, non-bold letters, such as a, while vectors in $\mathbb{R}^d$ for $d \geq 2$ are written in boldface lowercase, such as $\mathbf{a}$. Constant vectors are denoted using bold numerals, such as $\mathbf{1}$ and $\mathbf{0}$. The ith element of $\mathbf{a}$ is denoted by a_i. For indexed vector such as $\mathbf{a}_j$, its ith element is denoted as a_{ji}. Unless otherwise specified, random variables are written in uppercase non-bold (for example, X), while random vectors use bold uppercase (for example, $\mathbf{X}$). The ith element of $\mathbf{X}$ and $\mathbf{X}_j$ are denoted as X_i and X_{ji}, respectively. The set of non-negative integers is denoted by $\mathbb{N}_0 := \{0, 1, 2, \ldots\}$. For positive integers u and v, $[u]$ represents $\{1, \ldots, u\}$, and $[u]^v$ denotes its Cartesian product taken v times. A set of u elements indexed by i is written as $\{a_i\}_{i \in [u]} := \{a_1, \ldots, a_u\}$. For any real $s > 0$, $\lfloor s \rfloor$ denotes the largest integer strictly smaller than s. For $a, b \in \mathbb{R}$, we define $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$. Given $\mathbf{w} \in \mathbb{R}^d$ and $p \geq 1$, its ℓ_p-norm and ℓ_∞-norm are defined by $\|\mathbf{w}\|_p := \big(\sum_{i=1}^d |w_i|^p\big)^{1/p}$ and $\|\mathbf{w}\|_\infty := \max\big\{|w_1|, \ldots, |w_d|\big\}$, respectively. For a function supported on $[0,1]^d$, its $\mathbb{L}_p$-norm and sup-norm are defined as follows:

$$|||f|||_{\mathbb{L}_p} := \left(\int_{[0,1]^d} |f(\mathbf{x})|^p d\mathbf{x} \right)^{1/p}, \quad 1 \leq p < \infty, \quad |||f|||_\infty := \sup \Big\{ |f(\mathbf{x})| : \mathbf{x} \in [0,1]^d \Big\}.$$

For any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$, with $u_j \leq v_j$ for $j = 1, \ldots, d$, a hyperrectangle $[\mathbf{u}, \mathbf{v}]$ is defined as follows:

$$[\mathbf{u}, \mathbf{v}] := \prod_{j=1}^d [u_j, v_j].$$

Given the privacy parameter $\alpha > 0$, we write $z_\alpha := e^{2\alpha} - e^{-2\alpha} = 2\sinh(2\alpha)$. A constant that only depends on parameters $\theta_1, \theta_2, \ldots$ is denoted as $C(\theta_1, \theta_2, \ldots)$. The indicator function $\mathbb{1}(\mathcal{A})$ takes value 1 if the event $\mathcal{A}$ is true and 0 otherwise.

1.4 Outline of the Paper

The remainder of the paper is organized as follows: Section 2 introduces the necessary background on LDP, the minimax framework, and the permutation procedure. Section 3 illustrates the minimax analysis for multinomial testing under LDP and the optimal permutation testing procedure. Building on this result, Section 4 presents the minimax analysis for multivariate two-sample density testing under LDP and an optimal permutation testing procedure. Finally, Section 5 presents numerical validation of the procedures proposed in Sections 3 and 4. All proofs and additional simulation results are deferred to the appendix.

2. Background

This section introduces the notion of local differential privacy and two-sample testing under LDP. We then explain the minimax framework for two-sample testing under LDP, along with the permutation test procedure considered throughout this paper.

2.1 Two-Sample Testing under Local Differential Privacy Constraint

Let $\mathcal{P}$ denote the space of pairs of distributions of interest: multinomial distributions taking values in $[k]$ or continuous distributions taking values in $[0,1]^d$. Given a pair of distributions $(P_Y, P_Z) \in \mathcal{P}$, for each $i \in [n_1]$, the ith data owner draws Y_i from P_Y independently from the others. Similarly, for each $j \in [n_2]$, the jth data owner independently draws Z_j from P_Z. We allow the sample sizes n_1 and n_2 to differ, and assume $n_1 \leq n_2$, without loss of generality, throughout this paper. We denote the pooled sample size as $n := n_1 + n_2$. Under the LDP constraint, each owner releases only a randomized transformation of their raw sample, formally referred to as an LDP view:

Definition 1 (Local differential privacy) *Given a privacy level $\alpha > 0$, let X_i and $\tilde{X}_i$ be random elements mapped to measurable spaces $(\mathcal{X}, \mathcal{F})$ and $(\tilde{\mathcal{X}}_i, \tilde{\mathcal{F}}_i)$, respectively, for each $i \in [n]$. Then $\tilde{X}_i$ is an α-local differentially private (α-LDP) view of X_i if there exists a bivariate function $Q_i(\cdot \,|\, \cdot) : \tilde{\mathcal{F}}_i \times \mathcal{X} \mapsto [0,1]$ such that:*

1. *For any $x \in \mathcal{X}$, $Q_i(\cdot \,|\, x)$ is a conditional distribution of $\tilde{X}_i$ given $X_i = x$,*
2. *For any $A \in \tilde{\mathcal{F}}_i$, $x \mapsto Q_i(A \,|\, x)$ is a measurable function on $\mathcal{X}$, and*
3. *For any $x, x' \in \mathcal{X} \times \mathcal{X}$ and $A \in \tilde{\mathcal{F}}_i$, the inequality $Q_i(A \,|\, x) \leq e^{\alpha} Q_i(A \,|\, x')$ holds.*

Let $\mathcal{Q}_\alpha$ denote the collection of function sets $\{Q_i\}_{i=1}^n$ satisfying the above properties. Then $Q \in \mathcal{Q}_\alpha$ is called an α-LDP mechanism (or channel) associated with $\{X_i\}_{i=1}^n$.

The curator, aware of the privacy level α and the mechanism Q, only receives the α-LDP views $\{\tilde{X}_i\}_{i\in[n]}$, consisting of $\{\tilde{Y}_i\}_{i\in[n_1]}$ and $\{\tilde{Z}_j\}_{j\in[n_2]}$, and uses them to decide whether

$$H_0 : P_Y = P_Z \quad \text{or} \quad H_1 : P_Y \neq P_Z. \tag{1}$$

The definition of LDP above is non-interactive in a sense that the ith conditional distribution Q_i is assumed to be independent of other private views $\tilde{X}_1, \ldots, \tilde{X}_{i-1}, \tilde{X}_{i+1}, \ldots, \tilde{X}_n$. It has been pointed out that allowing Q_i to be interactive with private views $\{\tilde{X}_{i'}\}_{i'\in[n]\setminus i}$ can yield more efficient statistical procedures (for example, Acharya et al., 2021a; Berrett and Butucea, 2020; Kasiviswanathan et al., 2011). However, the non-interactive approach requires less communication, making it more suitable for large-scale inference than the interactive alternative (Berrett et al., 2021; Joseph et al., 2019). Accordingly, we focus on the non-interactive privacy mechanisms throughout this paper.

Unless otherwise stated, all expectations and variances in this paper are taken with respect to the distributions of the α-LDP views. Formally, for $i \in [n_1]$, $j \in [n_2]$, $A \in \tilde{\mathcal{F}}_i$, and $B \in \tilde{\mathcal{F}}_j$, we define the distributions of the α-LDP views as:

$$P_{\tilde{Y}_i}(A) := \int_{\mathcal{X}_i} Q_i(A \,|\, y) P_Y(dy) \quad \text{and} \quad P_{\tilde{Z}_j}(B) := \int_{\mathcal{X}_j} Q_i(B \,|\, z) P_Z(dz),$$

where the distributions of the raw samples Y_i and Z_j are marginalized out.

2.2 Non-Private and LDP Minimax Framework for Two-Sample Testing

Let $\mathcal{P}_0$ and $\mathcal{P}_1$ denote the collections of null and alternative distributions, respectively, corresponding to H_0 and H_1 introduced in (1). In the minimax framework, we focus on a subset of $\mathcal{P}_1$, denoted as $\mathcal{P}_1(\rho_{n_1,n_2})$, where ρ_{n_1,n_2} indicates a minimum separation between $\mathcal{P}_1(\rho_{n_1,n_2})$ and $\mathcal{P}_0$.

Given a privacy mechanism Q, the curator evaluates a test function $\Delta_Q : \prod_{i=1}^{n} \tilde{\mathcal{X}}_i \mapsto \{0,1\}$ on the α-LDP views and rejects H_0 if $\Delta_Q = 1$ and accepts H_0 otherwise. Our objective is to design the private test Δ_Q that controls the type I and II errors uniformly over distributions in $\mathcal{P}_0$ and $\mathcal{P}_1(\rho_{n_1,n_2})$, respectively. In particular, for fixed $\gamma, \beta \in (0,1)$, we aim for the private test Δ_Q to satisfy the following conditions:

$$\text{Type I error:} \sup_{(P_Y,P_Z)\in\mathcal{P}_0} \mathbb{E}[\Delta_Q] \le \gamma \quad \text{and} \quad \text{Type II error:} \sup_{(P_Y,P_Z)\in\mathcal{P}_1(\rho_{n_1,n_2})} \mathbb{E}[1-\Delta_Q] \le \beta. \tag{2}$$

Let Φ_γ^α be the set of α-LDP level-γ tests, which take α-LDP views as inputs and control the type I error as in (2). The quality of a level-γ test $\Delta_{Q,\gamma} \in \Phi_\gamma^\alpha$ is measured by its uniform separation, which quantifies how close two hypotheses can be while still being reliably distinguished by $\Delta_{Q,\gamma}$. In more technical terms, the uniform separation $\tilde{\rho}_{n_1,n_2}(\Delta_{Q,\gamma})$ is the smallest separation ρ_{n_1,n_2} which accomplishes the type II error control as in (2), namely,

$$\tilde{\rho}_{n_1,n_2}(\Delta_{Q,\gamma}) := \inf\Big\{\rho_{n_1,n_2} > 0 : \sup_{(P_Y,P_Z)\in\mathcal{P}_1(\rho_{n_1,n_2})} \mathbb{E}[1-\Delta_{Q,\gamma}] \le \beta\Big\}. \tag{3}$$

Since different α-LDP mechanism Q may assume different $\tilde{\mathcal{X}}_i$'s, a test $\Delta_{Q,\gamma}$ depends on a particular LDP mechanism Q. Therefore, an optimal private level-γ test can be described as a pair of an α-LDP mechanism and a test function, designed to achieve the minimal uniform separation (3). We define this minimal uniform separation as the optimal testing separation under LDP, which we refer to as the minimax testing separation:

Definition 2 (α-LDP non-asymptotic minimax testing separation) *For a fixed privacy level $\alpha > 0$, the α-LDP non-asymptotic minimax separation of testing is defined as*

$$\rho^*_{n_1,n_2,\alpha} := \inf_{Q\in\mathcal{Q}_\alpha} \inf_{\Delta_{Q,\gamma}\in\Phi_\gamma^\alpha} \tilde{\rho}_{n_1,n_2}(\Delta_{Q,\gamma}). \tag{4}$$

We aim to quantify the price to pay for privacy by comparing the private minimax separation (4) with the non-private (unconstrained) one $\rho^*_{n_1,n_2} := \inf_{\Delta_\gamma\in\Phi_\gamma} \tilde{\rho}_{n_1,n_2}(\Delta_\gamma)$, where Φ_γ denotes the set of all level-γ tests without privacy constraints. Since obtaining a test achieving the exact minimax separation is mostly infeasible in a nonparametric setting, we follow the convention (Ingster, 1994, 2000; Baraud, 2002) and focus on minimax *rate* optimality. In particular, a test is minimax rate optimal if its uniform separation is upper bounded by the minimax separation up to some constant.

It is worth noting that the minimax separation depends crucially on the choice of distance used to distinguish between the null from the alternative. In this paper, we adopt the ℓ_2 distance for discrete testing and the $\mathbb{L}_2$ distance for continuous testing. The ℓ_2 (and $\mathbb{L}_2$) distance is widely used in the literature because of its mathematical tractability and interpretability, and our results build on prior work in this direction (Lam-Weil et al., 2022; Kim et al., 2022a). Another commonly studied metric in minimax testing is the ℓ_1 (total variation) distance, which enjoys appealing properties such as invariance under bijective transformations (Balakrishnan and Wasserman, 2019). However, establishing minimax separation with respect to the ℓ_1 distance requires a more delicate bias–variance trade-off, as there is no natural unbiased estimator of the ℓ_1 distance. Subsequent to our work, Kent et al. (2025) explored this topic in the context of locally private two-sample testing, demonstrating that minimax separations under ℓ_1 and ℓ_2 distances differ for multinomial distributions but align in the continuous setting.

2.3 Permutation Testing Procedure

The permutation test is a simple yet powerful method to calibrate a test statistic, yielding a valid level-γ test under exchangeability of the samples, meaning that when the distribution of the permuted samples is the same as the distribution of the original samples. Note that some α-LDP mechanisms might violate exchangeability even when the raw samples are i.i.d. In order to guarantee exchangeability, throughout this paper, we only consider α-LDP mechanisms with identical marginals. Below, we briefly explain the permutation test in the two-sample setting.

Let $\mathbb{X}_n$ denote the pooled sample $\{\xi_1,\ldots,\xi_n\} := \{Y_1,\ldots,Y_{n_1},Z_1,\ldots,Z_{n_2}\}$. Let $\boldsymbol{\pi}_n$ be the set of all possible permutations of $[n]$, and denote its cardinality as $|\boldsymbol{\pi}_n|$. Given a permutation $\pi := (\pi_1,\ldots,\pi_n)$ sampled from a uniform distribution over $\boldsymbol{\pi}_n$, the permuted version of $\mathbb{X}_n$ is denoted as $\mathbb{X}_n^\pi := \{\xi_{\pi_1},\ldots,\xi_{\pi_n}\}$. For a two-sample statistic $T_n(\mathbb{X}_n)$, its permutation distribution function conditional on $\mathbb{X}_n$ is $F_T(t) := \sum_{\pi\in\boldsymbol{\pi}_n} \mathbb{1}\big(T(\mathbb{X}_n^\pi) \le t\big)/|\boldsymbol{\pi}_n|$. The permutation testing procedure rejects the null hypothesis if $T(\mathbb{X}_n) > \inf\{t : F_T(t) \ge 1-\gamma\}$. If the exchangeability assumption of $\mathbb{X}_n \overset{d}{=} \mathbb{X}_n^\pi$ for any $\pi \in \boldsymbol{\pi}_n$ is satisfied under H_0, the resulting test controls the type I error non-asymptotically (see, for example, Ramdas et al., 2023).

In practice, however, it is computationally infeasible to consider all $|\boldsymbol{\pi}_n|$ permutations. Therefore, it is now a standard practice to consider a Monte Carlo-based (MC-based) permutation test which uses much smaller number of permutations. To explain, for a given $B \in \mathbb{N}$, let $\pi_1,\ldots,\pi_B$ be B independent random permutations sampled from uniform distribution over $\boldsymbol{\pi}_n$. These permutations are used to calculate the following MC-based permutation p-value:

$$\hat{p}_B := \frac{1}{B+1}\left[1 + \sum_{b=1}^{B} \mathbb{1}\big\{T_n(\mathbb{X}_n^{\pi_b}) \ge T_n(\mathbb{X}_n)\big\}\right], \tag{5}$$

which controls the type I error non-asymptotically under exchangeability for any $B \in \mathbb{N}$ and any test statistic (see, for example, Hemerik and Goeman, 2018; Ramdas et al., 2023, for details). Regarding type II error, the MC-based permutation test using the U-statistic in (10) achieves power comparable to the full permutation test if B is sufficiently large. The required B is independent of sample size and data dimension, and is much smaller than $|\boldsymbol{\pi}_n|$ (see Proposition I.1 of Kim et al., 2022a, for details).

3. Two-Sample Testing for Multinomials under LDP

Having introduced the background, we now turn to the main results of this paper. In Section 3.1, we consider the problem of comparing two multinomial distributions, and establish the corresponding minimax testing separation under the LDP constraint. The upper bound for this separation is achieved by the permutation tests proposed in Section 3.2. These tests and rate play a pivotal role developing minimax optimal tests for multivariate continuous data, as discussed in Section 4.

3.1 Private Minimax Separation for Two-Sample Multinomial Testing

The problem of interest is formulated as follows. Let $\mathcal{P}^{(k)}_{\mathrm{multi}}$ denote the set of pairs of probability vectors with k categories. Suppose the raw sample sets $\{Y_i\}_{i\in[n_1]}$ and $\{Z_j\}_{j\in[n_2]}$ are drawn from multinomial distributions with probability vectors $(\mathbf{p}_Y, \mathbf{p}_Z) \in \mathcal{P}^{(k)}_{\mathrm{multi}}$. The curator receives two sets of α-LDP views, $\{\tilde{Y}_i\}_{i\in[n_1]}$ and $\{\tilde{Z}_j\}_{j\in[n_2]}$, and determines whether $(\mathbf{p}_Y, \mathbf{p}_Z)$ belongs to $\mathcal{P}_{0,\mathrm{multi}} := \{(\mathbf{p}_Y, \mathbf{p}_Z) \in \mathcal{P}^{(k)}_{\mathrm{multi}} : \mathbf{p}_Y = \mathbf{p}_Z\}$ or to the alternative hypothesis set defined as:

$$\mathcal{P}_{1,\mathrm{multi}}(\rho_{n_1,n_2}) := \big\{(\mathbf{p}_Y, \mathbf{p}_Z) \in \mathcal{P}^{(k)}_{\mathrm{multi}} : \|\mathbf{p}_Y - \mathbf{p}_Z\|_2 \geq \rho_{n_1,n_2}\big\}. \tag{6}$$

Let us fix the type I error γ and the type II error β such that $2\gamma + \beta < 1$. This constraint arises from Ingster's minimax lower bounding method, as similarly considered in Lam-Weil et al. (2022). The main result of this section, stated below, establishes a lower bound as well as an upper bound for the minimax separation for this multinomial problem under LDP in terms of the ℓ_2 distance.

Theorem 3 (Minimax separation for multinomial testing) *There exist constants $C_\ell(\gamma,\beta) > 0$ and $C_u(\gamma,\beta) > 0$ such that the α-LDP minimax testing separation $\rho^*_{n_1,n_2,\alpha}$ over the class of alternatives $\mathcal{P}_{1,\mathrm{multi}}$ in (6) is bounded as*

$$C_\ell(\gamma,\beta)\left[\left(\frac{k^{1/4}}{(n_1 z_\alpha^2)^{1/2}} \wedge (k\log k)^{-1/2}\right) \vee \frac{1}{{n_1}^{1/2}}\right] \leq \rho^*_{n_1,n_2,\alpha} \leq C_u(\gamma,\beta)\left[\frac{k^{1/4}}{(n_1\alpha^2)^{1/2}} \vee \frac{1}{{n_1}^{1/2}}\right], \tag{7}$$

where we recall $z_\alpha = 2\sinh(2\alpha)$.

Theorem 3 states that the private minimax separation for two-sample multinomial testing is notably different from its non-private counterpart $n_1^{-1/2}$ with respect to the ℓ_2 distance (Chan et al., 2014; Kim et al., 2022a). In particular, in the high privacy regime, we observe additional dependence on k and α. The result also indicates that the privacy guarantee can be obtained at no additional cost in the low privacy regime where $n_1^{-1/2}$ dominates the other term.

We point out that while the lower bound and the upper bound do not exactly match, the gap is notably small. For instance, in the regime where the $(k\log k)^{-1/2}$ term is negligible, the only different terms are α and z_α. We note that these two terms are the same, up to a constant factor, as long as α is bounded. Hence, in most practical scenarios where a small value of α is of interest, the upper bound matches the lower bound.

We now discuss the proof of Theorem 3, and a detailed analysis can be found in Appendix D.

- **Lower bound.** We can obtain the lower bound almost for free by observing that the two-sample problem is more difficult than the one-sample problem. In particular, one can always turn the one-sample problem into the two-sample problem by drawing additional samples from the target distribution. Therefore, a minimax lower bound for the one-sample problem does not exceed that of the two-sample problem (see Lemma 1 of Arias-Castro et al., 2018, for a formal argument). Given this insight, the lower bound follows by combining lower bound results of Theorem 3.2 in Lam-Weil et al. (2022) for the one-sample problem under LDP, as well as Chan et al. (2014) and Kim et al. (2022a) for the two-sample problem without privacy constraints. We point out that the lower bound in Chan et al. (2014) and Kim et al. (2022a) depends on $\max\{\|\mathbf{p}_Y\|_2, \|\mathbf{p}_Z\|_2\}$. This quantity becomes a constant at the worst case scenario and it can be thereby disregarded in the lower bound for the global minimax separation.

- **Upper bound.** We build on the ℓ_2 permutation test in Kim et al. (2022a), which achieves minimax optimality under non-private setting. The test statistic we consider is essentially a U-statistic of $\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2$ constructed from private views of indicator functions generated by privacy mechanisms. Using this privatized U-statistic detailed in Section 3.2, we apply the two moments method (Kim et al., 2022a), which provides a sufficient condition for significant power of a permutation test based solely on the first two moments of the statistic. The next subsection introduces the privacy mechanisms, specifies the form of the test statistic, and presents the testing procedure that achieves the upper bound.

3.2 Privacy Mechanisms and Testing Procedure for the Upper Bound

This subsection is dedicated to explaining the private permutation testing procedure which achieves the upper bound stated in (7). As mentioned earlier, this procedure builds on the U-statistic (Kim et al., 2022a) estimating the squared ℓ_2 distance $\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2$. In a similar way, we deal with the U-statistic of $k\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2$, but a notable difference is that ours is based on α-LDP views generated by privacy mechanisms. Before formally introducing the test statistic, let us explain the considered LDP mechanisms: Laplace, discrete Laplace, and `RAPPOR` mechanisms.

Privacy mechanisms. The first one is the standard Laplace mechanism, also considered in Lam-Weil et al. (2022) and Berrett and Butucea (2020), which adds independent Laplace noises to the original data represented as indicator variables.

Definition 4 (Laplace mechanism for multinomial data: `LapU`) *Consider a pooled raw multinomial sample $\{X_i\}_{i\in[n]}$ with k categories. Fix the privacy level $\alpha > 0$. Each data owner encodes their data point as a one-hot vector and adds Laplace noises, with the noise variance parameterized by*

$$\sigma_\alpha := \frac{2\sqrt{2k}}{\alpha}.$$

Formally, the resulting locally privatized sample $\{\tilde{\mathbf{X}}_i\}_{i\in[n]} = \{\tilde{\mathbf{Y}}_i\}_{i\in[n_1]} \cup \{\tilde{\mathbf{Z}}_i\}_{i\in[n_2]}$ *is a set of* k*-dimensional random vectors whose* m*th element* $\tilde{X}_{im}$ *is defined as follows:*

$$\tilde{X}_{im} := \sqrt{k}\, \mathbb{1}\big(X_i = m\big) + \sigma_\alpha W_{im}. \tag{8}$$

Here, $\{W_{im}\}_{i\in[n],m\in[k]} \overset{i.i.d.}{\sim} \mathrm{Lap}(1/\sqrt{2})$ *is independent of* $\{X_i\}_{i\in[n]}$*, and* $\mathrm{Lap}(1/\sqrt{2})$ *denotes the centered Laplace distribution with variance one.*

The second mechanism is based on discrete Laplace noise. We say that a random variable W follows a discrete Laplace distribution with parameter $\zeta \in (0,1)$, denoted as $W \sim \mathrm{DL}(\zeta)$, if its probability mass function satisfies

$$\mathbb{P}(W = w) = \frac{1-\zeta}{1+\zeta}\zeta^{|w|}, \quad \text{for all } w \in \mathbb{Z}. \tag{9}$$

The second mechanism is analogous to the first, but it replaces continuous Laplace noise with discrete Laplace noise.

Definition 5 (Discrete Laplace mechanism for multinomial data: `DiscLapU`**)** *Consider a pooled raw multinomial sample* $\{X_i\}_{i\in[n]}$ *with* k *categories. Fix the privacy level* $\alpha > 0$*. Each data owner encodes their data point as a one-hot vector and adds discrete Laplace noises, with the noise distribution parametrized by*

$$\zeta_\alpha := e^{-\frac{\alpha}{2\sqrt{k}}},$$

where $\zeta_\alpha \in (0,1)$ *for any value of* $\alpha > 0$ *and* $k \geq 2$*. Formally, the resulting locally privatized sample* $\{\tilde{\mathbf{X}}_i\}_{i\in[n]} = \{\tilde{\mathbf{Y}}_i\}_{i\in[n_1]} \cup \{\tilde{\mathbf{Z}}_i\}_{i\in[n_2]}$ *is a set of* k*-dimensional random vectors whose* m*th element* $\tilde{X}_{im}$ *is defined as follows:*

$$\tilde{X}_{im} := \sqrt{k}\, \mathbb{1}\big(X_i = m\big) + W_{im}.$$

Here, $\{W_{im}\}_{i\in[n],m\in[k]} \overset{i.i.d.}{\sim} DL(\zeta_\alpha)$*.*

The third mechanism that we consider privatizes multinomial data vectors by randomly flipping individual components, instead of injecting additive random noise. This mechanism, proposed by Duchi et al. (2013), is equivalent to Google's basic one-time RAPPOR (randomized aggregatable privacy-preserving ordinal response; Erlingsson et al., 2014), which is deployed in the Chrome browser.

Definition 6 (Basic one-time RAPPOR mechanism for multinomial data: `RAPPOR`**)** *Consider a pooled raw multinomial sample* $\{X_i\}_{i\in[n]}$ *with* k *categories. Fix the privacy level* $\alpha > 0$*. Each data owner encodes their data point as a one-hot vector and randomly flips its bits. Formally, the resulting locally privatized sample* $\{\tilde{\mathbf{X}}_i\}_{i\in[n]} = \{\tilde{\mathbf{Y}}_i\}_{i\in[n_1]} \cup \{\tilde{\mathbf{Z}}_i\}_{i\in[n_2]}$ *is a set of* k*-dimensional random vectors whose* m*th element* $\tilde{X}_{im}$ *is defined as follows:*

$$\tilde{X}_{im} := \begin{cases} \mathbb{1}\big(X_i = m\big) & \textit{with probability} \quad \dfrac{e^{\alpha/2}}{e^{\alpha/2}+1}, \\ \mathbb{1}\big(X_i \neq m\big) & \textit{with probability} \quad \dfrac{1}{e^{\alpha/2}+1}. \end{cases}$$

The next lemma proves the α-LDP guarantee for Laplace, discrete Laplace, and `RAPPOR` mechanisms. The proof for discrete Laplace mechanism can be found in Appendix C, whereas the proof for Laplace and `RAPPOR` mechanisms is provided in Lemma 4.2 of Lam-Weil et al. (2022) and Section 3.2 of Duchi et al. (2013), respectively.

Lemma 7 (LDP guarantee) *Given a raw data set $\{X_i\}_{i\in[n]}$ and privacy level $\alpha > 0$, if $\{\tilde{\mathbf{X}}_i\}_{i\in[n]}$ is generated by one of the $\{\mathtt{LapU}, \mathtt{DiscLapU}, \mathtt{RAPPOR}\}$ mechanisms, it is an α-LDP view of $\{X_i\}_{i\in[n]}$.*

As defined below, our test statistic for multinomial testing builds on the α-LDP views $\{\tilde{\mathbf{X}}_i\}_{i\in[n]}$ from one of the $\{\mathtt{LapU}, \mathtt{DiscLapU}, \mathtt{RAPPOR}\}$ mechanisms. All of these mechanisms rigorously maintain local differential privacy and return a test that achieves the uniform separation as in Theorem 3.

Testing procedure. Given an α-LDP view from one of the $\{\mathtt{LapU}, \mathtt{DiscLapU}, \mathtt{RAPPOR}\}$ mechanisms, we use the following U-statistic:

$$U_{n_1,n_2} := \frac{1}{n_1(n_1-1)} \sum_{1\leq i_1\neq i_2\leq n_1} \tilde{\mathbf{Y}}_{i_1}^\top \tilde{\mathbf{Y}}_{i_2} + \frac{1}{n_2(n_2-1)} \sum_{1\leq j_1\neq j_2\leq n_2} \tilde{\mathbf{Z}}_{j_1}^\top \tilde{\mathbf{Z}}_{j_2} - \frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \tilde{\mathbf{Y}}_i^\top \tilde{\mathbf{Z}}_j. \tag{10}$$

To carry out a test, the test statistic is calibrated by the permutation procedure described in Section 2.3 with the pooled α-LDP view $\{\tilde{\mathbf{X}}_i\}_{i\in[n]}$. Specifically, we reject the null when the p-value based on the test statistic (10) is smaller than or equal to significance level γ. The type I error of the resulting test is controlled at γ, as $\{\tilde{\mathbf{X}}_i\}_{i\in[n]}$ are i.i.d. random vectors under the null hypothesis. Our technical effort lies in studying the type II error guarantee of the proposed test, and in turn proving the upper bound in Theorem 3. We refer to Appendix D for details.

The statistic (10) based on either Laplace or discrete Laplace mechanism is an unbiased estimator of a scaled and squared ℓ_2 distance between probability vectors:

$$\mathbb{E}[U_{n_1,n_2}] = k\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2. \tag{11}$$

On the other hand, the test statistic based on the `RAPPOR` mechanism does not maintain the unbiasedness property, with its expectation shrinking to zero as α decreases (see Lemma B.2). It therefore requires a more careful and, indeed, more challenging analysis compared to the other two mechanisms.

Despite the fact that all three mechanisms ensure the minimax optimality, their finite-sample power performance may differ in various scenarios as illustrated in Section 5. In particular, our numerical studies demonstrate that tests based on Laplace or discrete Laplace mechanisms tend to underperform compared to those based on the `RAPPOR` mechanism. This underperformance is partly because the private views from Laplace and discrete Laplace mechanism can take extreme values due to their unbounded support, whereas those from `RAPPOR` mechanism are always bounded. Accordingly, we advocate for using the `RAPPOR` mechanism over the Laplace and discrete Laplace mechanisms, even though they present the same theoretical guarantee in terms of the uniform separation.

Another common LDP mechanism for multinomial testing is the generalized randomized response mechanism (`GenRR`; see Appendix G.1 for details), suggested by Gaboardi et al. (2016). In contrast to `RAPPOR` mechanism, which destroys the structure of one-hot vectors, the generalized randomized response mechanism maintains the one-hot vector format while randomly altering the position of the non-zero component. It turns out, however, that the test based on the generalized randomized response mechanism is notoriously suboptimal in terms of the uniform separation, as we show in Appendix H. Demonstrating this negative result requires a careful asymptotic analysis of the U-statistic, which we want to highlight as our technical contribution.

4. Two-Sample Testing for Hölder and Besov Densities under LDP

In this section, we switch gears to testing for equality between two multivariate densities under the LDP constraint. To this end, we consider two classes of smooth densities, namely the Hölder ball and Besov ball, and establish the minimax testing separation in terms of the $\mathbb{L}_2$ distance for each class. Especially, we derive the upper bound for the minimax testing separation by building on the multinomial permutation test introduced in Section 3.2 with a careful discretization scheme. We also introduce an aggregated test, which is adaptive to the unknown smoothness parameter.

4.1 Hölder and Besov Smootheness Classes

We start by formally defining the Hölder ball and Besov ball. The Hölder ball generalizes Lipschitz continuity and can be thought of as functions with bounded fractional derivatives. The following definition of the Hölder ball rephrases the one stated in Section 2.1 of Arias-Castro et al. (2018).

Definition 8 (Hölder ball) *The Hölder ball with smoothness parameter $s > 0$ and radius $R > 0$, denoted as $\mathcal{B}^{\mathrm{H}}_{d,s}(R)$, is the class of functions $f : [0,1]^d \mapsto \mathbb{R}$ satisfying the following conditions:*

1. $\left|f^{(\lfloor s \rfloor)}(\mathbf{x}) - f^{(\lfloor s \rfloor)}(\mathbf{x}')\right| \le R\,\|\mathbf{x}-\mathbf{x}'\|^{s-\lfloor s \rfloor}$, *for all* $\mathbf{x}, \mathbf{x}' \in [0,1]^d$.
2. $\|f^{(s')}\|_\infty \le R$, *for each* $s' \in \big[\lfloor s \rfloor\big]$,

where $f^{(\lfloor s \rfloor)}$ denotes the $\lfloor s \rfloor$-order derivative of f.

The Besov ball, on the other hand, measures the smoothness of a function by capturing its abrupt oscillations through wavelets. In this respect, it can address spatially inhomogeneous functions whose smoothness can vary across their domain. To elaborate, we consider an orthonormal wavelet basis of $\mathbb{L}_2([0,1]^d)$ at a fixed prime resolution level $J \in \mathbb{N}_0$. We denote this basis as $\overline{\Phi}_J \cup (\cup_{j \ge J} \overline{\Psi}_j)$, which are classified into two distinct categories. For $\phi \in \overline{\Phi}_J$, the scaling coefficient $\theta_\phi(f) := \int_{[0,1]^d} f(\mathbf{x})\phi(\mathbf{x})\,d\mathbf{x}$ detects an overall trend of f. On the other hand, for $\psi \in \overline{\Psi}_j$ with its resolution level $j \ge J$, the wavelet coefficient $\theta_\psi(f) := \int_{[0,1]^d} f(\mathbf{x})\psi(\mathbf{x})\,d\mathbf{x}$ captures abrupt oscillations from the general trend. As J and j increase, the coefficients capture more detailed behaviors. Among many existing types of wavelet basis, this paper focuses on Haar multivariate wavelet basis. It is useful because projecting densities onto a subset of this basis is equivalent to applying equal-sized binning, as used in our test proposed in Section 4.3. Consequently, the discretization error inherent in the test can be characterized by the corresponding wavelet coefficients. This basis is constructed by taking tensor products of many rescaled and shifted versions of two basic univariate functions (see Appendix B.3 for the details).

The Besov ball can be defined through the magnitudes of wavelet coefficients. The following definition paraphrases the one presented in Section 3 of Tang and Yang (2023).

Definition 9 (Besov seminorm and Besov ball) *Fix a smoothness parameter $s > 0$, a microscopic parameter $1 \leq q \leq \infty$, and a wavelet basis $\overline{\Phi}_0 \cup (\bigcup_{j \geq 0} \overline{\Psi}_j)$. The Besov seminorm of $f \in \mathbb{L}_2([0,1]^d)$ is defined using the sequences of its wavelet coefficients $\big(\theta_\psi(f)\big)_{\psi \in \overline{\Psi}_j}$ as follows:*

$$\|\!|f|\!\|_{s,2,q} := \begin{cases} \left[\sum_{j=0}^{\infty} 2^{jsq} \big(\sum_{\psi \in \overline{\Psi}_j} |\theta_\psi(f)|^2\big)^{\frac{q}{2}}\right]^{\frac{1}{q}}, & 1 \leq q < \infty, \\ \sup_{j \in \mathbb{N}_0} 2^{js} \big(\sum_{\psi \in \overline{\Psi}_j} |\theta_\psi(f)|^2\big)^{\frac{1}{2}}, & q = \infty. \end{cases}$$

For a radius $R > 0$, we define the Besov ball $\mathcal{B}^{\mathrm{B}}_{d,s,q}(R)$ as

$$\mathcal{B}^{\mathrm{B}}_{d,s,q}(R) := \{f \in \mathbb{L}_2([0,1]^d) : \|\!|f|\!\|_{s,2,q} \leq R\}.$$

Neither Definition 8 nor Definition 9 is restricted to density functions. Instead, we construct our models by defining classes of distribution pairs where the differences in their density functions lie within these smooth function classes.

Definition 10 (Smooth distribution pair classes) *Let $\mathcal{P}^{\mathrm{H},2}_{d,s}(R)$ denote the set of pairs of distributions $(P_{\mathbf{Y}}, P_{\mathbf{Z}})$ that satisfy the following conditions:*

1. *Both $P_{\mathbf{Y}}$ and $P_{\mathbf{Z}}$ have densities $f_{\mathbf{Y}}$ and $f_{\mathbf{Z}}$, respectively, with their sup-norms bounded by R.*

2. *The difference of these densities $(f_{\mathbf{Y}} - f_{\mathbf{Z}})$ lies in $\mathcal{B}^{\mathrm{H}}_{d,s}(R)$.*

Similarly, let $\mathcal{P}^{\mathrm{B},2}_{d,s,q}(R)$ denote the set of pairs of distributions that satisfy the two conditions above, with $\mathcal{B}^{\mathrm{H}}_{d,s}(R)$ replaced by $\mathcal{B}^{\mathrm{B}}_{d,s,q}(R)$.

