Differentially Private Permutation Tests: Applications to Kernel Methods

Ilmun Kim, Antonin Schrab

Differentially Private Permutation Tests: Applications to Kernel Methods

Ilmun Kim[†,◇] Antonin Schrab[‡,◇]

[†]Department of Statistics and Data Science, Yonsei University

[‡]Centre for Artificial Intelligence, Gatsby Computational Neuroscience Unit, University College London & Inria London

[◇]Both authors contributed equally to this work

October 31, 2023

Abstract

Recent years have witnessed growing concerns about the privacy of sensitive data. In response to these concerns, differential privacy has emerged as a rigorous framework for privacy protection, gaining widespread recognition in both academic and industrial circles. While substantial progress has been made in private data analysis, existing methods often suffer from impracticality or a significant loss of statistical efficiency. This paper aims to alleviate these concerns in the context of hypothesis testing by introducing differentially private permutation tests. The proposed framework extends classical non-private permutation tests to private settings, maintaining both finite-sample validity and differential privacy in a rigorous manner. The power of the proposed test depends on the choice of a test statistic, and we establish general conditions for consistency and non-asymptotic uniform power. To demonstrate the utility and practicality of our framework, we focus on reproducing kernel-based test statistics and introduce differentially private kernel tests for two-sample and independence testing: dpMMD and dpHSIC. The proposed kernel tests are straightforward to implement, applicable to various types of data, and attain minimax optimal power across different privacy regimes. Our empirical evaluations further highlight their competitive power under various synthetic and real-world scenarios, emphasizing their practical value. The code is publicly available to facilitate the implementation of our framework.

1 Introduction

Ensuring the privacy of sensitive data has become a critical concern in modern data analysis. As organizations collect and analyze vast amounts of personal information, safeguarding individual privacy has emerged as a crucial ethical and legal imperative. In response to these challenges, differential privacy (DP), introduced by Dwork et al. (2006b), has emerged as a rigorous framework for addressing privacy concerns, and has gained widespread recognition not only in academia but also in industrial companies. For instance, major industry players such as Apple (Apple, 2017), Google (Erlingsson et al., 2014) and Microsoft (Ding et al., 2017) have embraced differential privacy as a robust definition of privacy. This growing trend has sparked a recent surge of research in statistics and related fields, aiming at integrating differential privacy and its variants (Dwork et al., 2006a; Bun and Steinke, 2016; Mironov, 2017; Dong et al., 2022) into data analysis and developing privacy-preserving methodologies. In this line of work, a major challenge is to strike

a balance between privacy guarantees and statistical efficiency. Notably, a high privacy guarantee requires substantial data perturbation, which in turn degrades statistical performance. Conversely, releasing less perturbed data can improve statistical efficiency but at the expense of reduced privacy guarantees. Therefore, balancing this trade-off between privacy and efficiency has been a central topic in the existing literature (*e.g.*, Duchi et al., 2018; Cai et al., 2021; Kamath and Ullman, 2020).

Broadly, there are two major statistical problems tackled under privacy constraints: estimation and hypothesis testing (Kamath and Ullman, 2020, for a recent review). This paper focuses on the latter problem, which requires access to the null distribution of a test statistic in order to effectively calibrate a test statistic. Analyzing the distribution of a test statistic becomes particularly more challenging in private settings due to additional random sources arising from privacy mechanisms. As we review in Section 1.1, substantial efforts have been made to address private testing problems. These efforts involve adapting classical hypothesis tests to private settings or developing new testing procedures that achieve an optimal balance between privacy and statistical power.

Despite the significant progress made over the last decade, there are still several areas where further improvements can be made. One such area includes the reliance on asymptotic methods for determining the critical value of a test statistic. The practical quality of this asymptotic approach depends on the convergence rate of a privatized statistic to the limiting distribution. This convergence rate is often slow in private settings, and more importantly, the limiting distribution may vary depending on the delicate interplay between privacy and other parameters. This issue puts practitioners in a bind as it is unclear which limiting distribution should be considered a priori. All these concerns lead to the unreliability of asymptotic private tests in real-world applications.

Another area of concern is the limited practicality of existing methods. Many private statistical tests are designed specifically for discrete data, not directly applicable to handling continuous or mixed-type data. Moreover, existing methods often rely on unspecified constants and heuristics, making them less user-friendly and potentially undermining their reliability. We also point out that the majority of research has concentrated on theoretical aspects of private testing, and only a handful of papers are equipped with thorough empirical evaluations and open-source code.

In this work, we aim to tackle the aforementioned concerns by introducing differentially private permutation tests. The primary goal is to extend classical non-private permutation tests to differentially private settings, applicable to any test statistic with finite global sensitivity. The proposed private permutation test inherits the finite-sample validity of the classical permutation test under the exchangeability condition, while ensuring differential privacy. The power of the proposed test depends on the choice of a test statistic, and we establish sufficient conditions for consistency and non-asymptotic uniform power. To demonstrate the effectiveness of our framework, we focus on the two-sample and independence testing problems and propose differentially private versions of the maximum mean discrepancy (MMD) and Hilbert–Schmidt independence criterion (HSIC), which we coin as "dpMMD" and "dpHSIC", respectively. On the theoretical side, we prove minimax optimality of dpMMD and dpHSIC tests over the entire privacy regimes in terms of kernel metrics. On the empirical side, we showcase the competitive power performance of the proposed tests across various practical scenarios. The code that implements our methods is publicly available at `https://github.com/antoninschrab/dpkernel` to allow practitioners to build on our findings.

1.1 Related Work

In recent years, there has been a growing body of research on hypothesis testing problems under privacy constraints. Since the early work by Vu and Slavkovic (2009) and Fienberg et al. (2011), numerous attempts have been made to extend classical non-private tests to their private counterparts. Examples include ANOVA (Campbell et al., 2018; Swanberg et al., 2019), likelihood ratio tests (Canonne et al., 2019), tests for regression coefficients (Sheffet, 2017; Alabi and Vadhan, 2022), rank or sign-based nonparametric tests (Task and Clifton, 2016; Couch et al., 2019), conditional independence tests (Kalemaj et al., 2023) and χ^2-tests (Fienberg et al., 2011; Wang et al., 2015; Gaboardi et al., 2016; Rogers and Kifer, 2017; Kakizaki et al., 2017; Friedberg and Rogers, 2023). While most of the aforementioned work focuses on asymptotic settings where the sample size goes to infinity, a recent line of work within computer science has placed greater emphasis on finite-sample analysis. In particular, Cai et al. (2017) propose a two-step algorithm for identity testing for discrete distributions, and studies the sample complexity under DP. The work of Acharya et al. (2018) explores both identity (goodness-of-fit) testing and closeness (two-sample) testing in finite-sample settings, and improves the upper bound result by Cai et al. (2017) and Aliakbarpour et al. (2018) for identity testing. Aliakbarpour et al. (2019) privatize the non-private test proposed by Diakonikolas and Kane (2016), and investigate the sample complexity for closeness testing and independence testing. In line with these advancements, our work develops private permutation tests, and studies their non-asymptotic performance under DP settings.

Despite the extensive body of literature, the majority of research has focused on private tests designed for discrete or bounded data. There are a few notable exceptions that have explored other data types. For example, Canonne et al. (2020) and Narayanan (2022) have investigated goodness-of-fit testing for high-dimensional Gaussian distributions. In addition, Raj et al. (2020) have proposed private two-sample tests based on finite dimensional approximations of kernel mean embeddings. The flexibility offered by kernels methods enables these tests to handle a wide variety of data types. However, their tests are asymptotic in nature, which may introduce reliability concerns when working with small sample sizes. Moreover, their analysis requires the number of features to be fixed. Such requirement potentially limits the power of the test when dealing with alternatives which are not well-represented by these fixed numbers of features.

Another line of work aims to develop a generic way to create private tests from non-private ones. The subsample-and-aggregate idea (Nissim et al., 2007) has emerged as a useful tool for this purpose. In particular, it offers a strategy to convert non-private sample complexity results into private ones in a black-box manner as pointed out by Cai et al. (2017); Canonne et al. (2019, 2020). Recent studies by Peña and Barrientos (2022) and Kazan et al. (2023) have focused specifically on the practical implementation of the subsample-and-aggregate approach. However, it is worth mentioning that this generality typically comes at the cost of suboptimal power, and often fails to recover the optimal sample complexity (Canonne et al., 2019, 2020). Moreover, the performance of this subsample-and-aggregate approach is sensitive to the number of subsamples, and determining the optimal value of this parameter remains an open problem.

Beyond global differential privacy, there has been a substantial amount of work on hypothesis testing under local differential privacy. Some of the notable works include Liao et al. (2017);

Gaboardi and Rogers (2018); Sheffet (2018); Acharya et al. (2019); Berrett and Butucea (2020); Dubois et al. (2023); Lam-Weil et al. (2022) and see the references therein. Local differential privacy requires data perturbation at the individual level, proving particularly useful in settings where data providers lack trust in data analysts. This individual-wise approach demands different analyses than global differential privacy, and the results under global and local differential privacy are not directly comparable.

Our work is also related to recent advances in kernel-based minimax testing (Li and Yuan, 2019; Albert et al., 2022; Schrab et al., 2023; Kim et al., 2022a). Specifically, we extend the non-private minimax testing rates established in this line of work to private counterparts. To achieve this, we leverage the techniques therein, such as the two moments method and exponential inequalities for permuted statistics in Kim et al. (2022a), and adapt them to private settings.

1.2 An Overview of Our Results

The main contributions of this work are summarized below.

- **DP Permutation Tests (Section 3).** We introduce differentially private permutation tests in Algorithm 1, and establish their theoretical properties. A naive way of extending the classical permutation tests to private settings is to first make the original test statistic and its permuted counterparts differentially private, and then carry out the permutation test based on these individually privatized statistics. However, this naive approach results in an unnecessary power loss by adding more noise as the number of permutations increases. The proposed framework addresses this issue by utilizing the quantile representation of a permutation test. This strategy leads to a substantial power gain over the naive approach, while being finite-sample valid. We present sufficient conditions for pointwise consistency (Theorem 3) and non-asymptotic uniform power (Theorem 4) of the proposed tests. The latter uniform power condition can be regarded as an extension of the two moments method (Kim et al., 2022a) to private settings.

- **DP Kernel Tests (Section 4).** We showcase the versatility of our framework by applying it to two specific tasks: differentially private two-sample and independence testing based on reproducing kernel-based test statistics. We consider the plug-in estimators of the MMD and HSIC, and privatize them through the proposed method employing the standard Laplace mechanism. The practical performance of the resulting differentially private kernel tests heavily depends on the global sensitivity used in the Laplace mechanism. To boost the empirical performance, we put significant effort into establishing sharp upper bounds for the global sensitivity of the plug-in estimators of MMD and HSIC, as well as matching lower bounds for popular kernels (Lemma 5 and Lemma 6). We then establish key properties of the proposed kernel tests, including non-asymptotic validity and consistency against any fixed alternatives in Theorem 5 and Theorem 6.

- **Uniform Power and Optimality (Section 5).** We characterize the trade-off between differential privacy and statistical power through the lens of minimax analysis. To this end,

we analyze the minimum separation required for the differentially private MMD test to achieve significant power in terms of the MMD metric. We derive an upper bound on this minimum separation in Theorem 7, and a lower bound in Theorem 8, which matches in all relevant parameters including testing error rates and privacy levels. Our minimax results suggest that there is an unavoidable loss of power when the privacy parameter is smaller than a certain threshold (*i.e.*, high privacy). On the other hand, the privacy guarantee comes for free in terms of separation rate when the privacy parameter exceeds this threshold (*i.e.*, low privacy). We also derive the minimum separation in terms of the L_2 metric in Theorem 9 that extends the prior work (Li and Yuan, 2019; Schrab et al., 2023) on non-private minimax testing to private settings. In Appendix B.5 and Appendix B.6, we present analogous findings for the differentially private HSIC test, which closely resemble the results obtained for the MMD test.

- **Negative Results of U-statistics (Section 5.3).** We also derive perhaps unexpected negative results of U-statistics in private settings. U-statistics have played an important role, arguably more popular than V-statistics, in deriving non-private minimax rates for various testing problems (Li and Yuan, 2019; Albert et al., 2022; Kim et al., 2022a; Berrett et al., 2021; Schrab et al., 2023). Given this trend, it is natural to consider U-statistics as an initial building block for obtaining private minimax rates. However, it turns out that the U-statistics suffer from higher global sensitivity than the corresponding V-statistics, requiring a higher level of noise to effectively privatize the resulting procedure. We formalize this observation in the context of kernel testing and show that the private tests based on U-statistics have sub-optimal power in high privacy regimes. This negative result naturally justifies our approach based on the V-statistics (equivalently, plug-in estimators) of the MMD and HSIC.

- **Empirical Validation (Section 6).** A significant portion of the prior work on differentially private testing has focused on theoretical aspects, often lacking practical values. On the other hand, practical approaches to differentially private testing frequently rely on heuristics without proper theoretical validation. Our work serves a role in bridging the gap between theory and practice by balancing both theoretical and practical aspects. In particular, we highlight that our method is simple to use and comes with strong theoretical guarantees as we demonstrate throughout the paper. The empirical results in Section 6 and Appendix C also illustrate the competitive performance of the proposed method across diverse scenarios, highlighting its practical value.