The superscript 2 in $\mathcal{P}^{\mathrm{H},2}_{d,s}(R)$ and $\mathcal{P}^{\mathrm{B},2}_{d,s,q}(R)$ indicates that these sets consist of pairs of distributions, distinguishing them from the sets of single distributions used in Appendix E.2. For $\mathcal{P}^{\mathrm{B},2}_{d,s,q}(R)$, we extend the analysis of Lam-Weil et al. (2022) into a multivariate setting, focusing on the Besov ball defined using a multivariate Haar wavelet basis and $s < 1$. Details of the basis functions are provided in Appendix B.3. Notably, there is no restriction on the microscopic parameter q.

4.2 Private Minimax Separation for Two-Sample Density Testing

Building on the smooth distribution classes defined in Definition 10, we formally define the two density testing problems of interest and present the minimax testing rate applicable to both. In this subsection, we assume that the smoothness parameter s for Hölder ball and Besov ball is known, addressing the case of unknown s in Section 4.4.

Assume that the pair of data-generating distributions $(P_{\mathbf{Y}}, P_{\mathbf{Z}})$ is contained in a class $\mathcal{P}$. The curator uses two sets of α-LDP views $\{\tilde{\mathbf{Y}}_i\}_{i\in[n_1]}$ and $\{\tilde{\mathbf{Z}}_j\}_{j\in[n_2]}$, privatized as described in Section 2.1, to decide whether $(P_{\mathbf{Y}}, P_{\mathbf{Z}})$ came from

$$\mathcal{P}_0 := \{(P_{\mathbf{Y}}, P_{\mathbf{Z}}) \in \mathcal{P} : f_{\mathbf{Y}} = f_{\mathbf{Z}}\} \text{ or } \mathcal{P}_1(\rho_{n_1,n_2}) := \big\{(P_{\mathbf{Y}}, P_{\mathbf{Z}}) \in \mathcal{P} : \|\!|f_{\mathbf{Y}} - f_{\mathbf{Z}}|\!\|_{\mathbb{L}_2} \geq \rho_{n_1,n_2}\big\}, \tag{12}$$

where $f_{\mathbf{Y}}$ and $f_{\mathbf{Z}}$ are densities of $P_{\mathbf{Y}}$ and $P_{\mathbf{Z}}$, respectively. We consider the problems of $\mathcal{P} = \mathcal{P}^{\mathrm{H},2}_{d,s}(R)$ and $\mathcal{P} = \mathcal{P}^{\mathrm{B},2}_{d,s,q}(R)$. Notably, while we focus on the class of either $\mathcal{P}$ being Hölder or Besov smooth distributions, the permutation-based test presented in Section 4.3 guarantees type I error control over a broader class of null distributions beyond those defined over Hölder or Besov balls.

Let us fix the type I error $\gamma \in (0,1)$ and the type II error $\beta \in (0,1)$ such that $2\gamma + \beta < 1$ as in Section 3.1, and assume further that $n_1\alpha^2 \geq 1$, similarly considered in Lam-Weil et al. (2022). The main result of this section, stated below, establishes a lower bound as well as an upper bound for the minimax separation for two-sample density testing under the LDP constraint.

Theorem 11 (Minimax separation for density testing) *Assume $n_1\alpha^2 \geq 1$. For the testing problem of distinguishing between the null and alternatives as in* (12)*, where the distribution class $\mathcal{P}$ is $\mathcal{P}^{\mathrm{H},2}_{d,s}(R)$, there exist positive constants $C_\ell(\gamma,\beta,R,s) \equiv C_\ell$ and $C_u(\gamma,\beta,R,s,d) \equiv C_u$ such that the α-LDP minimax testing separation $\rho^*_{n_1,n_2,\alpha}$ is bounded as*

$$C_\ell\left[\frac{(n_1 z_\alpha^2)^{\frac{-2s}{4s+3d}}}{\sqrt{\log(n_1 z_\alpha^2)}} \vee n_1^{\frac{-2s}{4s+d}}\right] \leq \rho^*_{n_1,n_2,\alpha} \leq C_u\big[(n_1\alpha^2)^{\frac{-2s}{4s+3d}} \vee n_1^{\frac{-2s}{4s+d}}\big], \tag{13}$$

where we recall $z_\alpha = 2\sinh(2\alpha)$. Similarly, for the testing problem of distinguishing between the null and alternatives as in (12)*, where the distribution class $\mathcal{P}$ is $\mathcal{P}^{\mathrm{B},2}_{d,s,q}(R)$, the minimax testing separation is also bounded as* (13)*.*

Theorem 11 indicates that the private minimax separation for two-sample multivariate Hölder and Besov densities is noticeably different from its non-private counterpart $n_1^{-2s/(4s+d)}$ with respect to the $\mathbb{L}_2$ distance (Arias-Castro et al., 2018; Kim et al., 2022a). We point out that in the high privacy regime, a polynomial degradation on the minimax separation is observed, and this degradation becomes worse as the data dimension d increases. The result also implies the privacy guarantee is secured at no additional charge in the low privacy regime where $n_1^{-2s/(4s+d)}$ dominates the other term.

The bounds (13) are tight up to a logarithmic factor in the denominator of the lower bound, which can be omitted when $n_1 z_\alpha^2 \geq 1$. As already noted in Theorem 3, α and z_α are the same, up to a constant factor, as long as α is bounded. Hence, in most practical scenarios where a small value of α is of interest, the upper bound matches the lower bound.

We now discuss the proof of Theorem 11, and a detailed analysis can be found in Appendix E.

- **Lower bound.** We once again use the observation that two-sample testing is harder than goodness-of-fit testing. Based on this observation, we employ the lower bound result for goodness-of-fit testing, mirroring the approach employed in Theorem 3 for multinomial testing. We then extend the strategy presented in Lam-Weil et al. (2022) from the univariate case to the multivariate case. The same proof strategy also applies to the multivariate Hölder ball with minor modifications, and details can be found in Appendix E.2.
- **Upper bound.** We leverage the private multinomial permutation test in Section 3.2, which achieves the minimax optimality. For that purpose, we divide each side of the support $[0,1]^d$ into κ equally-sized subintervals, effectively transforming the initially continuous observations into multinomial observations. The detailed procedure is outlined in Section 4.3. We then apply the same privacy mechanism and permutation testing procedure as outlined in Section 3.2. We exploit the Hölder and Besov smoothness conditions to find the optimal number of bins κ that effectively controls the discretization error and thus leads to a tight upper bound.

4.3 Privacy Mechanisms and Testing Procedure for the Upper Bound

Our proposed density test applies our proposed private multinomial test to data discretized by binning the density support into a certain sample-size dependent number of bins, defined as follows:

$$\kappa := \begin{cases} \lfloor (n_1^{2/(4s+d)}) \wedge (n_1\alpha^2)^{2/(4s+3d)} \rfloor & \text{if } (P_{\mathbf{Y}}, P_{\mathbf{Z}}) \in \mathcal{P}_{d,s}^{\mathrm{H},2}(R), \\ \sup\{2^J : J \in \mathbb{N}_0 \quad \text{and} \quad 2^J \le (n_1^{2/(4s+d)}) \wedge (n_1\alpha^2)^{2/(4s+3d)}\} & \text{if } (P_{\mathbf{Y}}, P_{\mathbf{Z}}) \in \mathcal{P}_{d,s,q}^{\mathrm{B},2}(R). \end{cases} \tag{14}$$

Let $\{B_1, ..., B_{\kappa^d}\}$ be an enumeration of d-dimensional hypercubes whose length is set to $1/\kappa$. Each data owner bins their raw sample using the equal-sized binning function $D_\kappa : [0,1]^d \mapsto [\kappa^d]$, such that $D_\kappa(\mathbf{x}) = m$ if and only if $\mathbf{x} \in B_m$. Then, based on the discretized data $\{D_\kappa(\mathbf{Y}_1), \ldots, D_\kappa(\mathbf{Y}_{n_1})\}$ and $\{D_\kappa(\mathbf{Z}_1), \ldots, D_\kappa(\mathbf{Z}_{n_2})\}$, we carry out our proposed private multinomial test in Section 3.2; the permutation test with test statistic (10). The resulting test achieves the tight upper bound in (13).

We now outline how this test maintains the privacy and testing error guarantees. Since the discretized data remain i.i.d. under the null, the permutation procedure maintains the type I error at γ. As for the privacy guarantee, the reduced distinguishability between samples due to discretization, combined with any of $\{\texttt{LapU}, \texttt{DiscLapU}, \texttt{RAPPOR}\}$, guarantees α-LDP. Regarding the type II error, first note that the test compares two multinomial distributions with κ^d categories. The corresponding probability vectors, $\mathbf{p_Y}$ and $\mathbf{p_Z}$, are defined as $\mathbf{p_Y}(m) := \int_{B_m} f_{\mathbf{Y}}(\mathbf{t})\, d\mathbf{t}$ and $\mathbf{p_Z}(m) := \int_{B_m} f_{\mathbf{Z}}(\mathbf{t})\, d\mathbf{t}$ for $m \in [\kappa^d]$. The boldface subscripts in $\mathbf{p_Y}$ and $\mathbf{p_Z}$ indicate that these vectors correspond to multivariate data, distinguishing them from $\mathbf{p}_Y$ and $\mathbf{p}_Z$ in Section 3. Thus the test statistic (10) estimates $\kappa^d \|\mathbf{p_Y} - \mathbf{p_Z}\|_2^2$ instead of $\|\!|f_{\mathbf{Y}} - f_{\mathbf{Z}}|\!\|_{\mathbb{L}_2}^2$, introducing an approximation error that depends on the number of bins κ. This error is controlled using smoothness conditions on densities and techniques from Arias-Castro et al. (2018) and Lam-Weil et al. (2022) (see Appendix B.4 for details). We show that when the number of bins κ is chosen as in (14) and the $\mathbb{L}_2$ distance between $f_{\mathbf{Y}}$ and $f_{\mathbf{Z}}$ exceeds the threshold in (13), the multinomial test applied to the binned data controls the type II error to be at most β. A detailed proof is provided in Appendix E.1.

We note that our choice of κ in (14) depends on the smoothness parameter s, typically unknown. In the next subsection, we introduce an adaptive procedure that accommodates this unknown s without significant loss of power.

4.4 Adaptive Procedure for Private Two-Sample Density Testing

To achieve the tight upper bound in Theorem 11 using our test proposed in Section 4.3, the number of bins must be determined based on the unknown smoothness parameter s. To circumvent this requirement, this section introduces a multiscale permutation testing procedure that adapts to the unknown s. Following Ingster (2000), it aggregates test results from different bin numbers using a Bonferroni-corrected significance level. This procedure does not significantly sacrifice power compared to the one relying on the true value of s.

Fix the privacy parameter $\alpha > 0$. Denote the number of the test for the Bonferroni-type procedure as:

$$\mathcal{N} := \left\lceil \left\{ \frac{2}{d} \log_2 \left(\frac{n_1}{\log\log n_1} \right) \right\} \wedge \left\{ \frac{2}{3d} \log_2 \left(\frac{n_1 \alpha^2}{(\log n_1)^2 \log\log n_1} \right) \right\} \right\rceil . \tag{15}$$

For each $t \in [\mathcal{N}]$, let $\Delta^t_{\gamma/\mathcal{N}}$ denote the test function of our proposed method in Section 4.3, using 2^t bins, significance level $\gamma/\mathcal{N}$, and $(\alpha/\mathcal{N})$-LDP guarantee. The α-LDP adaptive test is formally defined as follows:

$$\Delta^{\mathrm{adapt}}_{\gamma} := \max_{t \in [\mathcal{N}]} \Delta^t_{\gamma/\mathcal{N}}. \tag{16}$$

This adaptive procedure queries $\mathcal{N}$ number of $(\alpha/\mathcal{N})$-LDP views per observation, each using a different number of bins for discretization. By the composition theorem of differential privacy (McSherry and Talwar, 2007), releasing $\mathcal{N}$ number of $(\alpha/\mathcal{N})$-LDP views satisfies the LDP constraint with a privacy level of $\mathcal{N} \times \alpha/\mathcal{N} = \alpha$. The type I error is at most γ by the union bound, and for the type II error, Theorem 12 states that the adaptive procedure achieves the same uniform separation up to logarithmic factors.

Theorem 12 (Uniform separation for the adaptive procedure) *For the problems and conditions stated in Theorem 11, further assume that $n_1 \asymp n_2$, $\gamma \leq e^{-1}, n_1 > e^e$ and $\alpha \leq n_1 \leq n_2$. Then there exists a positive constant $C_u \equiv C(s, d, R, \gamma, \beta)$ such that the condition*

$$\rho_{n_1,n_2} \geq C_u \left[\left(\frac{n}{\log\log n_1} \right)^{\frac{-2s}{4s+d}} \vee \left(\frac{n_1 \alpha^2}{(\log^2 n_1) \log\log n_1} \right)^{\frac{-2s}{4s+3d}} \right]$$

implies that the testing errors of the adaptive test $\Delta^{\mathrm{adapt}}_{\gamma}$ in (16) are uniformly bounded as in (2).

Comparing Theorems 12 and 11, in the high privacy regime, we find that the adaptive procedure incurs an additional cost of $\big((\log^2 n_1)\log\log n_1\big)^{2s/(4s+3d)}$. In the low privacy regime, the additional cost $(\log\log n_1)^{2s/(4s+d)}$ matches the adaptivity cost for non-private testing rates found in Fromont and Laurent (2006) and Kim et al. (2022a). Whether these additional logarithmic factors are necessary or can be improved upon remains an open question for future work. We briefly discuss the proof of Theorem 12; a detailed analysis is provided in Appendix F.

- **Type I error control.** The overall type I error is at most γ by the union bound, though this approach usually is conservative in practice. Introducing an additional layer of calibration, as in Schrab et al. (2023), could mitigate this issue but would increase the noise level of privacy mechanisms. Developing an adaptive test with precise type I error control and robust power guarantee is an interesting avenue for future research.

- **Type II error control.** Since the significance level now depends on d and n_1 via $\mathcal{N}$, we use a refined version of the two moments method presented in Kim et al. (2022a). It improves the dependence on the significance level γ from $\sqrt{1/\gamma}$ of Theorem 21 to $\log(1/\gamma)$, at the cost of an additional requirement of $n_1 \asymp n_2$ on the sample sizes. At the heart of this refinement is an exponential concentration inequality for permuted U-statistics. This technique allows us to improve the adaptivity result of Lam-Weil et al. (2022) in their Theorem 5.2, replacing logarithmic factors with iterated logarithmic factors.

5. Numerical Results

In this section, we conduct a series of simulation studies to illustrate the finite sample performance of our proposed tests. Specifically, we investigate the privacy-utility trade-offs by varying the privacy level parameter α. These simulation studies aim to confirm our theoretical results regarding these trade-offs, and also to demonstrate the rate at which the power diminishes as the privacy parameter α decreases in practical scenarios.

It is worth pointing out that there is no baseline method available in the literature for the problem we tackle. Therefore, we create the baseline methods by extending the LDP goodness-of-fit tests (Gaboardi and Rogers, 2018) into a two-sample setting. The first method combines the generalized randomized response (`GenRR`; Gaboardi and Rogers, 2018) privacy mechanism and the classical chi-square statistic (`Chi`). The second method combines `RAPPOR` privacy mechanism and projected chi-square statistic (`ProjChi`; Gaboardi and Rogers, 2018). The third method combines the generalized randomized response mechanism with the ℓ_2-type U-statistic in (10). A formal description and asymptotic properties of these extensions can be found in Appendix G.1. From now on, we refer to the LDP two-sample testing methods as "privacy mechanism+test statistic", for example, `RAPPOR`+`ProjChi` or `GenRR`+`Chi` for ease of reference.

Recall that the proposed method for density testing, defined as a multinomial test applied to equal-sized binned data, requires the original data to lie within the unit hypercube $[0,1]^d$. In order to apply the multinomial test to continuous data with larger and potentially unbounded domains, we transform the data through a map $T : \mathbb{R}^d \mapsto [0,1]^d$, which is applied on a component-wise basis. A specific transformation that we focus on in this simulation is given as:

$$T(\mathbf{x}) = \big(\Phi(x_1), \Phi(x_2), \ldots, \Phi(x_d)\big), \tag{17}$$

where $\Phi(x) := (2\pi)^{-1/2} \int_{-\infty}^{x} e^{-t^2/2}\, dt$ is the standard normal cumulative distribution function. We then apply the procedure from Section 4.3 to the transformed observations $\{T(\mathbf{Y}_i)\}_{i=1}^{n_1}$ and $\{T(\mathbf{Z}_j)\}_{j=1}^{n_2}$. Recall that the optimal value of κ in (14) depends on the unknown smoothness parameter s. Since this value is not directly usable, practitioners must select κ to balance two competing effects: a large κ reduces discretization error, but potentially weakens the signal and becomes vulnerable to the impact of added noise due to the privacy mechanism. Our simulation results indicate that $\kappa = 4$ is a reasonable choice in most of the scenarios considered in this section, and thus we stick with the equal-sized binning scheme with $\kappa = 4$ for density testing.

In all simulation scenarios, we consider equal sample sizes, denoted as $n := n_1 = n_2$, and fix the significance level at $\gamma = 0.05$. We estimate the power by independently repeating the test 2000 times, and calculating the rejection ratio of the null hypothesis. The permutation procedure employs the Monte Carlo p-value given in (5) with $B = 999$. The code for replicating the numerical results is available at `https://github.com/Jong-Min-Moon/optimal-local-dp-two-sample.git`.

Type I error control. First, we compare the type I error rates of three methods for multinomial testing: LapU+ℓ_2, GenRR+Chi, and RAPPOR+ProjChi, to highlight the advantages of the permutation approach. We consider two null distributions, where for $m \in [k]$, the mth elements of $\mathbf{p}_Y$ and $\mathbf{p}_Z$ are defined as:

$$\text{(a) power law: } p_{Ym} = p_{Zm} \propto 1/m \quad \text{and} \quad \text{(b) uniform law: } p_{Ym} = p_{Zm} = 1/k. \tag{18}$$

We set $k = 500$ and $\alpha = 0.1$, and investigate how the type I error rate varies with sample sizes $n \in \{500, 1000, 1500\}$. The results, displayed in Figure 2, indicate that the type I error of the permutation test, LapU+ℓ_2, is well controlled at any sample size and significance level (up to a small numerical error) as expected. In contrast, we see that the asymptotic tests, namely RAPPOR+ProjChi and GenRR+Chi, have the size significantly deviated from the straight baseline, especially when the sample size is small or moderate. This indicates that the resulting asymptotic test can be either conservative or anti-conservative depending on the significance level.

Simulation settings for power comparison in multinomial testing. We next compare the power of our proposed methods (RAPPOR+ℓ_2, LapU+ℓ_2, and DiscLapU+ℓ_2) with baseline methods (GenRR+Chi, RAPPOR+ProjChi, and GenRR+ℓ_2) for distinguishing between two multinomial distributions. As observed in Figure 2, both GenRR+Chi and RAPPOR+ProjChi can be miscalibrated when their thresholds are determined by the asymptotic null distributions. In order to ensure a fair power comparison, GenRR+Chi and RAPPOR+ProjChi were calibrated using permutation procedures in this power simulation. We aim to illustrate how the testing power varies with changes in the number of categories k and the privacy parameter α. The analysis is conducted under a perturbed uniform distribution scenario, where

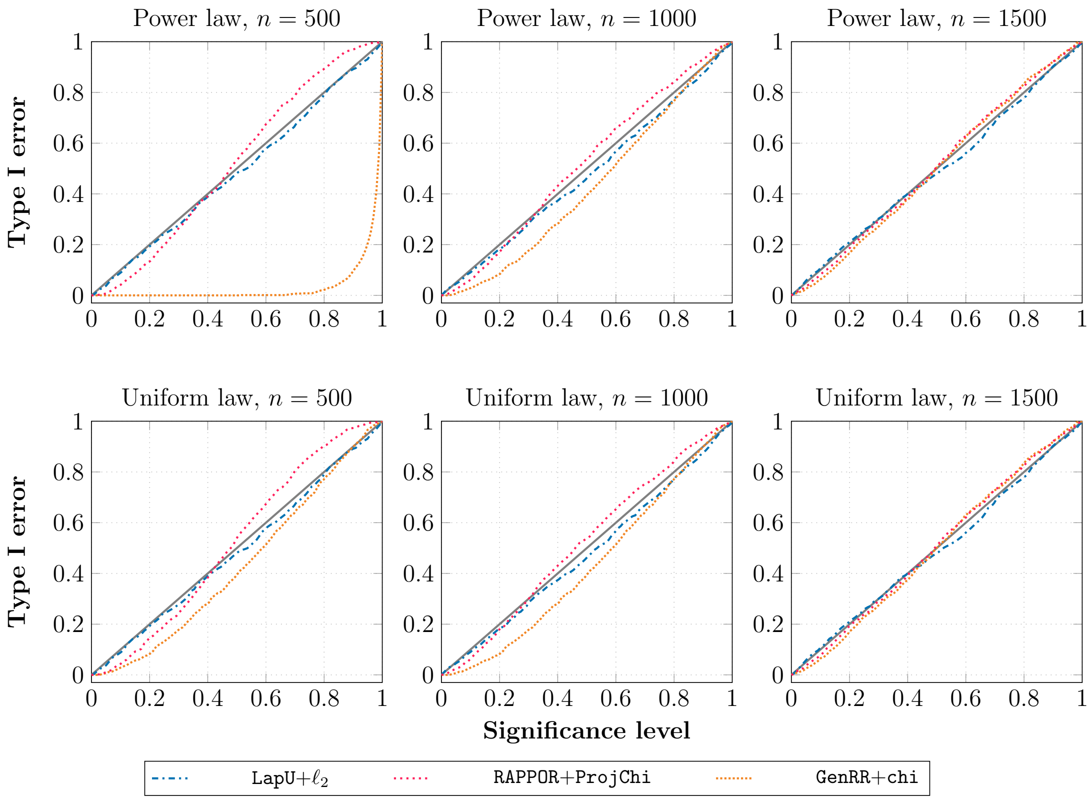


Figure 2: Comparison of type I error control between a permutation-calibrated method (`LapU`+ℓ_2) and methods calibrated through asymptotic chi-square null distribution (`RAPPOR`+`ProjChi` and `GenRR`+`Chi`), across configurations of uniform (top row) and power-law (bottom row) null distributions as described in (18). The gray solid lines represent the $y = x$ line, indicating perfect type I error control.

for $m \in [k]$, the mth elements of $\mathbf{p}_Y$ and $\mathbf{p}_Z$ are defined as:

$$p_{Ym} = \frac{1}{k} + (-1)^m \eta \quad \text{and} \quad p_{Zm} = \frac{1}{k} + (-1)^{(m+1)} \eta, \quad \text{for } m \in [k]. \tag{19}$$

The simulation considers the following combinations of three parameters, namely the number of categories k, the perturbation size η and the privacy parameter α:

$$(k, \eta, \alpha) \in \{(4, 0.04), (40, 0.015), (400, 0.002)\} \times \{2, 1, 0.5\}.$$

The simulation results for this setting are provided in Figure 3.

Simulation settings for power comparison in density testing. We also evaluate the density testing power of the same methods used in the simulations for multinomial testing. We consider two scenarios where two density functions differ in their location parameters or scale parameters. Since the results for scale difference show trend similar to that of location

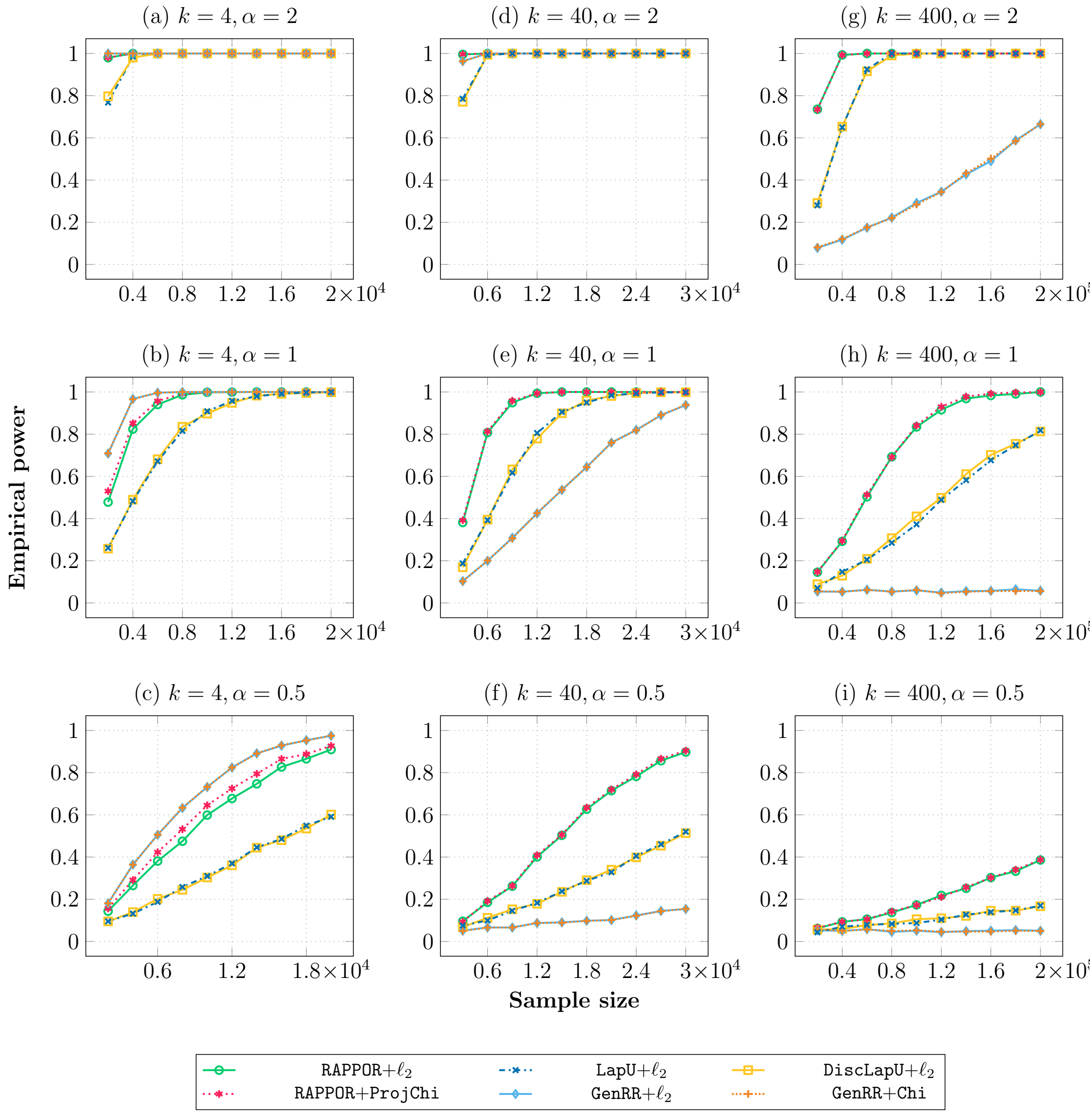


Figure 3: Comparison of the testing power between our proposed methods (first row in the legend) and baseline methods (second row in the legend) under the perturbed uniform alternatives (19). To ensure a fair comparison, all methods are calibrated using permutation procedures at level $\gamma = 0.05$.

difference, we present these results in Appendix G.3. For the location difference, we analyze scenarios involving mean differences between two d-dimensional Gaussian distributions $P_{\mathbf{Y}} = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{Y}}, \Sigma_{\mathbf{Y}})$ and $P_{\mathbf{Z}} = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{Z}}, \Sigma_{\mathbf{Z}})$. Let $\mathbf{1}_d := (1, \ldots, 1)^\top \in \mathbb{R}^d$, $\mathbf{0}_d := (0, \ldots, 0)^\top \in \mathbb{R}^d$, $\mathbf{J}_d := \mathbf{1}_d {\mathbf{1}_d}^\top \in \mathbb{R}^{d \times d}$, and $\mathbf{I}_d$ denote the identity matrix in $\mathbb{R}^{d \times d}$. We set the mean vectors and covariance matrices of the Gaussian distributions as:

$$\boldsymbol{\mu}_{\mathbf{Y}} = 0.5 \times \mathbf{1}_d, \ \boldsymbol{\mu}_{\mathbf{Z}} = -\boldsymbol{\mu}_{\mathbf{Y}}, \quad \text{and} \quad \Sigma_{\mathbf{Y}} = \Sigma_{\mathbf{Z}} = 0.5 \times \mathbf{J}_d + 0.5 \times \mathbf{I}_d. \tag{20}$$

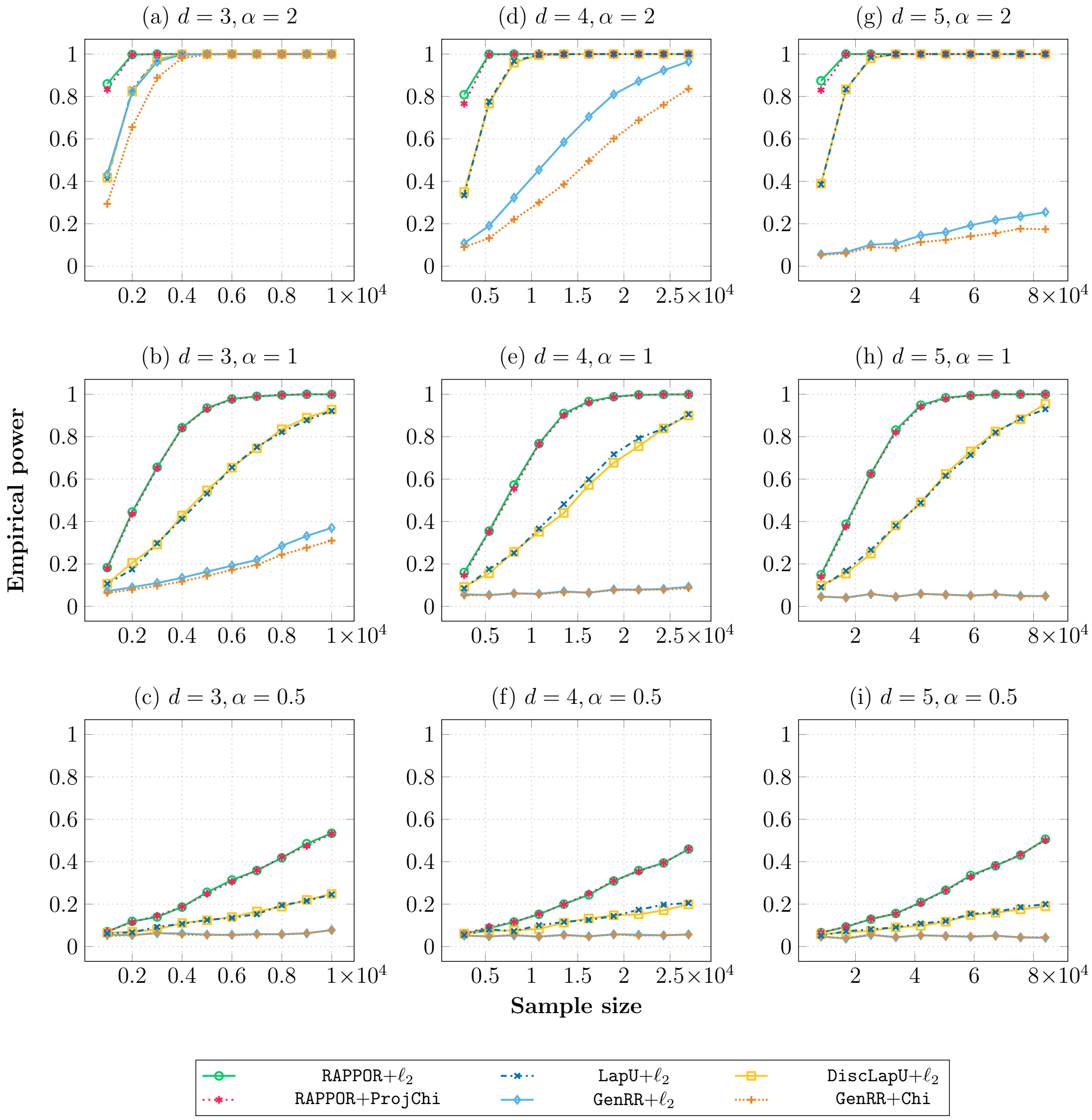


Figure 4: Comparison of the density testing power between our proposed methods (first row in the legend) and baseline methods (second row in the legend) under the location alternatives in (20). To ensure a fair comparison, all methods are calibrated using permutation procedures at level $\gamma = 0.05$.

In simulations, the dimension of the original data is chosen as $d \in \{3, 4, 5\}$, and after binning through the map T in (17) with $\kappa = 4$, the number of categories becomes $k \in \{64, 256, 1024\}$.

Simulation results for power comparisons. The simulation results in this section are illustrated in Figures 3 and 4. We first consistently observe in all of the figures that the power tends to decrease as the privacy parameter α decreases meaning a stronger privacy guarantee.

These trade-offs are all predictable from the minimax separations in (7) and (13). We next highlight the differences in trends related to the number of categories. For multinomial distributions with a small number of categories ($k = 4$), the generalized randomized response, a natural extension of the classical mechanism (Warner, 1965), outperforms all other methods. Following the generalized randomized response mechanism is the `RAPPOR` mechanism, while Laplace-noise-based mechanisms performing the least. However, in scenarios with a larger number of categories, or in density testing scenarios (which also correspond to a large number of categories), the testing power of the generalized randomized response diminishes, and `RAPPOR` emerges as the method with the highest power. The suboptimal performance of the generalized randomized response in high-dimensional settings, as theoretically explored in Appendix H and numerically observed by Gaboardi and Rogers (2018), aligns with a simple intuition: since the generalized randomized response modifies data by shifting it from one category to another, the difference between the original sample and its corresponding α-LDP view becomes more pronounced as the number of categories increases. These simulation results prove the superiority of `RAPPOR` over the other mechanisms, especially in multinomial testing with large k and density testing, and we therefore recommend using `RAPPOR` in practical applications. Within the tests built upon `RAPPOR`, we observe in Figures 3 and 4 that `RAPPOR`+`ProjChi` and `RAPPOR`+ℓ_2 perform comparably to each other, and in some cases, `RAPPOR`+`ProjChi` is slightly more powerful. It is also possible that the difference between `RAPPOR`+`ProjChi` and `RAPPOR`+ℓ_2 is more pronounced in some settings. For example, if the signal is large in terms of chi-square divergence but relatively small in terms of the ℓ_2 distance, we would expect `RAPPOR`+`ProjChi` to perform better than `RAPPOR`+ℓ_2. Conversely, if the signal is large in the ℓ_2 distance, the opposite holds true; we present the numerical results that confirm this in Appendix G.2. This suggests that when practitioners have insights into the nature of the deviation between two distributions, selecting a statistic that aligns with that specific deviation might be more effective. We leave it as future work to conduct more extensive simulations in diverse settings and using various types of test statistics.

6. Discussion

In this work, we studied minimax separation rates for two-sample testing under LDP constraint. Moving beyond the univariate Besov ball with $q = \infty$ considered in Lam-Weil et al. (2022), our work encompasses a larger Besov class of densities in a multivariate setting without restriction on q. We also considered the Hölder class and extended the non-private results of Kim et al. (2022a) to a locally private setting. By noting the equivalence between the binning approach in Kim et al. (2022a) and the projection approach in Lam-Weil et al. (2022), we proposed an integrated private testing framework that provides an optimal test for a large class of smooth densities. We proved our results using three distinct LDP mechanisms, thereby extending the toolkit available for practitioners. Additionally, an adaptive test is introduced that retains optimality up to a log factor without the knowledge of the smoothness parameter. Echoing prior work (Aliakbarpour et al., 2018, 2019; Cai et al., 2017; Lam-Weil et al., 2022), our results reaffirm that there exists an inevitable trade-off between data privacy and statistical efficiency that data analysts should bear in mind.

Our paper leaves several open questions for future investigation. Throughout the paper, we focused on equal-sized binning scheme that returns minimax optimal procedures. However, this framework may be problematic in high-dimensional settings as many bins would be empty. To address this issue, one can develop a data-adaptive binning scheme and improve the high-dimensional performance. In terms of smoothness classes, future work can be dedicated to extending our minimax result to a more general Besov class and other smoothness classes. Another interesting direction of future work is to develop optimal tests of conditional independence under privacy constraints, building on the recent work of Neykov et al. (2021) and Kim et al. (2022b). Finally, one can attempt to improve the $\log^2 n_1 \log\log n_1$ cost for adaptivity or find the matching lower bound. We leave all of these interesting but challenging problems for future work.

Acknowledgments

Ilmun Kim acknowledges support from the Yonsei University Research Fund of 2023-22-0419 as well as support from the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2022R1A4A1033384), and the Korea government (MSIT) RS-2023-00211073. The authors acknowledge the Center for Advanced Research Computing (CARC) at the University of Southern California for providing computing resources that have contributed to the research results reported within this publication. URL: `https://carc.usc.edu`.