We further make contributions by presenting asymptotic distributions of privatized kernel statistics (Appendix B.1), general consistency results for resampling-based tests (Appendix B.2) as well as other technical innovations (Appendix G). We also introduce private kernel tests obtained through the subsample-and-aggregate idea (Appendix D). Due to space constraints, we relegate these additional results to the appendix.

1.3 Organization

The rest of this paper is organized as follows. We begin in Section 2 by providing a brief overview of the fundamental concepts of differential privacy. Section 3 presents our main proposal, namely the

differentially private permutation test, and investigates its finite-sample validity and consistency in power. Moving forward to Section 4, we apply our proposed permutation framework to specific scenarios, focusing on differentially private kernel testing. In particular, we explore the privatization of kernel MMD and HSIC tests, and delve into minimum separation rates of the resulting tests in Section 5. To validate our theoretical findings, Section 6 presents empirical evaluations of the proposed algorithms, comparing their performance with existing methods. Finally, we conclude in Section 7 with a discussion outlining potential directions for future work. All the proofs and additional results are deferred to the appendix.

1.4 Notation

Given two datasets $\mathcal{X}_n := (X_1, \ldots, X_n)$ and $\tilde{\mathcal{X}}_n := (\tilde{X}_1, \ldots, \tilde{X}_n)$, we denote the Hamming distance between $\mathcal{X}_n$ and $\tilde{\mathcal{X}}_n$ by $d_{\mathrm{ham}}(\mathcal{X}_n, \tilde{\mathcal{X}}_n) := \sum_{i=1}^n \mathbb{1}(X_i \neq \tilde{X}_i)$. For two sequences of real numbers a_n, b_n, we write $a_n \lesssim b_n$ (and similarly $a_n \gtrsim b_n$) if there exists some positive constant $C > 0$ independent of n such that $a_n \leq C b_n$ for all $n \geq 1$. We also write $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$. For $x \in \mathbb{R}$, $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x. For a natural number $k \in \mathbb{N}$, we use $[k]$ to denote the set $\{1, \ldots, k\}$. We let $\mathbf{\Pi}_n$ denote the set of all permutations of $[n]$. For a continuous function $f : \mathbb{R}^d \mapsto \mathbb{R}$, the L_2 and L_∞ norms of f are given as $\|f\|_{L_2} = \{\int_{\mathbb{R}^d} f^2(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}\}^{1/2}$ and $\|f\|_{L_\infty} = \sup_{\boldsymbol{x} \in \mathbb{R}^d} |f(\boldsymbol{x})|$, respectively. We say $X \sim \mathsf{Laplace}(0, 1)$ if X follows a Laplace distribution with local and scale parameters $(0, 1)$. We often denote

$$\xi_{\varepsilon,\delta} := \varepsilon + \log\big(1/(1-\delta)\big) \tag{1}$$

to simplify the notation in various places.

2 Background: Differential Privacy

This section presents a brief overview of the basic concepts and properties regarding differential privacy. For a comprehensive treatment, we refer the readers to Dwork et al. (2014). In our work, we adhere to the definition of differential privacy (Dwork et al., 2014, page 25), allowing for the inclusion of additional auxiliary variables. This extended definition requires that the standard differential privacy condition holds for every possible value of the auxiliary variable. In our permutation testing framework, we treat random permutations as the auxiliary variables independent of the dataset.

Definition 1 (Differential Privacy). *Consider a randomized algorithm $\mathcal{A}$, which takes as input a dataset $\mathcal{X}_n$ and an additional auxiliary variable $w \in \mathcal{W}$. For $\varepsilon > 0$ and $\delta \in [0, 1)$, the algorithm $\mathcal{A}$ is said to be (ε, δ)-differentially private if for (i) all $S \in \mathrm{range}(\mathcal{A})$, (ii) all $w \in \mathcal{W}$ and (iii) all two datasets $\mathcal{X}_n$ and $\tilde{\mathcal{X}}_n$ with $d_{\mathrm{ham}}(\mathcal{X}_n, \tilde{\mathcal{X}}_n) \leq 1$, the following inequality holds:*

$$\mathbb{P}\big(\mathcal{A}(\mathcal{X}_n; w) \in S \,|\, \mathcal{X}_n, w\big) \;\leq\; e^{\varepsilon} \mathbb{P}\big(\mathcal{A}(\tilde{\mathcal{X}}_n; w) \in S \,|\, \tilde{\mathcal{X}}_n, w\big) + \delta.$$

We note that $(\varepsilon, 0)$-differential privacy is often simply referred to as ε-DP or pure-DP. On the other hand, (ε, δ)-differential privacy with $\delta \in (0, 1)$ is referred to as approximate-DP, considered as a relaxation of pure-DP. As mentioned by Dwork et al. (2014, page 18), employing a large value

of δ may lead to serious privacy breaches, potentially exposing the complete information of a small number of individuals with a non-trivial probability. Hence, it is generally desirable to choose small values of δ such as $\delta \lesssim \varepsilon^2 n^{-1}$. Nevertheless, our interest lies in exploring entire privacy regimes, and developing comprehensive results applicable in a variety of settings. Consequently, we do not place restrictions on privacy parameters other than $\varepsilon > 0$ and $\delta \in [0, 1)$.

We collect several fundamental properties of differential privacy that are useful in our contexts. The first property is called post-processing (Dwork et al., 2014, Proposition 2.1), which asserts that any arbitrary post-processing applied to the outcome of a differentially private algorithm preserves the same level of privacy.

Lemma 1 (Post-Processing). *Suppose that an algorithm $\mathcal{A}$ is (ε, δ)-differentially private. Then for an arbitrary randomized function f, the composition $f \circ \mathcal{A}$ also preserves (ε, δ)-differentially privacy.*

Another important property of differential privacy, called the composition theorem (Dwork et al., 2014, Theorem 3.16), presents the overall privacy guarantee for a composition of multiple DP mechanisms.

Lemma 2 (Composition). *Suppose that each algorithm $\mathcal{A}_i$ is (ε, δ)-differentially private for $i \in [m]$. Then, the composed algorithm $\mathcal{A}_{1:m}$ defined as $\mathcal{A}_{1:m} \coloneqq (\mathcal{A}_1, \ldots, \mathcal{A}_m)$ is $(\sum_{i=1}^m \varepsilon_i, \sum_{i=1}^m \delta_i)$-differentially private.*

The definition of (ε, δ)-DP immediately leads to the following group property (Acharya et al., 2021, Lemma 19), which plays an important role in constructing minimax lower bounds under DP.

Lemma 3 (Group Privacy). *Suppose that an algorithm $\mathcal{A}$ is (ε, δ)-differentially private. Then for (i) all $S \in \mathrm{range}(\mathcal{A})$, (ii) all $w \in \mathcal{W}$, and (iii) all two datasets $\mathcal{X}_n$ and $\tilde{\mathcal{X}}_n$ with $d_{\mathrm{ham}}(\mathcal{X}_n, \tilde{\mathcal{X}}_n) \leq m$, the following inequality holds:*

$$\mathbb{P}\big(\mathcal{A}(\mathcal{X}_n; w) \in S \,|\, \mathcal{X}_n, w\big) \;\leq\; e^{m\varepsilon}\mathbb{P}\big(\mathcal{A}(\tilde{\mathcal{X}}_n; w) \in S \,|\, \tilde{\mathcal{X}}_n, w\big) + me^{(m-1)\varepsilon}\delta. \tag{2}$$

Several mechanisms have been developed to safeguard differential privacy, with the Laplace mechanism (Dwork et al., 2006b) standing out as one of the most commonly used approaches. To formally state the Laplace mechanism, we first describe the global sensitivity, which is a keystone in the differential privacy framework. We stress that the definition presented below allows us to take into account an additional auxiliary variable w, and thus it is more general than the definition commonly encountered in the DP literature, such as in Dwork et al. (2006b).

Definition 2 (Global ℓ_p-Sensitivity). *Consider a function f taking as input a dataset $\mathcal{X}_n$ and an additional auxiliary variable $w \in \mathcal{W}$, and assume that the output of f lies in $\mathbb{R}^r$. For $p \geq 1$, the global ℓ_p-sensitivity of f is defined as*

$$\Delta_f^p \;\coloneqq\; \sup_{w \in \mathcal{W}} \sup_{\substack{\mathcal{X}_n, \tilde{\mathcal{X}}_n: \\ d_{\mathrm{ham}}(\mathcal{X}_n, \tilde{\mathcal{X}}_n) \leq 1}} \big\| f(\mathcal{X}_n; w) - f(\tilde{\mathcal{X}}_n; w) \big\|_p,$$

where $\|x\|_p$ denotes the ℓ_p norm of a vector x.

The Laplace mechanism works with the ℓ_1-sensitivity, which determines the scaling factor of the Laplace noise injected into the outputs of the function. Formally, the Laplace mechanism is given as follows.

Definition 3 (Laplace Mechanism). *Consider a function f with the ℓ_1-sensitivity Δ_f^1 described in Definition 2. For a given privacy parameter $\xi > 0$, the Laplace mechanism is defined as the random function:*

$$\mathcal{M}_f^{\xi}(\mathcal{X}_n; w) \coloneqq f(\mathcal{X}_n; w) + \frac{\Delta_f^1}{\xi}(\zeta_1, \ldots, \zeta_r)^\top,$$

where $\zeta_1, \ldots, \zeta_r \overset{\text{i.i.d.}}{\sim} \mathsf{Laplace}(0, 1)$ generated independent of $\mathcal{X}_n$ and w.

The privacy guarantee of the Laplace mechanism depends crucially on the choice of privacy parameter ξ. It is well-known that the Laplace mechanism is $(\varepsilon, 0)$-DP when $\xi = \varepsilon$ (Dwork et al., 2014, Theorem 3.6). In general, Acharya et al. (2018, Lemma 5) shows that any $(\varepsilon + \delta, 0)$-DP algorithm is also (ε, δ)-DP. While this strategy is effective for small values of δ, it returns a suboptimal result when δ is close to one. Concretely, when δ approaches one, we are entering the non-private regime where adding noise is unnecessary. However, the Laplace mechanism with $\xi = \varepsilon + \delta$ injects a non-negligible amount of noise to the algorithm. The refined calibration result proposed by Holohan et al. (2015) avoids such issue, proving that $\mathcal{M}_f^{\xi_{\varepsilon,\delta}}$ with $\xi_{\varepsilon,\delta} = \varepsilon + \log(1/(1-\delta))$ is also (ε, δ)-DP. We record this guarantee in the following lemma.

Lemma 4 (Differential Privacy of Laplace Mechanism). *Let $\varepsilon > 0$ and $\delta \in [0, 1)$. The Laplace mechanism $\mathcal{M}_f^{\xi_{\varepsilon,\delta}}$ in Definition 3 with $\xi_{\varepsilon,\delta} = \varepsilon + \log\big(1/(1-\delta)\big)$ is (ε, δ)-differentially private.*

It is worth noting that Holohan et al. (2015) consider typical differential privacy without considering an auxiliary variable w. Nevertheless, the same proof can be applied to differential privacy involving auxiliary variables, provided that we consider the global sensitivity holding uniformly over $w \in \mathcal{W}$ as in Definition 2.

Remark 1 (Gaussian Mechanism). For (ε, δ)-DP, one can also consider the Gaussian mechanism with the ℓ_2-sensitivity, another common method for preserving privacy (Dwork et al., 2006b, 2014). The Gaussian mechanism can be beneficial over the Laplace mechanism when the ℓ_2-sensitivity is significantly smaller than the ℓ_1-sensitivity. However, such benefit is not immediately clear when the outcome of f is one-dimensional where the ℓ_p sensitivity remains the same for any $p \geq 1$. As our framework is mainly concerned with one-dimensional numeric outcomes, we simply focus on the Laplace mechanism and refer to the *global ℓ_1-sensitivity* as the *global sensitivity* whenever it is clear from the context, and simply denote it by Δ_f.

3 Differentially Private Permutation Tests

In this section, we introduce a general framework for constructing a differentially private permutation test. To begin, consider a class of distributions $\mathcal{P}$, which is formed by the union of two disjoint subclasses: $\mathcal{P}_0$ and $\mathcal{P}_1$. Suppose that we observe a random sample $\mathcal{X}_n$ of size n drawn from $P \in \mathcal{P}$.