Appendix A. Overview of Appendix

This supplementary material provides the technical proofs deferred in the main text, along with some additional results of interest. The content is organized as follows:

- Appendix B presents the technical lemmas, constructions, and calculations used in the main proofs of our results.
- Appendix C provides the proof of the α-LDP guarantees for our proposed privacy mechanisms.
- Appendix D proves the minimax separation result for multinomial testing, as presented in Theorem 3.
- Appendix E proves the minimax separation result for density testing, as presented in Theorem 11.
- Appendix F establishes the uniform separation for adaptive density testing, as shown in Theorem 12.
- Appendix G provides additional information on the numerical studies presented in Section 5.
- Appendix H discusses the asymptotic suboptimality of the `GenRR`+ℓ_2 multinomial test mentioned in Section 5.

Appendix B. Preliminary Results

This section presents the technical lemmas, constructions, and calculations used in the main proofs.

B.1 First Two Moments of Discrete Laplace Noise

We analyze the first two moments of a discrete Laplace noise random variable to establish the upper bound for the separation rate of our private test in Theorem 3. The next lemma shows that the discrete noise in discrete Laplace mechanism (Definition 5) has mean zero and variance at most $8k/\alpha^2$, which matches the variance of the continuous noise in Laplace mechanism (Definition 4) with the same privacy guarantee.

Lemma 13 *If W follows $\mathrm{DL}(\zeta_\alpha)$ defined in (9) and Definition 5, we have*

$$\mathbb{E}(W) = 0 \quad \textit{and} \quad \mathrm{Var}(W) \leq \frac{8k}{\alpha^2}.$$

Proof *From Proposition 2.2 of Inusah and Kozubowski (2006), we have $\mathbb{E}(W) = 0$ and $\mathrm{Var}(W) = 2\zeta_\alpha/(1-\zeta_\alpha)^2$. Therefore, it suffices to show that*

$$\frac{2\zeta_\alpha}{(1-\zeta_\alpha)^2} \leq \frac{8k}{\alpha^2}.$$

For notational convenience, let $v := \alpha/\sqrt{4k} > 0$ *for* $\alpha > 0$, *so that we have* $\zeta_\alpha = \exp(-\alpha/\sqrt{4k}) = \exp(-v)$. *The proof then reduces to showing that*

$$v^2 \le \frac{(1-\zeta_\alpha)^2}{\zeta_\alpha} = \big(e^{v/2} - e^{-v/2}\big)^2.$$

Since $v > 0$, *both sides of the above inequality is positive. Therefore it suffices to show that*

$$v \le e^{v/2} - e^{-v/2}. \tag{21}$$

Note that Taylor expansions of $\exp(v/2)$ *and* $\exp(-v/2)$ *have the same even order terms, and their odd order terms have the same absolute values with opposite signs. Therefore, it holds that*

$$(e^{v/2} - e^{-v/2}) = 2 \cdot \sum_{n=0}^{\infty} \frac{(v/2)^{2n+1}}{(2n+1)!} = 2 \cdot \left((v/2) + \frac{(v/2)^3}{3!} + \frac{(v/2)^5}{5!} + \frac{(v/2)^7}{7!} + \cdots \right).$$

Since $v > 0$, *all the terms of above are positive. Thus, the condition* (21) *is satisfied. This concludes the proof of Lemma 13.* ■

B.2 Summary Statistics of `RAPPOR` Mechanism

This section provides calculations for the summary statistics of the `RAPPOR` private views, which are used in Appendix D.1.2. First, Lemma 14 presents the exact calculation of the expectation of the U-statistic in (10), as well as the variance and covariance of the entries of a `RAPPOR` private view.

Lemma 14 (Summary statistics of `RAPPOR` private views) *Let* $\{Y_i\}_{i\in[n_1]}$ *be i.i.d. multinomial sample with* k *categories. Let* $\{\tilde{\mathbf{Y}}_i\}_{i\in[n_1]}$ *represent the corresponding* α*-LDP views obtained through the* `RAPPOR` *mechanism, where for each* $m \in [k]$, *the* m*th entry is distributed as*

$$\tilde{Y}_{im} \sim \mathrm{Ber}\big(\alpha_{\mathrm{bf}} \mathbb{1}(Y_i = m) + \delta_{\mathrm{bf}}\big), \text{ where } \alpha_{\mathrm{bf}} := \frac{e^{\alpha/2}-1}{e^{\alpha/2}+1} \quad \text{and} \quad \delta_{\mathrm{bf}} := \frac{1}{e^{\alpha/2}+1}.$$

Then we have $\mathbb{E}[U_{n_1,n_2}] = \alpha_{\mathrm{bf}}^2 \|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2$. *Also, we have*

$$\mathrm{Var}(\tilde{Y}_{1m}) = \big(\alpha_{\mathrm{bf}} p_{Ym} + \delta_{\mathrm{bf}}\big)\big(1 - \alpha_{\mathrm{bf}} p_{Ym} - \delta_{\mathrm{bf}}\big) \text{ and } \mathrm{Cov}(\tilde{Y}_{1m}, \tilde{Y}_{1m'}) = -\alpha_{\mathrm{bf}}^2 p_{Ym} p_{Ym'},$$

for each $m \in [k]$ *and for* $m, m' \in [k]$ *such that* $m \ne m'$, *respectively.*

Proof The expectation is verified as follows:

$$\begin{aligned}
\mathbb{E}[U_{n_1,n_2}] &= \mathbb{E}[(\tilde{\mathbf{Y}}_1 - \tilde{\mathbf{Z}}_1)^\top (\tilde{\mathbf{Y}}_2 - \tilde{\mathbf{Z}}_2)] \\
&= \mathbb{E}[\tilde{\mathbf{Y}}_1 - \tilde{\mathbf{Z}}_1]^\top \mathbb{E}[\tilde{\mathbf{Y}}_2 - \tilde{\mathbf{Z}}_2] \\
&= (\alpha_{\mathrm{bf}}\, \mathbf{p}_Y - \alpha_{\mathrm{bf}} \mathbf{p}_Z)^\top (\alpha_{\mathrm{bf}}\, \mathbf{p}_Y - \alpha_{\mathrm{bf}} \mathbf{p}_Z)
\end{aligned}$$

$$= \alpha_{\mathrm{bf}}^2 \|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2.$$

The variance is verified as follows:

$$\mathrm{Var}(\tilde{Y}_{1m}) = \mathbb{E}[\tilde{Y}_{1m}^2] - \mathbb{E}[\tilde{Y}_{1m}]^2 \overset{(a)}{=} \mathbb{E}[\tilde{Y}_{1m}] - \mathbb{E}[\tilde{Y}_{1m}]^2 = \big(\alpha_{\mathrm{bf}} p_{Ym} + \delta_{\mathrm{bf}}\big)\big(1 - \alpha_{\mathrm{bf}} p_{Ym} - \delta_{\mathrm{bf}}\big), \quad (22)$$

where step (a) uses the fact that $\tilde{Y}_{1m}^2 = \tilde{Y}_{1m}$ since $\tilde{Y}_{1m}$ is either 0 or 1. Finally, for $m, m' \in [k]$ such that $m \neq m'$, the covariance is calculated as:

$$\begin{aligned}
\mathrm{Cov}(\tilde{Y}_{1m}, \tilde{Y}_{1m'}) &= \mathbb{E}[\tilde{Y}_{1m}\tilde{Y}_{1m'}] - \mathbb{E}[\tilde{Y}_{1m}]\mathbb{E}[\tilde{Y}_{1m'}] \\
&\overset{(a)}{=} \mathbb{P}[\tilde{Y}_{1m} = 1, \tilde{Y}_{1m'} = 1] - \mathbb{E}[\tilde{Y}_{1m}]\mathbb{E}[\tilde{Y}_{1m'}] \\
&\overset{(b)}{=} \big(\alpha_{\mathrm{bf}} p_{Ym} + \delta_{\mathrm{bf}}\big)\big(\alpha_{\mathrm{bf}} p_{Ym'} + \delta_{\mathrm{bf}}\big) - \alpha_{\mathrm{bf}}^2 p_{Ym} p_{Ym'} - \big(\alpha_{\mathrm{bf}} p_{Ym} + \delta_{\mathrm{bf}}\big)\big(\alpha_{\mathrm{bf}} p_{Ym'} + \delta_{\mathrm{bf}}\big) \\
&= -\alpha_{\mathrm{bf}}^2 p_{Ym} p_{Ym'},
\end{aligned}$$

where step (a) uses the fact that each entry is 0 or 1, and step (b) is from Fact 1 of Acharya et al. (2019), which holds only when $m \neq m'$. ■

We next give upper bounds for the sum of entrywise variances and covariances of an `RAPPOR` α-LDP view.

Lemma 15 *For any $m, m' \in [k]$ such that $m' \neq m'$, the following inequalities hold:*

$$\sum_{m=1}^{k} \mathrm{Var}(\tilde{Y}_{1m})\big(p_{Ym} - p_{Zm}\big)^2 \leq \alpha_{\mathrm{bf}} b^{1/2} \|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 + \alpha_{\mathrm{bf}}\delta_{\mathrm{bf}} \|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2, \quad (23)$$

$$\sum_{m=1}^{k} \mathrm{Var}(\tilde{Y}_{1m})^2 \leq 2\alpha_{\mathrm{bf}}^2 b + 2\delta_{\mathrm{bf}}^2 k, \ \textit{and} \quad (24)$$

$$\sum_{1 \leq m \neq m' \leq k} \mathrm{Cov}(\tilde{Y}_{1m}, \tilde{Y}_{1m'})^2 \leq \alpha_{\mathrm{bf}}^2 b, \quad (25)$$

where $b = \max\{\|\mathbf{p}_Y\|_2^2, \|\mathbf{p}_Z\|_2^2\}$. The same type of inequalities also hold for $\tilde{Z}_{1m}$ and $\tilde{Z}_{1m'}$.

Proof For (23), we have

$$\begin{aligned}
\sum_{m=1}^{k} \mathrm{Var}(\tilde{Y}_{1m})\big(p_{Ym} - p_{Zm}\big)^2 &\overset{(a)}{=} \sum_{m=1}^{k} \big(\alpha_{\mathrm{bf}} p_{Ym} + \delta_{\mathrm{bf}}\big)\big\{1 - \big(\alpha_{\mathrm{bf}} p_{Ym} + \delta_{\mathrm{bf}}\big)\big\}\big(p_{Ym} - p_{Zm}\big)^2 \\
&\overset{(b)}{\leq} \sum_{m=1}^{k} \big(\alpha_{\mathrm{bf}} p_{Ym} + \delta_{\mathrm{bf}}\big)\big(p_{Ym} - p_{Zm}\big)^2 \\
&\overset{(c)}{\leq} \alpha_{\mathrm{bf}} b^{1/2} \|\mathbf{p}_Y - \mathbf{p}_Z\|_4^2 + \delta_{\mathrm{bf}} \|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 \\
&\overset{(d)}{\leq} \alpha_{\mathrm{bf}} b^{1/2} \|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 + \delta_{\mathrm{bf}} \|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2,
\end{aligned}$$

where step (a) uses Lemma 14, step (b) uses the fact that $0 < \alpha_{\mathrm{bf}} p_{Ym} + \delta_{\mathrm{bf}} < 1$ for any $m \in [k]$, step (c) uses the Cauchy–Schwarz inequality, and step (d) uses the monotonicity of the ℓ_p norm, specifically $\ell_4 \leq \ell_2$.

Next, for (24), we have

$$
\begin{aligned}
\sum_{m=1}^{k} \mathrm{Var}(\tilde{Y}_{1m})^2 \leq\ & \sum_{m=1}^{k} \big(\alpha_{\mathrm{bf}} p_{Ym} + \delta_{\mathrm{bf}}\big)^2 \big\{1 - \big(\alpha_{\mathrm{bf}} p_{Ym} + \delta_{\mathrm{bf}}\big)\big\}^2 \\
\overset{(a)}{\leq}\ & \sum_{m=1}^{k} \big(\alpha_{\mathrm{bf}} p_{Ym} + \delta_{\mathrm{bf}}\big)^2 \\
\overset{(b)}{\leq}\ & \sum_{m=1}^{k} \big(2\alpha_{\mathrm{bf}} p_{Ym}^2 + 2\delta_{\mathrm{bf}}^2\big) \\
=\ & 2\alpha_{\mathrm{bf}}^2 \|\mathbf{p}_Y\|_2^2 + 2\delta_{\mathrm{bf}}^2 k \\
\leq\ & 2\alpha_{\mathrm{bf}}^2 b + 2\delta_{\mathrm{bf}}^2 k,
\end{aligned}
$$

where step (a) uses the fact that $0 < \alpha_{\mathrm{bf}} p_{Ym} + \delta_{\mathrm{bf}} < 1$ for any $m \in [k]$, and step (b) uses the inequality $(a+b)^2 \leq 2a^2 + 2b^2$. Finally, for (25), we have

$$
\begin{aligned}
\sum_{1 \leq m \neq m' \leq k} \mathrm{Cov}(\tilde{Y}_{1m}, \tilde{Y}_{1m'})^2 \overset{(a)}{=}\ & \sum_{1 \leq m \neq m' \leq k} \alpha_{\mathrm{bf}}^4\, p_{Ym}^2 p_{Ym'}^2 \\
\overset{(b)}{\leq}\ & \alpha_{\mathrm{bf}}^2 \sum_{m=1}^{k} p_{Ym}^2 \sum_{m'=1}^{k} p_{Ym'}^2 \\
=\ & \alpha_{\mathrm{bf}}^2 \|\mathbf{p}_Y\|_2^4 \\
\overset{(c)}{\leq}\ & \alpha_{\mathrm{bf}}^2 \|\mathbf{p}_Y\|_2^2 \\
\leq\ & \alpha_{\mathrm{bf}}^2 b,
\end{aligned}
$$

where step (a) uses Lemma 14, step (b) uses the fact that $0 < \alpha_{\mathrm{bf}} < 1$ for any $\alpha > 0$, and step (c) uses the fact that $\|\mathbf{p}_Y\|_2^2 \leq 1$. ■

B.3 Construction of Multivariate Haar Wavelet Basis

We outline the construction of multivariate Haar wavelets, following Giné and Nickl (2015), Section 4.3.6, and Autin et al. (2010), Section 2. For $f : [0,1] \to \mathbb{R}$ and integers u, v, define the re-scaled and shifted version as $f_{u,v}(x) := 2^{u/2} f(2^u x - v)$. Define the univariate Haar scaling and wavelet functions from $[0,1]$ to $\mathbb{R}$ as

$$
\phi(x) := \mathbb{1}(0 \leq x < 1) \quad \text{and} \quad \psi(x) := \mathbb{1}(0 \leq x < 1/2) - \mathbb{1}(1/2 \leq x < 1),
$$

respectively. A multivariate Haar wavelet basis is indexed by its prime resolution level $J \in \mathbb{N}_0$, with each element characterized by up to three (multi) indices that range over:

1. Re-scaling: $\mathbb{N} \setminus [J-1]$,
2. Shifting: $\Lambda(j) := \{0, 1, \ldots, (2^j - 1)\}^d$, for an integer $j \geq J$,
3. On-off: $\mathcal{I} := \{0,1\}^d \setminus \mathbf{0}$,

respectively. For any $\boldsymbol{k} = (k_1, \ldots, k_d) \in \Lambda(J)$, let $\phi_{J,\boldsymbol{k}} : [0,1]^d \to \mathbb{R}$ be a tensor product of the re-scaled and shifted ϕ's, evaluated at $\mathbf{x} = (x_1, \ldots, x_d) \in [0,1]^d$:

$$\phi_{J,\boldsymbol{k}}(\mathbf{x}) := \prod_{p=1}^{d} \phi_{J,k_p}(x_p).$$

For a resolution level $j \geq J$, and for any $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_d) \in \Lambda(j)$, let $\psi^{\epsilon}_{j,\boldsymbol{\ell}} : [0,1]^d \to \mathbb{R}$ be a mixed tensor product of re-scaled and shifted ϕ's and ψ's, evaluated at $\mathbf{x} = (x_1, \ldots, x_d) \in [0,1]^d$:

$$\psi^{\epsilon}_{j,\boldsymbol{\ell}}(\mathbf{x}) := \prod_{p=1}^{d} \bigl\{\phi_{j,\ell_p}(x_p)\bigr\}^{1-\epsilon_p} \bigl\{\psi_{j,\ell_p}(x_p)\bigr\}^{\epsilon_p}.$$

The Haar multivariate wavelet basis at prime resolution level J is defined as

$$\Phi_J \cup \bigl(\bigcup_{j \geq J} \Psi_j\bigr), \text{ where } \Phi_J := \{\phi_{J,\boldsymbol{k}}\}_{\boldsymbol{k} \in \Lambda(J)} \text{ and } \Psi_j := \{\psi^{\epsilon}_{j,\boldsymbol{\ell}}\}_{\boldsymbol{\ell} \in \Lambda(j), \epsilon \in \mathcal{I}}. \tag{26}$$

B.4 Density Discretization Error Analysis

This section analyzes the error arising from comparing discretized multinomial probability vectors instead of the original multivariate densities. Given the number of bins κ, let $B_1, \ldots, B_{\kappa^d}$ enumerate d-dimensional hypercubes with side length $1/\kappa$. For a density $f_{\mathbf{Y}}$, its step function approximation is defined as:

$$\hat{f}_{\mathbf{Y}}(\mathbf{y}) := \sum_{m \in [\kappa^d]} \mathbb{1}(\mathbf{y} \in B_m) \kappa^d \int_{B_m} f_{\mathbf{Y}}(\mathbf{t})\, d\mathbf{t}. \tag{27}$$

Similarly, define $\hat{f}_{\mathbf{Z}}$ from $f_{\mathbf{Z}}$. Then it holds that $\|\mathbf{p}_{\mathbf{Y}} - \mathbf{p}_{\mathbf{Z}}\|_2^2 = \kappa^{-d} |||\hat{f}_{\mathbf{Y}} - \hat{f}_{\mathbf{Z}}|||^2_{\mathbb{L}_2}$. We analyze the difference between $|||\hat{f}_{\mathbf{Y}} - \hat{f}_{\mathbf{Z}}|||_{\mathbb{L}_2}$ and $|||f_{\mathbf{Y}} - f_{\mathbf{Z}}|||_{\mathbb{L}_2}$ when $(f_{\mathbf{Y}} - f_{\mathbf{Z}})$ lies in $\mathcal{B}^{\mathrm{H}}_{d,s}(R)$ or $\mathcal{B}^{\mathrm{B}}_{d,s,\infty}(R)$. For both cases, the discretization error scales as κ^{-s}, where we recall that s is the smoothness parameter. For the Hölder density case, the analysis is a simple application of Lemma 7.2 of Arias-Castro et al. (2018), rephrased below:

Lemma 16 *For a function $h \in \mathcal{B}^{\mathrm{H}}_{d,s}(R)$, let $\hat{h}$ be its step function approximation as in (27), with κ bins. Then there exist positive constants C_1 and C_2 depending only on d, s and R, but not on h, such that*

$$|||\hat{h}_{\kappa}|||_{\mathbb{L}_2} \geq C_1 |||h|||_{\mathbb{L}_2} - C_2 \kappa^{-s}.$$

Substituting $(f_{\mathbf{Y}} - f_{\mathbf{Z}}) \in \mathcal{B}^{\mathrm{H}}_{d,s}(R)$ into h in Lemma 16, the discretization error is characterized as follows:

$$\|\mathbf{p}_{\mathbf{Y}} - \mathbf{p}_{\mathbf{Z}}\|_2 \geq \kappa^{-d/2} \bigl(C_1 |||f_{\mathbf{Y}} - f_{\mathbf{Z}}|||_{\mathbb{L}_2} - C_2 \kappa^{-s}\bigr). \tag{28}$$

For the Besov density case, we derive a similar lemma as follows:

Lemma 17 *Define $\mathcal{B}^{\mathrm{B}}_{d,s,\infty}(R)$ using Haar multivariate wavelet basis at prime resolution level J constructed in Appendix B.3. For a function $(f_{\mathbf{Y}} - f_{\mathbf{Z}}) \in \mathcal{B}^{\mathrm{B}}_{d,s,\infty}(R)$, the following error bound holds:*

$$\|\mathbf{p}_{\mathbf{Y}} - \mathbf{p}_{\mathbf{Z}}\|_2 \geq \kappa^{-d/2}\big(\|\!|f_{\mathbf{Y}} - f_{\mathbf{Z}}\|\!|_{\mathbb{L}_2} - R\kappa^{-s}\big), \tag{29}$$

where $\kappa = 2^J$ and $\mathbf{p}_{\mathbf{Y}}$ and $\mathbf{p}_{\mathbf{Z}}$ are binned with side length $1/\kappa$.

Proof Let $\Phi_J \cup (\bigcup_{j\geq J}\Psi_j)$ denote the Haar multivariate wavelet basis that defines $\mathcal{B}^{\mathrm{B}}_{d,s,\infty}(R)$. The analysis proceeds in two steps. First, we show that the sum of squared coefficients from projecting $(f_{\mathbf{Y}} - f_{\mathbf{Z}})$ onto $\mathrm{span}(\Phi_J)$ equals the scaled ℓ_2 distance between the probability vectors:

$$\begin{aligned}
\sum_{\phi\in\Phi_J} \theta_\phi^2(f_{\mathbf{Y}} - f_{\mathbf{Z}}) &= \sum_{(k_1,\dots k_d)\in\Lambda(J)} \left(\int_{[0,1]^d} \big(f_{\mathbf{Y}}(\mathbf{x}) - f_{\mathbf{Z}}(\mathbf{x})\big) \prod_{p=1}^d \phi_{J,k_p}(x_p)\, d\mathbf{x}\right)^2 \\
&\overset{(a)}{=} \kappa^d \sum_{(k_1,\dots k_d)\in\Lambda(J)} \left(\int_{[0,1]^d} \big(f_{\mathbf{Y}}(\mathbf{x}) - f_{\mathbf{Z}}(\mathbf{x})\big) \prod_{p=1}^d \mathbb{1}\left(\frac{k_p}{\kappa} \leq x_p < \frac{k_p+1}{\kappa}\right) d\mathbf{x}\right)^2 \\
&= \kappa^d \sum_{m\in[\kappa^d]} \left(\int_{B_m} \big(f_{\mathbf{Y}}(\mathbf{x}) - f_{\mathbf{Z}}(\mathbf{x})\big)\, d\mathbf{x}\right)^2 \\
&= \kappa^d \|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2,
\end{aligned} \tag{30}$$

where step (a) uses the definition $\phi_{J,\boldsymbol{k}}(\mathbf{x}) = \prod_{p=1}^d \phi_{J,k_p}(x_p)$, with $\phi : [0,1] \to \{0,1\}$ defined as $\phi(x) = \mathbb{1}(0 \leq x < 1)$, and $\phi_{J,k_p}(x_p) = 2^{J/2} f(2^J x_p - k_p)$ indicates its re-scaled and shifted version. See Appendix B.3 for details. Since $\Phi_J \cup \big(\bigcup_{j\geq J}\Psi_j\big)$ forms an orthonormal basis of $\mathbb{L}_2([0,1]^d)$, the approximation error is given by the sum of squared projection coefficients of $(f_{\mathbf{Y}} - f_{\mathbf{Z}})$ onto $\mathrm{span}\big(\bigcup_{j\geq J}\Psi_j\big)$. Bounding this term is the second step of our analysis, which begins by noting that from Definition 9, for any $j \in \mathbb{N}_0$, the sum of squared wavelet coefficients is bounded as

$$\sum_{\psi\in\Psi_j} \theta_\psi^2(f_{\mathbf{Y}} - f_{\mathbf{Z}}) \leq 2^{-2js}R^2. \tag{31}$$

Then the approximation error is bounded as follows:

$$\begin{aligned}
\|\!|f_{\mathbf{Y}} - f_{\mathbf{Z}}\|\!|_{\mathbb{L}_2}^2 - \kappa^d\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 &= \sum_{j=J}^{\infty}\sum_{\psi\in\Psi_j} \theta_\psi^2(f_{\mathbf{Y}} - f_{\mathbf{Z}}) \\
&\overset{(a)}{\leq} \sum_{j=J}^{\infty} 2^{-2js}R^2 \\
&\overset{(b)}{=} R^2\frac{2^{-2Js}}{1-2^{-2s}} \leq R^2 2^{-2Js} = R^2\kappa^{-2s},
\end{aligned}$$

where step (a) uses (31), and step (b) uses an infinite geometric series. Applying $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ to the inequality above, we get

$$\|\!|f_{\mathbf{Y}} - f_{\mathbf{Z}}\|\!|_{\mathbb{L}_2} = \kappa^{d/2}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2 + R2^{-Js}\kappa^{d/2}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2 + R\kappa^{-s}.$$

■

B.5 Gaussian Approximation of One-Sample U-statistic

The theorems presented here are used to establish a negative result for our statistic in (10) when applied with the generalized randomized response mechanism. First, Theorem 18 restates a result from Kim (2020), which demonstrates the asymptotic normality of a one-sample U-statistic, which is a special case of our statistic in (10) with either $n_1 = \infty$ or $n_2 = \infty$ under the uniform null hypothesis.

Theorem 18 (Kim, 2020, Corollary 3.3) *Consider a multinomial goodness-of-fit test where the null hypothesis is a discrete uniform distribution with k categories, whose probability vector is denoted as $\boldsymbol{\pi}_0$. Let $\{\mathbf{Y}_i\}_{i=1}^n$ be a random sample represented in one-hot vector form drawn from $\boldsymbol{\pi}_0$. Define the U-statistic as:*

$$U_I := \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} (\mathbf{Y}_i - \boldsymbol{\pi}_0)^\top (\mathbf{Y}_j - \boldsymbol{\pi}_0). \tag{32}$$

If $n/\sqrt{k} \to \infty$, then the test statistic U_I has the following asymptotic normality:

$$\sqrt{\binom{n}{2}} \frac{U_I}{\sqrt{(1-1/k)/k}} \xrightarrow{d} \mathcal{N}(0,1).$$

The next theorem studies the limiting distribution of U_I under the alternative hypothesis.

Theorem 19 (Kim, 2020, Theorem 3.3) *Consider the same testing problem as Theorem 18. Let $\Sigma := \mathrm{Cov}(\mathbf{Y}_1)$. Under the alternative where the data is generated from a multinomial distribution with a probability vector $\boldsymbol{\pi} \neq \boldsymbol{\pi}_0$, suppose that the following conditions hold as $n, k \to \infty$:*

C1. $\dfrac{\mathrm{tr}(\Sigma^4)}{\{\mathrm{tr}(\Sigma^2)\}^2} \to 0.$

C2. $\dfrac{\mathbb{E}[\{(\mathbf{Y}_1 - \boldsymbol{\pi}_0)^\top (\mathbf{Y}_2 - \boldsymbol{\pi}_0)\}^4] + n_1 \mathbb{E}[\{(\mathbf{Y}_1 - \mathbf{p}_Z)^\top (\mathbf{Y}_2 - \mathbf{p}_Z\}^2 \{(\mathbf{Y}_1 - \mathbf{p}_Z)^\top (\mathbf{Y}_3 - \mathbf{p}_Z)\}^2]}{n^2 \{\mathrm{tr}(\Sigma^2)\}^2} \to 0.$

C3. $(\boldsymbol{\pi} - \boldsymbol{\pi}_0)^\top \Sigma (\boldsymbol{\pi} - \boldsymbol{\pi}_0) < \infty.$

Then the test statistic U_I in (32) has the following asymptotic normality:

$$\sqrt{\frac{n(n-1)}{2}} \frac{U_I - \|\boldsymbol{\pi} - \boldsymbol{\pi}_0\|_2^2}{\sqrt{\mathrm{tr}(\Sigma^2) + 2(n-1)(\boldsymbol{\pi} - \boldsymbol{\pi}_0)^\top \Sigma (\boldsymbol{\pi} - \boldsymbol{\pi}_0)}} \xrightarrow{d} \mathcal{N}(0,1). \tag{33}$$

Appendix C. Proof of Lemma 7

This section provides privacy guarantee proof for the mechanisms proposed in Section 3.2. The privacy proof for the `RAPPOR` mechanism is given in Section 3.2 of Duchi et al. (2013). The privacy proof for the Laplace mechanism is a minor modification of the proof of Lemma 4.2 from Lam-Weil et al. (2022), adjusting the domain of the raw samples. Thus, we only present the privacy proof for the discrete Laplace mechanism, defined in Definition 5. Its outline resembles the proof of Lemma 4.2 from Lam-Weil et al. (2022).

Proof We note that the conditional distribution of the mth entry of its α-LDP view, denoted as $\tilde{X}_{im}$, is a discrete Laplace distribution with parameter $\zeta_\alpha = \exp(-\alpha/(2\sqrt{k}))$, shifted by $\sqrt{k}\mathbb{1}(X_i = m)$. Note again that for $m \neq m'$, $\tilde{X}_{im}$ and $\tilde{X}_{im'}$ are independent. With slight abuse of notation, we denote the conditional density function of $\tilde{\mathbf{X}}_i$ as $q_i(\cdot \,|\, \cdot)$, which can be written as

$$q_i(\tilde{\mathbf{x}} \,|\, x) = \prod_{m=1}^{k} \frac{1-\zeta_\alpha}{1+\zeta_\alpha} \zeta_\alpha^{|\tilde{x}_m - \sqrt{k}\mathbb{1}(x=m)|}.$$

Then for all $\tilde{\mathbf{x}} = (\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_k)^\top \in \mathbb{R}^k$ and all $x, x' \in [k]$, we have

$$\begin{aligned}
\frac{q_i(\tilde{\mathbf{x}} \,|\, x)}{q_i(\tilde{\mathbf{x}} \,|\, x')} &= \prod_{m=1}^{k} \zeta_\alpha^{|\tilde{x}_m - \sqrt{k}\mathbb{1}(x=m)| - |\tilde{x}_m - \sqrt{k}\mathbb{1}(x'=m)|} \\
&= \zeta_\alpha^{\sum_{m=1}^{k} |\tilde{x}_m - \sqrt{k}\mathbb{1}(x=m)| - |\tilde{x}_m - \sqrt{k}\mathbb{1}(x'=m)|} \\
&\overset{(a)}{\leq} \left(\zeta_\alpha^{-1}\right)^{\sum_{m=1}^{k} \left| |\tilde{x}_m - \sqrt{k}\mathbb{1}(x=m)| - |\tilde{x}_m - \sqrt{k}\mathbb{1}(x'=m)| \right|} \\
&\overset{(b)}{\leq} \left(\zeta_\alpha^{-1}\right)^{\sum_{m=1}^{k} \left\{ \sqrt{k}\mathbb{1}(x=m) + \sqrt{k}\mathbb{1}(x'=m) \right\}} \\
&\overset{(c)}{\leq} \left(\zeta_\alpha^{-1}\right)^{2\sqrt{k}} = e^{\alpha},
\end{aligned}$$

where step (a) uses the fact that $\zeta_\alpha^x \leq \zeta_\alpha^{-|x|}$ for all $x \in \mathbb{R}$ since $\zeta_\alpha \in (0,1)$, step (b) uses the reverse triangle inequality, and step (c) holds since $\mathbb{1}(x = m) \neq 0$ for only a single value of m. Using the inequality above, for any Borel set $A \in \mathcal{B}(\mathbb{R}^d)$, and for any $x, x' \in [k]$, we have:

$$\frac{Q_i(A \,|\, x)}{Q_i(A \,|\, x')} = \frac{\int_A q_i(\tilde{\mathbf{x}} \,|\, x)\, d\tilde{\mathbf{x}}}{\int_A q_i(\tilde{\mathbf{x}} \,|\, x')\, d\tilde{\mathbf{x}}} \leq \frac{\int_A q_i(\tilde{\mathbf{x}} \,|\, x') e^{\alpha}\, d\tilde{\mathbf{x}}}{\int_A q_i(\tilde{\mathbf{x}} \,|\, x) e^{-\alpha}\, d\tilde{\mathbf{x}}} = \frac{Q_i(A \,|\, x')}{Q_i(A \,|\, x)} e^{2\alpha}.$$

This completes the proof of the guarantee for the discrete Laplace mechanism. ∎

Appendix D. Proof of Theorem 3

In this section, we prove the minimax separation result for multinomial testing presented in Theorem 3, first focusing on the upper bound result followed by the lower bound result.

D.1 Upper Bound

This section demonstrates that the permutation test with the U-statistic proposed in (10), in conjunction with one of the mechanisms $\{\texttt{LapU}, \texttt{DiscLapU}, \texttt{RAPPOR}\}$, achieves a tight upper bound presented in Theorem 3. Since the permutation procedure controls the type I error, it suffices to prove that the condition in Theorem 3 guarantees type II error control. Let

$$b := \max\{\|\mathbf{p}_Y\|_2^2, \|\mathbf{p}_Z\|_2^2\}. \tag{34}$$

The control of the type II error is then rephrased in the following lemma.

Lemma 20 *Assume the settings of Theorem 3. For each of the mechanisms* $\{\texttt{LapU}, \texttt{DiscLapU}, \texttt{RAPPOR}\}$, *there exists a constant* $C_u(\gamma, \beta)$ *such that the type II error of the permutation test of size* γ *with U-statistic in* (10) *is uniformly controlled by* β *over* $\mathcal{P}_{1,\mathrm{multi}}$ *if*

$$\rho_{n_1,n_2} \geq C_u(\gamma, \beta) \left(\frac{k^{1/4}}{(n_1\alpha^2)^{1/2}} \vee \frac{b^{1/4}}{n_1^{1/2}} \right). \tag{35}$$

Since $b \leq 1$, Lemma 20 proves the upper bound result of Theorem 3. To prove Lemma 20, we leverage the two moments method (Theorem 4.1 of Kim et al., 2022a), rephrased in Lemma 21. This method states that if the test statistic's expectation is sufficiently larger than a variance proxy, then uniform type II error control is achieved for the permutation test with a two-sample U-statistic of order 2. Its key advantage is that the condition bypasses the randomness arising from permutations. It lets us avoid the complex analysis typically required for permuted statistics.

To proceed, consider the following kernel:

$$h_{ts}(\mathbf{y}_1, \mathbf{y}_2; \mathbf{z}_1, \mathbf{z}_2) = \mathbf{y}_1^\top \mathbf{y}_2 + \mathbf{z}_1^\top \mathbf{z}_2 - \mathbf{z}_1^\top \mathbf{y}_2 - \mathbf{z}_2^\top \mathbf{y}_1, \tag{36}$$

which defines our two-sample U-statistic U_{n_1,n_2} in (10). Its symmetrized version, denoted h_{ts}, is defined as:

$$\check{h}_{ts}(\mathbf{y}_1, \mathbf{y}_2; \mathbf{z}_1, \mathbf{z}_2) := \frac{1}{2!2!} \sum_{1 \leq i_1 \neq i_2 \leq n_1} \sum_{1 \leq j_1 \neq j_2 \leq n_2} h_{ts}(\mathbf{y}_{i_1}, \mathbf{y}_{i_2}; \mathbf{z}_{j_1}, \mathbf{z}_{j_2}).$$

For the upper bound analysis, we use a U-statistic represented by $\check{h}_{ts}$, which is equivalent to our original statistic in U_{n_1,n_2} (10). Recall that under the LDP constraint, the raw samples $\{Y_i\}_{i\in[n_1]}$ and $\{Z_j\}_{j\in[n_2]}$ are generated from $P = (P_Y, P_Z)$ and then transformed into LDP-views $\{\tilde{\mathbf{Y}}_i\}_{i\in[n_1]}$ and $\{\tilde{\mathbf{Z}}_j\}_{j\in[n_2]}$ through an LDP mechanism Q. Let us denote the associated moments as

$$\begin{aligned}
M_{Y,1}(P,Q) &:= \mathrm{Var}_{P,Q}[\mathbb{E}\{\check{h}_{ts}(\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2; \tilde{\mathbf{Z}}_1, \tilde{\mathbf{Z}}_2) \,|\, \tilde{\mathbf{Y}}_1\}],\\
M_{Z,1}(P,Q) &:= \mathrm{Var}_{P,Q}[\mathbb{E}\{\check{h}_{ts}(\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2; \tilde{\mathbf{Z}}_1, \tilde{\mathbf{Z}}_2) \,|\, \tilde{\mathbf{Z}}_1\}],\\
M_{YZ,2}(P,Q) &:= \max\{\mathbb{E}[(\tilde{\mathbf{Y}}_1^\top \tilde{\mathbf{Y}}_2)^2],\, \mathbb{E}[(\tilde{\mathbf{Y}}_1^\top \tilde{\mathbf{Z}}_1)^2],\, \mathbb{E}[(\tilde{\mathbf{Z}}_1^\top \tilde{\mathbf{Z}}_2)^2]\}.
\end{aligned} \tag{37}$$

Using these moments, we rephrase the two moments method under the setting of LDP.