Given $\mathcal{X}_n$, our ultimate goal is to test whether $H_0 : P \in \mathcal{P}_0$ or $H_1 : P \in \mathcal{P}_1$, while preserving differential privacy. Consider a test statistic $T : \mathcal{X}_n \mapsto \mathbb{R}$, which is assumed to take a large value under the alternative hypothesis H_1. To build on the permutation principle (Lehmann and Romano, 2005, Chapter 15.2), we make the assumption that $\mathcal{X}_n$ is exchangeable under the null H_0. That is, for any permutation $\boldsymbol{\pi} \coloneqq (\pi_1, \ldots, \pi_n) \in \boldsymbol{\Pi}_n$, the joint distribution of $\mathcal{X}_n$ is the same as that of $\mathcal{X}_n^{\boldsymbol{\pi}} \coloneqq (X_{\pi_1}, \ldots, X_{\pi_n})$. Under the exchangeability assumption, the permutation test rejects the null when T is significantly larger than the permuted counterparts. More formally, let $\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_B$ be i.i.d. random permutations of $[n]$, and denote by $T(\mathcal{X}_n^{\boldsymbol{\pi}_1}), \ldots, T(\mathcal{X}_n^{\boldsymbol{\pi}_B})$, the test statistics computed based on $\mathcal{X}_n^{\boldsymbol{\pi}_1}, \ldots, \mathcal{X}_n^{\boldsymbol{\pi}_B}$, respectively. The (Monte Carlo) permutation test then rejects the null hypothesis when the permutation p-value is less than or equal to significance level α, *i.e.*,

$$\widehat{p} \coloneqq \frac{1}{B+1}\left\{\sum_{i=1}^{B} \mathbb{1}\big(T(\mathcal{X}_n^{\boldsymbol{\pi}_i}) \geq T(\mathcal{X}_n)\big) + 1\right\} \leq \alpha. \tag{3}$$

It is well-known that $\widehat{p}$ is super-uniform, *i.e.*, $\mathbb{P}(\widehat{p} \leq t) \leq t$ for all $t \in [0,1]$, under exchangeability of $\mathcal{X}_n$ (*e.g.*, Lemma 15). Therefore the permutation test $\mathbb{1}(\widehat{p} \leq \alpha)$ controls the type I error for any finite sample size n. Our aim is to privatize the permutation test under the DP constraint, while maintaining finite-sample validity and achieving competitive (potentially optimal) power.

For notational convenience, we often write $T_0 = T(\mathcal{X}_n)$ and $T_i = T(\mathcal{X}_n^{\boldsymbol{\pi}_i})$ for $i \in [B]$, and set $\boldsymbol{\pi}_0 = (1, 2, \ldots, n)$ in what follows.

3.1 Proposed Privatization Method

To describe the proposed method, suppose that the test statistic T has the global sensitivity (Definition 2) with the permutation $\boldsymbol{\pi}$ as an auxiliary variable:

$$\Delta_T \coloneqq \sup_{\boldsymbol{\pi} \in \boldsymbol{\Pi}_n} \sup_{\substack{\mathcal{X}_n, \tilde{\mathcal{X}}_n: \\ d_{\mathrm{ham}}(\mathcal{X}_n, \tilde{\mathcal{X}}_n) \leq 1}} \big|T(\mathcal{X}_n^{\boldsymbol{\pi}}) - T(\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}})\big|. \tag{4}$$

Our tailored definition of global sensitivity to permutation tests above is stronger than the usual one since it measures the sensitivity over all possible permutations. However, this additional requirement is not overly restrictive as we demonstrate below for integral probability metrics.

Example 1 (Sensitivity of Integral Probability Metric)**.** Consider a two-sample setting where we observe random variables $\mathcal{Y}_n = \{Y_1, \ldots, Y_n\}$ and $\mathcal{Z}_m = \{Z_1, \ldots, Z_m\}$, each supported on $\mathbb{S}$. Let $\mathcal{F}$ be a class of real-valued functions on $\mathbb{S}$. A plug-in estimator of the corresponding integral probability metric (IPM) is given as

$$T = \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} f(Y_i) - \frac{1}{m} \sum_{i=1}^{m} f(Z_i) \right|. \tag{5}$$

As detailed in Appendix B.3, the global sensitivity of T is precisely equal to

$$\Delta_T = \frac{1}{\min\{n, m\}} \sup_{X, X' \in \mathbb{S}} \sup_{f \in \mathcal{F}} |f(X) - f(X')|.$$

The IPM includes several metrics commonly used in the literature (Sriperumbudur et al., 2012) and their sensitivity can be analyzed as follows.

(a) *Mean difference in ℓ_p*: Let $\mathbb{S} \subset \mathbb{R}^d$, and set $1 \le p, q \le \infty$ such that $1/p + 1/q = 1$. Choosing $\mathcal{F} = \{f : \mathbb{S} \mapsto \mathbb{R} \,|\, f(x) = a^\top x, \|a\|_q \le 1\}$, the IPM becomes the pth norm of the sample mean difference between $\mathcal{Y}_n$ and $\mathcal{Z}_m$. In this case, we have $\sup_{X,X' \in \mathbb{S}} \sup_{f \in \mathcal{F}} |f(X) - f(X')| = \sup_{X,X' \in \mathbb{S}} \|X - X'\|_p$.

(b) *Wasserstein distance*: When $\mathcal{F} = \{f : \mathbb{S} \mapsto \mathbb{R} \,|\, \|f\|_{\mathrm{Lip}} \le 1\}$ where $\|f\|_{\mathrm{Lip}}$ denotes the minimal Lipschitz constant for f on a metric space $(\mathbb{S}, \|\cdot\|)$, the IPM corresponds to the Wasserstein 1-distance. By the Lipschitz property of f, we have $\sup_{X,X' \in \mathbb{S}} \sup_{f \in \mathcal{F}} |f(X) - f(X')| \le \sup_{X,X' \in \mathbb{S}} \|X - X'\|$.

(c) *Total variation distance*: Let $\mathcal{F} = \{f : \mathbb{S} \mapsto \mathbb{R} \,|\, \sup_{x \in \mathbb{S}} |f(x)| \le 1\}$. The corresponding IPM is the total variation distance for which we have $\sup_{X,X' \in \mathbb{S}} \sup_{f \in \mathcal{F}} |f(X) - f(X')| \le 2$.

(d) *Kolmogorov distance*: When $\mathcal{F} = \{\mathbb{1}(-\infty, x] : x \in \mathbb{R}^d\}$, the IPM is called the Kolmogorov distance. In this case, we have $\sup_{X,X' \in \mathbb{S}} \sup_{f \in \mathcal{F}} |f(X) - f(X')| \le 1$ with $\mathbb{S} = \mathbb{R}^d$.

(e) *Maximum mean discrepancy*: Let $\|f\|_{\mathcal{H}_k}$ be the norm of a function f in a reproducing kernel Hilbert space $\mathcal{H}_k$ equipped with kernel k. When $\mathcal{F} = \{f : \mathbb{S} \mapsto \mathbb{R} \,|\, \|f\|_{\mathcal{H}_k} \le 1\}$, the IPM corresponds to the maximum mean discrepancy and it satisfies that $\sup_{X,X' \in \mathbb{S}} \sup_{f \in \mathcal{F}} |f(X) - f(X')| \le \sqrt{2K}$ where K is the maximum value of a non-negative kernel k. See Lemma 5 for details.

In general, obtaining the exact value of the global sensitivity Δ_T can be challenging, and thus we often work with an upper bound for Δ_T. We remark that the differential privacy guarantee of the Laplace mechanism in Lemma 4 remains valid when we replace the sensitivity in the Laplace mechanism with any upper bound. With an abuse of notation, we also use Δ_T to denote an upper bound for the global sensitivity, when the exact value of the global sensitivity is not available.

Naive Approach. Given the global sensitivity of T, one naive attempt to privatize the permutation test is to apply the basic composition theorem (Lemma 2). To depict the idea, let $\{\zeta_i\}_{i=0}^B$ be a sequence of i.i.d. $\mathsf{Laplace}(0,1)$ random variables, and define

$$\widetilde{M}_i \coloneqq T_i + \frac{\Delta_T}{\varepsilon(B+1)^{-1} + \log\big(1/\{1 - \delta(B+1)^{-1}\}\big)} \zeta_i, \quad \text{for } i \in \{0\} \cup [B].$$

By the Laplace mechanism, each $\widetilde{M}_i$ is $\big(\varepsilon/(B+1), \delta/(B+1)\big)$-DP and the composition theorem in Lemma 2 then ensures that the permutation p-value given as

$$\widehat{p}_{\mathrm{dp}}^{\,\mathrm{naive}} \coloneqq \frac{1}{B+1} \left\{ \sum_{i=1}^{B} \mathbb{1}\big(\widetilde{M}_i \ge \widetilde{M}_0\big) + 1 \right\} \tag{6}$$

is (ε, δ)-DP. Moreover, $\{\widetilde{M}_i\}_{i=0}^B$ are exchangeable under the null, which in turn yields that $\widehat{p}_{\mathrm{dp}}^{\,\mathrm{naive}}$ is a valid p-value by Lemma 15. While this naive approach returns rigorous guarantees on both

Algorithm 1 Differentially Private Permutation Test

Input: Data $\mathcal{X}_n$, significance level $\alpha \in (0,1)$, privacy parameters $\varepsilon > 0$ and $\delta \in [0,1)$, test statistic T, global sensitivity (or its upper bound) Δ_T, number of permutations $B \in \mathbb{N}$.

For $i \in [B]$ **do**

Generate a random permutation $\boldsymbol{\pi}_i$ of $[n]$.

Generate $\zeta_i \sim \mathsf{Laplace}(0,1)$.

Set $M_i \leftarrow T(\mathcal{X}_n^{\boldsymbol{\pi}_i}) + 2\Delta_T \xi_{\varepsilon,\delta}^{-1} \zeta_i$ where $\xi_{\varepsilon,\delta} \coloneqq \varepsilon + \log(1/(1-\delta))$.

End For

Generate $\zeta_0 \sim \mathsf{Laplace}(0,1)$ and set $M_0 \leftarrow T(\mathcal{X}_n) + 2\Delta_T \xi_{\varepsilon,\delta}^{-1} \zeta_0$.

Compute the permutation p-value $\widehat{p}_{\mathrm{dp}}$ as in (7).

Output: Reject H_0 if $\widehat{p}_{\mathrm{dp}} \leq \alpha$.

privacy and validity, it has room for improvement in regard to the power performance. Observe that the noise level grows linearly in the number of permutations B. This means that the Laplace noise overwhelms the signal when B is significantly large, leading to a loss of power. It is also worth noting that the permutation p-value is lower bounded by $(B+1)^{-1}$. This means that in order to have non-zero power, the number of permutations B must be greater than $\alpha^{-1} - 1$. Hence, one cannot take B to be arbitrarily small. This issue serves as the motivation for our proposal, which is described below.

Refined Approach. The factor of $B+1$ arises from an application of the composition theorem (Lemma 2), which cannot be improved in general (*e.g.*, Section 2.1 of Steinke, 2022). As one of our key contributions, we remove this unpleasant dependence on B via the quantile representation of the permutation test (Lemma 17). To describe our proposal, define

$$M_i \coloneqq T_i + \frac{2\Delta_T}{\xi_{\varepsilon,\delta}} \zeta_i,$$

for $i \in \{0\} \cup [B]$, where $\xi_{\varepsilon,\delta}$ can be recalled in (1). Notably, the noise level $2\Delta_T \xi_{\varepsilon,\delta}^{-1}$ is independent of B and strictly smaller than that of the naive approach for any $B > 1$. Given $\{M_i\}_{i=0}^B$, we define the private permutation p-value as

$$\widehat{p}_{\mathrm{dp}} \coloneqq \frac{1}{B+1} \left\{ \sum_{i=1}^{B} \mathbb{1}\big(M_i \geq M_0\big) + 1 \right\}, \tag{7}$$

and reject the null when $\widehat{p}_{\mathrm{dp}} \leq \alpha$. We summarize the proposed method in Algorithm 1.

3.2 Validity and Privacy Guarantee

Having introduced our method, we next investigate its theoretical guarantees and provide intuition behind our proposal. We start with the validity of the private test, which follows immediately from Lemma 15.

Theorem 1 (Validity Guarantee). *Suppose that $\mathcal{X}_n$ are exchangeable under the null $H_0 : P \in \mathcal{P}_0$. Then for any $\alpha \in (0,1)$ and $B, n \geq 1$, the type I error of the test $\mathbb{1}(\widehat{p}_{\mathrm{dp}} \leq \alpha)$ from Algorithm 1 satisfies*

$$\sup_{P \in \mathcal{P}_0} \mathbb{P}_P(\widehat{p}_{\mathrm{dp}} \leq \alpha) = \frac{\lfloor (B+1)\alpha \rfloor}{B+1} \leq \alpha.$$

It is worth emphasizing that type I error control of the proposed test is both non-asymptotic and uniform over the entire class of null distributions $\mathcal{P}_0$. Another distinct feature is that the type I error is equal to $\lfloor (B+1)\alpha \rfloor / (B+1)$, which can be strictly smaller than α. If this small gap is a concern, one can make the type I error exactly equal to α through randomization (Lemma 16). We also remark that even if we replace the global sensitivity Δ_T in the procedure with any other value, type I error control remains valid. In other words, the validity of the proposed test is not affected by the noise level of the Laplace mechanism.

Next we turn to the privacy guarantee of the proposed test and show that it is (ε, δ)-DP.