Lemma 21 (Two moments method) *Let U_{n_1,n_2} be a two-sample U-statistic based on the kernel given in* (36)*. Assume that the samples are privatized through an α-LDP mechanism Q. Then there exists a sufficiently large constant $C > 0$ such that if*

$$\mathbb{E}[U_{n_1,n_2}] \geq C\sqrt{\max\left\{\frac{M_{Y,1}(P,Q)}{\beta n_1}, \frac{M_{Z,1}(P,Q)}{\beta n_2}, \frac{M_{YZ,2}(P,Q)}{\gamma\beta}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^2\right\}} \tag{38}$$

for all pairs of distributions $P = (P_Y, P_Z) \in \mathcal{P}_{1,\mathrm{multi}}(\rho_{n_1,n_2})$, then the type II error of the permutation test over $\mathcal{P}_{1,\mathrm{multi}}(\rho_{n_1,n_2})$ is uniformly bounded by β as in (2)*.*

Having stated the two moments method, our goal is to verify that inequality (38) holds under the separation conditions described in Section 3.2. We provide separate proofs for the Laplace-noise based mechanisms (`LapU` and `DiscLapU`) and the `RAPPOR` mechanism. However, both proofs follow the same two steps:

1. Derive upper bounds for the moments $M_{Y,1}(P,Q)$, $M_{Z,1}(P,Q)$, and $M_{YZ,2}(P,Q)$ in (37).
2. Using the upper bounds established in Step 1, show that condition (38) in the two moments method is fulfilled as long as inequality (35) in Lemma 20 holds.

We now verify the previous preliminary steps in order, first assuming the Laplace-noise based mechanisms.

D.1.1 Proof of Lemma 20 Through Laplace or Discrete Laplace Mechanism

Since the analysis for Laplace and discrete Laplace mechanism is similar, we present the proof using the former (see Remark 22 for details).

Proof We follow the two steps mentioned above.

Step 1: Bound the moments from above. We start by bounding the variance of conditional expectation terms. Recall the notation $M_{Y,1}(P,Q) = \mathrm{Var}_{P,Q}[\mathbb{E}\{\check{h}_{ts}(\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2; \tilde{\mathbf{Z}}_1, \tilde{\mathbf{Z}}_2) \mid \tilde{\mathbf{Y}}_1\}]$. To upper bound $M_{Y,1}(P,Q)$, we first calculate the conditional expectation of the kernel function, namely $A := \mathbb{E}\{\check{h}_{ts}(\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2; \tilde{\mathbf{Z}}_1, \tilde{\mathbf{Z}}_2) \mid \tilde{\mathbf{Y}}_1\}$. Then we bound the variance of A.

For the conditional expectation, recall the mth component of $\mathbf{p}_Y, \mathbf{p}_Z, \tilde{\mathbf{Y}}_i, \tilde{\mathbf{Z}}_j$ are denoted as $p_{Ym}, p_{Zm}, \tilde{Y}_{im}$ and $\tilde{Z}_{j,m}$, respectively. Due to i.i.d. assumptions, regardless of privacy mechanism, we have the following equalities and inequalities. First we calculate the conditional expectation:

$$A := \mathbb{E}\big[\check{h}_{ts}(\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2; \tilde{\mathbf{Z}}_1, \tilde{\mathbf{Z}}_2) \mid \tilde{\mathbf{Y}}_1\big] = \big(\tilde{\mathbf{Y}}_1 - \mathbb{E}[\tilde{\mathbf{Z}}_1]\big)^\top \big(\mathbb{E}[\tilde{\mathbf{Y}}_1] - \mathbb{E}[\tilde{\mathbf{Z}}_1]\big).$$

Then we calculate the unconditional expectation:

$$\mathbb{E}[A] = \big(\mathbb{E}[\tilde{\mathbf{Y}}_1] - \mathbb{E}[\tilde{\mathbf{Z}}_1]\big)^\top \big(\mathbb{E}[\tilde{\mathbf{Y}}_1] - \mathbb{E}[\tilde{\mathbf{Z}}_1]\big) = \|\mathbb{E}[\tilde{\mathbf{Y}}_1] - \mathbb{E}[\tilde{\mathbf{Z}}_1]\|_2^2.$$

Then the squared and centered conditional expectation is calculated as follows:

$$A - \mathbb{E}[A] = \big(\tilde{\mathbf{Y}}_1 - \mathbb{E}[\tilde{\mathbf{Y}}_1]\big)^\top \big(\mathbb{E}[\tilde{\mathbf{Y}}_1] - \mathbb{E}[\tilde{\mathbf{Z}}_1]\big).$$

Based on this, the variance is calculated as

$$\begin{aligned}\mathbb{E}\big[\big(A-\mathbb{E}[A]\big)^2\big] &= \mathbb{E}\big[\big\{\big(\tilde{\mathbf{Y}}_1-\mathbb{E}[\tilde{\mathbf{Y}}_1]\big)^\top\big(\mathbb{E}[\tilde{\mathbf{Y}}_1]-\mathbb{E}[\tilde{\mathbf{Z}}_1]\big)\big\}^2\big] \quad (39)\\ &= \mathbb{E}\big[\big\{\tilde{\mathbf{Y}}_1^\top\big(\mathbb{E}[\tilde{\mathbf{Y}}_1]-\mathbb{E}[\tilde{\mathbf{Z}}_1]\big)-\mathbb{E}[\tilde{\mathbf{Y}}_1]^\top\big(\mathbb{E}[\tilde{\mathbf{Y}}_1]-\mathbb{E}[\tilde{\mathbf{Z}}_1]\big)\big\}^2\big]\\ &\le 2\mathbb{E}\big[\big\{\tilde{\mathbf{Y}}_1^\top\big(\mathbb{E}[\tilde{\mathbf{Y}}_1]-\mathbb{E}[\tilde{\mathbf{Z}}_1]\big)\big\}^2\big]+2\big[\big\{\mathbb{E}[\tilde{\mathbf{Y}}_1]^\top\big(\mathbb{E}[\tilde{\mathbf{Y}}_1]-\mathbb{E}[\tilde{\mathbf{Z}}_1]\big)\big\}^2\big], \quad (40)\end{aligned}$$

where the last inequality uses $(x+y)^2 \le 2x^2+2y^2$. Using $\|\mathbf{p}_Y\|_2 \le 1$, we bound the second term in (40) by

$$2k^2\|\mathbf{p}_Y\|_2\|\mathbf{p}_Y-\mathbf{p}_Z\|_2^2. \tag{41}$$

For the first term, we leverage the structure of our privacy mechanism, which adds independent and centered noises. Applying the Cauchy–Schwarz inequality then provides an upper bound, reducing the order of k compared to a direct application of Cauchy–Schwarz. More formally,

$$\begin{aligned}&2\mathbb{E}\Big\{\tilde{\mathbf{Y}}_1^\top\big(\mathbb{E}[\tilde{\mathbf{Y}}_1]-\mathbb{E}[\tilde{\mathbf{Z}}_1]\big)\Big\}^2\\ &= 2\sum_{m=1}^k\sum_{m'=1}^k \mathbb{E}[\tilde{Y}_{1m}\tilde{Y}_{1m'}]\big(\mathbb{E}[\tilde{Y}_{1m}]-\mathbb{E}[\tilde{Z}_{1m}]\big)\big(\mathbb{E}[\tilde{Y}_{1m'}]-\mathbb{E}[\tilde{Z}_{1m'}]\big)\\ &\overset{(a)}{=} 2\sum_{m=1}^k \mathbb{E}[\tilde{Y}_{1m}^2]\big(\mathbb{E}[\tilde{Y}_{1m}]-\mathbb{E}[\tilde{Z}_{1m}]\big)^2\\ &\overset{(b)}{=} 2\sum_{m=1}^k \big\{kp_{Ym}+\sigma_\alpha^2\big\}\big(\mathbb{E}[\tilde{Y}_{1m}]-\mathbb{E}[\tilde{Z}_{1m}]\big)^2\\ &= 2\sum_{m=1}^k kp_{Ym}\big(\mathbb{E}[\tilde{Y}_{1m}]-\mathbb{E}[\tilde{Z}_{1m}]\big)^2+2k\sigma_\alpha^2\|\mathbf{p}_Y-\mathbf{p}_Z\|_2^2\\ &\overset{(c)}{\le} 2k^2\|\mathbf{p}_Y\|_2\|\mathbf{p}_Y-\mathbf{p}_Z\|_2^2+2k\sigma_\alpha^2\|\mathbf{p}_Y-\mathbf{p}_Z\|_2^2, \quad (42)\end{aligned}$$

where step (a) uses $\mathbb{E}[\tilde{Y}_{1m}\tilde{Y}_{1m'}]=0$ for $m\ne m'$, step (b) uses $\mathbb{E}[\tilde{Y}_{1m}^2]=kp_{Ym}+\sigma_\alpha^2$, and step (c) uses the Cauchy–Schwarz inequality, $\|\mathbf{p}_Y\|_2^2 \le \|\mathbf{p}_Y\|_2$, and monotonicity of ℓ_p norm, specifically $\ell_4 \le \ell_2$. Combining the results in (41) and (42), we achieve an upper bound for $M_{Y,1}(P,Q)$ given as

$$M_{Y,1}(P,Q) \le \big(4k^2\|\mathbf{p}_Y\|_2+2k\sigma_\alpha^2\big)\|\mathbf{p}_Y-\mathbf{p}_Z\|_2^2. \tag{43}$$

By symmetry, we also have

$$M_{Z,1}(P,Q) \le \big(4k^2\|\mathbf{p}_Z\|_2+2k\sigma_\alpha^2\big)\|\mathbf{p}_Y-\mathbf{p}_Z\|_2^2. \tag{44}$$

Combining the upper bound results of (43) and (44), and keeping in mind that $b=\max\{\|\mathbf{p}_Y\|_2^2,\|\mathbf{p}_Z\|_2^2\}$, we obtain the following upper bound:

$$\max\{M_{Y,1}(P,Q),M_{Z,1}(P,Q)\} \le (4k^2b^{1/2}+2k\sigma_\alpha^2)\|\mathbf{p}_Y-\mathbf{p}_Z\|_2^2. \tag{45}$$

We now turn our attention to the expectation of square terms:

$$M_{YZ,2}(P,Q) := \max\big\{\mathbb{E}\big[(\tilde{\mathbf{Y}}_1^\top \tilde{\mathbf{Y}}_2)^2\big],\ \mathbb{E}\big[(\tilde{\mathbf{Z}}_1^\top \tilde{\mathbf{Z}}_2)^2\big],\ \mathbb{E}\big[(\tilde{\mathbf{Y}}_1^\top \tilde{\mathbf{Z}}_1)^2\big]\big\}.$$

First, we calculate $\mathbb{E}\big[(\tilde{\mathbf{Y}}_1^\top \tilde{\mathbf{Y}}_2)^2\big]$ and $\mathbb{E}\big[(\tilde{\mathbf{Z}}_1^\top \tilde{\mathbf{Z}}_2)^2\big]$. Note that

$$\begin{aligned}
\mathbb{E}\big[(\tilde{\mathbf{Y}}_1^\top \tilde{\mathbf{Y}}_2)^2\big] = \mathbb{E}\Bigg[\bigg(\sum_{m=1}^k \tilde{Y}_{1m}\tilde{Y}_{2m}\bigg)^2\Bigg] &\overset{(a)}{=} \sum_{m=1}^k \sum_{m'=1}^k \mathbb{E}\big[\tilde{Y}_{1m}\tilde{Y}_{1m'}\big]\mathbb{E}\big[\tilde{Y}_{2m}\tilde{Y}_{2m'}\big] \\
&\overset{(b)}{=} \sum_{m=1}^k \big\{kp_{Ym} + \sigma_\alpha^2\big\}^2 \\
&= \sum_{m=1}^k \big\{k^2 p_Y^2(m) + 2k\sigma_\alpha^2 p_{Ym} + \sigma_\alpha^4\big\} \\
&= k^2\|\mathbf{p}_Y\|_2^2 \ + \ 2k\sigma_\alpha^2 \ + \ k\sigma_\alpha^4, \qquad (46)
\end{aligned}$$

where step (a) uses the independence between observations, and step (b) uses $\mathbb{E}\big[\tilde{Y}_{1m}\tilde{Y}_{1m'}\big] = \big(kp_{Ym} + \sigma_\alpha^2\big)\mathbb{1}(m = m')$ (and similar equalitiy for $\tilde{\mathbf{Y}}_2$). By symmetry, we also have:

$$\mathbb{E}\big[(\tilde{\mathbf{Z}}_1^\top \tilde{\mathbf{Z}}_2)^2\big] = k^2\|\mathbf{p}_Z\|_2^2 \ + \ 2k\sigma_\alpha^2 \ + \ k\sigma_\alpha^4. \qquad (47)$$

Moving on, we upper bound $\mathbb{E}\big[(\tilde{\mathbf{Y}}_1^\top \tilde{\mathbf{Z}}_1)^2\big]$ as follows:

$$\begin{aligned}
\mathbb{E}\big[(\tilde{Y}_{1m}^\top \tilde{Z}_{1m})^2\big] &= \ \mathbb{E}\Bigg[\bigg(\sum_{m=1}^k \tilde{Y}_{1m}\tilde{Z}_{1m}\bigg)^2\Bigg] \\
&\overset{(a)}{=} \sum_{m=1}^k \sum_{m'=1}^k \mathbb{E}[\tilde{Y}_{1m}\tilde{Y}_{1m'}]\mathbb{E}[\tilde{Z}_{1m}\tilde{Z}_{1m'}] \\
&\overset{(b)}{=} \sum_{m=1}^k \big\{kp_{Ym} + \sigma_\alpha^2\big\}\big\{kp_{Zm} + \sigma_\alpha^2\big\} \\
&\overset{(c)}{\leq} \frac{1}{2}\sum_{m=1}^k \Big[\big\{kp_{Ym} + \sigma_\alpha^2\big\}^2 + \big\{kp_{Zm} + \sigma_\alpha^2\big\}^2\Big] \\
&= \frac{k^2}{2}\|\mathbf{p}_Y\|_2^2 \ + \ \frac{k^2}{2}\|\mathbf{p}_Z\|_2^2 \ + \ 2k\sigma_\alpha^2 \ + \ k\sigma_\alpha^4, \qquad (48)
\end{aligned}$$

where step (a) uses the independence between $\tilde{\mathbf{Y}}_1$ and $\tilde{\mathbf{Z}}_1$, step (b) uses $\mathbb{E}\big[\tilde{Y}_{1m}\tilde{Y}_{1m'}\big] = \big(kp_{Ym} + \sigma_\alpha^2\big)\mathbb{1}(m = m')$ (and similar equality for $\tilde{\mathbf{Z}}_1$), and step (c) applies the inequality $xy \leq x^2/2 + y^2/2$.

Finally, by combining (46), (47) and (48), we obtain the following upper bound:

$$M_{YZ,2}(P,Q) \leq 2(k^2 b + k\sigma_\alpha^2 + k\sigma_\alpha^4). \qquad (49)$$

Remark 22 (Proving the upper bound using `DiscLapU`) *In our proof with continuous Laplace noise from* `LapU`*, we use the independence and the equality* $\mathbb{E}\big[\tilde{Y}_{1m}\tilde{Y}_{1m'}\big] = \big(kp_{Ym} + \sigma_\alpha^2\big)\mathbb{1}(m = m')$*, which holds due to the Laplace noise's moments: mean zero and variance* $\sigma_\alpha^2 = 8k/\alpha^2$*. The discrete Laplace noise of* `DiscLapU` *also satisfies these independence and moment conditions, with variance upper bounded by* $8k/\alpha^2$ *(Lemma 13). Due to these properties, the use of* `DiscLapU` *also leads to* $\mathbb{E}\big[\tilde{Y}_{1m}\tilde{Y}_{1m'}\big] \leq \big(kp_{Ym} + \sigma_\alpha^2\big)\mathbb{1}(m = m')$*. Given this inequality, the entire proof of this section remains valid for* `DiscLapU` *as well.*

Step 2: Apply the two moments method. Using the bounds (45) and (49), we show that condition (38) in the two moments method holds if the separation condition (35) in Lemma 20 is met. Since $\mathbb{E}[U_{n_1,n_2}] = k\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2$, assuming $n_1 \leq n_2$, condition (38) of Lemma 21 is satisfied when

$$k\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 \geq C_1\sqrt{\frac{(k^2b^{1/2} + k\sigma_\alpha^2)\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2}{\beta n_1}} \quad \text{from (45), and} \tag{50}$$

$$k\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 \geq C_2\sqrt{\frac{k^2b + k\sigma_\alpha^2 + k\sigma_\alpha^4}{\gamma\beta n_1^2}} \quad \text{from (49),} \tag{51}$$

for any $P = (P_Y, P_Z) \in \mathcal{P}_{1,\text{multi}}(\rho_{n_1,n_2})$. Since $\sigma_\alpha = 2\sqrt{2k}/\alpha$, condition (50) is satisfied when

$$k\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 \geq C_3\sqrt{\frac{(k^2b^{1/2} + k^2/\alpha^2)\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2}{\beta n_1}}.$$

Using $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y} \leq 2\max\{\sqrt{x}, \sqrt{y}\}$ for $x, y \geq 0$, the condition above is implied by:

$$\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 \geq C_4(\beta)\,\frac{\max\big\{b^{1/4}, 1/\alpha\big\}}{n_1^{1/2}}. \tag{52}$$

On the other hand, condition (51) is satisfied when

$$k\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 \geq C_5\sqrt{\frac{k^2b + k^2/\alpha^2 + k^3/\alpha^4}{\gamma\beta n_1^2}}.$$

Using $\sqrt{x+y+z} \leq 3\max\{\sqrt{x}, \sqrt{y}, \sqrt{z}\}$ for $x, y, z \geq 0$, the condition above is implied by:

$$\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 \geq C_6(\gamma, \beta)\,\frac{\max\{b^{1/4}, 1/\sqrt{\alpha}, k^{1/4}/\alpha\}}{n_1^{1/2}}. \tag{53}$$

Then, by combining (52) and (53), the condition (38) of Lemma 21 is satisfied when

$$
\begin{aligned}
\rho_{n_1,n_2} \geq\ & C_6(\gamma,\beta)\frac{1}{n_1^{1/2}}\max\biggl\{\max\biggl(b^{1/4}, \frac{1}{\alpha}\biggr), \max\biggl(b^{1/4}, \frac{1}{\sqrt{\alpha}}, \frac{k^{1/4}}{\alpha}\biggr)\biggr\} \\
=\ & C_6(\gamma,\beta)\frac{1}{n_1^{1/2}}\max\biggl\{b^{1/4}, \max\biggl(\frac{1}{\alpha}, \frac{1}{\sqrt{\alpha}}, \frac{k^{1/4}}{\alpha}\biggr)\biggr\} \\
=\ & \begin{cases} C_6(\gamma,\beta)\dfrac{1}{n_1^{1/2}}\max\biggl(b^{1/4}, \dfrac{k^{1/4}}{\alpha}\biggr) & \text{if } \alpha \leq k^{1/2}, \\ C_6(\gamma,\beta)\dfrac{1}{n_1^{1/2}}\max\biggl(b^{1/4}, \dfrac{1}{\sqrt{\alpha}}\biggr) & \text{if } \alpha \geq k^{1/2} \end{cases} \\
\overset{(a)}{=}\ & C_6(\gamma,\beta)\frac{1}{n_1^{1/2}}\max\biggl(b^{1/4}, \frac{k^{1/4}}{\alpha}\biggr) \\
=\ & C_6(\gamma,\beta)\max\left(\frac{b^{1/4}}{n_1^{1/2}}, \frac{k^{1/4}}{(n_1\alpha^2)^{1/2}}\right),
\end{aligned}
$$

where the step (a) holds because if $\alpha \geq k^{1/2}$, then we have $b^{1/4} \geq 1/k^{1/4} \geq 1/\sqrt{\alpha} \geq k^{1/4}/\alpha$. This completes the proof of the upper bound through the (discrete) Laplace mechanism. ■

D.1.2 Proof of Lemma 20 Through `RAPPOR` Mechanism

While we follow the same two main steps as in Appendix D.1.1, proving a tight upper bound under the `RAPPOR` mechanism requires a more delicate analysis due to the dependence and bias inherent in private views. To elaborate, recall from Lemma 14 that the entries of an α-LDP view under the `RAPPOR` mechanism are dependent Bernoulli random variables. Specifically, the mth entry of $\tilde{\mathbf{Y}}_i$, denoted as $\tilde{Y}_{im}$, and the m'th entry of $\tilde{\mathbf{Z}}_j$, denoted as $\tilde{Z}_{jm'}$, follow the following distributions:

$$
\tilde{Y}_{im} \sim \mathrm{Ber}\bigl(\alpha_{\mathrm{bf}}\mathbb{1}(Y_i = m) + \delta_{\mathrm{bf}}\bigr) \quad \text{and} \quad \tilde{Z}_{jm'} \sim \mathrm{Ber}\bigl(\alpha_{\mathrm{bf}}\mathbb{1}(Z_j = m') + \delta_{\mathrm{bf}}\bigr),
$$

where

$$
\alpha_{\mathrm{bf}} = \frac{e^{\alpha/2}-1}{e^{\alpha/2}+1} \quad \text{and} \quad \delta_{\mathrm{bf}} = \frac{1}{e^{\alpha/2}+1}.
$$

As in Appendix D.1.1, we employ the two moments method (Lemma 21), which compares the the expectation of the U-statistic with a variance proxy. The challenge is that, unlike `LapU`, for each $m \in [k]$, the mth entry of each α-LDP view in `RAPPOR` is not centered at the scaled multinomial probability:

$$
\mathbb{E}\bigl[\tilde{Y}_{im}\bigr] = \mathbb{E}\bigl[\mathbb{E}\bigl[\tilde{Y}_{im} \,\big|\, Y_{im}\bigr]\bigr] = \alpha_{\mathrm{bf}}p_{Ym} + \delta_{\mathrm{bf}}, \tag{54}
$$

and pairwise negatively correlated for $m, m' \in [k]$ such that $m \neq m'$:

$$
\mathrm{Cov}(\tilde{Y}_{1m}, \tilde{Y}_{1m'}) = -\alpha_{\mathrm{bf}}^2 p_{Ym}p_{Ym'}, \tag{55}
$$

which is proved in Lemma 14. Keeping these facts in mind, we proceed to the upper bound proof.

Proof We follow the two steps of Appendix D.1.1.

Step 1: Bound the moments from above. We start by bounding the variance of conditional expectation terms. We begin with the intermediate form of $M_{Y,1}(P,Q)$ (39) found in Appendix D.1.1:

$$
\begin{aligned}
M_{Y,1}(P,Q) =\ & \mathbb{E}\Big[\big\{\big(\tilde{\mathbf{Y}}_1 - \mathbb{E}[\tilde{\mathbf{Y}}_1]\big)^\top\big(\mathbb{E}[\tilde{\mathbf{Y}}_1] - \mathbb{E}[\tilde{\mathbf{Z}}_1]\big)\big\}^2\Big] \\
\overset{(a)}{=}\ & \mathbb{E}\Big[\big\{\big(\tilde{\mathbf{Y}}_1 - \mathbb{E}[\tilde{\mathbf{Y}}_1]\big)^\top \alpha_{\mathrm{bf}}\big(\mathbf{p}_Y - \mathbf{p}_Z\big)\big\}^2\Big] \\
=\ & \alpha_{\mathrm{bf}}^2 \mathbb{E}\Bigg[\bigg\{\sum_{m=1}^{k}\big(\tilde{Y}_{1m} - \mathbb{E}[\tilde{Y}_{1m}]\big)\big(p_{Ym} - p_{Zm}\big)\bigg\}^2\Bigg] \\
=\ & \alpha_{\mathrm{bf}}^2 \mathbb{E}\Bigg[\sum_{m=1}^{k}\big(\tilde{Y}_{1m} - \mathbb{E}[\tilde{Y}_{1m}]\big)^2\big(p_{Ym} - p_{Zm}\big)^2\Bigg] \\
& + \alpha_{\mathrm{bf}}^2 \mathbb{E}\Bigg[\sum_{1\le m\neq m'\le k}\big(\tilde{Y}_{1m} - \mathbb{E}[\tilde{Y}_{1m}]\big)\big(\tilde{Y}_{1m'} - \mathbb{E}[\tilde{Y}_{1m'}]\big)\big(p_{Ym} - p_{Zm}\big)\big(p_{Ym'} - p_{Zm'}\big)\Bigg] \\
=\ & \alpha_{\mathrm{bf}}^2 \Bigg[\sum_{m=1}^{k}\mathrm{Var}(\tilde{Y}_{1m})\big(p_{Ym} - p_{Zm}\big)^2\Bigg] \\
& + \alpha_{\mathrm{bf}}^2 \Bigg[\sum_{1\le m\neq m'\le k}\mathrm{Cov}(\tilde{Y}_{1m}, \tilde{Y}_{1m'})\big(p_{Ym} - p_{Zm}\big)\big(p_{Ym'} - p_{Zm'}\big)\Bigg] \\
\overset{(b)}{\le}\ & \alpha_{\mathrm{bf}}^3 b^{1/2}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 + \alpha_{\mathrm{bf}}^2\delta_{\mathrm{bf}}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2,
\end{aligned}
\tag{56}
$$

where step (a) uses (54), and step (b) uses (23) in Lemma 15 for the first term and (55) for the second term. Therefore we have

$$
M_{Y,1}(P,Q) \le \alpha_{\mathrm{bf}}^3 b^{1/2}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 + \alpha_{\mathrm{bf}}^2\delta_{\mathrm{bf}}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2, \tag{57}
$$

and by symmetry, we also have

$$
M_{Z,1}(P,Q) \le \alpha_{\mathrm{bf}}^3 b^{1/2}\|\mathbf{p}_Z - \mathbf{p}_Z\|_2^2 + \alpha_{\mathrm{bf}}^2\delta_{\mathrm{bf}}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2. \tag{58}
$$

Next, we examine the expectation of square terms $M_{YZ,2}(P,Q)$. The proof's key technique involves rewriting the kernel $(\tilde{\mathbf{Y}}_1 - \tilde{\mathbf{Z}}_1)^\top(\tilde{\mathbf{Y}}_2 - \tilde{\mathbf{Z}}_2)$ (36) of the U-statistic (10) using $\bar{\mathbf{Y}}_1 := \tilde{\mathbf{Y}}_1 - \mathbb{E}[\tilde{\mathbf{Y}}_1]$, $\bar{\mathbf{Z}}_1 := \tilde{\mathbf{Z}}_1 - \mathbb{E}[\tilde{\mathbf{Y}}_1]$, $\bar{\mathbf{Y}}_2 := \tilde{\mathbf{Y}}_2 - \mathbb{E}[\tilde{\mathbf{Y}}_2]$, and $\bar{\mathbf{Z}}_2 := \tilde{\mathbf{Z}}_2 - \mathbb{E}[\tilde{\mathbf{Y}}_2]$. Formally, we re-express the kernel (36) as the following:

$$
\begin{aligned}
(\tilde{\mathbf{Y}}_1 - \tilde{\mathbf{Z}}_1)^\top(\tilde{\mathbf{Y}}_2 - \tilde{\mathbf{Z}}_2) &= (\tilde{\mathbf{Y}}_1 - \mathbb{E}[\tilde{\mathbf{Y}}_1] + \mathbb{E}[\tilde{\mathbf{Y}}_1] - \tilde{\mathbf{Z}}_1)^\top(\tilde{\mathbf{Y}}_2 - \mathbb{E}[\tilde{\mathbf{Y}}_2] + \mathbb{E}[\tilde{\mathbf{Y}}_2] - \tilde{\mathbf{Z}}_2) \\
&= \big\{(\tilde{\mathbf{Y}}_1 - \mathbb{E}[\tilde{\mathbf{Y}}_1]) - (\tilde{\mathbf{Z}}_1 - \mathbb{E}[\tilde{\mathbf{Y}}_1])\big\}^\top\big\{(\tilde{\mathbf{Y}}_2 - \mathbb{E}[\tilde{\mathbf{Y}}_2]) - (\tilde{\mathbf{Z}}_2 - \mathbb{E}[\tilde{\mathbf{Y}}_2])\big\} \\
&= (\bar{\mathbf{Y}}_1 - \bar{\mathbf{Z}}_1)^\top(\bar{\mathbf{Y}}_2 - \bar{\mathbf{Z}}_2)
\end{aligned}
\tag{59}
$$

Let $\bar{U}_{n_1,n_2}$ denote the U-statistic defined by the kernel $(\bar{\mathbf{Y}}_1 - \bar{\mathbf{Z}}_1)^\top (\bar{\mathbf{Y}}_2 - \bar{\mathbf{Z}}_2)$ in (59). Then $\bar{U}_{n_1,n_2}$ is equal to our original U-statistic (10). Thus, it suffices to prove Lemma 20 with respect to $\bar{U}_{n_1,n_2}$. Let $\bar{M}_{Y,1}(P,Q)$, $\bar{M}_{Z,1}(P,Q)$, and $\bar{M}_{YZ,2}(P,Q)$ be the moments defined as in (37) using the kernel $(\bar{\mathbf{Y}}_1 - \bar{\mathbf{Z}}_1)^\top (\bar{\mathbf{Y}}_2 - \bar{\mathbf{Z}}_2)$. Since $\bar{M}_{Y,1}(P,Q) = M_{Y,1}(P,Q)$ and $\bar{M}_{Z,1}(P,Q) = M_{Z,1}(P,Q)$, the bounds (57) and (58) are also valid for $\bar{M}_{Y,1}(P,Q)$ and $\bar{M}_{Z,1}(P,Q)$. Now, we move on to bounding $\bar{M}_{YZ,2}(P,Q)$ given as

$$\bar{M}_{YZ,2}(P,Q) = \max\{\mathbb{E}[(\bar{\mathbf{Y}}_1^\top \bar{\mathbf{Y}}_2)^2],\, \mathbb{E}[(\bar{\mathbf{Y}}_1^\top \bar{\mathbf{Z}}_1)^2],\, \mathbb{E}[(\bar{\mathbf{Z}}_1^\top \bar{\mathbf{Z}}_2)^2]\}.$$

Let us start by upper bounding $\mathbb{E}[(\bar{\mathbf{Y}}_1^\top \bar{\mathbf{Y}}_2)^2]$:

$$\begin{aligned}
\mathbb{E}[(\bar{\mathbf{Y}}_1^\top \bar{\mathbf{Y}}_2)^2] &= \mathbb{E}\bigg[\bigg\{\sum_{m=1}^{k} \big(\tilde{Y}_{1m} - \mathbb{E}\big[\tilde{Y}_{1m}\big]\big)\big(\tilde{Y}_{2m} - \mathbb{E}\big[\tilde{Y}_{2m}\big]\big)\bigg\}^2\bigg] \\
&= \mathbb{E}\bigg[\sum_{m=1}^{k} \big(\tilde{Y}_{1m} - \mathbb{E}\big[\tilde{Y}_{1m}\big]\big)^2\big(\tilde{Y}_{2m} - \mathbb{E}\big[\tilde{Y}_{2m}\big]\big)^2\bigg] \\
&\quad + \mathbb{E}\bigg[\sum_{1\le m\neq m'\le k} \big(\tilde{Y}_{1m} - \mathbb{E}\big[\tilde{Y}_{1m}\big]\big)\big(\tilde{Y}_{1m'} - \mathbb{E}\big[\tilde{Y}_{1m'}\big]\big)\big(\tilde{Y}_{2m} - \mathbb{E}\big[\tilde{Y}_{2m}\big]\big)\big(\tilde{Y}_{2m'} - \mathbb{E}\big[\tilde{Y}_{2m'}\big]\big)\bigg] \\
&= \sum_{m=1}^{k} \mathrm{Var}(\tilde{Y}_{1m})^2 + \sum_{1\le m\neq m'\le k} \mathrm{Cov}(\tilde{Y}_{1m}, \tilde{Y}_{1m'})^2 \\
&\le 3\alpha_{\mathrm{bf}}^2 b + 2\delta_{\mathrm{bf}}^2 k, \qquad (60)
\end{aligned}$$

where the last inequality uses (24) and (25) in Lemma 15. By symmetry, we also have:

$$\mathbb{E}[(\bar{\mathbf{Z}}_1^\top \bar{\mathbf{Z}}_2)^2] \le 3\alpha_{\mathrm{bf}}^2 b + 2\delta_{\mathrm{bf}}^2 k. \qquad (61)$$