Theorem 2 (Privacy Guarantee). *For any $\alpha \in (0,1)$, the permutation test $\mathbb{1}(\widehat{p}_{\mathrm{dp}} \leq \alpha)$ from Algorithm 1 is (ε, δ)-differentially private.*

It is worth highlighting that the privacy guarantee does not require the exchangeability of $\mathcal{X}_n$. Hence the proposed test is (ε, δ)-DP under both the null and the alternative. As mentioned before, we prove the privacy guarantee of the proposed test via the quantile representation of the permutation test (Lemma 17). That is, rejecting the null when $\widehat{p} \leq \alpha$ where $\widehat{p}$ is given in (3) is equivalent to rejecting the null when $T_0 > Q_{1-\alpha}$ where $Q_{1-\alpha}$ is the $1-\alpha$ quantile of $\{T_i\}_{i=0}^B$. Roughly speaking, our proof proceeds by privatizing T_0 and $Q_{1-\alpha}$, separately, which raises the factor of 2 in $2\Delta_T \xi_{\varepsilon,\delta}^{-1}$. However, a direct application of the Laplace mechanism to T_0 and $Q_{1-\alpha}$ destroys the exchangeability of random variables, thereby type I error control is no longer guaranteed. Making both T_0 and $Q_{1-\alpha}$ private while ensuring the finite-sample validity of the resulting test is non-trivial, and thus we highlight it as our main contribution. Along the way, we develop a general sensitivity result of quantiles in Lemma 19, which may be of independent interest. We also point out that the factor of 2 in the noise level is a price to pay for not knowing the null distribution of T. When T is distribution-free under the null, then it is possible to sharpen the constant factor from two to one.

3.3 Power Analysis

Moving our focus to the power property, we aim to provide tractable conditions for pointwise consistency and non-asymptotic uniform power. Starting with pointwise consistency, the following result provides conditions under which the power converges to one as the sample size increases against a fixed alternative. Below, we add the subscript n to B_n to indicate that the number of permutations can vary with the sample size.

Theorem 3 (Pointwise Consistency). *Let $\alpha \in (0,1)$ be a fixed constant. For a given alternative distribution P, suppose that $\lim_{n\to\infty} \mathbb{P}_P(M_0 \leq M_1) = 0$. Then for any positive sequence of B_n such that $\min_{n\geq 1} B_n + 1 > \alpha^{-1}$, the differentially private permutation test is consistent in power as $\lim_{n\to\infty} \mathbb{P}_P(\widehat{p}_{\mathrm{dp}} \leq \alpha) = 1$.*

In view of the above result, proving consistency of the permutation test essentially boils down to verifying the condition $M_0 > M_1$, *i.e.*, the original statistic is greater than a permuted statistic, with probability approaching one. In Section 4, we showcase the consistency results based on kernel-based methods for two-sample and independence testing. We note that Theorem 3 can be proven in a straightforward manner via a union bound when B_n is fixed. A similar result for fixed B_n can be found in Dobriban (2022, Lemma 5.2) and Rindt et al. (2021, Theorem 6). Extending this result to any arbitrary sequence of B_n requires a different technique that exploits the conditional i.i.d. structure of given variables. To broaden the scope of our paper, we develop a consistency result for general resampling-based tests in Lemma 8 of Appendix B.2, from which we can derive Theorem 3 as a corollary.

While pointwise consistency is a useful property, it is often regarded as a relatively weak guarantee. We now shift our focus to the second result of this subsection, providing a non-asymptotic, uniform guarantee on the power under stronger assumptions. In particular, we identify the moment conditions under which the proposed test has significant power. These conditions can be regarded as the private extension of Kim et al. (2022a, Lemma 3.1). Below, the symbols $\mathbb{E}_{P,\boldsymbol{\pi}}$ and $\mathrm{Var}_{P,\boldsymbol{\pi}}$ denote the expectation and the variance, respectively, taken over both $\mathcal{X}_n$ and $\boldsymbol{\pi}$.

Theorem 4 (Uniform Power)**.** *For $\alpha \in (0,1)$, $\beta \in (0, 1-\alpha)$ and $\xi_{\varepsilon,\delta} > 0$, assume that $B \geq 16\alpha^{-2}\log(8/\beta)$ and for any $P \in \mathcal{P}_1$,*

$$\begin{aligned} \mathbb{E}_P[T(\mathcal{X}_n)] - \mathbb{E}_{P,\boldsymbol{\pi}}[T(\mathcal{X}_n^{\boldsymbol{\pi}})] \;\geq\; & C_1\sqrt{\frac{\mathrm{Var}_P[T(\mathcal{X}_n)] + \mathrm{Var}_{P,\boldsymbol{\pi}}[T(\mathcal{X}_n^{\boldsymbol{\pi}})]}{\alpha\beta}} \\ & + C_2\frac{\Delta_T}{\xi_{\varepsilon,\delta}}\max\left\{\log\left(\frac{1}{\alpha}\right), \log\left(\frac{1}{\beta}\right)\right\}, \end{aligned} \tag{8}$$

where C_1 and C_2 are universal constants. Then the uniform power of the private permutation test is bounded below by $1-\beta$ as

$$\inf_{P\in\mathcal{P}_1} \mathbb{P}_P(\widehat{p}_{\mathrm{dp}} \leq \alpha) \geq 1-\beta.$$

A few remarks are in order.

- The above theorem ensures that the private permutation test has significant power as long as the signal of the problem, namely the difference between the expected values of the original test statistic and of the permuted test statistic, is larger than the noise of the problem, namely the square root of the variances and the noise level of the Laplace mechanism.

- The proof of Theorem 4, given in Appendix E.2, builds on the proof of Kim et al. (2022a, Lemma 3.1) where the key idea is to replace the random permutation threshold with a deterministic one using concentration inequalities. The main distinction from Kim et al. (2022a) is the incorporation of Laplace noises in the analysis, which results in the second line of the condition (8). We also note that Theorem 4 concerns the Monte Carlo permutation test, which is computationally more efficient than the full permutation test analyzed in Kim et al. (2022a, Lemma 3.1).

- Notably, the condition on B is independent of the sample size, which is a consequence of the Dvoretzky–Kiefer–Wolfowitz (DKW) inequality (Massart, 1990), and similar conditions can be found in Schrab et al. (2023) and Schrab et al. (2022). One can improve this restriction, especially constant factors, with more technical effort or by using other techniques, such as the one based on order statistics (Domingo-Enrich et al., 2023, Lemma 6).
- It is worth mentioning that the first line of the condition (8) relies on a polynomial dependence on α and β, which arise from the application of Chebyshev's and Markov's inequalities. If the considered test statistic has an exponential tail bound, these polynomial factors can be improved to logarithmic ones as we illustrate in Section 5.

Before moving on, let us briefly illustrate Theorem 4 based on the plug-in IPM statistic considered in Example 1.

Example 2 (Power Analysis against IPM alternatives)**.** Continuing our discussion from Example 1, denote the IPM between P and Q with a class of functions $\mathcal{F}$ as

$$\mathrm{IPM}_{\mathcal{F}}(P, Q) = \sup_{f \in \mathcal{F}} \left| \mathbb{E}_P[f(Y)] - \mathbb{E}_Q[f(Z)] \right|.$$

Without loss of generality, assume $n \leq m$ and write the pooled sample as $\mathcal{X}_{n+m} = \mathcal{Y}_n \cup \mathcal{Z}_m = \{X_1, \ldots, X_{n+m}\}$. Consider the maximum Rademacher complexity of $\mathcal{F}$ over all possible permuted samples given as

$$\mathcal{R}_n(\mathcal{F}) = \sup_{\boldsymbol{\pi} \in \boldsymbol{\Pi}_{n+m}} \mathbb{E}\left[\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} \omega_i f(X_{\pi_i}) \right| \right],$$

where $\{\omega_i\}_{i=1}^n$ are i.i.d. Rademacher random variables independent of $\mathcal{X}_{n+m}$. Suppose that we implement Algorithm 1 using the plug-in IPM estimator $T(\mathcal{X}_{n+m})$ in (5). Then Theorem 4 yields that the resulting permutation test has the power greater than $1 - \beta$ if

$$\mathrm{IPM}_{\mathcal{F}}(P, Q) \geq C_1 \frac{\mathcal{R}_n(\mathcal{F})}{\sqrt{\alpha\beta}} + C_2 \frac{\sqrt{n}\Delta_T}{\sqrt{\alpha\beta}} + C_3 \frac{\Delta_T}{\xi_{\varepsilon,\delta}} \max\left\{ \log\left(\frac{1}{\alpha}\right), \log\left(\frac{1}{\beta}\right) \right\},$$

where C_1, C_2, C_3 are some positive constants. We defer a detailed analysis that leads to the above result to Appendix B.4. We also refer to van der Vaart and Wellner (1996); Bartlett and Mendelson (2002); Wainwright (2019) for additional information on the Rademacher complexity and illustrative examples.

So far we have examined the properties of the private permutation test in a general context. In the next section, we will apply our framework to the specific problem of kernel testing, and provide a detailed analysis.

4 Application: Differentially Private Kernel Tests

In recent years, there has been a growing trend in employing kernel-based methods for hypothesis testing problems, such as the MMD and the HSIC. This popularity is partly due to their ability to

capture complex, non-linear relationships and to their straightforward implementation. Equipped with such benefits, the MMD (Gretton et al., 2012) is used to measure the difference between two probability distributions, while the HSIC (Gretton et al., 2005) is used to quantify the dependence between two random variables. In this and subsequent sections, we propose differentially private tests based on these two kernel-based measures, and provide an in-depth analysis of their theoretical properties.

Terminology. Before we begin, let us establish the terminology related to kernels. Consider a reproducing kernel $k : \mathbb{S} \times \mathbb{S} \mapsto \mathbb{R}$ defined on a separable topological space $\mathbb{S}$. Let $\mathcal{H}_k$ be a reproducing kernel Hilbert space (RKHS) endowed with kernel k. A kernel k is said to be *characteristic* if the kernel mean embedding

$$\mu_P = \int_{\mathbb{S}} k(\cdot, x) \mathrm{d}P(x) \in \mathcal{H}_k$$

is injective. In addition, a kernel $k : \mathbb{S} \times \mathbb{S} \mapsto \mathbb{R}$ is said to be *translation invariant* if there exists a symmetric positive definite function κ such that $k(x, y) = \kappa(x - y)$ for all $x, y \in \mathbb{S}$. Assuming that $0 \leq k(x, y) \leq K$ for all $x, y \in \mathbb{S}$, we say that the kernel k has *non-empty level sets* on $\mathbb{S}$ if, for any $\epsilon \in (0, K)$, there exist $x, y \in \mathbb{S}$, such that $k(x, y) \leq \epsilon$. Some popular examples of kernels include the Gaussian kernel $k(x, y) = e^{-\sigma \|x-y\|_2^2}$ and the Laplacian kernel $k(x, y) = e^{-\sigma \|x-y\|_1}$ for $\sigma > 0$. These two kernels are translation invariant and known to be characteristic on $\mathbb{R}^d$ (*e.g.*, Sriperumbudur et al., 2011). They also have non-empty level sets on $\mathbb{R}^d$, which can be deduced from the continuity of the kernel function.

4.1 Differentially Private MMD Test

Starting with the MMD, suppose we are given mutually independent samples $\mathcal{Y}_n \coloneqq \{Y_1, \ldots, Y_n\} \overset{\text{i.i.d.}}{\sim} P$ and $\mathcal{Z}_m \coloneqq \{Z_1, \ldots, Z_m\} \overset{\text{i.i.d.}}{\sim} Q$ on a domain $\mathbb{S}$. Without loss of generality, assume $n \leq m$ throughout the rest of this paper. Based on these samples, the two-sample problem aims to determine whether two probability distributions P and Q coincide. A majority of two-sample methodologies target a certain metric between P and Q, and use their empirical counterpart as a test statistic. One such method is the non-private MMD test (Gretton et al., 2012) where the difference between P and Q is quantified in terms of MMD. To elaborate, consider the unit ball in a RKHS $\mathcal{H}_k$ denoted by $\mathcal{F}_k \coloneqq \{f \in \mathcal{H}_k : \|f\|_{\mathcal{H}_k} \leq 1\}$. The maximum mean discrepancy between P and Q is defined as

$$\mathrm{MMD}_k(P, Q) \coloneqq \sup_{f \in \mathcal{F}_k} \big\{ \mathbb{E}_P[f(Y)] - \mathbb{E}_Q[f(Z)] \big\}.$$

The empirical MMD is a plug-in estimator of MMD that replaces P and Q with the corresponding empirical probability measures. Formally, letting $\mathcal{X}_{n+m} = \mathcal{Y}_n \cup \mathcal{Z}_m$ be the pooled sample as before, the empirical MMD is given as

$$\widehat{\mathrm{MMD}}(\mathcal{X}_{n+m}) \coloneqq \sup_{f \in \mathcal{F}_k} \left\{ \frac{1}{n} \sum_{i=1}^{n} f(Y_i) - \frac{1}{m} \sum_{j=1}^{m} f(Z_i) \right\}. \tag{9}$$

Thanks to the reproducing kernel property, the empirical MMD can be computed straightforwardly. In particular, the squared empirical MMD can be calculated in quadratic time using the kernel-based expression

$$\widehat{\mathrm{MMD}}^2(\mathcal{X}_{n+m}) = \frac{1}{n^2}\sum_{i,j=1}^{n} k(Y_i, Y_j) + \frac{1}{m^2}\sum_{i,j=1}^{m} k(Z_i, Z_j) - \frac{2}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} k(Y_i, Z_j). \tag{10}$$

In order to propose a private version of the MMD test, we begin with the global sensitivity of the empirical MMD.