Now we move on to bounding $\mathbb{E}[(\bar{\mathbf{Y}}_1^\top \bar{\mathbf{Z}}_1)^2]$, which is expanded as:

$$\begin{aligned}
\mathbb{E}[(\bar{\mathbf{Y}}_1^\top \bar{\mathbf{Z}}_1)^2] &= \mathbb{E}\bigg[\bigg\{\sum_{m=1}^{k} \big(\tilde{Y}_{1m} - \mathbb{E}\big[\tilde{Y}_{1m}\big]\big)\big(\tilde{Z}_{1m} - \mathbb{E}\big[\tilde{Y}_{1m}\big]\big)\bigg\}^2\bigg] \\
&= \mathbb{E}\bigg[\sum_{m=1}^{k} \big(\tilde{Y}_{1m} - \mathbb{E}\big[\tilde{Y}_{1m}\big]\big)^2\big(\tilde{Z}_{1m} - \mathbb{E}\big[\tilde{Y}_{1m}\big]\big)^2\bigg] \\
&\quad + \mathbb{E}\bigg[\sum_{1\le m\neq m'\le k} \big(\tilde{Y}_{1m} - \mathbb{E}\big[\tilde{Y}_{1m}\big]\big)\big(\tilde{Y}_{1m'} - \mathbb{E}\big[\tilde{Y}_{1m'}\big]\big)\big(\tilde{Z}_{1m} - \mathbb{E}\big[\tilde{Y}_{1m}\big]\big)\big(\tilde{Z}_{1m'} - \mathbb{E}\big[\tilde{Y}_{1m'}\big]\big)\bigg] \\
&= \sum_{m=1}^{k} \mathrm{Var}(\tilde{Y}_{1m})\mathbb{E}\big(\tilde{Z}_{1m} - \mathbb{E}\big[\tilde{Y}_{1m}\big]\big)^2 \qquad (62) \\
&\quad + \sum_{1\le m\neq m'\le k} \mathrm{Cov}(\tilde{Y}_{1m}, \tilde{Y}_{1m'})\mathbb{E}\big(\tilde{Z}_{1m} - \mathbb{E}\big[\tilde{Y}_{1m}\big]\big)\big(\tilde{Z}_{1m'} - \mathbb{E}\big[\tilde{Y}_{1m'}\big]\big), \qquad (63)
\end{aligned}$$

where the last equality uses Lemma 14. We bound the terms (62) and (63) separately. For the term (62), we use the following equality that holds for each of $m \in [k]$:

$$
\begin{aligned}
&\mathbb{E}\big[\tilde{Z}_{1m} - \mathbb{E}[\tilde{Y}_{1m}]\big]^2 \\
&\qquad = \mathbb{E}\big[\tilde{Z}_{1m} - \mathbb{E}[\tilde{Z}_{1m}] + \mathbb{E}[\tilde{Z}_{1m}] - \mathbb{E}[\tilde{Y}_{1m}]\big]^2 \\
&\qquad = \mathbb{E}\big[\tilde{Z}_{1m} - \mathbb{E}[\tilde{Z}_{1m}]\big]^2 + \big[\mathbb{E}[\tilde{Z}_{1m}] - \mathbb{E}[\tilde{Y}_{1m}]\big]^2 + 2\mathbb{E}\big[\tilde{Z}_{1m} - \mathbb{E}[\tilde{Z}_{1m}]\big]\big(\mathbb{E}[\tilde{Z}_{1m}] - \mathbb{E}[\tilde{Y}_{1m}]\big) \\
&\qquad = \mathrm{Var}(\tilde{Z}_{1m}) + \big[\mathbb{E}[\tilde{Z}_{1m}] - \mathbb{E}[\tilde{Y}_{1m}]\big]^2 \\
&\qquad = \mathrm{Var}(\tilde{Z}_{1m}) + \alpha_{\mathrm{bf}}^2\big(p_{Ym} - p_{Zm}\big)^2, \qquad (64)
\end{aligned}
$$

where the last equality uses (54). Using this equality, the term (62) is bounded as:

$$
\begin{aligned}
&\sum_{m=1}^{k}\mathrm{Var}(\tilde{Y}_{1m})\mathbb{E}\big(\tilde{Z}_{1m} - \mathbb{E}\big[\tilde{Y}_{1m}\big]\big)^2 \\
&= \sum_{m=1}^{k}\mathrm{Var}(\tilde{Y}_{1m})\mathrm{Var}(\tilde{Z}_{1m}) + \sum_{m=1}^{k}\mathrm{Var}(\tilde{Y}_{1m})\alpha_{\mathrm{bf}}^2\big(p_{Ym} - p_{Zm}\big)^2 \\
&\overset{(a)}{\leq} \frac{1}{2}\sum_{m=1}^{k}\mathrm{Var}(\tilde{Y}_{1m})^2 + \frac{1}{2}\sum_{m=1}^{k}\mathrm{Var}(\tilde{Z}_{1m})^2 + \alpha_{\mathrm{bf}}^2\sum_{m=1}^{k}\mathrm{Var}(\tilde{Y}_{1m})\big(p_{Ym} - p_{Zm}\big)^2 \\
&\overset{(b)}{\leq} \alpha_{\mathrm{bf}}^2 b + \alpha_{\mathrm{bf}}^3 b^{1/2}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 + \alpha_{\mathrm{bf}}^2\delta_{\mathrm{bf}}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2,
\end{aligned}
$$

where step (a) uses $2ab \leq a^2 + b^2$, and step (b) uses Lemma 15. For the term (63), note that:

$$
\begin{aligned}
&\mathbb{E}\big[(\tilde{Z}_{1m} - \mathbb{E}[\tilde{Y}_{1m}])(\tilde{Z}_{1m'} - \mathbb{E}[\tilde{Y}_{1m'}])\big] \\
&\quad = \mathbb{E}\big[(\tilde{Z}_{1m} - \mathbb{E}[\tilde{Z}_{1m}] + \mathbb{E}[\tilde{Z}_{1m}] - \mathbb{E}[\tilde{Y}_{1m}])(\tilde{Z}_{1m'} - \mathbb{E}[\tilde{Z}_{1m'}] + \mathbb{E}[\tilde{Z}_{1m'}] - \mathbb{E}[\tilde{Y}_{1m'}])\big] \\
&\quad = \mathbb{E}\big[(\tilde{Z}_{1m} - \mathbb{E}[\tilde{Z}_{1m}])(\tilde{Z}_{1m'} - \mathbb{E}[\tilde{Z}_{1m'}])\big] + (\mathbb{E}[\tilde{Z}_{1m}] - \mathbb{E}[\tilde{Y}_{1m}])(\mathbb{E}[\tilde{Z}_{1m'}] - \mathbb{E}[\tilde{Y}_{1m'}]) \\
&\quad = \mathrm{Cov}(\tilde{Z}_{1m}, \tilde{Z}_{1m'}) + \alpha_{\mathrm{bf}}^2\big(p_{Ym} - p_{Zm}\big)\big(p_{Ym'} - p_{Zm'}\big),
\end{aligned}
$$

for each $m \in [k]$. Using this equality, the term (63) is bounded as:

$$
\begin{aligned}
&\sum_{1\leq m\neq m'\leq k}\mathrm{Cov}(\tilde{Y}_{1m}, \tilde{Y}_{1m'})\mathbb{E}\big(\tilde{Z}_{1m} - \mathbb{E}\big[\tilde{Y}_{1m}\big]\big)\big(\tilde{Z}_{1m'} - \mathbb{E}\big[\tilde{Y}_{1m'}\big]\big) \\
&\quad = \sum_{1\leq m\neq m'\leq k}\mathrm{Cov}(\tilde{Y}_{1m}, \tilde{Y}_{1m'})\mathrm{Cov}(\tilde{Z}_{1m}, \tilde{Z}_{1m'}) \\
&\qquad\quad + \sum_{1\leq m\neq m'\leq k}\mathrm{Cov}(\tilde{Y}_{1m}, \tilde{Y}_{1m'})\alpha_{\mathrm{bf}}^2\big(p_{Ym} - p_{Zm}\big)\big(p_{Ym'} - p_{Zm'}\big) \\
&\quad \overset{(a)}{\leq} \frac{1}{2}\sum_{1\leq m\neq m'\leq k}\mathrm{Cov}^2(\tilde{Y}_{1m}, \tilde{Y}_{1m'}) + \frac{1}{2}\sum_{1\leq m\neq m'\leq k}\mathrm{Cov}^2(\tilde{Z}_{1m}, \tilde{Z}_{1m'}) \\
&\qquad\quad + \frac{1}{2}\sum_{1\leq m\neq m'\leq k}\mathrm{Cov}^2(\tilde{Y}_{1m}, \tilde{Y}_{1m'}) + \frac{1}{2}\sum_{1\leq m\neq m'\leq k}\alpha_{\mathrm{bf}}^4\big(p_{Ym} - p_{Zm}\big)^2\big(p_{Ym'} - p_{Zm'}\big)^2
\end{aligned}
$$

$$\stackrel{(b)}{\leq} \frac{3}{2}\alpha_{\mathrm{bf}}^2 b + \frac{\alpha_{\mathrm{bf}}^4}{2}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2$$
$$\stackrel{(c)}{\leq} \frac{5}{2}\alpha_{\mathrm{bf}}^2 b,$$

where step (a) uses $ab \leq a^2/2 + b^2/2$, step (b) uses Lemma 15, and step (c) uses the fact that $0 < \alpha_{\mathrm{bf}} < 1$ for any $\alpha > 0$ and $\mathbf{p}_Y^\top \mathbf{p}_Z > 0$. Collecting the bounds for (62) and (63), we finally bound $\mathbb{E}[(\bar{\mathbf{Y}}_1^\top \bar{\mathbf{Z}}_1)^2]$ as

$$\mathbb{E}[(\bar{\mathbf{Y}}_1^\top \bar{\mathbf{Z}}_1)^2] \leq \frac{7}{2}\alpha_{\mathrm{bf}}^2 b + \alpha_{\mathrm{bf}}^3 b^{1/2}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 + \alpha_{\mathrm{bf}}^2 \delta_{\mathrm{bf}}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2. \tag{65}$$

Collecting the bounds (60), (61), and (65), we finally bound $\bar{M}_{YZ,2}(P,Q)$ as

$$\bar{M}_{YZ,2}(P,Q) \leq \frac{7}{2}\alpha_{\mathrm{bf}}^2 b + \alpha_{\mathrm{bf}}^3 b^{1/2}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 + \alpha_{\mathrm{bf}}^2 \delta_{\mathrm{bf}}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 + 2\delta_{\mathrm{bf}}^2 k. \tag{66}$$

Step 2: Apply the two moments method. Using the bounds in (57), (58) and (66), we show that condition (38) in the two moments method holds if the separation condition (35) in Lemma 20 is met. Since $\mathbb{E}[U_{n_1,n_2}] = \alpha_{\mathrm{bf}}^2\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2$, assuming $n_1 \leq n_2$ and $\gamma = \beta$ for simplicity, the condition (38) of Lemma 21 is satisfied if, for all pairs of distributions $P = (P_Y, P_Z) \in \mathcal{P}_{1,\mathrm{multi}}(\rho_{n_1,n_2})$, the following conditions hold:

$$\alpha_{\mathrm{bf}}^2\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 \geq \sqrt{\frac{\alpha_{\mathrm{bf}}^3 b^{1/2}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 + \alpha_{\mathrm{bf}}^2\delta_{\mathrm{bf}}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2}{\beta n_1}} \ \big(\text{from } (57),(58)\big), \text{ and}$$
$$\alpha_{\mathrm{bf}}^2\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 \geq \sqrt{\frac{\frac{7}{2}\alpha_{\mathrm{bf}}^2 b + \alpha_{\mathrm{bf}}^3 b^{1/2}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 + \alpha_{\mathrm{bf}}^2\delta_{\mathrm{bf}}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 + \delta_{\mathrm{bf}}^2 k}{\gamma\beta n_1^2}} \ \big(\text{from } (66)\big).$$

Since $\sqrt{a} + \sqrt{b} \geq \sqrt{a+b}$ for any nonnegative a and b, the conditions above are satisfied when

$$\alpha_{\mathrm{bf}}^2\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 \geq \sqrt{\frac{\alpha_{\mathrm{bf}}^3 b^{1/2}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 + \alpha_{\mathrm{bf}}^2\delta_{\mathrm{bf}}\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2}{\beta n_1}}, \text{ and}$$
$$\alpha_{\mathrm{bf}}^2\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 \geq \sqrt{\frac{\frac{7}{2}\alpha_{\mathrm{bf}}^2 b + \delta_{\mathrm{bf}}^2 k}{\gamma\beta n_1^2}}.$$

Since $k \geq 2$, the above inequalities are further satisfied when

$$\|\mathbf{p}_Y - \mathbf{p}_Z\|_2^2 \geq C_1(\beta)\max\left\{\frac{b^{1/4}}{\alpha_{\mathrm{bf}}^{1/2} n_1^{1/2}}, \frac{k^{1/4}\delta_{\mathrm{bf}}^{1/2}}{\alpha_{\mathrm{bf}} n_1^{1/2}}\right\}.$$

Therefore, to obtain our desire result, it suffices to show that

$$\max\left\{\frac{b^{1/4}}{\alpha_{\mathrm{bf}}^{1/2} n_1^{1/2}}, \frac{k^{1/4}\delta_{\mathrm{bf}}^{1/2}}{\alpha_{\mathrm{bf}} n_1^{1/2}}\right\} \leq 6\max\left\{\frac{b^{1/4}}{{n_1}^{1/2}}, \frac{k^{1/4}}{n_1^{1/2}\alpha}\right\}, \tag{67}$$

and set $C_u(\gamma,\beta) = 6C_1(\beta)$. Using an indicator function, we separately consider the cases of high privacy ($\alpha \le 1$) and low privacy ($\alpha > 1$). Then we derive (67) for both cases. First, assuming $0 < \alpha \le 1$, we have

$$\max\left\{\frac{b^{1/4}}{\alpha_{\mathrm{bf}}^{1/2} n_1^{1/2}}, \frac{k^{1/4}\delta_{\mathrm{bf}}^{1/2}}{\alpha_{\mathrm{bf}} n_1^{1/2}}\right\} \le \max\left\{\frac{b^{1/4}}{n_1^{1/2}}, \frac{k^{1/4}\delta_{\mathrm{bf}}^{1/2}}{\alpha_{\mathrm{bf}} n_1^{1/2}}\right\} \le \max\left\{\frac{b^{1/4}}{n_1^{1/2}}, \frac{6k^{1/4}}{\alpha n_1^{1/2}}\right\}, \tag{68}$$

where the last inequality uses the fact that for $0 \le \alpha \le 1$, we have

$$\frac{\delta_{\mathrm{bf}}^{1/2}}{\alpha_{\mathrm{bf}}} = \frac{\sqrt{e^{\alpha/2}+1}}{(e^{\alpha/2}-1)} < \frac{6}{\alpha}.$$

Next, assuming $\alpha > 1$, we have

$$\begin{aligned}
\max\left\{\frac{b^{1/4}}{\alpha_{\mathrm{bf}}^{1/2} n_1^{1/2}}, \frac{k^{1/4}\delta_{\mathrm{bf}}^{1/2}}{\alpha_{\mathrm{bf}} n_1^{1/2}}\right\} &= \max\left\{\frac{b^{1/4}}{n_1^{1/2}}\sqrt{\frac{e^{\alpha/2}+1}{e^{\alpha/2}-1}} + \frac{k^{1/4}}{n_1^{1/2}} \frac{1}{\sqrt{e^{\alpha/2}+1}} \frac{e^{\alpha/2}+1}{e^{\alpha/2}-1}\right\} \\
&= \sqrt{\frac{e^{\alpha/2}+1}{e^{\alpha/2}-1}} \max\left\{\frac{b^{1/4}}{n_1^{1/2}}, \frac{k^{1/4}}{n^{1/2}\sqrt{e^{\alpha/2}-1}}\right\} \\
&\overset{(a)}{\le} 3\max\left\{\frac{b^{1/4}}{n_1^{1/2}}, \frac{k^{1/4}}{n^{1/2}\sqrt{e^{\alpha/2}-1}}\right\} \\
&\overset{(b)}{\le} 3\max\left\{\frac{b^{1/4}}{n_1^{1/2}}, \frac{2k^{1/4}}{n^{1/2}\alpha}\right\},
\end{aligned} \tag{69}$$

where step (a) and step (b) use the following inequalities:

$$\sqrt{\frac{e^{\alpha/2}+1}{e^{\alpha/2}-1}} \le 3, \text{ and } \frac{1}{\sqrt{e^{\alpha/2}-1}} < \frac{2}{\alpha},$$

respectively, both of which holds for $\alpha > 1$. Combining (68) and (69), we obtain the following inequality:

$$\begin{aligned}
&\max\left\{\frac{b^{1/4}}{\alpha_{\mathrm{bf}}^{1/2} n_1^{1/2}}, \frac{k^{1/4}\delta_{\mathrm{bf}}^{1/2}}{\alpha_{\mathrm{bf}} n_1^{1/2}}\right\} \\
&\quad = \mathbb{1}(\alpha \le 1)\cdot \max\left\{\frac{b^{1/4}}{\alpha_{\mathrm{bf}}^{1/2} n_1^{1/2}}, \frac{k^{1/4}\delta_{\mathrm{bf}}^{1/2}}{\alpha_{\mathrm{bf}} n_1^{1/2}}\right\} + \mathbb{1}(\alpha > 1)\cdot \max\left\{\frac{b^{1/4}}{\alpha_{\mathrm{bf}}^{1/2} n_1^{1/2}}, \frac{k^{1/4}\delta_{\mathrm{bf}}^{1/2}}{\alpha_{\mathrm{bf}} n_1^{1/2}}\right\} \\
&\quad \le \mathbb{1}(\alpha \le 1)\cdot 6\max\left\{\frac{b^{1/4}}{n_1^{1/2}}, \frac{k^{1/4}}{n^{1/2}\alpha}\right\} + \mathbb{1}(\alpha > 1)\cdot 6\max\left\{\frac{b^{1/4}}{n_1^{1/2}}, \frac{k^{1/4}}{n^{1/2}\alpha}\right\} \\
&\quad = 6\max\left\{\frac{b^{1/4}}{n_1^{1/2}}, \frac{k^{1/4}}{n^{1/2}\alpha}\right\}.
\end{aligned}$$

This completes the proof of the upper bound through `RAPPOR` mechanism. ■

D.2 Lower Bound

The lower bound result follows by combining the lower bound results for the one-sample problem under LDP (Theorem 3.2 in Lam-Weil et al., 2022) with results for the two-sample problem without privacy constraints (Chan et al., 2014; Kim et al., 2022a). Hence we omit the details.

Appendix E. Proof of Theorem 11

Here we prove the upper bound and lower bound results for density two-sample testing. We start with the upper bound (Appendix E.1) and move onto the lower bound (Appendix E.2).

E.1 Upper Bound

The proof of the upper bound proceeds in two steps, where step 1 leverage the result from multinomial testing (Lemma 20), and step 2 uses the discretization error analysis (Appendix B.4).

Proof We verify the two steps mentioned above in order.

Step 1: Separation condition for the probability vectors. We derive a separation condition for the probability vectors $\mathbf{p}_\mathbf{Y}$ and $\mathbf{p}_\mathbf{Z}$ of κ^d categories, obtained by binning the densities $f_\mathbf{Y}$ and $f_\mathbf{Z}$, that ensures our multinomial test distinguishes between them using the samples. By substituting $(\mathbf{p}_Y, \mathbf{p}_Z)$ with $(\mathbf{p}_\mathbf{Y}, \mathbf{p}_\mathbf{Z})$ and k with κ^d in the separation condition (35) of Lemma 20, we obtain the following separation condition:

$$\|\mathbf{p}_\mathbf{Y} - \mathbf{p}_\mathbf{Z}\|_2 \geq C_u(\gamma, \beta)\biggl(\frac{\kappa^{d/4}}{(n_1\alpha^2)^{1/2}} \vee \frac{\max\{\|\mathbf{p}_\mathbf{Y}\|_2^{1/2}, \|\mathbf{p}_\mathbf{Z}\|_2^{1/2}\}}{n_1{}^{1/2}}\biggr). \tag{70}$$

Since we assume that $\|\!|f_\mathbf{Y}|\!\|_\infty < R$ and $\|\!|f_\mathbf{Z}|\!\|_\infty < R$ in Definition 10, $\|\mathbf{p}_\mathbf{Y}\|_2^2$ is upper bounded as

$$\|\mathbf{p}_\mathbf{Y}\|_2^2 = \sum_{m \in [\kappa^d]} \biggl(\int_{B_m} f_\mathbf{Y}(\mathbf{t})\, d\mathbf{t}\biggr)^2 \leq R \sum_{\boldsymbol{k} \in [\kappa]^d} \biggl(\int_{B_m} f_\mathbf{Y}(\mathbf{t})\, d\mathbf{t}\biggr)^2 = R\kappa^d(\kappa^{-d})^2 = R\kappa^{-d},$$

and in a similar manner, we also have $\|\mathbf{p}_\mathbf{Z}\|_2^2 \leq R\kappa^{-d}$. Thus, the condition (70) is implied by:

$$\|\mathbf{p}_\mathbf{Y} - \mathbf{p}_\mathbf{Z}\|_2 \geq C_1(R, \gamma, \beta)\biggl(\frac{\kappa^{d/4}}{(n_1\alpha^2)^{1/2}} \vee \frac{\kappa^{-d/4}}{n_1{}^{1/2}}\biggr). \tag{71}$$

Step 2: Separation condition for the densities. Now we find a density separation condition that ensures the probability vector separation condition (71). From the discretization error analysis (28) and (29) in Appendix B.4, both for the Hölder and Besov case, we have the following inequality:

$$\|\mathbf{p}_\mathbf{Y} - \mathbf{p}_\mathbf{Z}\|_2 \geq C_2(s, R, d, \gamma, \beta)\, \kappa^{-d/2}\bigl(\|\!|f_\mathbf{Y} - f_\mathbf{Z}|\!\|_{\mathbb{L}_2} - \kappa^{-s}\bigr).$$

Using the above inequality, we derive the following sufficient condition for (71):

$$|||f_{\mathbf{Y}} - f_{\mathbf{Z}}|||_{\mathbb{L}_2} \geq C_3(s, R, d, \gamma, \beta) \max\bigg\{ \frac{\kappa^{3d/4}}{(n_1\alpha^2)^{1/2}} + \kappa^{-s}, \frac{\kappa^{d/4}}{{n_1}^{1/2}} + \kappa^{-s} \bigg\}.$$

Recalling from (14) that $\kappa \leq \big(n^{2/(4s+d)} \wedge (n\alpha^2)^{2/(4s+3d)}\big)$, and using

$$\frac{2}{4s+3d}\frac{3d}{4} - \frac{1}{2} = \frac{-2s}{4s+3d} \text{ and } \frac{2}{4s+d}\frac{d}{4} - \frac{1}{2} = \frac{-2s}{4s+d},$$

the condition above is implied by:

$$|||f_{\mathbf{Y}} - f_{\mathbf{Z}}|||_{\mathbb{L}_2} \geq C_3(s, R, d, \gamma, \beta) \max\Big\{ (n_1\alpha^2)^{\frac{-2s}{4s+3d}} + \kappa^{-s}, {n_1}^{\frac{-2s}{4s+d}} + \kappa^{-s} \Big\}.$$

Since $\big(n^{2/(4s+d)} \wedge (n\alpha^2)^{2/(4s+3d)}\big) = \kappa + \delta$, where $0 \leq \delta < 1$ and $\kappa \geq 1$, we have $2\kappa \geq \big(n^{2/(4s+d)} \wedge (n\alpha^2)^{2/(4s+3d)}\big)$. Therefore we have $\kappa^{-s} \leq 2^s\big(n^{2/(4s+d)} \wedge (n\alpha^2)^{2/(4s+3d)}\big)$. Thus the condition above is implied by:

$$|||f_{\mathbf{Y}} - f_{\mathbf{Z}}|||_{\mathbb{L}_2} \geq C_4(s, R, d, \gamma, \beta) \max\Big\{ (n_1\alpha^2)^{\frac{-2s}{4s+3d}} + {n_1}^{\frac{-2s}{4s+d}} \Big\}.$$

This concludes the proof of the upper bound result of Theorem 11. ■

E.2 Lower Bound

By the argument in Arias-Castro et al. (2018), the lower bound for the minimax testing rate in the one-sample problem, denoted as $\rho^*_{n_1,\alpha}$, also provides a lower bound for the minimax separation in two-sample case, denoted as $\rho^*_{n_1,n_2,\alpha}$. Thus, we focus on bounding $\rho^*_{n_1,\alpha}$ in the following proof.

Proof In the one-sample setting, let f_0 denote the known density $f_{\mathbf{Z}}$, and assume that the data-generating distribution $P_{\mathbf{Y}}$ lies within the following class:

Definition 23 (Smooth distribution classes) *Let $\mathcal{P}^{\mathrm{H},1}_{d,s}(R)$ denote the set of distributions $P_{\mathbf{Y}}$ whose density function $f_{\mathbf{Y}}$ satisfies $(f_{\mathbf{Y}} - f_0) \in \mathcal{B}^{\mathrm{H}}_{d,s}(R)$ and $|||f_{\mathbf{Y}} - f_0|||_{\mathbb{L}_2} \leq R$. Similarly, define $\mathcal{P}^{\mathrm{B},1}_{d,s,q}(R)$ by replacing $\mathcal{B}^{\mathrm{H}}_{d,s}(R)$ with $\mathcal{B}^{\mathrm{B}}_{d,s,q}(R)$.*

The one-sample density testing problem is defined as follows: Given $\{\tilde{\mathbf{Y}}_i\}_{i\in[n_1]}$ generated from $P_{\mathbf{Y}}$ and privatized via α-LDP mechanism Q, decide whether $P_{\mathbf{Y}}$ came from

$$\mathcal{P}_0 = \{P_{\mathbf{Y}} \in \mathcal{P} : f_{\mathbf{Y}} = f_0\} \quad \text{or} \quad \mathcal{P}_1(\rho_{n_1}) = \{P_{\mathbf{Y}} \in \mathcal{P} : |||f_{\mathbf{Y}} - f_0|||_{\mathbb{L}_2} \geq \rho_{n_1}\}. \tag{72}$$

We consider two problems: (i) $\mathcal{P} = \mathcal{P}^{\mathrm{H},1}_{d,s}(R)$, and (ii) $\mathcal{P} = \mathcal{P}^{\mathrm{B},1}_{d,s,q}(R)$, with significant overlap in most of the proof steps.

We derive the lower bound by testing the uniform null hypothesis against a carefully constructed mixture alternative. This alternative balances two objectives regarding the distances between the distributions: achieving sufficient separation in $\mathbb{L}_2$ distance without the α-LDP constraint, and ensuring indistinguishability in the total variation distance under the α-LDP constraint. The mixture construction follows the approach in Appendix A.2.2 of Lam-Weil et al. (2022), where uniform densities are perturbed by eigenfunctions of the integral operator created from the α-LDP mechanism. We generalize this construction to multivariate Besov and Hölder smoothness classes.

As defined in Definition 1, let $Q \in \mathcal{Q}_\alpha$ be a non-interactive α-LDP mechanism, and $\{Q_i\}_{i\in[n_1]}$ be its marginals. By Lemma B.1 of Lam-Weil et al. (2022), for each $i \in [n_1]$, there exists a probability measure μ_i such that $Q_i(\cdot \,|\, \mathbf{y}_i)$ is absolutely continuous with respect to μ_i for all $\mathbf{y}_i \in [0,1]^d$. Denote its density with respect to μ_i by $q_i(\tilde{\mathbf{y}}_i \,|\, \mathbf{y}_i)$. Using this, we define the private counterpart of a density function on $[0,1]^d$:

Definition 24 (Private counterpart of density) *Let f be a probability density on $[0,1]^d$ and Q be an α-LDP mechanism with marginal conditional densities $q_i(\tilde{\mathbf{y}}_i \mid \mathbf{y})$ with respect to μ_i. The private counterpart of f is defined as*

$$\tilde{f}_i(\tilde{\mathbf{y}}_i) := \int_{[0,1]^d} q_i(\tilde{\mathbf{y}}_i \mid \mathbf{y}) f(\mathbf{y})\, d\mathbf{y}.$$

Let f_0 denote the uniform density supported on $[0,1]^d$. Using its private counterpart, we introduce the following integral operator whose kernel resembles the likelihood ratio of the privatized densities:

Definition 25 (Privacy mechanism integral operator) *For each $i \in [n_1]$, define an operator $L_i : [0,1]^d \times \tilde{\mathcal{X}} \to \mathbb{R}$ as:*

$$L_i(\mathbf{s}, \mathbf{t}) := \frac{q_i(\mathbf{t}|\mathbf{s})}{\sqrt{\tilde{f}_{0,i}(\mathbf{t})}}.$$

Using L_i as a kernel function, we define an integral operator $K_i : \mathbb{L}_2([0,1]^d) \to \mathbb{L}_2(\tilde{\mathcal{X}}, d\mu_i)$ as:

$$(K_i f)(\cdot) := \int_{[0,1]^d} L_i(\mathbf{y}, \cdot) f(\mathbf{y}) d\mathbf{y}.$$

Let K_i^* denote the adjoint of K_i. Finally, by aggregating all the operators for $i \in [n_1]$, we define a symmetric and positive semidefinite integral operator:

$$K := \sum_{i=1}^{n_1} \frac{1}{n_1} K_i^* K_i.$$

For $f \in \mathbb{L}_2\big([0,1]^d\big)$, the operator K yields the expected squared privatized likelihood ratio:

$$\begin{aligned}\langle Kf, f\rangle &= \frac{1}{n_1}\sum_{i=1}^{n} \int_{\tilde{\mathcal{X}}_i} \frac{\left(\int_{[0,1]^d} q_i(\tilde{\mathbf{y}}_i|\mathbf{y}) f(\mathbf{y}) d\mathbf{y}\right)^2}{\tilde{f}_{0,i}(\tilde{\mathbf{y}}_i)} d\mu_i(\tilde{\mathbf{y}}_i) \\ &= \frac{1}{n_1}\sum_{i=1}^{n} \int_{\tilde{\mathcal{X}}_i} \frac{\tilde{f}_i(\tilde{\mathbf{y}}_i)^2}{\tilde{f}_{0,i}(\tilde{\mathbf{y}}_i)} \frac{dQ_{f_0}}{\tilde{f}_{0,i}(\tilde{\mathbf{y}}_i)} = \frac{1}{n_1}\sum_{i=1}^{n} \mathbb{E}_{Q_{f_0}}\left[\frac{\tilde{f}_i^2(\tilde{\mathbf{y}}_i)}{\tilde{f}_{0,i}^2(\tilde{\mathbf{y}}_i)}\right].\end{aligned} \tag{73}$$

We construct each mixture element by adding carefully chosen eigenfunctions of K as bumps to f_0, ensuring the resulting densities belong to the Hölder or Besov classes. Specifically, we define two orthonormal sets of smoothness-inducing functions—one for Besov smoothness and one for Hölder smoothness. For a positive integer J and $\omega>0$, we set $\kappa = 2^J$ and determine the number of functions as κ^d. The parameter ω controls the perturbation height. The value of J and ω will be optimized later to achieve a tight lower bound. Both orthonormal sets have cardinality κ^d, the supports of their elements partition $[0,1]^d$ without overlap, each element has the $\mathbb{L}_\infty$ norm bounded by $\kappa^{d/2}$, and is orthogonal to f_0. We start by defining the orthonormal set for the Besov ball, consisting of bounded step functions from the multivariate Haar basis.

Definition 26 (Orthonormal set for the Besov ball) *Consider the Haar wavelet basis on $[0,1]^d$ with a prime resolution level $J \in \mathbb{N}$, as defined in Appendix B.3. Set $\kappa = 2^J$ and define $\Lambda(J) := \{0, 1, \ldots, (\kappa-1)\}^d$. Fix a d-dimensional on-off multi-index $\epsilon^* := \{1, 0, 0, \ldots, 0\}$. Let $\mathbb{B}_B$ represent the elements with this multi-index, drawn from the mixed tensor product set at the lowest resolution level J. Specifically, we define:*

$$\mathbb{B}_B := \{\psi^{\epsilon^*}_{J,\boldsymbol{\ell}}\}_{\boldsymbol{\ell}\in\Lambda(J)}.$$

Since all results ahead hold for any non-zero ϵ^*, the specific choice in Definition 26 is arbitrary. Each element of $\mathbb{B}_B$ is orthogonal to f_0 because its multi-index has at least one non-zero entry. In that dimension of non-zero entry, the scaled and shifted Haar wavelet function $\psi(x) = \mathbb{1}_{[0,1/2)}(x) - \mathbb{1}_{[1/2,1)}(x)$, which integrates to zero, contributes to the integral. By Fubini's theorem, this ensures that the entire integral equals zero. Next we define the orthonormal set for Hölder ball:

Definition 27 (Orthonormal set for the Hölder ball) *For a smoothness parameter $s > 0$, let $\bar{\varphi} : \mathbb{R}^d \to \mathbb{R}$ be an infinitely differentiable function supported on $[0,1]^d$ satisfying $\|\bar{\varphi}\|_{\mathbb{L}_2} = 1$, $\|\bar{\varphi}\|_{\mathbb{L}_\infty} < \infty$, $\int_{[0,1]^d} \bar{\varphi} = 0$, and $\|\bar{\varphi}^{(s')}\|_\infty < \infty$ for $s' \in \{1, \ldots, (\lfloor s \rfloor + 1)\}$. For $J \in \mathbb{N}$, set $\kappa = 2^J$ and define $\Lambda(J) := \{0, 1, \ldots, (\kappa - 1)\}^d$, $\varphi_{J,\boldsymbol{\ell}}(\mathbf{y}) := \kappa^{d/2}\bar{\varphi}(\kappa\mathbf{y} - \boldsymbol{\ell})$ for $\boldsymbol{\ell} \in \Lambda(J)$, and $\mathbb{B}_H := \{\varphi_{J,\boldsymbol{\ell}}\}_{\boldsymbol{\ell}\in\Lambda(J)}$, meaning scaled and shifted copies of $\bar{\varphi}$.*

Now we are ready to define the mixtures, one for Besov and one for Hölder ball.

Definition 28 (Mixture alternatives) *For the Hölder ball, let $V = \mathrm{Span}(\{f_0\} \cup \mathbb{B}_H)$. Construct an orthonormal basis of V as $\{f_0\} \cup \{u_t\}_{t=1}^{\kappa^d}$, where each u_t is an eigenfunction of K with eigenvalue λ_t. For each $t \in [\kappa^d]$, since u_t is orthogonal to f_0, it integrates to zero. Recall that $z_\alpha = e^{2\alpha} - e^{-2\alpha}$ for $\alpha > 0$, and set $\tilde{\lambda}_t = (\lambda_t/z_\alpha^2) \vee \kappa^{-d}$. Fix $\omega > 0$. For any $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_{\kappa^d}) \in \{-1,1\}^{\kappa^d}$, define the perturbed density as:*

$$f^{\boldsymbol{\eta}}_{\omega,\kappa} := f_0 + \omega \sum_{t=1}^{\kappa^d} \eta_t \tilde{\lambda}_t^{-1/2} u_t. \tag{74}$$

Since $u_t \in V$ for each $t \in [\kappa^d]$, we can also write:

$$f^{\boldsymbol{\eta}}_{\omega,\kappa} = f_0 + \omega \sum_{t=1}^{\kappa^d} \eta_t \tilde{\lambda}_t^{-1/2} \sum_{\psi\in\mathbb{B}_H} \theta_\psi(u_t)\, \psi, \tag{75}$$

where $\theta_\psi(u_t)$ is the Fourier coefficient of u_t with respect to ψ. The mixture for Hölder ball, denoted as ν_ρ^H, is defined as the uniform probability measure over $\{f_{\omega,\kappa}^{\boldsymbol{\eta}} : \boldsymbol{\eta} \in \{-1,1\}^{\kappa^d}\}$. The mixture for the Besov ball, denoted as ν_ρ^B, is defined through the same procedure as above, with $\mathbb{B}_H$ replaced by $\mathbb{B}_B$.

We now outline the properties of these mixtures, beginning with their $\mathbb{L}_2$ separation from the null:

Lemma 29 ($\mathbb{L}_2$ separation) *For any density $f_{\omega,\kappa}^{\boldsymbol{\eta}}$ drawn from ν_ρ, the $\mathbb{L}_2$ distance from f_0 is bounded as:*

$$\left\vert\kern-0.25ex\left\vert\kern-0.25ex\left\vert f_{\omega,\kappa}^{\boldsymbol{\eta}} - f_0 \right\vert\kern-0.25ex\right\vert\kern-0.25ex\right\vert_{\mathbb{L}_2} \geq \kappa^d \omega \sqrt{3/4}. \tag{76}$$

We next establish a series of conditions on κ and ω to ensure that the mixture elements belong to the smooth density class with high probability.

Lemma 30 (Nonnegativity) *For any $f_{\omega,\kappa}^{\boldsymbol{\eta}}$ drawn from either the distribution ν_ρ^H or ν_ρ^B, we always have $\int_{[0,1]^d} f_{\omega,\kappa}^{\boldsymbol{\eta}}(\mathbf{y})d\mathbf{y} = 1$. Also, $f_{\omega,\kappa}^{\boldsymbol{\eta}} \geq 0$ with a probability greater than $1-\gamma$ if*

$$\omega \leq \frac{\kappa^{-d}}{\sqrt{2\log(2\kappa^d/\gamma)}}. \tag{77}$$

Lemma 31 (Hölder class) *With probability at least $1-\gamma$, a function drawn from ν_ρ^H lies in $\mathcal{B}_{d,s}^{\mathrm{H}}(R)$ if*

$$\omega \leq \frac{R\kappa^{-(d+s)}}{C(s,\bar{\varphi})\sqrt{2\log(2\kappa^d/\gamma)}},$$

where $C(s,\bar{\varphi}) := \max_{s' \in \{1,\ldots,(\lfloor s\rfloor+1)\}} 4|\!|\!|\bar{\varphi}^{(s')}|\!|\!|_\infty$.

Lemma 32 (Besov class) *With probability at least $1-\gamma$, a function drawn from ν_ρ^{B} lies in $\mathcal{B}_{d,s,q}^{\mathrm{B}}(R)$ if*

$$\omega \leq \frac{R\kappa^{-(s+d)}}{\sqrt{2\log(\kappa^d/\gamma)}}. \tag{78}$$

Finally, we characterize the condition for κ and ω that ensures any valid private test fails.

Lemma 33 (Indistinguishability) *Under the conditions of Lemmas 30 and 31, if*

$$\omega \leq (nz_\alpha^2)^{-1/2} \left(\frac{\log\left[1 + 4(1-2\gamma-\beta)^2\right]}{\kappa^d} \right)^{1/4}, \tag{79}$$

then no valid α-LDP test of level γ can distinguish between the uniform density f_0 and the alternatives in (72) *(with $\mathcal{P} = \mathcal{P}_{d,s}^{\mathrm{H},1}(R)$) with type II error less than β. The same conclusion holds when $\mathcal{P}_{d,s}^{\mathrm{H},1}(R)$ and the condition in 31 is replaced by $\mathcal{P}_{d,s,q}^{\mathrm{B},1}(R)$ and the condition in 32, respectively.*

We now gather the results to form a conclusion. Recall from Lemma 29 that $\left|\left|\left| f_{\omega,\kappa}^{\boldsymbol{\eta}} - f_0 \right|\right|\right|_{\mathbb{L}_2} \geq \kappa^d \omega \sqrt{3/4}$. To maximize this separation while satisfying the conditions of Lemmas 30, 31, 32, and 33, we choose ω as the largest possible value:

$$\omega = (nz_\alpha^2)^{-1/2} \left(\frac{\log\left[1 + 4(1 - 2\gamma - \beta)^2\right]}{\kappa^d} \right)^{1/4} \wedge \frac{(R \wedge 1)\kappa^{-(d+s)}}{C(s, \bar{\varphi})\sqrt{2\log(2\kappa^d/\gamma)}}.$$

Substituting this ω into the $\mathbb{L}_2$ separation expression, and collecting the constant terms, we can show that any α-LDP test fails if the $\mathbb{L}_2$ separation between the hypotheses is larger than:

$$C_1(\gamma, \beta, R, s) \left(\frac{\kappa^{3d/4}}{(nz_\alpha^2)^{1/2}} \wedge \frac{\kappa^{-s}}{\sqrt{\log(\kappa^d)}} \right).$$

Taking J as the largest integer such that $2^J \leq C_1(\gamma, \beta, R, s)(n_1 z_\alpha^2)^{2/(4s+3d)}$ and taking $\kappa = 2^J$ leads to the following lower bound for the minimax testing rate:

$$\rho_{n_1,\alpha}^* \geq C_2(\gamma, \beta, R, s) \left[\frac{(nz_\alpha^2)^{-2s/(4s+3d)}}{\sqrt{\log(nz_\alpha^2)}} \right].$$

As justified in Lam-Weil et al. (2022), the α-LDP minimax testing rate is lower bounded by its non-private counterpart. Therefore, combining the result above with the non-private minimax testing rate $n_1^{-2s/(4s+d)}$ of Arias-Castro et al. (2018), we obtain the final result:

$$\rho_{n_1,\alpha}^* \geq C_3(\gamma, \beta, R, s) \left[\frac{(n_1 z_\alpha^2)^{-2s/(4s+3d)}}{\sqrt{\log(n_1 z_\alpha^2)}} \vee n^{-2s/(4s+d)} \right].$$

This completes the proof for the lower bound result in Theorem 11. ■

E.3 Proofs of the Lemmas in Appendix E.2

This section provides proofs for the lemmas used in the lower bound proof given in Appendix E.2.