Lemma 5 (Sensitivity of Empirical MMD)**.** *Assume that the kernel k is bounded as $0 \le k(x,y) \le K$ for all $x, y \in \mathbb{S}$. Then the global sensitivity of the empirical MMD satisfies*

$$\sup_{\boldsymbol{\pi}\in\boldsymbol{\Pi}_{n+m}} \sup_{\substack{\mathcal{X}_{n+m},\tilde{\mathcal{X}}_{n+m}:\\ d_{\mathrm{ham}}(\mathcal{X}_{n+m},\tilde{\mathcal{X}}_{n+m})\le 1}} \left|\widehat{\mathrm{MMD}}(\mathcal{X}^{\boldsymbol{\pi}}_{n+m}) - \widehat{\mathrm{MMD}}(\tilde{\mathcal{X}}^{\boldsymbol{\pi}}_{n+m})\right| \le \frac{\sqrt{2K}}{n}.$$

Moreover assume that k is translation invariant, and has non-empty level sets in $\mathbb{S}$. Then the inequality becomes an equality.

The proof of Lemma 5 can be found in Appendix E.5. Note that the sensitivity of the empirical MMD in Lemma 5 can be equivalently defined without the supremum over the permutations $\boldsymbol{\pi} \in \boldsymbol{\Pi}_{n+m}$ as $d_{\mathrm{ham}}(\mathcal{X}_{n+m}, \tilde{\mathcal{X}}_{n+m})$ is the same as $d_{\mathrm{ham}}(\mathcal{X}^{\boldsymbol{\pi}}_{n+m}, \tilde{\mathcal{X}}^{\boldsymbol{\pi}}_{n+m})$ for any $\boldsymbol{\pi} \in \boldsymbol{\Pi}_{n+m}$. As we will see in Section 4.2, however, this property does not hold for independence testing. We also highlight our lower bound result, indicating that the upper bound $\sqrt{2K}/n$ cannot be improved for translation invariant kernels with non-empty level sets on their domain.

With Lemma 5 in place, we set the sensitivity parameter $\Delta_T = \sqrt{2K}n^{-1}$ and run Algorithm 1 with the empirical MMD in (9) as the test statistic. We refer to the resulting private permutation test as the dpMMD test and denote it as ϕ_{dpMMD}. The dpMMD test has the following properties, which are proven in Appendix E.6.

Theorem 5 (Properties of dpMMD test)**.** *Let $\alpha \in (0,1)$ be a fixed constant and the kernel k be bounded as $0 \le k(x,y) \le K$ for all $x, y \in \mathbb{S}$. Then ϕ_{dpMMD} satisfies the following properties:*

P1. (Differential Privacy) For $\varepsilon > 0$ and $\delta \in [0,1)$, ϕ_{dpMMD} is (ε,δ)-differentially private.

P2. (Validity) The type I error of ϕ_{dpMMD} is controlled at level α non-asymptotically.

P3. (Consistency) Suppose that $\mathrm{MMD}_k(P,Q)$ is independent of the sample sizes and strictly positive for a fixed pair of (P,Q). Moreover assume that $n^{-1}\xi^{-1}_{\varepsilon,\delta} \to 0$ as $n \to \infty$. Then for any sequence B_n such that $\min_{n\ge 1} B_n + 1 > \alpha^{-1}$, we have $\lim_{n\to\infty} \mathbb{E}_{P,Q}[\phi_{\mathrm{dpMMD}}] = 1$.

The first two properties on differential privacy and validity are clear in view of Theorem 1 and Theorem 2. It is well-known that the population MMD is strictly positive under the alternative if the kernel is characteristic (Gretton et al., 2012). Hence the condition on the MMD metric in (*P3*) is satisfied for characteristic kernels against any fixed alternative. Another highlight is that

consistency holds irrespective of the relationship between n and m. The power converges to one as long as the minimum sample size goes to infinity. We also remark that the condition $n^{-1}\xi_{\varepsilon,\delta}^{-1} \to 0$ is critical for obtaining consistency. If not, the empirical MMD is overwhelmed by the Laplace noise, which leads to a significant loss of power.

We note in passing the recent work of Yang et al. (2023) that also utilizes the MMD in differentially private data analysis. Despite the fact that both Yang et al. (2023) and ours consider the differentially private MMD, their primary focus is on differentially private data generation, which is different from our focus on hypothesis testing.

4.2 Differentially Private HSIC Test

Turning to the second application, suppose that we are given an i.i.d. paired sample $\mathcal{X}_n = \{(Y_i, Z_i)\}_{i=1}^n$ from a joint distribution P_{YZ} on domain $\mathbb{Y} \times \mathbb{Z}$. Given $\mathcal{X}_n$, the aim of independence testing is to assess whether Y and Z are statistically independent or not. As a kernel dependence measure, the HSIC compares the joint probability measure P_{YZ} to the product of marginals $P_Y P_Z$. To formally define it, let k and ℓ be kernels on $\mathbb{Y}$ and $\mathbb{Z}$, and let $k \otimes \ell$ be the product kernel given as $k \otimes \ell\big((y,z),(y',z')\big) = k(y,y')\ell(z,z')$ for all $y, y' \in \mathbb{Y}$ and $z, z' \in \mathbb{Z}$. Further denoting the unit ball in the RKHS associated with $k \otimes \ell$ by $\mathcal{F}_{k\otimes\ell}$, HSIC is defined as[1]

$$\mathrm{HSIC}_{k\otimes\ell}(P_{YZ}) \coloneqq \sup_{f \in \mathcal{F}_{k\otimes\ell}} \big\{\mathbb{E}_{P_{YZ}}[f(Y,Z)] - \mathbb{E}_{P_Y P_Z}[f(Y,Z)]\big\}.$$

In other words, the HSIC of Y and Z is simply the MMD between P_{YZ} and $P_Y P_Z$ with the product kernel $k \otimes \ell$. The empirical HSIC is a plug-in estimator given as

$$\widehat{\mathrm{HSIC}}(\mathcal{X}_n) \coloneqq \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left\{\frac{1}{n}\sum_{i=1}^n f(Y_i, Z_i) - \frac{1}{n^2}\sum_{i,j=1}^n f(Y_i, Z_j)\right\}. \tag{11}$$

Similarly to the empirical MMD, the squared empirical HSIC also has an explicit form in terms of the kernels k and ℓ as

$$\begin{aligned}\widehat{\mathrm{HSIC}}^2(\mathcal{X}_n) \;=\; & \frac{1}{n^2}\sum_{i,j=1}^n k(Y_i,Y_j)\ell(Z_i,Z_j) + \frac{1}{n^4}\sum_{i_1,i_2,j_1,j_2=1}^n k(Y_{i_1},Y_{j_1})\ell(Z_{i_2},Z_{j_2}) \\ & - \frac{2}{n^3}\sum_{i,j_1,j_2=1}^n k(Y_i,Y_{j_1})\ell(Z_i,Z_{j_2}),\end{aligned} \tag{12}$$

which can be computed in quadratic time as described in Song et al. (2012, Theorem 1). For independence testing, the permutation test proceeds by randomly permuting either the Y observations or the Z observations. Here, we permute the Z observations and denote $\mathcal{X}_n^{\boldsymbol{\pi}} = \{(Y_i, Z_{\pi_i})\}_{i=1}^n$. With this notation in place, the next lemma explores the global sensitivity of the empirical HSIC.

[1] We remark that in the literature HSIC is often defined as the square of this quantity. However, for consistency with MMD, we define it without the square.

Lemma 6 (Sensitivity of Empirical HSIC)**.** *Assume that the kernels k and ℓ are bounded as $0 \leq k(y, y') \leq K$ and $0 \leq \ell(z, z') \leq L$ for all $y, y' \in \mathbb{Y}$ and $z, z' \in \mathbb{Z}$. Then the global sensitivity of the empirical HSIC satisfies*

$$\sup_{\boldsymbol{\pi} \in \boldsymbol{\Pi}_n} \sup_{\substack{\mathcal{X}_n, \tilde{\mathcal{X}}_n: \\ d_{\mathrm{ham}}(\mathcal{X}_n, \tilde{\mathcal{X}}_n) \leq 1}} \left| \widehat{\mathrm{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}}) - \widehat{\mathrm{HSIC}}(\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}}) \right| \leq \frac{4(n-1)}{n^2} \sqrt{KL}.$$

Moreover assume that k and ℓ are translation invariant, and have non-empty level sets on $\mathbb{Y}$ and $\mathbb{Z}$, respectively. Then the global sensitivity is lower bounded by $4(n-2.5)n^{-2}\sqrt{KL}$.

The proof of Lemma 5 can be found in Appendix E.7. Contrary to the MMD case, we observe that the two hamming distances, namely $d_{\mathrm{ham}}(\mathcal{X}_n^{\boldsymbol{\pi}}, \tilde{\mathcal{X}}_n^{\boldsymbol{\pi}})$ and $d_{\mathrm{ham}}(\mathcal{X}_n, \tilde{\mathcal{X}}_n)$, can differ for independence testing. Consequently, the supremum over the permutations $\boldsymbol{\pi} \in \boldsymbol{\Pi}_n$ plays a non-trivial role in the sensitivity of the empirical HSIC. We mention the work of Kusner et al. (2016) that also examines the global sensitivity of the empirical HSIC. Our upper bound result improves theirs by replacing the constant factor 12 to 4 with a tighter analysis. In fact, as the lower bound result states, the proposed upper bound is asymptotically tight under mild conditions for k and ℓ.

In view of the above lemma, we set $\Delta_T = 4(n-1)n^{-2}\sqrt{KL}$ and run Algorithm 1 with the empirical HSIC in (11) as the test statistic. We refer to the resulting permutation test as the dpHSIC test and denote it as ϕ_{dpHSIC}. Similar to Theorem 5, the dpHSIC test has the following properties, which are proven in Appendix E.8.

Theorem 6 (Properties of dpHSIC test)**.** *Let $\alpha \in (0, 1)$ be a fixed constant and assume that the kernels k and ℓ are bounded as $0 \leq k(y, y') \leq K$ and $0 \leq \ell(z, z') \leq L$ for all $y, y' \in \mathbb{Y}$ and $z, z' \in \mathbb{Z}$. Then ϕ_{dpHSIC} satisfies the following properties:*

P1. (Differential Privacy) For $\epsilon > 0$ and $\delta \in [0, 1)$, ϕ_{dpHSIC} is (ε, δ)-differentially private.

P2. (Validity) The type I error of ϕ_{dpHSIC} is controlled at level α non-asymptotically.

P3. (Consistency) Suppose that $\mathrm{HSIC}_{k \otimes \ell}(P_{YZ})$ is independent of the sample sizes and strictly positive for a fixed distribution P_{YZ}. Moreover assume that $n^{-1}\xi_{\varepsilon,\delta}^{-1} \to 0$ as $n \to \infty$. Then for any sequence B_n such that $\min_{n \geq 1} B_n + 1 > \alpha^{-1}$, we have $\lim_{n \to \infty} \mathbb{E}_{P_{YZ}}[\phi_{\mathrm{dpHSIC}}] = 1$.

As for the dpMMD test, the first two properties on differential privacy and validity are direct consequences of Theorem 1 and Theorem 2. The condition for consistency is ensured under any alternative when the kernels are characteristic (Gretton, 2015). Therefore the dpHSIC test equipped with a characteristic kernel is pointwise consistent against any fixed alternative, provided that $n^{-1}\xi_{\varepsilon,\delta}^{-1} \to 0$ and $\min_{n \geq 1} B_n + 1 > \alpha^{-1}$.

Before moving on and studying uniform power properties, let us briefly remark on the asymptotic null distributions of private kernel test statistics.

Remark 2 (Asymptotic null distributions)**.** As mentioned earlier, when the null distribution is tractable, we can improve the power by eliminating the factor of 2 in the noise level. However, characterizing the limiting distribution is not a trivial task, even for non-private kernel statistics.

In Appendix B.1, we show that a private kernel statistic converges in distribution to a mixture of Gaussian chaos and Laplace distributions, which is even more intricate than the limiting distribution of a non-private kernel statistic. A recent line of work (Shekhar et al., 2022a,b) propose cross MMD and cross HSIC that have a tractable limiting distribution with competitive power. We leave the exploration of extending these variants to the private setting and comparing their power performance with our proposed methods as an avenue for future research.

5 Uniform Power and Optimality

In the previous section, we examined the fundamental properties of the private kernel tests, including their asymptotic power against fixed alternatives. This section delves into a more challenging setting where the alternative can shrink to the null as the sample size increases, and develops uniform power results. Moreover we highlight an intrinsic trade-off between privacy and statistical power through the lens of minimax analysis, and explore optimality of the proposed private tests under the differential privacy constraint. In the main text, we focus on the analysis of the dpMMD test, and defer analogous results for the dpHSIC test to Appendix B.5 and Appendix B.6.