E.3.1 Proof of Lemma 29

The proof follows the same reasoning as inequality (31), Lemma B.4, and inequality (32) in the Appendix of Lam-Weil et al. (2022), with the only difference being the number of orthonormal basis functions. Therefore, we only outline the key steps.

Proof Since $f_{\omega,\kappa}^{\boldsymbol{\eta}} - f_0$ is a weighted sum of orthonormal singular vectors of Q, its $\mathbb{L}_2$ norm $|||f_{\boldsymbol{\eta}} - f_0|||_{\mathbb{L}_2}$ is bounded below by a quantity involving the sum of the corresponding singular values of Q. Bounding this sum by $z_\alpha^2 = e^{2\alpha} - e^{-2\alpha}$, using the definition of the α-LDP constraint, completes the proof. ■

E.3.2 Proof of Lemma 33

As the proof follows Appendix B.4 and Lemma 3.1 in Lam-Weil et al. (2022), we outline only the key steps.

Proof The argument applies to both $(\mathcal{P}^{\mathrm{H},1}_{d,s}(R), \nu^H_\rho)$ and $(\mathcal{P}^{\mathrm{B},1}_{d,s,q}(R), \nu^B_\rho)$. For simplicity, denote the distribution class and mixture as $\mathcal{P}$ and ν_ρ, respectively. Let $L_{Q^n_{\nu_\rho}}(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{n_1})$ be the likelihood ratio between $Q^n_{\nu_\rho}$ and $Q^n_{f_0}$. Define the total variation distance as $\|\mathbb{P}_{Q^n_{\nu_\rho}} - \mathbb{P}_{Q^n_{f_0}}\|_{\mathrm{TV}} := \frac{1}{2}\int |LQ^n_{\nu_\rho} - 1| \, d\mathbb{P}_{Q^n_{f_0}}$, and chi-square divergence as $\chi^2(Q^n_{\nu_\rho} \| Q^n_{f_0}) := \frac{1}{2}\big(\mathbb{E}_{Q^n_{f_0}}\big[L_{Q^n_{\nu_\rho}}(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{n_1})^2 - 1\big]\big)^{1/2}$. The minimax type II error is lower bounded as:

$$
\begin{aligned}
\inf_{\Delta_{\gamma,Q}} \sup_{f \in \mathcal{P}} \mathbb{P}_{Q^n_f}\Big(\Delta_{\gamma,Q}(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{n_1}) = 0\Big) &\geq \inf_{\Delta_{\gamma,Q}} \mathbb{P}_{Q^n_{\nu_\rho}}\Big(\Delta_{\gamma,Q}(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{n_1}) = 0\Big) - \gamma \\
&\geq 1 - 2\gamma - \|Q^n_{\nu^H_\rho} - Q^n_{f_0}\|_{\mathrm{TV}} \\
&\geq 1 - 2\gamma - \frac{1}{2}\sqrt{\chi^2(Q^n_{\nu_\rho} \| Q^n_{f_0})}, \qquad (80)
\end{aligned}
$$

where the last inequality uses the inequality between chi-square divergence and total variation distance $\chi^2(Q^n_{\nu_\rho} \| Q^n_{f_0}) \geq 4\|Q^n_{\nu_\rho} - Q^n_{f_0}\|^2_{\mathrm{TV}}$ (see, for example, Lemma B.2 of Lam-Weil et al., 2022). Thus it suffices to verify that the condition in (79) implies that the last right-hand side term in (80) is lower bounded by β. Thus the proof is completed by bounding the chi-square divergence as $\chi^2(Q^n_{\nu_\rho} \| Q^n_{f_0}) \leq 1 + \exp(n_1^2 \omega^4 z_\alpha^4 \kappa^d)$, by using the equation (73) and the definition in (74) involving the orthonormal singular vectors of Q. ■

E.3.3 Proof of Lemma 30

The proof proceeds in two steps: we first verify that $f^{\boldsymbol{\eta}}_{\omega,\kappa}$ integrates to 1, and prove the nonnegativeness property, using Hoeffding's inequality.

Proof We verify the two steps mentioned above in order.

Verification of integration to 1. Using the expansion of $f^{\boldsymbol{\eta}}_{\omega,\kappa}$ in terms of the singular vectors from (74), which are orthogonal to f_0 for any $\boldsymbol{\eta}$, the integral is computed as follows:

$$
\int_{[0,1]^d} f^{\boldsymbol{\eta}}_{\omega,\kappa}(\mathbf{y}) \, d\mathbf{y} = \langle f^{\boldsymbol{\eta}}_{\omega,\kappa}, f_0 \rangle_{\mathbb{L}_2} = \langle f_0, f_0 \rangle_{\mathbb{L}_2} + \sum_{t=1}^{\kappa^d} \omega \eta_t \tilde{\lambda}_t^{-1/2} \langle f_0, u_t \rangle = 1,
$$

where the final equality follows from the orthogonality between f_0 and the u_t's. Note that this equation holds for both ν^H_ρ and ν^B_ρ.

Verification of nonnegativeness with high probability. The proof follows the approach of Lemma B.5 in Lam-Weil et al. (2022), with an adjustment for the change of basis. Since the proof is the same for ν^H_ρ and ν^B_ρ, we state if only for ν^B_ρ. For each $\psi \in \mathbb{B}_B$, define an event A_ψ as:

$$A_\psi := \left[\left|\omega \sum_{t=1}^{\kappa^d} \eta_t \tilde{\lambda}_t^{-1/2} \theta_\psi(u_t)\right| \geq \kappa^{-d/2}\right],$$

which belongs to σ-algebra generated by $\boldsymbol{\eta}$. Since the sum of independent scaled Rademacher variables inside A_ψ is sub-Gaussian with variance proxy $\omega^2 \tilde{\lambda}_t^{-1} \sum_{t=1}^{\kappa^d} \theta_\psi^2(u_t)$, by the union bound, we have:

$$\begin{aligned}
\mathbb{P}_{\nu_\rho^B}\left(\bigcup_{\psi \in \mathbb{B}_B} A_\psi\right) &\leq \sum_{\psi \in \mathbb{B}_B} 2\exp\left(\frac{-\kappa^{-d}}{2\omega^2 \tilde{\lambda}_t^{-1} \sum_{t=1}^{\kappa^d} \theta_\psi^2(u_t)}\right) \\
&\overset{(a)}{\leq} 2 \sum_{\psi \in \mathbb{B}_B} \exp\left(\frac{-\kappa^{-2d}}{2\omega^2 \sum_{t=1}^{\kappa^d} \theta_\psi^2(u_t)}\right) \\
&\overset{(b)}{=} 2\kappa^d \exp\left(\frac{-\kappa^{-2d}}{2\omega^2}\right), \qquad (81)
\end{aligned}$$

where step (a) uses $\tilde{\lambda}_t \leq \kappa^{-d}$ and step (b) uses the following equality that stems from the fact that u_t's are orthonormal eigenbasis of V and the orthonormality of $\mathbb{B}_B$:

$$\sum_{t=1}^{\kappa^d} \theta_\psi^2(u_t) = \langle \psi, \sum_{t=1}^{\kappa^d} \langle \psi, u_t \rangle_{\mathbb{L}_2}\, u_t \rangle_{\mathbb{L}_2} = \langle \psi, \psi \rangle_{\mathbb{L}_2} = 1.$$

Using the expansion in (75), the expression $f_{\omega,\kappa}^{\boldsymbol{\eta}} - f_0$ can be represented as:

$$f_{\omega,\kappa}^{\boldsymbol{\eta}} - f_0 = \sum_{\psi \in \mathbb{B}_B} \left(\omega \sum_{t=1}^{\kappa^d} \eta_t \tilde{\lambda}_t^{-1/2} \theta_\psi(u_t)\right) \psi.$$

Each $\psi \in \mathbb{B}_B$ has mutually disjoint support, and the maximum magnitude of ψ across its domain is $\kappa^{d/2}$. Therefore, $f_{\omega,\kappa}^{\boldsymbol{\eta}}$ is nonnegative if the event $\bigcap_{\psi \in \mathbb{B}_B} A_\psi^c$ holds, which occurs with a probability exceeding $1 - 2\kappa^d \exp(-\kappa^{-2d}/2\omega^2)$. Finally, solving $\gamma \leq 2\kappa^d \exp(-\kappa^{-2d}/2\omega^2)$ for ω confirms the condition (77). This completes the proof of Lemma 30. ■

E.3.4 Proof of Lemma 31

For a $f_{\omega,\kappa}^{\boldsymbol{\eta}}$ drawn from ν_ρ^H, we use the basis expansion form in (75) using $\mathbb{B}_H$ (defined in (27)):

$$f_{\omega,\kappa}^{\boldsymbol{\eta}} = f_0 + \omega \sum_{t=1}^{\kappa^d} \eta_t\, \tilde{\lambda}_t^{-1/2} \sum_{\varphi \in \mathbb{B}_H} \theta_\varphi(u_t)\, \varphi.$$

The proof proceeds in two steps: we first identify a sufficient condition for the Lipschitz property and then proceed to establish a sufficient condition for bounded derivatives.

Proof We verify the two steps mentioned above in order.

Hölder continuity of derivatives with high probability. Fix arbitrary $\mathbf{y}_1, \mathbf{y}_2 \in [0,1]^d$. Since each $\varphi \in \mathbb{B}_H$ has disjoint support, we can specify the element of $\mathbb{B}_H$ whose $\lfloor s \rfloor$-derivative has support containing $\mathbf{y}_1$ and $\mathbf{y}_2$, say $\varphi_{\mathbf{y}_1}$ and $\varphi_{\mathbf{y}_2}$. Then using the fact that $\varphi_{\mathbf{y}_1}^{(\lfloor s \rfloor)}(\mathbf{y}_2) - \varphi_{\mathbf{y}_2}^{(\lfloor s \rfloor)}(\mathbf{y}_1) = 0$, we can write the difference of the $\lfloor s \rfloor$-derivative of $f_{\omega,\kappa}^{\boldsymbol{\eta}}$ as

$$
\begin{aligned}
&|(f_{\omega,\kappa}^{\boldsymbol{\eta}})^{(\lfloor s \rfloor)}(\mathbf{y}_1) - (f_{\omega,\kappa}^{\boldsymbol{\eta}})^{(\lfloor s \rfloor)}(\mathbf{y}_2)| \\
&\qquad = \omega \left| \sum_{t=1}^{\kappa^d} \eta_t \tilde{\lambda}_t^{-1/2} \left\{ \theta_{\varphi_{\mathbf{y}_1}}(u_t) \varphi_{\mathbf{y}_1}^{(\lfloor s \rfloor)}(\mathbf{y}_1) - \theta_{\varphi_{\mathbf{y}_2}}(u_t) \varphi_{\mathbf{y}_2}^{(\lfloor s \rfloor)}(\mathbf{y}_2) \right\} \right| \\
&\qquad \leq \omega \left| \sum_{t=1}^{\kappa^d} \theta_{\varphi_{\mathbf{y}_1}}(u_t) \eta_t \tilde{\lambda}_t^{-1/2} \left(\varphi_{\mathbf{y}_1}^{(\lfloor s \rfloor)}(\mathbf{y}_1) - \varphi_{\mathbf{y}_1}^{(\lfloor s \rfloor)}(\mathbf{y}_2) \right) \right| \qquad (82) \\
&\qquad\quad + \omega \left| \sum_{t=1}^{\kappa^d} \theta_{\varphi_{\mathbf{y}_2}}(u_t) \eta_t \tilde{\lambda}_t^{-1/2} \left(\varphi_{\mathbf{y}_2}^{(\lfloor s \rfloor)}(\mathbf{y}_1) - \varphi_{\mathbf{y}_2}^{(\lfloor s \rfloor)}(\mathbf{y}_2) \right) \right|, \qquad (83)
\end{aligned}
$$

where the last inequality uses the triangle inequality. We separately bound the terms (82) and (83), starting with (82). For $\mathbf{y} \in [0,1]^d$, recall from (27) that $\varphi_{\mathbf{y}_1}$ has the form $\kappa^{d/2}\varphi(\kappa\mathbf{y} - \boldsymbol{\ell})$. By the chain rule, its $\lfloor s \rfloor$-mixed partial derivatives are written as

$$
\varphi_{\mathbf{y}_1}^{(\lfloor s \rfloor)}(\mathbf{y}) = \kappa^{(d/2)+\lfloor s \rfloor} \bar{\varphi}^{(\lfloor s \rfloor)}(\kappa\mathbf{y} - \boldsymbol{\ell}). \qquad (84)
$$

Using the above equality, the mean value theorem and the boundedness of all partial mixed derivatives specified in Definition 27, the term (82) is upper bounded by

$$
\omega\kappa^{(d/2)+\lfloor s \rfloor} \left| \sum_{t=1}^{\kappa^d} \theta_{\varphi_{\mathbf{y}_1}}(u_t) \eta_t \tilde{\lambda}_t^{-1/2} \right| C_1(s, \bar{\varphi})\, \kappa \|\mathbf{y}_1 - \mathbf{y}_2\|,
$$

where $C_1(s, \bar{\varphi}) := \left(2 \| \bar{\varphi}^{(\lfloor s \rfloor)} \|_\infty \wedge \| \bar{\varphi}^{(\lfloor s+1 \rfloor)} \|_\infty \right)$. Bounding the term (83) using the same argument, we obtain the following bound:

$$
\begin{aligned}
&|(f_{\omega,\kappa}^{\boldsymbol{\eta}})^{(\lfloor s \rfloor)}(\mathbf{y}_1) - (f_{\omega,\kappa}^{\boldsymbol{\eta}})^{(\lfloor s \rfloor)}(\mathbf{y}_2)| \\
&\leq \omega\kappa^{(d/2)+\lfloor s \rfloor} \left(\left| \sum_{t=1}^{\kappa^d} \theta_{\varphi_{\mathbf{y}_1}}(u_t) \eta_t \tilde{\lambda}_t^{-1/2} \right| + \left| \sum_{t=1}^{\kappa^d} \theta_{\varphi_{\mathbf{y}_2}}(u_t) \eta_t \tilde{\lambda}_t^{-1/2} \right| \right) C_1(s, \bar{\varphi})\, \kappa \|\mathbf{y}_1 - \mathbf{y}_2\|. \qquad (85)
\end{aligned}
$$

For each $\varphi \in \mathbb{B}_H$, let A_φ be an event within the σ-algebra generated by $\boldsymbol{\eta}$, defined as follows:

$$
A_\varphi := \left[\left| \sum_{t=1}^{\kappa^d} \eta_t \tilde{\lambda}_t^{-1/2} \theta_\varphi(u_t) \right| \geq \sqrt{2\kappa^d \log(2\kappa^d/\gamma)} \right]. \qquad (86)
$$

Since the sum of independent scaled Rademacher variables inside A_φ is sub-Gaussian with variance proxy $\tilde{\lambda}_t^{-1} \sum_{t=1}^{\kappa^d} \theta_\varphi^2(u_t)$, by the same technique as (81), we have $\mathbb{P}_{\nu_\rho^H}\left(\bigcap_{\varphi \in \mathbb{B}_H} A_\varphi^c \right) \geq$

$1-\gamma/2$. Under the event $\bigcap_{\varphi \in \mathbb{B}_H} A_\varphi^c$, the bound (85) continues to

$$\begin{aligned}
|(f_{\omega,\kappa}^{\boldsymbol{\eta}})^{(\lfloor s \rfloor)}(\mathbf{y}_1) - (f_{\omega,\kappa}^{\boldsymbol{\eta}})^{(\lfloor s \rfloor)}(\mathbf{y}_2)| &\overset{(a)}{\leq} \omega \kappa^{\frac{d}{2}+\lfloor s \rfloor} 2\sqrt{2\kappa^d \log(2\kappa^d/\gamma)}\, C_1(s,\bar{\varphi})\, \kappa \|\mathbf{y}_1 - \mathbf{y}_2\| \\
&\overset{(b)}{\leq} \omega \kappa^{\frac{d}{2}+\lfloor s \rfloor} \sqrt{2\kappa^d \log(2\kappa^d/\gamma)}\, C_2(s,\bar{\varphi})\; (1 \wedge \kappa \|\mathbf{y}_1 - \mathbf{y}_2\|) \\
&\overset{(c)}{\leq} \omega \kappa^{\frac{d}{2}+\lfloor s \rfloor} \sqrt{2\kappa^d \log(2\kappa^d/\gamma)}\, C_2(s,\bar{\varphi})\; \kappa^{(s-\lfloor s \rfloor)} \|\mathbf{y}_1 - \mathbf{y}_2\|^{s-\lfloor s \rfloor} \\
&= \omega \kappa^{d+s} \sqrt{2 \log(2\kappa^d/\gamma)}\, C_2(s,\bar{\varphi})\; \|\mathbf{y}_1 - \mathbf{y}_2\|^{s-\lfloor s \rfloor}, \qquad (87)
\end{aligned}$$

where $C_2(s,\bar{\varphi}) := \bigl(4\|\bar{\varphi}^{(\lfloor s \rfloor)}\|_{L_\infty} \vee 2\|\bar{\varphi}^{(\lfloor s+1 \rfloor)}\|_{L_\infty}\bigr)$, step (a) uses the event $\bigcap_{\varphi \in \mathbb{B}_H} A_\varphi^c$, step (b) uses the basic inequality that

$$(a \times 1) \wedge (b \times c) \leq (a \vee b) \times (1 \wedge c) \quad \text{for } a,b,c \geq 0,$$

and step (c) holds by the following fact: When $(1 \wedge c) \leq 1$ and $0 \leq s - \lfloor s \rfloor < 1$, it holds that

$$(1 \wedge c) \leq (1 \wedge c)^{s-\lfloor s \rfloor} \leq c^{s - \lfloor s \rfloor}.$$

Starting from inequality (87), we solve for ω in $\omega \kappa^{d+s} \sqrt{2\log(2\kappa^d/\gamma)}\, C_2(s,\bar{\varphi}) \leq R$. Solving for ω, we obtain

$$\omega \leq \frac{R\kappa^{-d+s}}{C_2(s,\bar{\varphi})\sqrt{2\log(2\kappa^d/\gamma)}.} \qquad (88)$$

Therefore, if ω satisfies this inequality, then with probability at least $1-\gamma/2$, the Lipschitz property from Definition 8 holds for a function $f_{\omega,\kappa}^{\boldsymbol{\eta}}$ drawn from ν_ρ^H.

Bounded derivatives with high probability. Fix an integer $s' \in \bigl[\lfloor s \rfloor\bigr]$ and an arbitrary $\mathbf{y} \in [0,1]^d$. Since each $\varphi \in \mathbb{B}_H$ has disjoint support, we can specify the element of $\mathbb{B}_H$ whose s'-derivative has support containing $\mathbf{y}$, say $\varphi_{\mathbf{y}}$. Using (84) to the basis expansion of $f_{\omega,\kappa}^{\boldsymbol{\eta}}$ with respect to $\mathbb{B}_H$, we compute the s'-derivative of $f_{\omega,\kappa}^{\boldsymbol{\eta}}$ as:

$$(f_{\omega,\kappa}^{\boldsymbol{\eta}})^{(s')}(\mathbf{y}) = \; \omega \kappa^{\frac{d}{2}+s'} \Biggl(\sum_{t=1}^{\kappa^d} \eta_t \tilde{\lambda_t}^{-1/2} \theta_{\varphi_{\mathbf{y}}}(u_t) \Biggr) \bar{\varphi}^{(s')}(\kappa \mathbf{y} - \boldsymbol{\ell}).$$

Under the event $\bigcap_{\varphi \in \mathbb{B}_H} A_\varphi^c$ as defined in (86), the derivative above is bounded by

$$(f_{\omega,\kappa}^{\boldsymbol{\eta}})^{(s')}(\mathbf{y}) \leq \kappa^{d+s'} \|\varphi^{(s')}\|_\infty \omega \sqrt{2\log(2\kappa^d/\gamma)}.$$

We solve for ω in $\kappa^{d+s'}\|\varphi^{(s')}\|_\infty \omega \sqrt{2\log(2\kappa^d/\gamma)} \leq R$. Solving for ω, we obtain

$$\omega \leq \frac{R\kappa^{-(d+s')}}{\|\varphi^{(s')}\|_\infty \sqrt{2\log(2\kappa^d/\gamma)}}. \qquad (89)$$

Therefore, if ω satisfies this inequality, then with probability at least $1-\gamma/2$, the s'-derivative of $f_{\omega,\kappa}^{\boldsymbol{\eta}}$ drawn from ν_ρ^H is bounded by R. Combining the above condition for $s' \in \bigl[\lfloor s \rfloor\bigr]$, we conclude that if

$$\omega \leq \frac{R\kappa^{-(d+s)}}{C_3(s,\bar{\varphi})\sqrt{2\log(2\kappa^d/\gamma)}}, \qquad (90)$$

where $C_3(s,\bar{\varphi}) := \max_{s'\in\{1,\ldots,\lfloor s\rfloor\}} |||\bar{\varphi}^{(s')}|||_\infty$, then with probability at least $1-\gamma/2$, the bounded derivatives property of Definition 8 holds for a function $f^{\boldsymbol{\eta}}_{\omega,\kappa}$ drawn from ν^H_ρ.

Conclusion. Combining the conditions (88) and (90), we conclude that if

$$\omega \leq \frac{R\kappa^{-(d+s)}}{C_4(s,\bar{\varphi})\sqrt{2\log(2\kappa^d/\gamma)}}, \tag{91}$$

where $C_4(s,\bar{\varphi}) := \max_{s'\in\{1,\ldots,(\lfloor s\rfloor+1)\}} 4|||\bar{\varphi}^{(s')}|||_\infty$, then with probability at least $1-\gamma$, for a function $f^{\boldsymbol{\eta}}_{\omega,\kappa}$ drawn from ν^H_ρ, we have $f^{\boldsymbol{\eta}}_{\omega,\kappa} \in \mathcal{B}^{\mathrm{H}}_{d,s}(R)$. This completes the proof Lemma 31. ∎

E.3.5 Proof of Lemma 32

Recall from Definition 26 that $\mathbb{B}_B$ is a subset of the multivariate Haar wavelet basis defined in Appendix B.3, specifically the mixed tensor products at resolution level J with on-off multi-index $\epsilon^* = \{1,0,0,\ldots,0\}$. Using the orthonormality of the Haar wavelet basis and the basis expansion of $f^{\boldsymbol{\eta}}_{\omega,\kappa}$ in (75), the Fourier coefficient of $f^{\boldsymbol{\eta}}_{\omega,\kappa} - f_0$ with respect to a Haar wavelet ψ is calculated as

$$\theta_\psi(f^{\boldsymbol{\eta}}_{\omega,\kappa} - f_0) = \begin{cases} \omega \sum_{t=1}^{\kappa^d} \eta_t \tilde{\lambda}_t^{-1/2} \theta_\psi(u_t), & \text{if } \psi \in \mathbb{B}_B, \\ 0, & \text{otherwise.} \end{cases}$$

due to the orthonormality of Haar wavelets (see Appendix B.3 for details). The following proof uses the calculation above.

Proof Recall that we set $\kappa = 2^J$. We proceed the analysis for the case of $1 \leq q < \infty$ and $q = \infty$.

The case of $1 \leq q < \infty$. The qth power of the Besov seminorm (Definition 9) of $f^{\boldsymbol{\eta}}_{\omega,\kappa} - f_0$ is:

$$\begin{aligned} \|f^{\boldsymbol{\eta}}_{\omega,\kappa} - f_0\|^q_{s,2,q} &= \kappa^{sq}\Bigg(\sum_{\psi\in\mathbb{B}_B} \theta^2_\psi(f^{\boldsymbol{\eta}}_{\omega,\kappa} - f_0)\Bigg)^{q/2} \\ &= \kappa^{sq}\Bigg\{\sum_{\psi\in\mathbb{B}_B}\Bigg(\omega\sum_{t=1}^{\kappa^d} \eta_t\ \tilde{\lambda}_t^{-1/2}\theta_\psi(u_t)\Bigg)^2\Bigg\}^{q/2}. \end{aligned} \tag{92}$$

For each $\psi \in \mathbb{B}_B$, let E_ψ denote an event within the σ-algebra generated by $\boldsymbol{\eta}$, defined as follows:

$$E_\psi := \Bigg[\Big|\sum_{t=1}^{\kappa^d} \eta_t \tilde{\lambda}_t^{-1/2}\theta_\psi(u_t)\Big| \geq \sqrt{2\kappa^d\log(\kappa^d/\gamma)}\Bigg].$$

Since the sum of independent scaled Rademacher variables is sub-Gaussian with variance proxy $\tilde{\lambda}_t^{-1}\sum_{t=1}^{\kappa^d}\theta^2_\psi(u_t)$, by the same technique as (81), we have $\mathbb{P}_{\nu^B_\rho}\big(\bigcap_{\psi\in\mathbb{B}_B} E^c_\psi\big) \geq 1-\gamma$.

Under the event $\bigcap_{\psi \in \mathbb{B}_B} E_\psi^c$, and noting that the cardinality of $\mathbb{B}_B$ is κ^d, we bound the quantity in (92) as:

$$\|f_{\omega,\kappa}^{\boldsymbol{\eta}} - f_0\|_{s,2,q}^q \;\le\; \kappa^{sq}\biggl(\sum_{\phi \in \mathbb{B}_B} \omega^2 2\kappa^d \log(\kappa^d/\gamma)\biggr)^{q/2} \;=\; \kappa^{(s+d)q}\omega^q\bigl(2\log(\kappa^d/\gamma)\bigr)^{q/2},$$

which is equivalent to

$$\|f_{\omega,\kappa}^{\boldsymbol{\eta}} - f_0\|_{s,2,q} \;\le\; \kappa^{(s+d)}\omega\sqrt{2\log(\kappa^d/\gamma)}.$$

The case of $q = \infty$. Under the event $\bigcap_{\psi \in \mathbb{B}_B} E_\psi^c$, and noting that the cardinality of $\mathbb{B}_B$ is κ^d, the Besov seminorm (Definition 9) of $(f_{\omega,\kappa}^{\boldsymbol{\eta}} - f_0)$ is bounded as:

$$\begin{aligned}
\|f_{\omega,\kappa}^{\boldsymbol{\eta}} - f_0\|_{s,2,q} &= \kappa^s\biggl(\sum_{\psi \in \mathbb{B}_B} \theta_\psi^2(f_{\omega,\kappa}^{\boldsymbol{\eta}} - f_0)\biggr)^{1/2} \\
&= \kappa^s\biggl\{\sum_{\psi \in \mathbb{B}_B}\biggl(\omega\sum_{t=1}^{\kappa^d} \eta_t\; \tilde{\lambda}_t^{-1/2}\theta_\psi(u_t)\biggr)^2\biggr\}^{1/2} \\
&\le\; \kappa^{(s+d)}\omega\sqrt{2\log(\kappa^d/\gamma)}.
\end{aligned} \tag{93}$$

Conclusion. From (92) and (93), we have $\|f_{\omega,\kappa}^{\boldsymbol{\eta}} - f_0\|_{s,2,q} \;\le\; \kappa^{(s+d)}\omega\sqrt{2\log(\kappa^d/\gamma)}$ for all possible values of q. We solve for ω in $\kappa^{(s+d)}\omega\sqrt{2\log(\kappa^d/\gamma)} \;\le\; R$. Solving for ω, we obtain

$$\omega \le \frac{R\kappa^{-(s+d)}}{\sqrt{2\log(\kappa^d/\gamma)}}. \tag{94}$$

Therefore, if ω satisfies this inequality, then with probability at least $1-\gamma$, a function $f_{\omega,\kappa}^{\boldsymbol{\eta}}$ drawn from ν_ρ^B lies in the Besov ball $\mathcal{B}_{d,s,q}^{\mathrm{B}}(R)$. This concludes the proof of Lemma 32. ∎

Appendix F. Proof of Theorem 12

This section proves the upper bound for the adaptive density test presented in Theorem 12.

Proof For any $(P_{\mathbf{Y}}, P_{\mathbf{Z}}) \in \mathcal{P}_0$, the type I error is controlled through the union bound:

$$\mathbb{E}[\Delta_\gamma^{\mathrm{adapt}}] \le \sum_{t \in [\mathcal{N}]} \mathbb{E}[\Delta_{\gamma/\mathcal{N}}^t] \le \mathcal{N} \cdot \frac{\gamma}{\mathcal{N}} = \gamma.$$

Now we move onto the type II error guarantee, following the proof strategy of Proposition 7.1 of Kim et al. (2022a) for non-private adaptive two-sample test. Recall from (15) that the number of tests is defined as

$$\mathcal{N} = \left\lceil \frac{2}{d}\log_2\left(\frac{n_1}{\log\log n_1}\right) \wedge \frac{2}{3d}\log_2\left(\frac{n_1\alpha^2}{(\log n_1)^2\log\log n_1}\right)\right\rceil.$$

Since $2/d > 2/(d+4s)$ and $2/3d > 2/(3d+4s)$ for $s > 0$, there exists an integer $t^* \in [\mathcal{N}]$ such that

$$\kappa^* := 2^{t^*} \leq 2\left[\left(\frac{n_1}{\log\log n_1}\right)^{2/(4s+d)} \wedge \left(\frac{n_1\alpha^2}{(\log n_1)^2 \log\log n_1}\right)^{2/(4s+3d)}\right] \leq 2^{t^*+1}. \tag{95}$$

Since the type II error of $\Delta_\gamma^{\mathrm{adapt}}$ is upper bounded by that of a single inner test $\Delta_{\gamma/\mathcal{N}}^{t}$, it suffices to control the type II error of $\Delta_{\gamma/\mathcal{N}}^{t^*}$. The dependence of the two-moments method on the significance level as $1/(\gamma/\mathcal{N})$ significantly affects the resulting rate. Therefore, we use the following improved two-moments method (Lemma 34), which has a logarithmic dependence on $1/(\gamma/\mathcal{N})$, at the cost of assuming $n_1 \asymp n_2$:

Lemma 34 (Kim et al., 2022a, Lemma C.1) *For $0 < \gamma < 1/e$, suppose that there is a sufficiently large constant $C > 0$ such that*

$$\mathbb{E}[U_{n_1,n_2}] \geq C \max\left\{\sqrt{\frac{M_{Y,1}(P,Q)}{\beta n_1}}, \sqrt{\frac{M_{Z,1}(P,Q)}{\beta n_2}}, \sqrt{\frac{M_{YZ,2}(P,Q)}{\beta}} \log\left(\frac{1}{\gamma}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right\}$$

for all pairs of distributions $P = (P_{\mathbf{Y}}, P_{\mathbf{Z}}) \in \mathcal{P}_1(\rho_{n_1,n_2})$. Then under the assumptions that $n_1 \asymp n_2$, the type II error of the permutation test over $\mathcal{P}_1(\rho_{n_1,n_2})$ is uniformly bounded by β.

Applying Lemma 34 on the upper bounds of moments in (45) and (49) (for `LapU` or `DiscLapU`) or the bounds in (57), (58) and (66) (for `RAPPOR`), one can verify that the type II error of the two-sample multinomial test with k categories is at most β if

$$\|\mathbf{p}_Y - \mathbf{p}_Z\|_2 \geq C_1(\beta)\sqrt{\log(\mathcal{N}/\gamma)}\left(\frac{k^{1/4}}{(n_1\alpha^2)^{1/2}} \vee \frac{\max\{\|\mathbf{p}_Y\|_2^{1/2}, \|\mathbf{p}_Z\|_2^{1/2}\}}{{n_1}^{1/2}}\right).$$

Note that since $\Delta_{\gamma/\mathcal{N}}^{t^*}$ is $(\alpha/\mathcal{N})$-LDP, the scaling factor for Laplace noise is multiplied by $\mathcal{N}$. As in Appendix E.1, we substitute $(\mathbf{p}_Y, \mathbf{p}_Z)$ with $(\mathbf{p}_{\mathbf{Y}}, \mathbf{p}_{\mathbf{Z}})$ and k with $(\kappa^*)^d$ and use the discretization error analysis result:

$$\|\mathbf{p}_{\mathbf{Y}} - \mathbf{p}_{\mathbf{Z}}\|_2 \geq C_2(s,R,d,\gamma,\beta)\,(\kappa^*)^{-d/2}\big(|||f_{\mathbf{Y}} - f_{\mathbf{Z}}|||_{\mathbb{L}_2} - (\kappa^*)^{-s}\big).$$

As a result, type II error of $\Delta_{\gamma/\mathcal{N}}^{t^*}$ is at most β if:

$$|||f_{\mathbf{Y}} - f_{\mathbf{Z}}|||_{\mathbb{L}_2} \geq C_3(s,R,d,\gamma,\beta)\max\left\{\mathcal{N}\sqrt{\log\mathcal{N}}\frac{(\kappa^*)^{3d/4}}{(n_1\alpha^2)^{1/2}} + (\kappa^*)^{-s}, \sqrt{\log\mathcal{N}}\frac{(\kappa^*)^{d/4}}{{n_1}^{1/2}} + (\kappa^*)^{-s}\right\}.$$

Since we have $\mathcal{N} \leq C(d)\log n_1$ for any $d \geq 1$ and $n_1 \geq e^e$, the condition above is satisfied if $|||f_{\mathbf{Y}} - f_{\mathbf{Z}}|||_{\mathbb{L}_2}$ is larger than:

$$C_4(s,R,d,\gamma,\beta)\max\left\{(\kappa^*)^{3d/4}\left(\frac{(\log n_1)^2\log\log n_1}{n_1\alpha^2}\right)^{1/2} + (\kappa^*)^{-s}, (\kappa^*)^{d/4}\left(\frac{\log\log n_1}{n_1}\right)^{1/2} + (\kappa^*)^{-s}\right\}.$$

Applying the left part of inequality (95), the condition above is implied by:

$$\|\|f_{\mathbf{Y}} - f_{\mathbf{Z}}\|\|_{\mathbb{L}_2} \geq C_5(s,R,d,\gamma,\beta)\bigg\{(\kappa^*)^{-s} + \bigg(\frac{n_1\alpha^2}{(\log n_1)^2 \log\log n_1}\bigg)^{\frac{-2s}{4s+3d}} + \bigg(\frac{n_1}{\log\log n_1}\bigg)^{\frac{-2s}{4s+d}}\bigg\}.$$

Applying the right part of inequality (95), the condition above is implied by:

$$\|\|f_{\mathbf{Y}} - f_{\mathbf{Z}}\|\|_{\mathbb{L}_2} \geq C_6(s,R,d,\gamma,\beta)\bigg\{\bigg(\frac{n_1\alpha^2}{(\log n_1)^2 \log\log n_1}\bigg)^{\frac{-2s}{4s+3d}} \vee \bigg(\frac{n_1}{\log\log n_1}\bigg)^{\frac{-2s}{4s+d}}\bigg\}.$$

This completes the proof of Theorem 12. ■

Appendix G. Supplementary Information for the Numerical Results

This section describes the baseline methods for evaluating our proposed methods in Section 5 and presents additional numerical results.

G.1 Baseline Methods

To the best of our knowledge, no methods exist for nonparametric two-sample testing under LDP. For a reliable evaluation, we adapt two one-sample testing methods from Gaboardi and Rogers (2018) to the two-sample setting for our simulations. The first combines the generalized randomized response (`GenRR`) with a two-sample chi-square statistic (`Chi`), while the second integrates the `RAPPOR` mechanism with a projected two-sample chi-square statistic (`ProjChi`), both calibrated using asymptotic chi-square null distributions. Lastly, we adopt the combination of the generalized randomized response mechanism and ℓ_2-type U-statistic (10) as the third baseline method.

Baseline method 1: `GenRR`+`Chi`. We begin by introducing the generalized randomized response mechanism.

Definition 35 (Generalized randomized response for multinomial data: `GenRR`) *Consider a pooled raw multinomial sample $\{X_i\}_{i\in[n]}$ with k categories. Fix the privacy level $\alpha > 0$. Each data owner perturbs their data point, resulting in the privatized sample $\{\tilde{X}_i\}_{i\in[n]} = \{\tilde{Y}_i\}_{i\in[n_1]} \cup \{\tilde{Z}_i\}_{i\in[n_2]}$; a multinomial random sample where the ith sample $\tilde{X}_i$ represents a modified version of the original ith sample X_i, with category change from $m \in [k]$ to $\tilde{m} \in [k]$ occurring according to the following conditional probabilities:*

$$\mathbb{P}\big(\tilde{X}_i = \tilde{m} \mid X_i = m\big) = \frac{\exp\big(\alpha \mathbb{1}\,(\tilde{m} = m)\big)}{\exp(\alpha) + k - 1}. \tag{96}$$

Gaboardi and Rogers (2018) invoke the exponential mechanism arguement to prove the privacy guarantee. We here present an alternative proof which directly uses the definition of LDP (Definition 1).