5.1 Separation in MMD Metric

Consider the setting described in Section 4.1, and denote by $\mathcal{P}_{\mathbb{S}}$ the class of distributions defined on $\mathbb{S}$. Our first goal is to determine the minimum separation for ϕ_{dpMMD} based on the MMD metric with kernel k. To this end, for $\rho > 0$, we define a class of paired distributions (P, Q) such that

$$\mathcal{P}_{\text{MMD}_k}(\rho) \coloneqq \big\{(P, Q) \in \mathcal{P}_{\mathbb{S}} \times \mathcal{P}_{\mathbb{S}} : \text{MMD}_k(P, Q) \geq \rho\big\}.$$

For a given target type II error $\beta \in (0, 1-\alpha)$, the minimum separation for the dpMMD test against $\mathcal{P}_{\text{MMD}_k}(\rho)$ is given by

$$\rho_{\phi_{\text{dpMMD}}}(\alpha, \beta, \varepsilon, \delta, m, n) \coloneqq \inf\bigg\{\rho > 0 : \sup_{(P,Q)\in\mathcal{P}_{\text{MMD}_k}(\rho)} \mathbb{E}_{P,Q}[1 - \phi_{\text{dpMMD}}] \leq \beta\bigg\}. \tag{13}$$

In simpler terms, the minimum separation $\rho_{\phi_{\text{dpMMD}}}$ refers to the smallest MMD metric between P and Q that can be correctly detected by the dpMMD test with probability at least $1-\beta$. The next theorem provides an upper bound for $\rho_{\phi_{\text{dpMMD}}}$ as a function of the parameters α, β, ε, δ, m, and n. The proof can be found in Appendix E.9.

Theorem 7 (Minimum Separation of dpMMD over $\mathcal{P}_{\text{MMD}_k}$)**.** *Assume that the kernel k is bounded as $0 \leq k(x, y) \leq K$ for all $x, y \in \mathbb{S}$, and $n \leq m \leq \tau n$ for some fixed constant $\tau \geq 1$. Then for all values of $\alpha \in (0, 1)$, $\beta \in (0, 1-\alpha)$, $\varepsilon > 0$, $\delta \in [0, 1)$ and $B \geq 16\alpha^{-2}\log\big(8/\beta\big)$, the minimum separation for ϕ_{dpMMD} satisfies*

$$\rho_{\phi_{\text{dpMMD}}} \leq C_{K,\tau} \max\Bigg\{\sqrt{\frac{\max\big\{\log(1/\alpha),\, \log(1/\beta)\big\}}{n}},\, \frac{\max\big\{\log(1/\alpha),\, \log(1/\beta)\big\}}{n\xi_{\varepsilon,\delta}}\Bigg\},$$

where $C_{K,\tau}$ is a positive constant that depends only on K and τ, and $\xi_{\varepsilon,\delta}$ can be recalled in (1).

We present several comments on the upper bound result.

- Theorem 7 states that the separation rate for the dpMMD test becomes $n^{-1/2}$ in low privacy regimes (*i.e.*, $\xi_{\varepsilon,\delta} \gtrsim n^{-1/2}$), whereas it becomes $n^{-1}\xi_{\varepsilon,\delta}^{-1}$ in high privacy regimes (*i.e.*, $\xi_{\varepsilon,\delta} \lesssim n^{-1/2}$). Notably, this upper bound result allows the parameters $\alpha, \beta, \xi_{\varepsilon,\delta}$ to vary freely within the constraints in the theorem statement. We also mention that the minimum separation is meaningful only when $n^{-1}\xi_{\varepsilon,\delta}^{-1} \to 0$, which coincides with the condition for consistency established in Theorem 5.

- We point out that ϕ_{dpMMD} is equivalent to the non-DP MMD test (Gretton et al., 2012) when $\varepsilon \to \infty$ or $\delta \to 1$ (*i.e.*, $\xi_{\varepsilon,\delta} \to \infty$). Thus, our result also yields the minimum separation rate for the non-DP MMD test as a byproduct.

- One can prove Theorem 7 by verifying the general conditions in Theorem 4. However this strategy results in polynomial factors of α and β instead of logarithmic ones. To obtain logarithmic dependence in both α and β, we modify the proof of Theorem 4 and utilize exponential concentration inequalities for the empirical MMD statistic (Lemma 13) and permuted MMD statistic (Lemma 10). As we will see in Theorem 8, these logarithmic factors cannot be improved further when $\alpha \asymp \beta$.

- The constraint on the sample size ratio can be completely removed by using the Markov inequality for a permuted MMD statistic (Lemma 11). Nevertheless, this alternative approach yields a polynomial factor of α instead of a logarithmic one. See Remark 3. It is currently unknown whether the constraint on m and n can be eliminated, while preserving the logarithmic factors.

We next investigate minimax optimality of ϕ_{dpMMD} under certain regimes in $\mathbb{S} = \mathbb{R}^d$. To set the stage, let $\phi : \mathcal{Y}_n \cup \mathcal{Z}_m \mapsto \{0, 1\}$ be a test function, and denote the set of (ε, δ)-DP level α tests as

$$\Phi_{\alpha,\varepsilon,\delta} \coloneqq \Big\{\phi : \sup_{P \in \mathcal{P}_{\mathbb{S}}} \mathbb{E}_{P,P}[\phi] \leq \alpha \text{ and } \phi \text{ is } (\varepsilon, \delta)\text{-DP}\Big\}.$$

From a theoretical point of view, it is of interest to figure out an information-theoretic lower bound on the minimum separation for any test. This is often called the minimax separation or critical radius in the literature (Ingster, 1994; Ingster et al., 2003; Baraud, 2002). Formally, the minimax separation in terms of the MMD metric is defined as

$$\rho^{\star}_{\mathrm{MMD}}(\alpha, \beta, \varepsilon, \delta, m, n) \coloneqq \inf\Big\{\rho > 0 : \inf_{\phi \in \Phi_{\alpha,\varepsilon,\delta}} \sup_{(P,Q) \in \mathcal{P}_{\mathrm{MMD}_k}(\rho)} \mathbb{E}_{P,Q}[1 - \phi] \leq \beta\Big\}.$$

In simpler terms, the minimax separation $\rho^{\star}_{\mathrm{MMD}}$ refers to the largest MMD metric between P and Q that cannot be correctly detected with probability at least $1 - \beta$ by any level α test. We say that a test ϕ is minimax rate optimal in terms of the MMD metric if the minimum separation of ϕ is equivalent to $\rho^{\star}_{\mathrm{MMD}}$ up to constant factors. The next theorem, proved in Appendix E.10, establishes a lower bound for the minimax separation under the DP constraint, from which we demonstrate minimax optimality of the dpMMD test.

Theorem 8 (Minimax Separation over $\mathcal{P}_{\mathrm{MMD}_k}$). *Let α and β be real numbers in the interval $(0, 1/5)$, $\varepsilon > 0$ and $\delta \in [0, 1)$. Assume that the kernel function k is translation invariant on $\mathbb{R}^d$. In particular there exists some function κ such that $k(x, y) = \kappa(x - y)$ for all $x, y \in \mathbb{R}^d$. Moreover, the kernel is non-constant in the sense that there exists a positive constant η such that $\kappa(0) - \kappa(z) \geq \eta$ for some $z \in \mathbb{R}^d$. Then the minimax separation over $\mathcal{P}_{\mathrm{MMD}_k}$ is lower bounded as*

$$\rho^{\star}_{\mathrm{MMD}} \geq C_\eta \max\left\{ \min\left(\sqrt{\frac{\log(1/(\alpha+\beta))}{n}}, 1 \right), \min\left(\frac{\log(1/\beta)}{n\xi_{\varepsilon,\delta}}, 1 \right) \right\},$$

where C_η is a positive constant that only depends on η, and $\xi_{\varepsilon,\delta}$ can be recalled in (1).

Several remarks are in order.

- First of all, the restriction on α and β is mild as we are typically interested in small values of α and β. In fact, the same result holds for any α, β such that $\alpha + \beta \leq C$ where C is some fixed constant strictly smaller than $1/2$. We also note that for a bounded kernel ranging from 0 and K, the MMD as well as the corresponding minimax separation cannot exceed $\sqrt{2K}$. Our lower bound result captures this restriction through the minimum operator.

- Second, our proof builds on the (ε, δ)-DP Le Cam's method outlined in Acharya et al. (2018, 2021). This technique generalizes classical Le Cam's two-point method (Le Cam, 1973) to private settings via coupling argument. As pointed out by Acharya et al. (2018, Lemma 5), one can obtain a lower bound result for (ε, δ)-DP by replacing ε with $\varepsilon + \delta$ in the lower bound result for ε-DP. However, this method fails to yield a tight lower bound in terms of β. Our approach differs from Acharya et al. (2018, Lemma 5) and returns a sharp lower bound for all parameters of interest, namely $\beta, n, \varepsilon, \delta$.

- Theorem 8 holds for translation invariant kernels. Indeed, many kernels commonly used in practice are translation invariant including the Gaussian, Laplacian, inverse multiquadrics and Matérn kernels. Moreover, as discussed in Tolstikhin et al. (2017), if we further assume that the kernel k is characteristic, it guarantees the existence of $z \in \mathbb{R}^d$ and $\eta > 0$ that satisfy the conditions of Theorem 8. For instance, for the Gaussian kernel $k(x, y) = e^{-\sigma\|x-y\|_2^2}$, one can take $\eta = \frac{\sigma}{2}\|z\|_2^2$ for any non-zero z such that $\|z\|_2^2 \leq \sigma^{-1}$.

- The last point worth highlighting is that our lower bound permits varying values of α and β, which is in contrast to most existing research on minimax testing. A notable exception is the recent work by Diakonikolas et al. (2021), which examines the sample complexity of testing for discrete distributions with high probability.

We now compare the results of Theorem 7 and Theorem 8, and observe that the lower bound for $\rho^{\star}_{\mathrm{MMD}}$ matches the upper bound for $\rho_{\phi_{\mathrm{dpMMD}}}$ in the regime where $m \asymp n$ and $\alpha \asymp \beta$ for $\mathbb{S} = \mathbb{R}^d$. This shows that the proposed dpMMD test is minimax rate optimal against the class of alternatives determined by the MMD metric in the considered regime. It is noteworthy that there is no restriction on the privacy parameters $\varepsilon > 0$ and $\delta \in [0, 1)$, and hence the dpMMD test achieves optimal separation rates in all privacy regimes.

5.2 Separation in L_2 Metric

We next investigate the minimum separation of the dpMMD test in terms of the L_2 metric. Let p and q denote the Lebesgue density functions of P and Q, respectively, defined on $\mathbb{R}^d$. As in Schrab et al. (2023) and Li and Yuan (2019), we restrict our attention to a smooth class of density functions defined over a Sobolev ball. In particular, for a smoothness parameter $s > 0$ and a radius $R > 0$, the Sobolev ball $\mathcal{S}_d^s(R)$ is given as

$$\mathcal{S}_d^s(R) := \left\{ f \in L_1(\mathbb{R}^d) \cap L_2(\mathbb{R}^d) : \int_{\mathbb{R}^d} \|w\|_2^{2s} |\widehat{f}(w)|^2 \mathrm{d}w \leq (2\pi)^d R^2 \right\}, \tag{14}$$

where $\widehat{f}$ is the Fourier transform of f, *i.e.*, $\widehat{f}(w) = \int_{\mathbb{R}^d} f(x) e^{-ix^\top w} \mathrm{d}x$ for $w \in \mathbb{R}^d$. The condition $f \in L_1(\mathbb{R}^d) \cap L_2(\mathbb{R}^d)$ simply requires the function $f \colon \mathbb{R}^d \to \mathbb{R}$ to be both integrable and square-integrable with respect to the Lebesgue measure. For $\rho > 0$, let $\mathcal{P}_{L_2}(\rho)$ be the collection of paired distributions (P, Q) on $\mathbb{R}^d \times \mathbb{R}^d$ where P and Q are equipped with the Lebesgue density functions p and q, respectively, such that $\|p - q\|_{L_2} \geq \rho$. The target class of distributions is a subset of $\mathcal{P}_{L_2}(\rho)$ defined as

$$\mathcal{P}_{L_2}^s(\rho) := \left\{ (P, Q) \in \mathcal{P}_{L_2}(\rho) : p - q \in \mathcal{S}_d^s(R),\ \max(\|p\|_{L_\infty}, \|q\|_{L_\infty}) \leq M \right\}.$$

The aim of this subsection is to characterize the minimum value of ρ for which the dpMMD test has significant power uniformly over $\mathcal{P}_{L_2}^s(\rho)$. For simplicity, we focus on the dpMMD test with a Gaussian kernel. This choice is motivated by the observation that the population MMD with the Gaussian kernel approximates $\|p - q\|_{L_2}^2$ for small bandwidth values (Li and Yuan, 2019), and a similar result can be derived using other kernels in view of Schrab et al. (2023). For $x = (x_1, \ldots, x_d)^\top \in \mathbb{R}^d$ and $y = (y_1, \ldots, y_d)^\top \in \mathbb{R}^d$, the Gaussian kernel with bandwidth $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)^\top \in (0, \infty)^d$ is given as

$$k_{\boldsymbol{\lambda}}(x, y) = \prod_{i=1}^d \frac{1}{\sqrt{2\pi}\lambda_i} e^{-\frac{(x_i - y_i)^2}{2\lambda_i^2}}.$$

Let us denote the minimum separation of the dpMMD test with the Gaussian kernel against L_2 alternatives as

$$\rho_{\phi_{\mathrm{dpMMD}}, L_2}(\alpha, \beta, \varepsilon, \delta, m, n, d, s, R, M) := \inf \left\{ \rho > 0 : \sup_{(P,Q) \in \mathcal{P}_{L_2}^s(\rho)} \mathbb{E}_{P,Q}[1 - \phi_{\mathrm{dpMMD}}] \leq \beta \right\}. \tag{15}$$

The next theorem, proved in Appendix E.11, provides an upper bound for $\rho_{\phi_{\mathrm{dpMMD}}, L_2}$ in terms of a set of parameters, including the bandwidth $\boldsymbol{\lambda}$ and sample sizes.

Theorem 9 (Minimum Separation of dpMMD over $\mathcal{P}_{L_2}^s$)**.** *Assume that $n \leq m \leq \tau n$ for some fixed constant $\tau \geq 1$, and that $\alpha \in (0, e^{-1})$, $\beta \in (0, 1 - \alpha)$, $\varepsilon > 0$, $\delta \in [0, 1)$, $B \geq 16\alpha^{-2} \log(8/\beta)$ and $\prod_{i=1}^d \lambda_i \leq 1$. The minimum separation of the* dpMMD *test with the Gaussian kernel over $\mathcal{P}_{L_2}^s$ is*

upper bounded as

$$\rho^2_{\phi_{\mathrm{dpMMD}},L_2} \le C_{\tau,\beta,s,R,M,d}\Bigg\{\sum_{i=1}^{d}\lambda_i^{2s} + \frac{\log(1/\alpha)}{n\sqrt{\lambda_1\cdots\lambda_d}} + \frac{\log(1/\alpha)}{n^{3/2}\lambda_1\cdots\lambda_d\xi_{\varepsilon,\delta}} + \frac{\log^2(1/\alpha)}{n^2\lambda_1\cdots\lambda_d\xi^2_{\varepsilon,\delta}} + \frac{\log^{3/2}(1/\alpha)}{n^{3/2}(\lambda_1\cdots\lambda_d)^{3/4}\xi_{\varepsilon,\delta}}\Bigg\},$$

where $C_{\tau,\beta,s,R,M,d}$ is a positive constant, depending only on τ,β,s,R,M,d, and $\xi_{\varepsilon,\delta}$ is as in (1).

There are several points needed to be highlighted. To facilitate our discussion, assume that α is a fixed number and write

$$\text{(I)} = \frac{\log(1/\alpha)}{n\sqrt{\lambda_1\cdots\lambda_d}},\ \text{(II)} = \frac{\log(1/\alpha)}{n^{3/2}\lambda_1\cdots\lambda_d\xi_{\varepsilon,\delta}},\ \text{(III)} = \frac{\log^2(1/\alpha)}{n^2\lambda_1\cdots\lambda_d\xi^2_{\varepsilon,\delta}},\ \text{(IV)} = \frac{\log^{3/2}(1/\alpha)}{n^{3/2}(\lambda_1\cdots\lambda_d)^{3/4}\xi_{\varepsilon,\delta}}.$$

When α is fixed, we can absorb the term (IV) into the term (II) as $\lambda_1\cdots\lambda_d \le 1$, and simplify the interpretation of the result as follows.

- In the low privacy regime where the first term (I) dominates the others, our result recovers Schrab et al. (2023, Theorem 6), which studies the minimum separation of the non-private MMD test against $\mathcal{P}^s_{L_2}$. In this low privacy regime, by setting bandwidths $\lambda_i = n^{-2/(4s+d)}$ for $i \in [d]$, we can achieve the optimal separation rate over the Sobolev ball, that is $n^{-2s/(4s+d)}$.

- In the mid privacy regime where the term (II) becomes a leading term, equating (II) with $\sum_{i=1}^d \lambda_i^{2s}$ yields the optimal choice of bandwidths $\lambda_i = n^{-3/(4s+2d)}\xi_{\varepsilon,\delta}^{-1/(2s+d)}$ for $i \in [d]$. The resulting separation rate is $n^{-3s/(4s+2d)}\xi_{\varepsilon,\delta}^{-s/(2s+d)}$. Similarly, in the high privacy regime where the term (III) dominates the others, equating (III) with $\sum_{i=1}^d \lambda_i^{2s}$ yields the optimal choice of bandwidths $\lambda_i = (n\xi_{\varepsilon,\delta})^{-2/(2s+d)}$ for $i \in [d]$. This returns the separation rate $(n\xi_{\varepsilon,\delta})^{-2s/(2s+d)}$. By tracking conditions for each term dominating the others, the minimum separation rate that one can achieve using different bandwidths is summarized as

$$\rho_{\phi_{\mathrm{dpMMD}},L_2} \lesssim \begin{cases} n^{-\frac{2s}{4s+d}}, & \text{if } n^{-\frac{2s-d/2}{4s+d}} \lesssim \xi_{\varepsilon,\delta} \text{ (low privacy)},\\ (n^{\frac{3}{2}}\xi_{\varepsilon,\delta})^{-\frac{s}{2s+d}}, & \text{if } n^{-\frac{1}{2}} \lesssim \xi_{\varepsilon,\delta} \lesssim n^{-\frac{2s-d/2}{4s+d}} \text{ (mid privacy)},\\ (n\xi_{\varepsilon,\delta})^{-\frac{2s}{2s+d}}, & \text{if } \xi_{\varepsilon,\delta} \lesssim n^{-\frac{1}{2}} \text{ (high privacy)}. \end{cases} \tag{16}$$

  In particular, when $\xi_{\varepsilon,\delta} \asymp n^{-1/2}$, the separation rate becomes $n^{-2/(2s+d)}$, which is known to be the minimax optimal rate of density estimation under the L_2 loss. We refer to Appendix B.8 for a detailed discussion of the separation rate.

- It is important to note that all these separation rates are achieved by using different bandwidths, which requires knowledge of the smoothness parameter s. Building on the idea of Ingster (2000); Schrab et al. (2023); Biggs et al. (2023), one can develop an aggregated dpMMD test that is adaptive to s without losing much power. The main idea would be to consider a wide range of private MMD statistics with different bandwidths and aggregate them properly. A detailed analysis of this approach is left for future research.

- While the dpMMD test can achieve the minimax rate in the low privacy regime, it remains unknown whether the derived separation rates are optimal in the mid/high privacy regimes. We believe that (ε, δ)-DP Le Cam's method (Acharya et al., 2021, Theorem 1) plays an important role in constructing a lower bound for the L_2 separation as well. The key challenge lies in finding a coupling between continuous distributions, yielding a small expected Hamming distance. We leave this important direction for future work.

In contrast to the prior work (Li and Yuan, 2019; Schrab et al., 2023) that utilize a U-statistic for minimax two-sample testing, our approach is based on a plug-in estimator, also known as a V-statistic, of the MMD. While plug-in estimators can often exhibit suboptimal performance in estimation problems due to their inherent bias, they can still achieve optimal results in testing problems. This can be explained by the interplay between the test statistic and the critical value in a testing procedure, where the bias terms in these components may offset each other. Theorem 9 demonstrates this phenomenon by showing that the test based on the plug-in estimator of the MMD attains the minimax separation rate over $\mathcal{P}^s_{L_2}$ in the low privacy regime. Perhaps more interestingly, the plug-in estimator can outperform the U-statistic by having lower sensitivity and thus leading to greater power in high privacy regimes. This aspect of plug-in estimators has not been noticed in the literature, and we provide a more detailed discussion in the next subsection.

5.3 Private Test based on the MMD U-statistic

It has been shown that kernel tests based on U-statistics often produce optimal separation rates in non-DP settings (Li and Yuan, 2019; Schrab et al., 2023; Albert et al., 2022; Kim et al., 2022a). Therefore, one can naturally expect that their private extensions perform similarly well across different privacy regimes. In this section, we prove that this is not necessarily the case. In particular, we illustrate that the private MMD permutation test based on a U-statistic is provably outperformed by our approach based on the plug-in MMD estimate in high privacy regimes. A similar result for HSIC can be found in Appendix B.7.

We begin with the explicit form of the MMD U-statistic, which is an unbiased estimator of MMD_k^2, given as

$$U_{\mathrm{MMD}}(\mathcal{X}_{n+m}) \coloneqq \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} k(Y_i, Y_j) + \frac{1}{m(m-1)} \sum_{1 \leq i \neq j \leq m} k(Z_i, Z_j) - \frac{2}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} k(Y_i, Z_j).$$

The following lemma calculates the global sensitivity of U_{MMD}, which is proved in Appendix E.12.

Lemma 7 (Global Sensitivity of U_{MMD})**.** *Assume that the kernel k is bounded as $0 \leq k(x, y) \leq K$ for all $x, y \in \mathbb{S}$. In addition, assume that k is translation invariant, and have non-empty level sets on $\mathbb{S}$. Then there exists a positive sequence $c_{m,n} \in [4, 8]$ such that for all $2 \leq n \leq m$,*

$$\sup_{\boldsymbol{\pi} \in \boldsymbol{\Pi}_{n+m}} \sup_{\substack{\mathcal{X}_{n+m}, \tilde{\mathcal{X}}_{n+m}: \\ d_{\mathrm{ham}}(\mathcal{X}_{n+m}, \tilde{\mathcal{X}}_{n+m}) \leq 1}} \left| U_{\mathrm{MMD}}(\mathcal{X}^{\boldsymbol{\pi}}_{n+m}) - U_{\mathrm{MMD}}(\tilde{\mathcal{X}}^{\boldsymbol{\pi}}_{n+m}) \right| = \frac{c_{m,n} K}{n}.$$

The lemma above indicates that the global sensitivity of the U-statistic has the same dependence on n as that of the plug-in MMD in Lemma 5. However, it is important to mention that their target parameters are different. The U-statistic is an estimator of MMD_k^2, whereas the plug-in estimator given in (9) estimates MMD_k without squaring. This key difference can lead to a significant gap in their power performance in privacy regimes as explored below.

Given the sensitivity of U_{MMD} in Lemma 7, we consider the private permutation test in Algorithm 1 using the test statistic U_{MMD} and the global sensitivity $\Delta_T = c_{m,n} K n^{-1}$. Let us denote the resulting private test by ϕ^u_{dpMMD}. We analyze the minimum separations of ϕ^u_{dpMMD} over $\mathcal{P}_{\mathrm{MMD}_k}$ and $\mathcal{P}^s_{L_2}$ in Theorem 10 and Theorem 11, respectively, and compare them with those of ϕ_{dpMMD} based on the plug-in estimator. Starting with the MMD alternative, the following theorem demonstrates that ϕ^u_{dpMMD} fails to achieve the minimax separation rate over $\mathcal{P}_{\mathrm{MMD}_k}$.

Theorem 10 (Suboptimality of ϕ^u_{dpMMD} against MMD Alternatives)**.** *Assume that the kernel k fulfills the conditions specified in Lemma 7. Moreover, assume that if $P, Q \in \mathcal{P}_{\mathbb{S}}$, then $wP + (1-w)Q \in \mathcal{P}_{\mathbb{S}}$ for all $w \in [0,1]$, and there exist $P_0, Q_0 \in \mathcal{P}_{\mathbb{S}}$ such that $\mathrm{MMD}_k(P_0, Q_0) = \varrho_0$ for some fixed $\varrho_0 > 0$. Let $\alpha \in \big((B+1)^{-1}, 1\big)$, $\beta \in (0, 1-\alpha)$ be fixed values. Consider the high privacy regime where $\xi_{\varepsilon,\delta} \asymp n^{-1/2-r}$ with fixed $r \in (0, 1/2)$, for $\xi_{\varepsilon,\delta}$ as in (1). Then the uniform power of ϕ^u_{dpMMD} is asymptotically at most α over $\mathcal{P}_{\mathrm{MMD}_k}(\rho)$ where*

$$\rho = \log(n) \times \max\left\{ \sqrt{\frac{\max\{\log(1/\alpha),\, \log(1/\beta)\}}{n}},\ \frac{\max\{\log(1/\alpha),\, \log(1/\beta)\}}{n\xi_{\varepsilon,\delta}} \right\}. \tag{17}$$

In other words, it holds that

$$\limsup_{n\to\infty} \inf_{(P,Q)\in\mathcal{P}_{\mathrm{MMD}_k}(\rho)} \mathbb{E}_{P,Q}[\phi^u_{\mathrm{dpMMD}}] \le \alpha.$$

Theorem 10, proven in Appendix E.13, clearly shows that ϕ^u_{dpMMD} is not minimax optimal in the MMD metric as $\rho/\rho^\star_{\mathrm{MMD}} \to \infty$ as $n \to \infty$. We also mention that the factor $\log(n)$ in ρ is chosen for convenience, and it can be replaced by any other positive sequence that increases slower than n^r for $r \in (0, 1/2)$. The suboptimal performance of ϕ^u_{dpMMD} primarily stems from the relatively high noise level associated with the Laplace mechanism. Intuitively, we expect that ϕ^u_{dpMMD} is powerful in the private regime when the target parameter $\mathrm{MMD}_k^2(P, Q)$ is larger than the Laplace noise level $(n\xi_{\varepsilon,\delta})^{-1}$, equivalently, $\mathrm{MMD}_k(P, Q)$ is larger than $(n\xi_{\varepsilon,\delta})^{-1/2}$. Otherwise, the test statistic will be dominated by the Laplace noise. Importantly, the minimax separation under high privacy regimes in Theorem 8 is associated with $\min\{(n\xi_{\varepsilon,\delta})^{-1}, 1\}$, which is smaller than $(n\xi_{\varepsilon,\delta})^{-1/2}$. This briefly explains the suboptimality of ϕ^u_{dpMMD} against the MMD alternative. Nevertheless, our analysis is limited to the U-statistic with the Laplace mechanism, and it is unknown whether the U-statistic in conjunction with other DP mechanisms can lead to optimality.