Lemma 36 *The random variables $\{\tilde{X}_i\}_{i\in[n]}$ generated by* `GenRR` *are α-LDP views of $\{X_i\}_{i\in[n]}$.*

Proof

$$\begin{aligned}
\sup_{\tilde{m}\in[k],(m,m')\in[k]^2} \frac{\mathbb{P}\big(\tilde{X}_i=\tilde{m}\mid X_i=m\big)}{\mathbb{P}\big(\tilde{X}_i=\tilde{m}\mid X_i=m'\big)} &= \sup_{\tilde{m}\in[k],(m,m')\in[k]^2} \frac{\exp\big(\alpha\mathbb{1}\left(\tilde{m}=m\right)\big)/\big(\exp(\alpha)+k-1\big)}{\exp\big(\alpha\mathbb{1}\left(\tilde{m}=m'\right)\big)/\big(\exp(\alpha)+k-1\big)}\\
&= \sup_{\tilde{m}\in[k],(m,m')\in[k]^2} \exp\Big(\alpha\Big(\mathbb{1}\left(\tilde{m}=m\right)-\mathbb{1}\left(\tilde{m}=m'\right)\Big)\Big)\\
&= \exp(\alpha).
\end{aligned}$$

■

Now we define a chi-square statistic, referred to as `Chi`, which takes the `GenRR` views as inputs:

$$T_{n_1,n_2} := \left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1} (\hat{\boldsymbol{\mu}}_{\tilde{Y}}-\hat{\boldsymbol{\mu}}_{\tilde{Y}})^\top \mathrm{diag}(\hat{\mathbf{p}})^{-1}(\hat{\boldsymbol{\mu}}_{\tilde{Y}}-\hat{\boldsymbol{\mu}}_{\tilde{Y}}), \tag{97}$$

where for $m\in[k]$, the mth elements of $\hat{\boldsymbol{\mu}}_{\tilde{Y}}, \hat{\boldsymbol{\mu}}_{\tilde{Y}}$ and $\hat{\mathbf{p}}$ are defined as:

$$\begin{aligned}
&\hat{\mu}_{\tilde{Y}m} := \frac{1}{n_1}\sum_{i=1}^{n_1}\mathbb{1}(\tilde{Y}_i=m), \quad \hat{\mu}_{\tilde{Z}m} := \frac{1}{n_2}\sum_{j=1}^{n_2}\mathbb{1}(\tilde{Z}_j=m), \text{ and}\\
&\hat{p}_m := \frac{\sum_{i=1}^{n_1}\mathbb{1}(\tilde{Y}_i=m)+\sum_{j=1}^{n_2}\mathbb{1}(\tilde{Z}_j=m)}{n_1+n_2},
\end{aligned}$$

respectively. The test statistic T_{n_1,n_2} (97) can be calibrated using their asymptotic null distributions:

Lemma 37 (Asymptotic null distribution of `GenRR+Chi`) *Fix the number of categories k and the multinomial probability vectors $\mathbf{p}_Y$ and $\mathbf{p}_Z$, each contained in the interior of the set $\Omega := \{\mathbf{p}\in\mathbb{R}^k : p_m>0, \sum_{m=1}^{k-1}p_m\leq 1, p_k = 1-\sum_{m=1}^{k-1}p_m\}$. Fix the privacy level $\alpha>0$. Under the null hypothesis $\mathbf{p}_Y=\mathbf{p}_Z$, for each pair of sample sizes (n_1,n_2), compute the statistic T_{n_1,n_2} based on the `GenRR` α-LDP views $\{\tilde{Y}_i\}_{i\in[n_1]}$ and $\{\tilde{Z}_i\}_{i\in[n_2]}$, generated from $\mathbf{p}_Y$ and $\mathbf{p}_Z$, respectively. Then, as $n_1,n_2\to\infty$, we have $T_{n_1,n_2}\xrightarrow{d}\chi^2_{k-1}$.*

The proof given in Appendix G.4.1 follows directly from Lindberg's central limit theorem (CLT), as outlined in Proposition 2.27 of van der Vaart (1998), and the fact that a chi-square random variable arises from an asymptotic normal distribution with a projection matrix as its covariance.

Baseline method 2: `RAPPOR+ProjChi`. We propose a projected Hotelling's-type statistic, referred to as `ProjChi`, which takes the outputs of `RAPPOR` as input, defined as follows:

$$T^{\mathrm{proj}}_{n_1,n_2} := \left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}(\check{\mathbf{Y}}-\check{\mathbf{Z}})^\top\Pi\hat{\Sigma}^{-1}\Pi(\check{\mathbf{Y}}-\check{\mathbf{Z}}), \tag{98}$$

where $\check{\mathbf{Y}} := \sum_{i=1}^{n_1}\tilde{\mathbf{Y}}_i/n_1$ and $\check{\mathbf{Z}} := \sum_{j=1}^{n_2}\tilde{\mathbf{Z}}_j/n_2$ are the sample means of the `RAPPOR` α-LDP views, $\hat{\Sigma}$ is the pooled empirical covariance matrix:

$$\hat{\Sigma} := \frac{(n_1-1)\hat{\Sigma}_1+(n_2-1)\hat{\Sigma}_2}{n_1+n_2-2}, \tag{99}$$

where

$$\hat{\Sigma}_1 := \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(\tilde{\mathbf{Y}}_i - \check{\mathbf{Y}})(\tilde{\mathbf{Y}}_i - \check{\mathbf{Y}})^\top, \text{ and}$$
$$\hat{\Sigma}_2 := \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(\tilde{\mathbf{Z}}_i - \check{\mathbf{Z}})(\tilde{\mathbf{Z}}_i - \check{\mathbf{Z}})^\top,$$

and $\Pi = I_k - \mathbf{1}\mathbf{1}^\top$ is the projection matrix to the subspace spanned by the one-vector $\mathbf{1}$. The test statistic $T^{\text{proj}}_{n_1,n_2}$ (98) can be calibrated using its asymptotic null distribution:

Lemma 38 (Asymptotic null distribution of `RAPPOR+ProjChi`) *Assume the same multinomial setting as in Lemma 37. Fix the privacy level $\alpha > 0$. Under the null hypothesis $\mathbf{p}_Y = \mathbf{p}_Z$, for each pair of sample sizes (n_1, n_2), compute the statistic T_{n_1,n_2} based on the* `RAPPOR` *α-LDP views $\{\tilde{\mathbf{Y}}_i\}_{i\in[n_1]}$ and $\{\tilde{\mathbf{Z}}_i\}_{i\in[n_2]}$, generated from $\mathbf{p}_Y$ and $\mathbf{p}_Z$, respectively. Then, as $n_1, n_2 \to \infty$, we have $T^{\text{proj}}_{n_1,n_2} \xrightarrow{d} \chi^2_{k-1}$.*

The proof, provided in Appendix G.4.2, follows a similar approach to the proof of Lemma 37. The key difference is that the `RAPPOR` α-LDP views are no longer multinomial samples.

Baseline method 3: `GenRR`+ℓ_2. This method applies the ℓ_2-type U-statistic in (10) to the `GenRR` α-LDP views, with calibration via permutation procedures.

G.2 Power Comparison of `RAPPOR`-Based Methods under Additional Scenarios

Within the tests based on the `RAPPOR` mechanism, as shown in Figures 3, 4, and 6, we observe that `RAPPOR`+`ProjChi` and `RAPPOR`+ℓ_2 exhibit comparable performance. In certain cases, `RAPPOR`+`ProjChi` demonstrates slightly greater power. The dominance of one test over the other depends on the specific scenario, particularly on the privacy level α and the relative signal strength of the chi-square divergence versus the ℓ_2 distance.

In the perturbed uniform distribution scenario considered for Figure 3, the chi-square divergence is relatively strong, while the ℓ_2 distance is relatively weak compared to other potential scenarios. To contrast this, we now present a multinomial testing scenario where the chi-square divergence is relatively weak. In this scenario, the probability vectors follow a power law as in (18), but with different powers. Specifically, we consider the following setup: for $m = 1, \ldots, k = 40$, the mth elements of $\mathbf{p}_Y$ and $\mathbf{p}_Z$ are defined as:

$$p_{Ym} \propto m^{2.45}, \text{ and } p_{Zm} \propto m^{2.3}. \tag{100}$$

We also examine the effect of varying the privacy parameter $\alpha \in \{4, 2, 1\}$ to assess how the privacy level influences the power difference between the two methods. The results presented in Figure 5 demonstrate that in this power law scenario, where the signal is strong in ℓ_2 distance and weak in chi-square divergence, `RAPPOR`+ℓ_2 outperforms `RAPPOR`+`ProjChi`. This difference is more pronounced in the low-privacy setting (meaning larger α values) and becomes more subtle as the privacy level increases (meaning smaller α values).

An intuition behind this phenomenon is as follows. For each $m \in [k]$, the mth entry of the `RAPPOR` output follows $\text{Ber}\left(\alpha_{\text{bf}}\mathbb{1}(Y_i = m) + \delta_{\text{bf}}\right)$, where $\alpha_{\text{bf}} = \alpha/4 + o(\alpha)$ and $\delta_{\text{bf}} =$

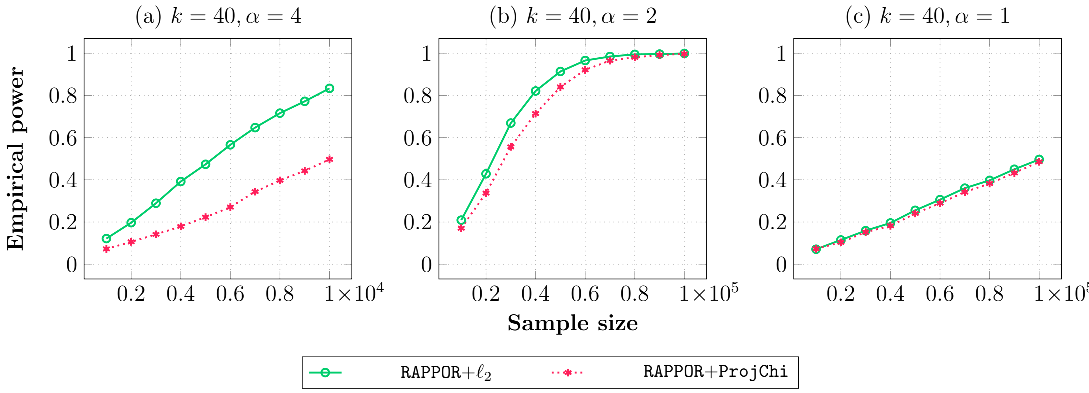


Figure 5: Comparison of the testing power between `RAPPOR`+ℓ_2 and `RAPPOR`+`ProjChi` under the power law alternatives in (100). To ensure a fair comparison, both tests are calibrated using permutation procedures at level $\gamma = 0.05$.

$1/2 + o(1)$, with $o(1)$ representing a term that vanishes as $\alpha \to 0$. Therefore, as α decreases, the distributions approach uniform multinomial distributions, diminishing the difference between chi-square divergence and ℓ_2 distance in the signal.

G.3 Numerical Result for Density Testing for Scale Alternatives

Similar to the location alternative scenario (20), we analyze scenarios involving covariance differences between two d-dimensional Gaussian distributions $P_{\mathbf{Y}} = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{Y}}, \Sigma_{\mathbf{Y}})$ and $P_{\mathbf{Z}} = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{Z}}, \Sigma_{\mathbf{Z}})$. We set the mean vectors and covariance matrices of the Gaussian distributions as

$$\boldsymbol{\mu}_{\mathbf{Y}} = \boldsymbol{\mu}_{\mathbf{Z}} = \mathbf{0}_d, \Sigma_{\mathbf{Y}} = 0.5 \times \mathbf{J}_d + 0.5 \times \mathbf{I}_d, \quad \text{and} \quad \Sigma_{\mathbf{Z}} = 5 \times \Sigma_{\mathbf{Y}}. \tag{101}$$

The power results against the scale alternatives, provided in Figure 6, shows similar patterns to Figure 4.

G.4 Proof of Asymptotic Null Distributions

This section presents the derivation of chi-square asymptotic null distributions of `GenRR+Chi` and `RAPPOR+Chi` α-LDP tests introduced in Appendix G.1, starting with `GenRR+Chi`.

G.4.1 Proof of Lemma 37

The proof is a straightforward application of Lindberg's CLT.

Proof Throughout the proof, assume that the null hypothesis $\mathbf{p}_Y = \mathbf{p}_Z$ holds. Thus, for each $m \in [k]$, we have $p_{\tilde{Y}m} = p_{\tilde{Z}m}$. Without loss of generality, we denote all $p_{\tilde{Z}m}$ as $p_{\tilde{Y}m}$. For each $i \in [n_1]$ and $j \in [n_2]$, we can view $\tilde{Y}_i$ and $\tilde{Z}_j$ as drawn independently from a

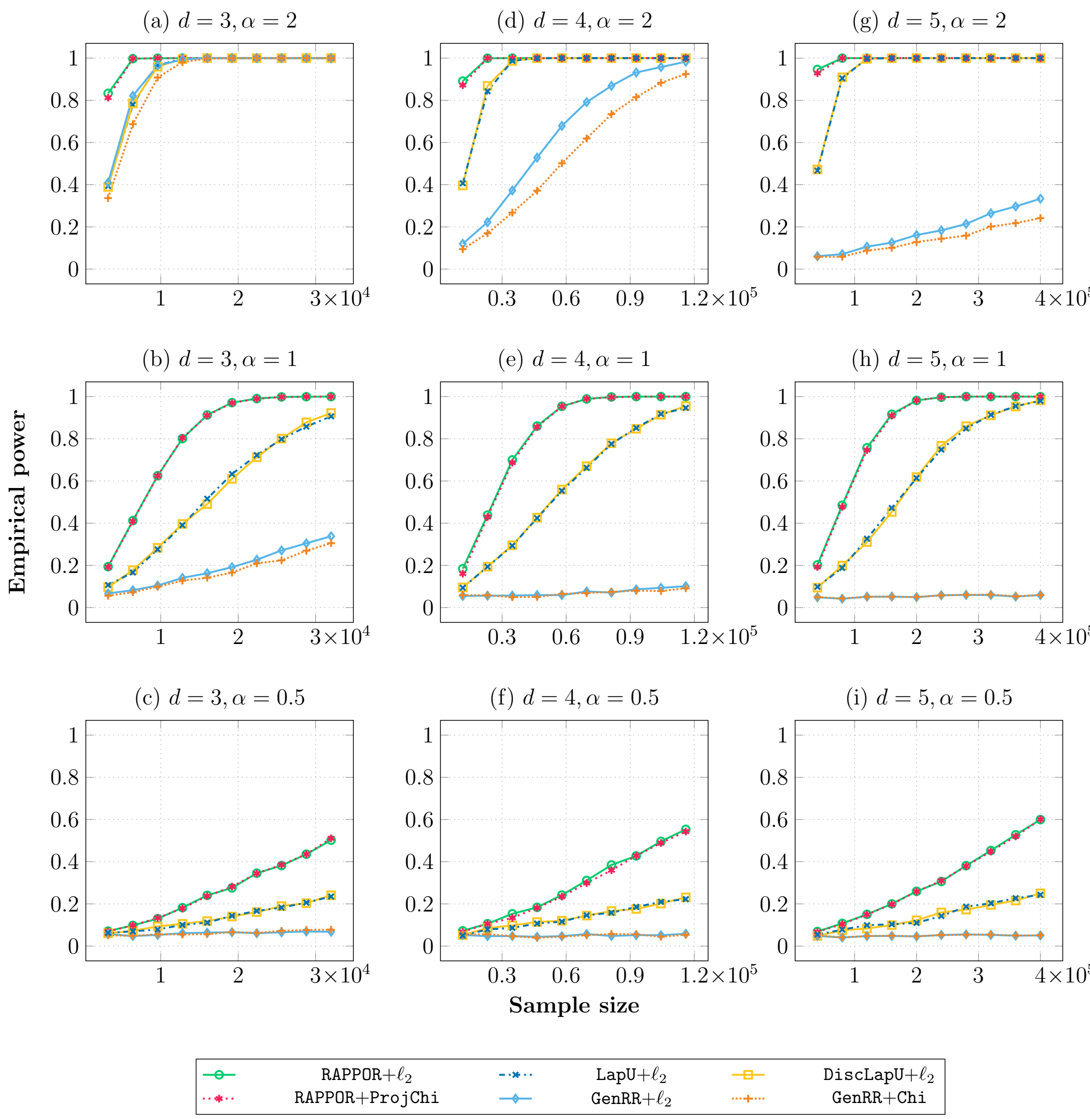


Figure 6: Comparison of the density testing power between our proposed methods (first row in the legend) and baseline methods (second row in the legend) under the scale alternatives in (101). To ensure a fair comparison, all methods are calibrated using permutation procedures at level $\gamma = 0.05$.

multinomial distribution with probability vector defined as:

$$\mathbf{p}_{\tilde{Y}} := \frac{\exp(\alpha)}{\exp(\alpha)+k-1}\mathbf{p}_Y + \frac{1}{\exp(\alpha)+k-1}(\mathbf{1}_k - \mathbf{p}_Y). \tag{102}$$

For each $m \in [k]$, the mth term of $\hat{\boldsymbol{\mu}}_{\tilde{Y}} - \hat{\boldsymbol{\mu}}_{\tilde{Y}}$ can be written as

$$\hat{\mu}_{\tilde{Y}m} - \hat{\mu}_{\tilde{Z}m} = \frac{1}{n_1}\sum_{i=1}^{n_1}\big(\mathbb{1}(\tilde{Y}_i = m) - p_{\tilde{Y}m}\big) - \frac{1}{n_2}\sum_{j=1}^{n_2}\big(\mathbb{1}(\tilde{Z}_j = m) - p_{\tilde{Y}m}\big).$$

To apply Lindeberg's CLT, we construct a triangular array sequence. Each row of the array corresponds to the pooled sample size $N := n_1 + n_2$. Within each row, the columns are indexed by $\ell \in [N]$. For each N and ℓ, we define a k-dimensional random vector $\tilde{\mathbf{X}}_{N[:,\ell]}$, whose mth element, denoted as $\tilde{\mathbf{X}}_{N[m,\ell]}$, is given by:

$$\tilde{\mathbf{X}}_{N[m,\ell]} := \begin{cases} \sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}}\frac{1}{n_1}\big(\mathbb{1}(\tilde{Y}_\ell = m) - p_{\tilde{Y}m}\big) & \text{if } 1 \le \ell \le n_1, \\ \sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}}\frac{1}{n_2}\big(p_{\tilde{Y}m} - \mathbb{1}(\tilde{Z}_{\ell-n_1} = m)\big) & \text{if } (n_1+1) \le \ell \le (n_1+n_2), \end{cases}$$

for $m \in [k]$. Then for each $\ell \in [N]$, we have $\mathbb{E}[\tilde{\mathbf{X}}_{N[:,\ell]}] = 0$ and

$$\sum_{\ell=1}^{N}\tilde{\mathbf{X}}_{N[:,\ell]} = \sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}}(\hat{\boldsymbol{\mu}}_{\tilde{Y}} - \hat{\boldsymbol{\mu}}_{\tilde{Z}}).$$

We also have:

$$\mathrm{Cov}(\tilde{\mathbf{X}}_{N[:,\ell]}) = \begin{cases} \left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}\frac{1}{n_1^2}\big(\mathrm{diag}(\mathbf{p}_{\tilde{Y}}) - \mathbf{p}_{\tilde{Y}}\mathbf{p}_{\tilde{Y}}^\top\big) & \text{if } 1 \le \ell \le n_1, \\ \left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}\frac{1}{n_2^2}\big(\mathrm{diag}(\mathbf{p}_{\tilde{Y}}) - \mathbf{p}_{\tilde{Y}}\mathbf{p}_{\tilde{Y}}^\top\big) & \text{if } (n_1+1) \le \ell \le (n_1+n_2). \end{cases}$$

Now we verify two conditions of Lindberg's CLT in order.

Condition 1: Convergence of the sum of covariances. We aim to verify the following condition:

$$\sum_{\ell=1}^{N}\mathrm{Cov}(\tilde{\mathbf{X}}_{N[:,\ell]}) \to \Sigma \text{ as } N \to \infty, \tag{103}$$

for some fixed covariance matrix Σ. The condition (103) is verified as follows. For each N,

$$\begin{aligned}\sum_{\ell=1}^{N}\mathrm{Cov}(\tilde{\mathbf{X}}_{N[:,\ell]}) &= \left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}\left(\sum_{i=1}^{n_1}\frac{1}{n_1^2}\big(\mathrm{diag}(\mathbf{p}_{\tilde{Y}}) - \mathbf{p}_{\tilde{Y}}\mathbf{p}_{\tilde{Y}}^\top\big) + \sum_{j=1}^{n_2}\frac{1}{n_2^2}\big(\mathrm{diag}(\mathbf{p}_{\tilde{Y}}) - \mathbf{p}_{\tilde{Y}}\mathbf{p}_{\tilde{Y}}^\top\big)\right) \\ &= \mathrm{diag}(\mathbf{p}_{\tilde{Y}}) - \mathbf{p}_{\tilde{Y}}\mathbf{p}_{\tilde{Y}}^\top.\end{aligned}$$

Since the covariance computed above depends on neither N nor ℓ, the condition (103) holds for $\Sigma = \mathrm{diag}(\mathbf{p}_{\tilde{Y}}) - \mathbf{p}_{\tilde{Y}}\mathbf{p}_{\tilde{Y}}^\top$. Now we check the second condition of Lindberg's CLT:

Condition 2: Lindberg's condition. We aim to verify the following condition:

$$\sum_{\ell=1}^{N} \mathbb{E}\big[\|\tilde{\mathbf{X}}_{N[:,\ell]}\|_2^2 \mathbb{1}\left(\|\tilde{\mathbf{X}}_{N[:,\ell]}\|_2 > \epsilon\right)\big] \to 0 \text{ as } N \to \infty, \text{ for any } \epsilon > 0. \tag{104}$$

The condition (104) is verified as follows. For any fixed ℓ, we have:

$$\|\tilde{\mathbf{X}}_{N[:,\ell]}\|_2 = \begin{cases} \sqrt{\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)^{-1}} \dfrac{1}{n_1}\big\{\sum_{m=1}^{k}\big(\mathbb{1}(\tilde{Y}_\ell = m) - p_{\tilde{Y}m}\big)^2\big\}^{1/2} & \text{if } \ell = 1, \ldots, n_1, \\ \sqrt{\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)^{-1}} \dfrac{1}{n_2}\big\{\sum_{m=1}^{k}\big(\mathbb{1}(\tilde{Z}_{\ell-n_1} = m) - p_{\tilde{Y}m}\big)^2\big\}^{1/2} & \text{if } (n_1+1) \leq \ell \leq (n_1+n_2). \end{cases} \tag{105}$$

Note that $|\mathbb{1}(\tilde{Y}_\ell = m) - p_{\tilde{Y}m}| \leq 1$ and $|\mathbb{1}(\tilde{Z}_{\ell-n_1} = m) - p_{\tilde{Y}m}| \leq 1$ for $m \in [k]$. Therefore, for any $\ell \in [N]$, we have the following upper bound:

$$\|\tilde{\mathbf{X}}_{N[:,\ell]}\|_2 \leq \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1}} \frac{\sqrt{k}}{\min(n_1, n_2)},$$

where the right-hand side is not random and does not depend on ℓ. Therefore, when verifying the condition (104), we can pull the indicator out of the expectation and the sum:

$$\begin{aligned} \sum_{\ell=1}^{N} \mathbb{E}\big[\|\tilde{\mathbf{X}}_{N[:,\ell]}\|_2^2 \mathbb{1}\left(\|\tilde{\mathbf{X}}_{N[:,\ell]}\|_2 > \epsilon\right)\big] &\leq \sum_{\ell=1}^{N} \mathbb{E}\left[\|\tilde{\mathbf{X}}_{N[:,\ell]}\|_2^2 \mathbb{1}\left(\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1}} \frac{\sqrt{k}}{\min(n_1, n_2)} > \epsilon\right)\right] \\ &= \mathbb{1}\left(\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1}} \frac{\sqrt{k}}{\min(n_1, n_2)} > \epsilon\right) \sum_{\ell=1}^{N} \mathbb{E}\big[\|\tilde{\mathbf{X}}_{N[:,\ell]}\|_2^2\big]. \end{aligned}$$

From (105), it can be shown that $\sum_{\ell=1}^{N} \mathbb{E}\big[\|\tilde{\mathbf{X}}_{N[:,\ell]}\|_2^2\big] = 2\mathrm{tr}\big(\mathrm{diag}(\mathbf{p}_{\tilde{Y}}) - \mathbf{p}_{\tilde{Y}}\mathbf{p}_{\tilde{Y}}^\top\big)$, where the right-hand side is finite and does not depend on N. For any n_1 and n_2 large enough, we have

$$\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1}} \frac{\sqrt{k}}{\min(n_1, n_2)} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\sqrt{k}}{\min(n_1, n_2)} < \epsilon.$$

Therefore, the condition (104) is satisfied when $n_1, n_2 \to \infty$. Therefore by Lindberg's CLT, we have

$$\sum_{\ell=1}^{N} \tilde{\mathbf{X}}_{N[:,\ell]} = \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1}} (\hat{\mu}_{\tilde{Y}} - \hat{\mu}_{\tilde{Z}}) \xrightarrow{d} \mathcal{N}\big(0, \mathrm{diag}(\mathbf{p}_{\tilde{Y}}) - \mathbf{p}_{\tilde{Y}}\mathbf{p}_{\tilde{Y}}^\top\big).$$

By the weak law of large numbers and Slutsky's theorem, it can be shown that

$$\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}}\mathrm{diag}(\hat{\mathbf{p}})^{-1/2}(\hat{\mu}_{\tilde{Y}}-\hat{\mu}_{\tilde{Z}}) \xrightarrow{d} \mathcal{N}\big(0, I_k - \mathrm{diag}(\mathbf{p}_{\tilde{Y}})^{-1/2}\mathbf{p}_{\tilde{Y}}\mathbf{p}_{\tilde{Y}}^\top \mathrm{diag}(\mathbf{p}_{\tilde{Y}})^{-1/2}\big). \tag{106}$$

Since the covariance matrix in (106) is an identity matrix minus a rank-one matrix formed by an orthonormal vector, its eigenvalues are 0 (with multiplicity 1) and 1 (with multiplicity $k-1$). Thus, the covariance matrix is a projection matrix of rank $k-1$. By a standard result (for example, Lemma 17.1 of van der Vaart, 1998), the test statistic in (97), equivalent to the squared ℓ_2 norm of the left-hand side of (106), converges to a chi-square distribution with $k-1$ degrees of freedom. This concludes the proof of Lemma 37. ■

G.4.2 Proof of Lemma 38

The proof is a straightforward application of Lindberg's CLT.

Proof Throughout the proof, assume that the null hypothesis $\mathbf{p}_Y = \mathbf{p}_Z$ holds. Thus, for each $m \in [k]$, we have $\mathbb{E}(\tilde{Y}_{1m}) = \mathbb{E}(\tilde{Z}_{1m})$. Without loss of generality, we denote all $\mathbb{E}(\tilde{Z}_{1m})$ as $\mathbb{E}(\tilde{Y}_{1m})$. The proof leverages the content and structure of the argument used in the proof for T_{n_1,n_2} in Appendex G.4.1. To apply Lindeberg's CLT, we construct a triangular array sequence. Each row of the array corresponds to the pooled sample size $N := n_1+n_2$. Within each row, the columns are indexed by $\ell \in [N]$. For each N and ℓ, we define a k-dimensional random vector $\tilde{\mathbf{X}}_{N[:,\ell]}$,whose mth element, denoted as $\tilde{\mathbf{X}}_{N[m,\ell]}$, is given by:

$$\tilde{\mathbf{X}}_{N[m,\ell]} := \begin{cases} \sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}}\frac{1}{n_1}\big(\tilde{Y}_{\ell m} - \mathbb{E}(\tilde{Y}_{1m})\big) & \text{if } 1 \le \ell \le n_1, \\ \sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}}\frac{1}{n_2}\big(\mathbb{E}(\tilde{Y}_{1m}) - \tilde{Z}_{(\ell-n_1)m}\big) & \text{if } (n_1+1) \le \ell \le (n_1+n_2), \end{cases}$$

for $m \in [k]$, where we recall from (54) that

$$\mathbb{E}(\tilde{Y}_{1m}) = \frac{(e^{\alpha/2}-1)p_{Ym}+1}{e^{\alpha/2}+1}.$$

Now we verify two conditions of Lindberg's CLT in order.

Condition 1: Convergence of the sum of covariances. To verify the condition (103), first note that:

$$\begin{aligned} \sum_{\ell=1}^{N}\mathrm{Cov}(\tilde{\mathbf{X}}_{N[:,\ell]}) &= \left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}\left(\sum_{i=1}^{n_1}\frac{1}{n_1^2}\mathrm{Cov}(\tilde{\mathbf{Y}}_i) + \sum_{j=1}^{n_2}\frac{1}{n_2^2}\mathrm{Cov}(\tilde{\mathbf{Z}}_j)\right) \\ &= \mathrm{Cov}(\tilde{\mathbf{Y}}_1) \\ &= \left(\frac{e^{\alpha/2}-1}{e^{\alpha/2}+1}\right)^2\big(\mathrm{diag}(\mathbf{p}_Y) - \mathbf{p}_Y\mathbf{p}_Y^\top\big) + \frac{e^{\alpha/2}}{(e^{\alpha/2}+1)^2}I_d, \end{aligned}$$

where the last calculation is from Lemma 14. Since the covariance computed above depends on neither N nor ℓ, the condition (103) is satisfied with $\Sigma = \mathrm{Cov}(\tilde{\mathbf{Y}}_1)$.

Condition 2: Lindberg's condition. To verify the condition (104) , first note that $0 \leq \tilde{Y}_{\ell m} - \mathbb{E}(\tilde{Y}_{1m}) \leq 1$ and $0 \leq \mathbb{E}(\tilde{Y}_{1m}) - \tilde{Z}_{(\ell-n_1)m} \leq 1$. Therefore by the same analysis as the proof of T_{n_1,n_2}, the condition (104) is also satisfied. Therefore, by Linbderg's CLT, we have:

$$\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}}(\check{\mathbf{Y}}-\check{\mathbf{Z}}) \xrightarrow{d} \mathcal{N}(0, \mathrm{Cov}(\tilde{\mathbf{Y}}_1)). \tag{107}$$

Note that $\mathrm{Cov}(\tilde{\mathbf{Y}}_1)$ has an eigenvector $\mathbf{1}$, which corresponds to an eigenvalue which is neither 0 nor 1. To turn the covariance matrix into a projection matrix, we delete the eigenvector $\mathbf{1}$ by pre-multiplying the random vector in the left-hand side of (107) by $\mathrm{Cov}(\tilde{\mathbf{Y}}_1)^{-1/2}\Pi$. Then we have:

$$\mathrm{Cov}(\tilde{\mathbf{Y}}_1)^{-1/2}\Pi\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}}(\check{\mathbf{Y}}-\check{\mathbf{Z}}) \xrightarrow{d} \mathcal{N}\big(0, \mathrm{Cov}(\tilde{\mathbf{Y}}_1)^{-1/2}\Pi\mathrm{Cov}(\tilde{\mathbf{Y}}_1)\Pi\mathrm{Cov}(\tilde{\mathbf{Y}}_1)^{-1/2}\big).$$

For a map $h : A \to A^{-1/2}$, define $\mathrm{Disc}(h) := \{A \in \mathbb{R}^{p\times p} : h \text{ is not continuous at } A\}$. By Lemma 5.6 of Gaboardi and Rogers (2018), $\mathrm{Cov}(\tilde{\mathbf{Y}}_1)$ is invertible for any $\alpha > 0$ and any $\mathbf{p}_Y \in \mathrm{int}(\Omega)$. Therefore we have $\mathrm{Cov}(\tilde{\mathbf{Y}}_1) \notin \mathrm{Disc}(h)$ almost surely. Thus by the weak law of large numbers and the continuous mapping theorem, we have $\hat{\Sigma}^{-1/2} \xrightarrow{p} \mathrm{Cov}(\tilde{\mathbf{Y}}_1)^{-1/2}$. Then by Slutsky's theorem, we have

$$\hat{\Sigma}^{-1/2}\Pi\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}}(\check{\mathbf{Y}}-\check{\mathbf{Z}}) \xrightarrow{d} \mathcal{N}(0, \mathrm{Cov}(\tilde{\mathbf{Y}}_1)^{-1/2}\Pi\mathrm{Cov}(\tilde{\mathbf{Y}}_1)\Pi\mathrm{Cov}(\tilde{\mathbf{Y}}_1)^{-1/2}). \tag{108}$$

By the analysis of Gaboardi and Rogers (2018), $\mathrm{Cov}(\tilde{\mathbf{Y}}_1)^{-1/2}\Pi\mathrm{Cov}(\tilde{\mathbf{Y}}_1)\Pi\mathrm{Cov}(\tilde{\mathbf{Y}}_1)^{-1/2}$ is an identity matrix except one of the entries on the diagonal being zero, thus a projection matrix with rank $k-1$. Therefore, by a standard result (for example, Lemma 17.1 van der Vaart, 1998), the test statistic (98), which is equivalent to squared ℓ_2 norm of the random vector on the left-hand side of (108), converges to a chi-square distribution with $k-1$ degrees of freedom. This concludes the proof of Lemma 38. ■

Appendix H. Suboptimality of Generalized Randomized Response

This section illustrates that the generalized randomized response mechanism in (96) can lead to suboptimal power. Specifically, we show that the `GenRR`+ℓ_2 test performs suboptimally in certain privacy regimes where the number of categories k increases with the sample size n_1, but the privacy level α is fixed. This negative result is further supported by our numerical studies in Section 5.

Theorem 39 (Asymptotic powerlessness of `GenRR`+ℓ_2 test) *Let $(\mathbf{p}_Y, \mathbf{p}_Z)$ be a pair of data-generating multinomial distributions with k categories, where $\mathbf{p}_Z$ is a uniform distribution and $\mathbf{p}_Y$ is a perturbed uniform distribution parametrized by $\epsilon > 0$. Formally, for each*

$m \in [k]$, the mth entry of $\mathbf{p}_Y$ and $\mathbf{p}_Z$ are defined as:

$$p_{Ym} := \frac{1}{k} + \frac{(-1)^m \epsilon}{\sqrt{k}}, \quad p_{Zm} := \frac{1}{k}, \tag{109}$$

ensuring that $\|\mathbf{p}_Y - \mathbf{p}_Z\|_2 = \epsilon$. Assume the privacy parameter $\alpha > 0$ is sufficiently small so that the minimax separation, given in equation (7), satisfies

$$\rho^*_{n_1,n_2,\alpha} \asymp \frac{k^{1/4}}{(n_1\alpha^2)^{1/2}}. \tag{110}$$

For a sufficiently large, but fixed constant $C > 0$, set

$$\epsilon = \frac{C \cdot k^{1/4}}{(n_1\alpha^2)^{1/2}}, \tag{111}$$

and consider the regime where

$$\sqrt{k}\epsilon = \frac{C \cdot k^{3/4}}{(n_1\alpha^2)^{1/2}} \to 0, \tag{112}$$

which guarantees that $0 < p_{Ym} < 1$ in (109) for each $m \in [k]$ and for sufficiently large n_1. Under this regime, the `GenRR`+ℓ_2 test is asymptotically powerless, meaning its asymptotic power becomes at most the size.

Given the setting in (111) from Theorem 39, an optimal test would achieve non-trivial power greater than the significance level by selecting a sufficiently large C. However, Theorem 39 proves that the `GenRR`+ℓ_2 test becomes asymptotically powerless for any fixed $C > 0$, thereby underscoring its suboptimality.

Proof For the proof, we analyze the power function of the test statistic. To achieve this, we use high-dimensional asymptotics for U-statistics, specifically Corollary 3.3 and Theorem 3.3 from Kim (2020), which are restated as Theorem 18 and Theorem 19, respectively.

Let $w := (e^\alpha - 1)/(e^\alpha + k - 1)$. Recall from (102) that the α-LDP view $\{\tilde{Y}_i\}_{i\in[n_1]}$ via `GenRR` mechanism is equivalent to a random sample drawn from a mixture of $\mathbf{p}_Y$ and $\mathbf{p}_Z$ defined as follows:

$$\begin{aligned}\mathbf{p}_{\tilde{Y}} := \frac{e^\alpha}{e^\alpha + k - 1}\mathbf{p}_Y + \frac{1}{e^\alpha + k - 1}(\mathbf{1}_k - \mathbf{p}_Y) &= \frac{e^\alpha - 1}{e^\alpha + k - 1}\mathbf{p}_Y + \frac{k}{e^\alpha + k - 1}\frac{1}{k}\mathbf{1}_k \\ &= w\mathbf{p}_Y + (1-w)\mathbf{p}_Z,\end{aligned} \tag{113}$$

Similarly, $\{\tilde{Z}_i\}_{i\in[n_2]}$ can be viewed as a random sample drawn from $\mathbf{p}_{\tilde{Z}} := w\mathbf{p}_Z + (1-w)\mathbf{p}_Z = \mathbf{p}_Z$. Note that the ℓ_2 distance between the probability vectors shrinks to $\|\mathbf{p}_{\tilde{Y}} - \mathbf{p}_{\tilde{Z}}\|_2 = w\epsilon$.

For convenience, let $\{\tilde{\mathbf{Y}}_i\}_{i\in[n_1]}$ be the one-hot vectorized version of $\{\tilde{Y}_i\}_{i\in[n_1]}$. We define a one-sample U-statistic for uniformity testing as follows:

$$U_{n_1} := \frac{2}{n_1(n_1-1)} \sum_{1\le i<j\le n_1} h(\tilde{\mathbf{Y}}_i, \tilde{\mathbf{Y}}_j), \tag{114}$$

where $h(\tilde{\mathbf{Y}}_i, \tilde{\mathbf{Y}}_j) := \big(\tilde{\mathbf{Y}}_i - \mathbf{p}_Z\big)^\top \big(\tilde{\mathbf{Y}}_j - \mathbf{p}_Z\big)$, which is a special case of the two-sample statistic U_{n_1,n_2} in (10) with $n_2 = \infty$. Now we analyze the power function of the test based on U_{n_1} by specifying the critical value, the distribution of the test statistic under the alternative, and the power function in order.

Critical value. We specify the critical value by establishing the asymptotic normality of the test statistic under the null by using Theorem 18. Since the statistic (114) is equivalent to the statistic U_I in (32) of Theorem 18, it suffices to show that $n_1/\sqrt{k} \to \infty$ to establish the asymptotic normality. Note that the assumption $\sqrt{k}\epsilon = Ck^{3/4}/(n_1\alpha^2)^{1/2} \to 0$ in (112) implies $n_1/\sqrt{k} \to \infty$ for fixed α. Therefore, under the null of $\mathbf{p}_Y = \mathbf{p}_Z$, which implies $\mathbf{p}_{\tilde{Y}} = w\mathbf{p}_Z + (1-w)\mathbf{p}_Z = \mathbf{p}_Z$, we have

$$\sqrt{\binom{n_1}{2}} \frac{U_{n_1}}{\sqrt{k^{-1}(1-k^{-1})}} \xrightarrow{d} \mathcal{N}(0,1).$$

Based on this asymptotic normality, an asymptotically exact critical value is specified as follows:

$$\text{Reject } H_0 \text{ if } U_{n_1} \geq z_\gamma \sqrt{\frac{2(1-1/k)}{kn_1(n_1-1)}}. \tag{115}$$

Here z_γ denotes the upper γ quantile of $W \sim \mathcal{N}(0,1)$ satisfying $\mathbb{P}(W > z_\gamma) = \gamma$.

Alternative distribution. Next we derive that the asymptotic distribution of the test statistic under the alternative specified in (109). By substituting $\|\mathbf{p}_{\tilde{Y}} - \mathbf{p}_Z\|_2^2 = w^2\epsilon^2$ into (33), we can establish the following asymptotic normality:

$$\sqrt{\frac{n_1(n_1-1)}{2}} \frac{U_{n_1} - w^2\epsilon^2}{\sqrt{\mathrm{tr}(\Sigma_{\tilde{Y}}^2) + 2(n_1-1)(\mathbf{p}_{\tilde{Y}} - \mathbf{p}_Z)^\top \Sigma_{\tilde{Y}} (\mathbf{p}_{\tilde{Y}} - \mathbf{p}_Z)}} \xrightarrow{d} \mathcal{N}(0,1), \tag{116}$$

provided that the conditions C1, C2, and C3 of Theorem 19 hold. We verify these conditions in order.

Verification of the condition C1. The condition C1 is restated as follows:

$$\frac{\mathrm{tr}(\Sigma_{\tilde{Y}}^4)}{\big\{\mathrm{tr}(\Sigma_{\tilde{Y}}^2)\big\}^2} \to 0.$$

We first analyze the denominator $\big\{\mathrm{tr}(\Sigma_{\tilde{Y}}^2)\big\}^2$. Since $\Sigma_{\tilde{Y}} = \mathrm{diag}(\mathbf{p}_{\tilde{Y}}) - \mathbf{p}_{\tilde{Y}}\mathbf{p}_{\tilde{Y}}^\top$, we have

$$\Sigma_{\tilde{Y}}^2 = \{\mathrm{diag}(\mathbf{p}_{\tilde{Y}})\}^2 + (\mathbf{p}_{\tilde{Y}}\mathbf{p}_{\tilde{Y}}^\top)^2 - \mathrm{diag}(\mathbf{p}_{\tilde{Y}})\mathbf{p}_{\tilde{Y}}\mathbf{p}_{\tilde{Y}}^\top - \mathbf{p}_{\tilde{Y}}\mathbf{p}_{\tilde{Y}}^\top \mathrm{diag}(\mathbf{p}_{\tilde{Y}}).$$

Using the equalities $\big(\mathbf{p}_{\tilde{Y}}\mathbf{p}_{\tilde{Y}}^\top\big)^2 = \|\mathbf{p}_{\tilde{Y}}\|_2^2 \mathbf{p}_{\tilde{Y}}\mathbf{p}_{\tilde{Y}}^\top$, $\mathrm{tr}\big(\mathbf{p}_{\tilde{Y}}\mathbf{p}_{\tilde{Y}}^\top\big) = \|\mathbf{p}_{\tilde{Y}}\|_2^2$, and $\mathrm{tr}\big(\mathrm{diag}(\mathbf{p}_{\tilde{Y}})\mathbf{p}_{\tilde{Y}}\mathbf{p}_{\tilde{Y}}^\top\big) = \mathbf{p}_{\tilde{Y}}^\top \mathrm{diag}(\mathbf{p}_{\tilde{Y}})\mathbf{p}_{\tilde{Y}}$, we can calculate $\mathrm{tr}(\Sigma_{\tilde{Y}}^2)$ as

$$\mathrm{tr}(\Sigma_{\tilde{Y}}^2) = \|\mathbf{p}_{\tilde{Y}}\|_2^2 + \|\mathbf{p}_{\tilde{Y}}\|_2^4 - 2\|\mathbf{p}_{\tilde{Y}}\|_3^3.$$

In a similar manner, we can expand the numerator term as follows:

$$\mathrm{tr}(\Sigma_{\tilde{Y}}^4) = \|\mathbf{p}_{\tilde{Y}}\|_4^4 + \|\mathbf{p}_{\tilde{Y}}\|_2^8 + 2\|\mathbf{p}_{\tilde{Y}}\|_3^6 + 4\|\mathbf{p}_{\tilde{Y}}\|_2^2\|\mathbf{p}_{\tilde{Y}}\|_4^4 - 4\|\mathbf{p}_{\tilde{Y}}\|_2^4\|\mathbf{p}_{\tilde{Y}}\|_3^3 - 4\|\mathbf{p}_{\tilde{Y}}\|_5^5.$$

Therefore, it suffices to study the asymptotic behavior of

$$\begin{aligned}&\mathrm{tr}(\Sigma_{\tilde{Y}}^4)/\big\{\mathrm{tr}(\Sigma_{\tilde{Y}}^2)\big\}^2\\ &= \frac{\big(\|\mathbf{p}_{\tilde{Y}}\|_2^8 - 4\|\mathbf{p}_{\tilde{Y}}\|_2^4\|\mathbf{p}_{\tilde{Y}}\|_3^3 + 2\|\mathbf{p}_{\tilde{Y}}\|_3^6\big) + \|\mathbf{p}_{\tilde{Y}}\|_4^4 + 4\|\mathbf{p}_{\tilde{Y}}\|_2^2\|\mathbf{p}_{\tilde{Y}}\|_4^4 - 4\|\mathbf{p}_{\tilde{Y}}\|_5^5}{\big(\|\mathbf{p}_{\tilde{Y}}\|_2^8 - 4\|\mathbf{p}_{\tilde{Y}}\|_2^4\|\mathbf{p}_{\tilde{Y}}\|_3^3 + 4\|\mathbf{p}_{\tilde{Y}}\|_3^6\big) + \|\mathbf{p}_{\tilde{Y}}\|_2^4 + 2\|\mathbf{p}_{\tilde{Y}}\|_2^2\|\mathbf{p}_{\tilde{Y}}\|_4^4 - 4\|\mathbf{p}_{\tilde{Y}}\|_2^2\|\mathbf{p}_{\tilde{Y}}\|_3^3},\end{aligned} \tag{117}$$

where parentheses group terms of the same order. Since $p_{Ym} = 1/k + (-1)^m\epsilon/\sqrt{k}$, we have

$$p_{\tilde{Y}m} = \frac{w}{k} + \frac{w(-1)^m\sqrt{k}\epsilon}{k} + \frac{1}{k} - \frac{w}{k} = \frac{1}{k} + \frac{(-1)^m w\epsilon}{\sqrt{k}}, \tag{118}$$

for $m \in [k]$. Using (118) and the fact that k is an even integer, the powers of norms of $\mathbf{p}_{\tilde{Y}}$ are calculated as follows:

$$\begin{aligned}&\|\mathbf{p}_{\tilde{Y}}\|_2^2 = \frac{1}{k} + w^2\epsilon^2,\ \|\mathbf{p}_{\tilde{Y}}\|_3^3 = \frac{1}{k^2} + \frac{3w^2\epsilon^2}{k},\\ &\|\mathbf{p}_{\tilde{Y}}\|_4^4 = \frac{1}{k^3} + 6\frac{w^2\epsilon^2}{k^2} + \frac{w^4\epsilon^4}{k},\ \|\mathbf{p}_{\tilde{Y}}\|_5^5 = \frac{1}{k^4} + 10\frac{w^2\epsilon^2}{k^3} + 5\frac{w^4\epsilon^4}{k^2}.\end{aligned} \tag{119}$$

By substituting (119) into (117), and recalling that $w\epsilon$ is of order $k^{-3/4}n_1^{-1/2}$, we find that among the terms in (117), the term that converges to 0 the most slowly is $-4/k^3$, which comes from $-4\|\mathbf{p}_{\tilde{Y}}\|_2^2\|\mathbf{p}_{\tilde{Y}}\|_3^3$ in the denominator. Consequently, the quantity in (119) converges to 0, confirming condition C1 of Theorem 19.

Verification of the condition C2. The condition C2 is recalled below as

$$\frac{\mathbb{E}[W_1] + n_1\mathbb{E}[W_2]}{n_1^2\big\{\mathrm{tr}(\Sigma_{\tilde{Y}}^2)\big\}^2} \to 0, \tag{120}$$

where

$$\begin{aligned}W_1 &:= \big\{(\tilde{\mathbf{Y}}_1 - \mathbf{p}_Z)^\top(\tilde{\mathbf{Y}}_2 - \mathbf{p}_Z)\big\}^4,\\ W_2 &:= \big\{(\tilde{\mathbf{Y}}_1 - \mathbf{p}_Z)^\top(\tilde{\mathbf{Y}}_2 - \mathbf{p}_Z)\big\}^2\big\{(\tilde{\mathbf{Y}}_1 - \mathbf{p}_Z)^\top(\tilde{\mathbf{Y}}_3 - \mathbf{p}_Z)\big\}^2.\end{aligned}$$

We separately analyze the terms in (120). First for $\mathbb{E}[W_1]$, the random variable W_1 takes two distinct values:

$$W_1 = \begin{cases}\left(1 - \dfrac{1}{k}\right)^4, & \tilde{\mathbf{Y}}_1 = \tilde{\mathbf{Y}}_2,\\ \dfrac{1}{k^4}, & \text{otherwise}.\end{cases}$$

Observe that, according to (118), we have

$$\mathbb{P}(\tilde{\mathbf{Y}}_1 = \tilde{\mathbf{Y}}_2) = \sum_{m=1}^k p_{\tilde{Y}m}^2 = \sum_{m=1}^k \left(\frac{1}{k} + \frac{(-1)^m w\epsilon}{\sqrt{k}}\right)^2 = \frac{1}{k} + w^2\epsilon^2. \tag{121}$$

Using the established probability (121), we calculate the target expected value as

$$\mathbb{E}[W_1] = -\frac{1}{k^5} + \frac{k}{k^5} - \frac{k\epsilon^2 w^2}{k^5} + \frac{(-1+k)^4}{k^5} + \frac{(-1+k)^4(1+k\epsilon^2 w^2)}{k^5}.$$

Among the terms in the expectation computed above, the term that converges to 0 most slowly is of order $1/k$, which arises from the fourth term of the numerator. Similarly, for $\mathbb{E}[W_2]$, the random variable W_2 takes on the values as:

$$W_2 = \begin{cases} \left(1 - \frac{1}{k}\right)^4, & \tilde{\mathbf{Y}}_1 = \tilde{\mathbf{Y}}_2 = \tilde{\mathbf{Y}}_3, \\ \frac{1}{k^2}\left(1 - \frac{1}{k}\right)^2, & [\tilde{\mathbf{Y}}_1 = \tilde{\mathbf{Y}}_2, \tilde{\mathbf{Y}}_1 \neq \tilde{\mathbf{Y}}_3] \cup [\tilde{\mathbf{Y}}_1 = \tilde{\mathbf{Y}}_3, \tilde{\mathbf{Y}}_1 \neq \tilde{\mathbf{Y}}_2] \\ \frac{1}{k^4}, & \text{otherwise.} \end{cases} \tag{122}$$

The corresponding probabilities are calculated as

$$\begin{aligned} \mathbb{P}(\tilde{\mathbf{Y}}_1 = \tilde{\mathbf{Y}}_2 = \tilde{\mathbf{Y}}_3) &= \sum_{m=1}^{k} p_{\tilde{Y}m}^3 = \|\mathbf{p}_{\tilde{Y}}\|_3^3 \overset{(a)}{=} \frac{1}{k^2} + \frac{3w^2\epsilon^2}{k}, \text{ and} \\ \mathbb{P}([\tilde{\mathbf{Y}}_1 = \tilde{\mathbf{Y}}_2, \tilde{\mathbf{Y}}_1 \neq \tilde{\mathbf{Y}}_3] \cup [\tilde{\mathbf{Y}}_1 = \tilde{\mathbf{Y}}_3, \tilde{\mathbf{Y}}_1 \neq \tilde{\mathbf{Y}}_2]) &\overset{(b)}{=} 2\mathbb{P}([\tilde{\mathbf{Y}}_1 = \tilde{\mathbf{Y}}_2, \tilde{\mathbf{Y}}_1 \neq \tilde{\mathbf{Y}}_3]) \\ &= 2\sum_{m=1}^{k} p_{\tilde{Y}m}^2(1 - p_{\tilde{Y}m}) \\ &= 2\|\mathbf{p}_{\tilde{Y}}\|_2^2 - 2\|\mathbf{p}_{\tilde{Y}}\|_3^3 \\ &\overset{(c)}{=} \frac{2}{k} + 2w^2\epsilon^2 - \frac{2}{k^2} - \frac{6w^2\epsilon^2}{k}, \end{aligned}$$

where steps (a) and (c) use the equations in (119) and step (b) uses the independence between samples, as well as the exclusiveness and symmetry of the events. Now $\mathbb{E}[W_2]$ is calculated as

$$\begin{aligned} \mathbb{E}[W_2] = {} & \left(1 - \frac{1}{k}\right)^4 \left(\frac{1}{k^2} + \frac{3w^2\epsilon^2}{k}\right) + \left\{\frac{1}{k^2}\left(1 - \frac{1}{k}\right)^2\right\}\left(\frac{2}{k} + 2w^2\epsilon^2 - \frac{2}{k^2} - \frac{6w^2\epsilon^2}{k}\right) \\ & + \frac{1}{k^4}\left(1 - \frac{1}{k^2} - \frac{3w^2\epsilon^2}{k} - \frac{2}{k} - 2w^2\epsilon^2 + \frac{2}{k^2} + \frac{6w^2\epsilon^2}{k}\right), \end{aligned} \tag{123}$$

where the term that converges to 0 the slowest is of order $1/k^2$, originating from the first term on the right-hand side of (123). Finally, the trace term in the denominator is calculated as

$$\begin{aligned} \mathrm{tr}(\Sigma_{\tilde{Y}}^2) = \|\mathbf{p}_{\tilde{Y}}\|_2^2 + \|\mathbf{p}_{\tilde{Y}}\|_2^4 - 2\|\mathbf{p}_{\tilde{Y}}\|_3^3 &= \frac{1}{k} + w^2\epsilon^2 + \frac{1}{k^2} + w^4\epsilon^4 + \frac{2w^2\epsilon^2}{k} - \frac{2}{k^2} - \frac{6w^2\epsilon^2}{k} \\ &= \frac{1}{k} - \frac{1}{k^2} + w^2\epsilon^2 + w^4\epsilon^4 - \frac{4w^2\epsilon^2}{k}, \end{aligned} \tag{124}$$

where the term that converges to 0 the slowest is $1/k$. Now to verify the condition C2, we separately analyze the asymptotic behaviors of $\mathbb{E}[W_1]/n_1^2\{\mathrm{tr}(\Sigma_{\tilde{Y}}^2)\}^2$ and $n_1\mathbb{E}[W_2]/n_1^2\{\mathrm{tr}(\Sigma_{\tilde{Y}}^2)\}^2$, focusing on the terms that converges the slowest to 0 in the numerator and denominator, respectively. For the first term, it suffices to investigate $(1/k)/(n_1^2/k^2)$, which converges to 0 as long as $k/n_1^2 \to 0$. Note that $k/n_1^2 \to 0$ is implied by our assumption (112). For the second term, it suffices to investigate $(n_1/k^2)/(n_1^2/k^2)$, which converges to 0. Therefore the convergence in (120) is verified, confirming the condition C2 of Theorem 19.

Verification of the condition C3. The condition C3 is restated as

$$(\mathbf{p}_{\tilde{Y}} - \mathbf{p}_Z)^\top \Sigma_{\tilde{Y}} (\mathbf{p}_{\tilde{Y}} - \mathbf{p}_Z) = o\left(n_1^{-1}\mathrm{tr}(\Sigma_{\tilde{Y}}^2)\right). \tag{125}$$

To calculate the quadratic form therein, note that since $\mathbf{p}_{\tilde{Y}} - \mathbf{p}_Z = w(\mathbf{p}_Y - \mathbf{p}_Z)$, we have

$$(\mathbf{p}_{\tilde{Y}} - \mathbf{p}_Z)^\top \Sigma_{\tilde{Y}} (\mathbf{p}_{\tilde{Y}} - \mathbf{p}_Z) = w^2(\mathbf{p}_Y - \mathbf{p}_Z)^\top \mathrm{diag}(\mathbf{p}_{\tilde{Y}})(\mathbf{p}_Y - \mathbf{p}_Z) - w^2(\mathbf{p}_Y - \mathbf{p}_Z)^\top \mathbf{p}_{\tilde{Y}}\mathbf{p}_{\tilde{Y}}^\top(\mathbf{p}_Y - \mathbf{p}_Z).$$

We separately analyze the two terms on the right-hand side. For the first term, we have

$$\begin{aligned}
w^2(\mathbf{p}_Y - \mathbf{p}_Z)^\top \mathrm{diag}(\mathbf{p}_{\tilde{Y}})(\mathbf{p}_Y - \mathbf{p}_Z) &= w^2 \sum_{m=1}^{k} \frac{(-1)^m\epsilon}{\sqrt{k}}\left(\frac{1}{k} + \frac{(-1)^m w\epsilon}{\sqrt{k}}\right)\frac{(-1)^m\epsilon}{\sqrt{k}} \\
&= \frac{w^2\epsilon^2}{k}\sum_{m=1}^{k}\left(\frac{1}{k} + \frac{(-1)^m w\epsilon}{\sqrt{k}}\right) \\
&= \frac{w^2\epsilon^2}{k},
\end{aligned}$$

where the last equality holds since k is even. The second term is computed as

$$\begin{aligned}
w^2(\mathbf{p}_Y - \mathbf{p}_Z)^\top \mathbf{p}_{\tilde{Y}}\mathbf{p}_{\tilde{Y}}^\top(\mathbf{p}_Y - \mathbf{p}_Z) = w^2\{(\mathbf{p}_Y - \mathbf{p}_Z)^\top \mathbf{p}_{\tilde{Y}}\}^2 &= w^2\left\{\sum_{m=1}^{k} \frac{(-1)^m\epsilon}{\sqrt{k}}\left(\frac{1}{k} + \frac{(-1)^m w\epsilon}{\sqrt{k}}\right)\right\}^2 \\
&= w^4\epsilon^4,
\end{aligned}$$

where the last equality again uses the fact that k is an even integer. Therefore we have

$$(\mathbf{p}_{\tilde{Y}} - \mathbf{p}_Z)^\top \Sigma_{\tilde{Y}} (\mathbf{p}_{\tilde{Y}} - \mathbf{p}_Z) = \frac{w^2\epsilon^2}{k} - w^4\epsilon^4, \tag{126}$$

where $w^2\epsilon^2/k$ is the term that converges to 0 most slowly. Using the trace calculation (124), we can verify the condition C3 (125) by studying the asymptotic behavior of $(w^2\epsilon^2/k)/(1/kn_1)$. Since this quantity converges to 0, the condition C3 is verified.

As all conditions of Theorem 19 are satisfied, the asymptotic distribution in (116) is confirmed.

Power function. Using the critical value in (115) and the asymptotic alternative distribution established in (116), under the asymptotic regime in (112), the power function is written as:

$$
\begin{aligned}
&\gamma_{n_1,k,\alpha}(\mathbf{p}_Y, \mathbf{p}_Z) \\
&\quad = \mathbb{P}_{H_1}\left(U_{n_1} \geq z_\gamma \sqrt{\frac{2(1-1/k)}{kn_1(n_1-1)}} \right) \\
&\quad = \mathbb{P}_{H_1}\left(W \geq \sqrt{\frac{n_1(n_1-1)}{2}} \frac{z_\gamma \sqrt{2\{1-(1/k)\}/\{kn_1(n_1-1)\}} - w^2\epsilon^2}{\sqrt{\mathrm{tr}(\Sigma_{\tilde{Y}}^2) + 2(n_1-1)(\mathbf{p}_{\tilde{Y}} - \mathbf{p}_Z)^\top \Sigma_{\tilde{Y}} (\mathbf{p}_{\tilde{Y}} - \mathbf{p}_Z)}} \right) + o(1) \\
&\quad = \Phi\left(\sqrt{\frac{n_1(n_1-1)}{2}} \frac{w^2\epsilon^2 - z_\gamma \sqrt{2\{1-(1/k)\}/\{kn_1(n_1-1)\}}}{\sqrt{\mathrm{tr}(\Sigma_{\tilde{Y}}^2) + 2(n_1-1)(\mathbf{p}_{\tilde{Y}} - \mathbf{p}_Z)^\top \Sigma_{\tilde{Y}} (\mathbf{p}_{\tilde{Y}} - \mathbf{p}_Z)}} \right) \\
&\quad = \Phi\left(\frac{\sqrt{n_1(n_1-1)/2}\, w^2\epsilon^2 - z_\gamma \sqrt{(1/k) - (1/k^2)}}{\sqrt{\{(1/k) - (1/k^2) + w^2\epsilon^2 + w^4\epsilon^4 - (4w^2\epsilon^2/k)\} + 2(n_1-1)\big\{(w^2\epsilon^2/k) - w^4\epsilon^4\big\}}} \right) \\
&\quad \to \Phi(-z_\gamma) = \gamma,
\end{aligned}
\tag{127}
$$

where $W \sim \mathcal{N}(0,1)$, Φ denotes the cumulative distribution function of W, and $o(1)$ refers to a quantity that goes to 0 as n_1 and k go to ∞, in the regime of (112). The last line in (127) uses the fact that given that $w\epsilon$ is of the order $k^{-3/4}n_1^{-1/2}$, the term in (127) that converges most slowly to 0 is $\sqrt{1/k}$, which appears both on the second term in the numerator and the first term in the denominator. It is important to note that the convergence result (112) holds for any sufficiently small α, indicating that the `GenRR`+ℓ_2 test becomes asymptotically powerless in the regime where the minimax testing rate is $k^{1/4}/(n_1\alpha^2)^{1/2}$. This completes the proof of Theorem 39. ■

References

Jayadev Acharya, Constantinos Daskalakis, and Gautam Kamath. Optimal testing for properties of distributions. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2015.

Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. Differentially private testing of identity and closeness of discrete distributions. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

Jayadev Acharya, Clément L. Canonne, Cody Freitag, and Himanshu Tyagi. Test without trust: Optimal locally private distribution testing. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2019.

Jayadev Acharya, Clément L. Canonne, and Himanshu Tyagi. Inference under information constraints I: Lower bounds from chi-square contraction. *IEEE Transactions on Information Theory*, 66:7835–7855, 2020.

Jayadev Acharya, Clément L. Canonne, Yuhan Liu, Ziteng Sun, and Himanshu Tyagi. Interactive inference under information constraints. In *IEEE International Symposium on Information Theory (ISIT)*, 2021a.

Jayadev Acharya, Clément L. Canonne, Cody Freitag, Ziteng Sun, and Himanshu Tyagi. Inference under information constraints III: Local privacy constraints. *IEEE Journal on Selected Areas in Information Theory*, 2:253–267, 2021b.

Jayadev Acharya, Peter Kairouz, Yuhan Liu, and Ziteng Sun. Estimating sparse discrete distributions under privacy and communication constraints. In *International Conference on Algorithmic Learning Theory (ALT)*, 2021c.

Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. Differentially private Assouad, Fano, and Le Cam. In *International Conference on Algorithmic Learning Theory (ALT)*, 2021d.

Maryam Aliakbarpour, Ilias Diakonikolas, and Ronitt Rubinfeld. Differentially private identity and equivalence testing of discrete distributions. In *International Conference on Machine Learning (ICML)*, 2018.

Maryam Aliakbarpour, Ilias Diakonikolas, Daniel Kane, and Ronitt Rubinfeld. Private testing of distributions via sample permutations. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Apple Differential Privacy Team. Learning with privacy at scale. *Apple Machine Learning Journal*, 1, 2017.

Ery Arias-Castro, Bruno Pelletier, and Venkatesh Saligrama. Remember the curse of dimensionality: The case of goodness-of-fit testing in arbitrary dimension. *Journal of Nonparametric Statistics*, 30(2):448–471, 2018.

Florent Autin, Erwan Le Pennec, and Karine Tribouley. Thresholding methods to estimate copula density. *Journal of Multivariate Analysis*, 101(1):200–222, 2010.

Sivaraman Balakrishnan and Larry Wasserman. Hypothesis testing for densities and high-dimensional multinomials: Sharp local minimax rates. *The Annals of Statistics*, 47:1893–1927, 2019.

Yannick Baraud. Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8 (5):577–606, 2002.

Tuğkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D. Smith, and Patrick White. Testing that distributions are close. *IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, 2000.

Thomas B. Berrett and Cristina Butucea. Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Thomas B. Berrett, László Györfi, and Harro Walk. Strongly universally consistent nonparametric regression and classification with privatised data. *Electronic Journal of Statistics*, 15:2430–2453, 2021.

Bryan Cai, Constantinos Daskalakis, and Gautam Kamath. Priv'it: Private and sample efficient identity testing. In *International Conference on Machine Learning (ICML)*, 2017.

T. Tony Cai, Yichen Wang, and Linjun Zhang. The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *The Annals of Statistics*, 49(5):2825–2850, 2021.

Siu-On Chan, Ilias Diakonikolas, Paul Valiant, and Gregory Valiant. Optimal algorithms for testing closeness of discrete distributions. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2014.

Simon Couch, Zeki Kazan, Kaiyan Shi, Andrew Bray, and Adam Groce. Differentially private nonparametric hypothesis testing. In *ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2019.

Ilias Diakonikolas and Daniel M Kane. A new approach for testing properties of discrete distributions. In *IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, 2016.

Ilias Diakonikolas, Themis Gouleakis, John Peebles, and Eric Price. Sample-optimal identity testing with high probability. In *International Colloquium on Automata, Languages, and Programming (ICALP)*, 2018.

Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. Collecting telemetry data privately. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Bolin Ding, Harsha Nori, Paul Li, and Joshua Allen. Comparing population means under local differential privacy: With significance and power. In *AAAI Conference on Artificial Intelligence*, 2018.

Amandine Dubois, Thomas B. Berrett, and Cristina Butucea. Goodness-of-fit testing for Hölder continuous densities under local differential privacy. In *Foundations of Modern Statistics*, pages 53–119, Cham, 2023. Springer International Publishing.

John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Local privacy and minimax bounds: sharp rates for probability estimation. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2013.

John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521): 182–201, 2018.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9, 2013.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*, pages 265–284. Springer Berlin Heidelberg, 2006.

Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2014.

Magalie Fromont and Béatrice Laurent. Adaptive goodness-of-fit tests in a density model. *The Annals of Statistics*, 34(2):680–720, 2006.

Marco Gaboardi and Ryan Rogers. Local private hypothesis testing: Chi-square tests. In *International Conference on Machine Learning (ICML)*, 2018.

Marco Gaboardi, Hyun Lim, Ryan Rogers, and Salil Vadhan. Differentially private chi-squared hypothesis testing: Goodness of fit and independence testing. In *International Conference on Machine Learning (ICML)*, 2016.

Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. In *Annual ACM Symposium on Theory of Computing (STOC)*, 2009.

Evarist Giné and Richard Nickl. *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press, 2015.

Oded Goldreich and Dana Ron. On testing expansion in bounded-degree graphs. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6650 LNCS, 2011.

Arthur Gretton, Kenji Fukumizu, Zaid Harchaoui, and Bharath K. Sriperumbudur. A fast, consistent kernel two-sample test. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2009.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13:723–773, 2012.

Zaid Harchaoui, Felicien Vallet, Alexandre Lung-Yut-Fong, and Olivier Cappe. A regularized kernel-based approach to unsupervised audio segmentation. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009.

Jesse Hemerik and Jelle Goeman. Exact testing with random permutations. *TEST*, 27(4): 811–825, 2018.

Nils Homer, Szabolcs Szelinger, Margot Redman, David Duggan, Waibhav Tembe, Jill Muehling, John V. Pearson, Dietrich A. Stephan, Stanley F. Nelson, and David W. Craig. Resolving individuals contributing trace amounts of dna to highly complex mixtures using high-density snp genotyping microarrays. *PLoS Genetics*, 4(8):e1000167, August 2008.

Yu Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. *Mathematical Methods of Statistics*, 2(2):85–114, 1993.

Yu Ingster. Adaptive chi-square tests. *Journal of Mathematical Sciences*, 99:1110–1119, 9 2000.

Yu I. Ingster. Minimax detection of a signal in ℓ_p metrics. *Journal of Mathematical Sciences*, 68:503–515, 1994.

Seidu Inusah and Tomasz J. Kozubowski. A discrete analogue of the Laplace distribution. *Journal of Statistical Planning and Inference*, 136(3):1090–1102, March 2006.

Aaron Johnson and Vitaly Shmatikov. Privacy-preserving data exploration in genome-wide association studies. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, 2013.

Matthew Joseph, Jieming Mao, Seth Neel, and Aaron Roth. The role of interactivity in local differential privacy. *IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, 2019.

Shiva Prasad Kasiviswanathan, Homin K Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011.

Alexander Kent, Thomas B. Berrett, and Yi Yu. Locally differentially private two-sample testing. *arXiv preprint arXiv:2505.24811*, 2025.

Ilmun Kim. Multinomial goodness-of-fit based on U-statistics: High-dimensional asymptotic and minimax optimality. *Journal of Statistical Planning and Inference*, 205:74–91, 2020.

Ilmun Kim and Antonin Schrab. Differentially private permutation tests: Applications to kernel methods. *arXiv preprint arXiv:2310.19043*, 2023.

Ilmun Kim, Ann B. Lee, and Jing Lei. Global and local two-sample tests via regression. *Electronic Journal of Statistics*, 13(2):5253–5305, 2019.

Ilmun Kim, Aaditya Ramdas, Aarti Singh, and Larry Wasserman. Classification accuracy as a proxy for two-sample testing. *The Annals of Statistics*, 49(1):411–434, 2021.

Ilmun Kim, Sivaraman Balakrishnan, and Larry Wasserman. Minimax optimality of permutation tests. *The Annals of Statistics*, 50(1):225–251, February 2022a.

Ilmun Kim, Matey Neykov, Sivaraman Balakrishnan, and Larry Wasserman. Local permutation tests for conditional independence. *The Annals of Statistics*, 50(6):3388–3414, 2022b.

Jan Kohout and Tomáš Pevný. Network traffic fingerprinting based on approximated kernel two-sample test. *IEEE Transactions on Information Forensics and Security*, 13:788–801, 2018.

Clément Lalanne, Aurélien Garivier, and Rémi Gribonval. On the statistical complexity of estimation and testing under privacy constraints. *Transactions on Machine Learning Research*, 2023.

Joseph Lam-Weil, Béatrice Laurent, and Jean-Michel Loubes. Minimax optimal goodness-of-fit testing for densities and multinomials under a local differential privacy constraint. *Bernoulli*, 28(1):579–600, 2022.

Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, 2007.

Matey Neykov, Sivaraman Balakrishnan, and Larry Wasserman. Minimax optimal conditional independence testing. *The Annals of Statistics*, 49(4):2151–2177, 2021.

N. Rabin, M. Golan, G. Singer, and D. Kleper. Modeling and analysis of students' performance trajectories using diffusion maps and kernel two-sample tests. *Engineering Applications of Artificial Intelligence*, 85, 2019.

Anant Raj, Ho Chung Leon Law, Dino Sejdinovic, and Mijung Park. A differentially private kernel two-sample test. In *Machine Learning and Knowledge Discovery in Databases*, pages 697–724, Cham, 2020.

Aaditya Ramdas, Rina Foygel Barber, Emmanuel J. Candès, and Ryan J. Tibshirani. Permutation tests using arbitrary permutation distributions. *Sankhya A*, 85(2):1156–1177, 2023.

Ryan Rogers and Daniel Kifer. A new class of private chi-square hypothesis tests. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.

Antonin Schrab, Ilmun Kim, Mélisande Albert, Béatrice Laurent, Benjamin Guedj, and Arthur Gretton. MMD Aggregated Two-Sample Test. *Journal of Machine Learning Research*, 24(194):1–81, 2023.

Or Sheffet. Locally private hypothesis testing. In *International Conference on Machine Learning (ICML)*, 2018.

Or Sheffet and Daniel Omer. Differentially private equivalence testing for continuous distributions and applications. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Student. The Probable Error of a Mean. *Biometrika*, 6(1):1–25, 1908.

Gábor J. Székely and Maria L. Rizzo. Testing for equal distributions in high dimension. *InterStat*, 5, 11 2004.

Gábor J. Székely and Maria L. Rizzo. A new test for multivariate normality. *Journal of Multivariate Analysis*, 93(1):58–80, March 2005.

Rong Tang and Yun Yang. Minimax nonparametric two-sample test under adversarial losses. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2023.

Christine Task and Chris Clifton. Differentially private significance testing on paired-sample data. In *SIAM International Conference on Data Mining (SDM)*, 2016.

Caroline Uhler, Aleksandra B. Slavkovic, and Stephen E. Fienberg. Privacy-preserving data sharing for genome-wide association studies. *Journal of Privacy and Confidentiality*, 5(1), 2013.

Gregory Valiant and Paul Valiant. An automatic inequality prover and instance optimal identity testing. In *IEEE Annual Symposium on Foundations of Computer Science (FOCS)*, 2014.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998.

Duy Vu and Aleksandra Slavkovic. Differential privacy for clinical trial data: Preliminary evaluations. In *IEEE International Conference on Data Mining Workshops*, pages 138–143, 2009.

Yue Wang, Jaewoo Lee, and Daniel Kifer. Revisiting differentially private hypothesis tests for categorical data. *arXiv preprint arXiv:1511.03376*, 2015.

Stanley L. Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.

Ian Waudby-Smith, Steven Wu, and Aaditya Ramdas. Nonparametric extensions of randomized response for private confidence sets. In *International Conference on Machine Learning (ICML)*, 2023.

Fei Yu, Stephen E. Fienberg, Aleksandra B. Slavković, and Caroline Uhler. Scalable privacy-preserving data sharing methodology for genome-wide association studies. *Journal of Biomedical Informatics*, 50:133–141, 2014.