Turning to the L_2 alternative, let us denote the minimum separation of ϕ^u_{dpMMD} with the Gaussian kernel over $\mathcal{P}^s_{L_2}$ as $\rho_{\phi^u_{\mathrm{dpMMD}}, L_2}$, which is similarly defined as (15). Our next concern is characterizing $\rho_{\phi^u_{\mathrm{dpMMD}}, L_2}$ and comparing it with the minimum separation of the dpMMD test established in Theorem 9.

Theorem 11 (Minimum Separation of ϕ^u_{dpMMD} over $\mathcal{P}^s_{L_2}$). *Assume that $n \leq m \leq \tau n$ for some fixed constant $\tau \geq 1$, and that $\alpha \in (0, e^{-1})$, $\beta \in (0,1)$, $\varepsilon > 0$, $\delta \in [0,1)$, $B \geq 16\alpha^{-2}\log(8/\beta)$ and $\prod_{i=1}^d \lambda_i \leq 1$. The minimum separation of ϕ^u_{dpMMD} with the Gaussian kernel over $\mathcal{P}^s_{L_2}$ is upper bounded as*

$$\rho^2_{\phi^u_{\text{dpMMD}}, L_2} \leq C_{\tau,\beta,s,R,M,d}\left\{\sum_{i=1}^d \lambda_i^{2s} + \frac{\log(1/\alpha)}{n\sqrt{\lambda_1 \cdots \lambda_d}} + \frac{\log(1/\alpha)}{n\lambda_1 \cdots \lambda_d \xi_{\varepsilon,\delta}}\right\},$$

where $C_{\tau,\beta,s,R,M,d}$ is a positive constant, depending only on τ, β, s, R, M, d, and $\xi_{\varepsilon,\delta}$ is as in (1).

The proof of Theorem 11 is given in Appendix E.14. To simplify our discussion, assume that α is a fixed constant. In this case, by comparing Theorem 11 with Theorem 9, the upper bound for $\rho^2_{\phi^u_{\text{dpMMD}}, L_2}$ is smaller than that for $\rho^2_{\phi_{\text{dpMMD}}, L_2}$, up to a constant, only when $n\xi_{\varepsilon,\delta} \lesssim 1$. Since $\prod_{i=1}^d \lambda_i \leq 1$ and α is fixed, the condition $n\xi_{\varepsilon,\delta} \lesssim 1$ essentially means that $\|p - q\|_{L_2}$ needs to be sufficiently larger than a specific constant for significant power. However, this condition may be infeasible as we assume that $\|p\|_{L_\infty}$ and $\|q\|_{L_\infty}$ are bounded by M. In fact, our earlier result in Theorem 5 suggests that the test is not even consistent in a pointwise sense when $n\xi_{\varepsilon,\delta} \lesssim 1$. Therefore, except for this boundary case, it is more beneficial to use ϕ_{dpMMD} than ϕ^u_{dpMMD} to achieve tighter separation rates over $\mathcal{P}^s_{L_2}$ in high privacy regimes.

As we mentioned before, it remains an open question whether there exist alternative privacy mechanisms that could potentially yield an optimal test based on U_{MMD} in high privacy regimes. While we still advocate using the plug-in MMD estimate over U_{MMD} due to its smaller sensitivity, it would be interesting to explore this question in future work.

6 Simulations

In this section, we compare the empirical power of dpMMD against other private two-sample tests, including the naive dpMMD introduced in Equation (6) and the U-statistic dpMMD studied in Section 5.3. We also implement two generic methods for privatizing the permutation MMD test, which we refer to as TOT MMD (Kazan et al., 2023) and SARRM MMD (Peña and Barrientos, 2022). Finally, we also compare against the differentially private kernel two-sample test, TCS-ME, proposed by Raj et al. (2020). A brief overview of these alternative methods is provided in Section 6.1, with detailed information available in Appendix D.

We empirically study the power attained by dpMMD on synthetic data sampled from perturbed uniform distributions in Section 6.2, and on real-world high-dimensional CelebA image data in Section 6.3. The latter scenario involves sensitive information on human faces, justifying the incorporation of differential privacy in the analysis. In both simulation settings, we observe consistent patterns in the power behavior, and a detailed discussion on the results is presented in Section 6.4. For all our experiments, we use a Gaussian kernel, $B = 2000$ permutations, and report power results averaged over 200 repetitions.

Due to space constraints, we defer the dpHSIC simulations to Appendix C, along with test-level analysis and additional low-privacy experiments. The code to run our tests and to reproduce the experiments is available at `https://github.com/antoninschrab/dpkernel`.

6.1 Alternative differentially private tests

We provide a brief introduction to the alternative differentially private tests, namely TCS-ME, TOT, and SARRM, with detailed implementation information available in Appendix D.

TOT (Kazan et al., 2023). TOT is constructed based on the subsample-and-aggregate idea outlined in Canonne et al. (2019). It is guaranteed to be differentially private and to correctly control the probability of type I error for any sample size, any number of partitioned subsets, and any sub-test significance level. However, for non-parametric testing, there is no principled way to choose the last two parameters and one has to rely on heuristics in practice. This heuristic aspect presents a notable disadvantage of TOT, being highly sensitive to the choice of these parameters.

SARRM (Peña and Barrientos, 2022). As another method based on the subsample-and-aggregate idea, SARRM also depends on the number of partitioned subsets and on the sub-test significance level. Peña and Barrientos (2022) propose a method to select these parameters, which can be implemented for MMD/HSIC tests. However, the differential privacy constraint and type I error control for SARRM are only guaranteed for sufficiently large sample sizes (determined by a minimum number of partitioned subsets to use), which means that SARRM simply cannot be run in some settings depending on the values of ε, α, and n.

TCS-ME (Raj et al., 2020). The TCS-ME test is a privatized version of the ME test (Jitkrittum et al., 2016) that utilizes kernel mean embeddings. This method builds on a Hotelling-type test statistic privatized by the Gaussian mechanism, and requires a careful choice of test locations. The resulting test is (ε, δ)-DP for $\delta > 0$ (run with $\delta = 10^{-5}$). A major limitation of TCS-ME is its potential for significant miscalibration, particularly in high-dimensional settings. See our discussion on type I error control in Section 6.5.

6.2 Perturbed Uniform Distributions

As recalled in Section 5.2, the minimax L_2 separation rate over the Sobolev ball $\mathcal{S}_d^s(R)$ is $n^{-2s/(4s+d)}$ in the non-privacy regime. A lower bound for this minimax separation rate is derived by constructing two densities whose difference lies in the Sobolev ball with a small L_2 norm. As explained by Schrab et al. (2023, Appendix D), the uniform and perturbed uniform densities meet the requirements in the lower bound construction, and we adopt this setting in our two-sample experiments. In more detail, we compare the uniform distribution with its perturbed counterpart, varying the amplitudes.

Specifically, the considered uniform distribution on $[0,1]^d$ has density $\mathbb{1}(x \in [0,1]^d)$ for $x \in \mathbb{R}^d$, while the perturbed uniform density on $[0,1]^d$ with a perturbation amplitude $a \in [0,1]$ is

$$\mathbb{1}\big(x \in [0,1]^d\big) + a \prod_{i=1}^{d} P(x_i), \quad \text{for } x = (x_1, \ldots, x_d)^\top \in \mathbb{R}^d, \tag{18}$$

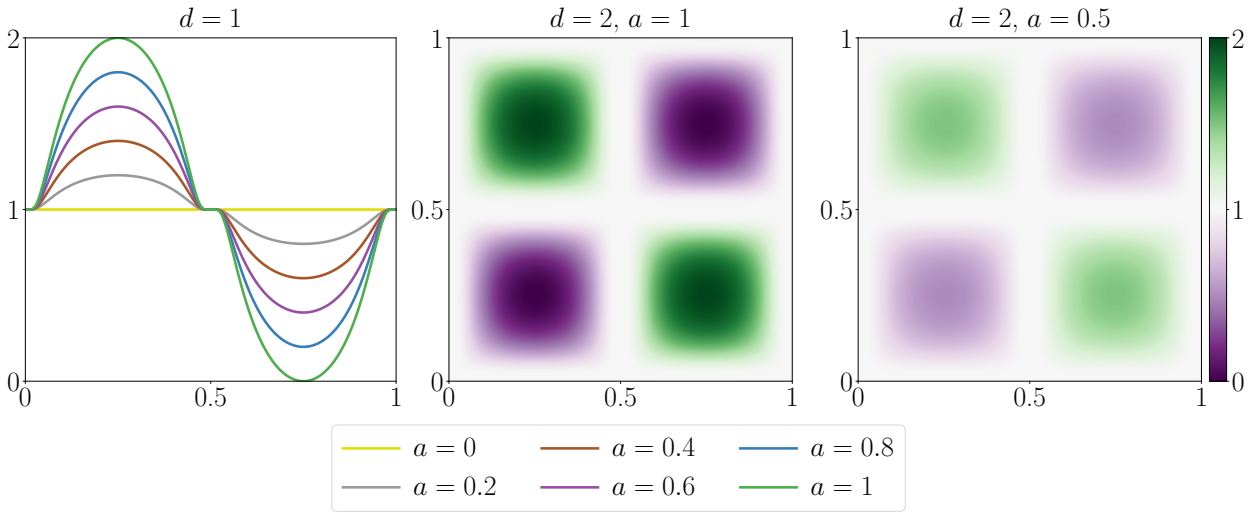

Figure 1: Perturbed uniform d-dimensional densities on $[0,1]^d$ with varying perturbation amplitude a.

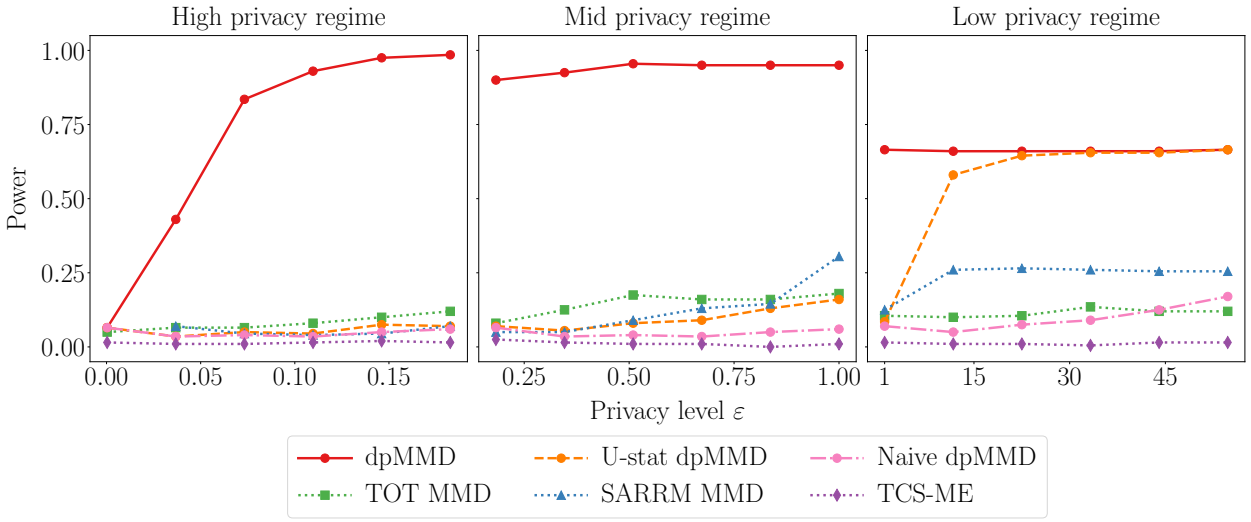

Figure 2: Comparing uniform vs. perturbed uniform while varying the privacy level ε. We set the sample sizes $m = n = 3000$ and dimension $d = 1$, and change the privacy level ε and perturbation amplitude a as follows: *(Left)* Privacy level ε from $1/n$ to $10/\sqrt{n}$, perturbation amplitude $a = 0.2$. *(Middle)* Privacy level ε from $10/\sqrt{n}$ to 1, perturbation amplitude $a = 0.15$. *(Right)* Privacy level ε from 1 to $\sqrt{n}$, perturbation amplitude $a = 0.1$.

where the one-dimensional perturbation is defined as

$$P(x_i) \coloneqq \exp\left(1 - \frac{1}{1-(4x_i-1)^2}\right)\mathbb{1}\big(x_i \in (0,1/2)\big) - \exp\left(1 - \frac{1}{1-(4x_i-3)^2}\right)\mathbb{1}\big(x_i \in (1/2,1)\big).$$

This definition matches the one of Schrab et al. (2023, Equation 17) with only one (scaled) perturbation per dimension. The one-dimensional and two-dimensional perturbed densities with various perturbation amplitudes are visualized in Figure 1.

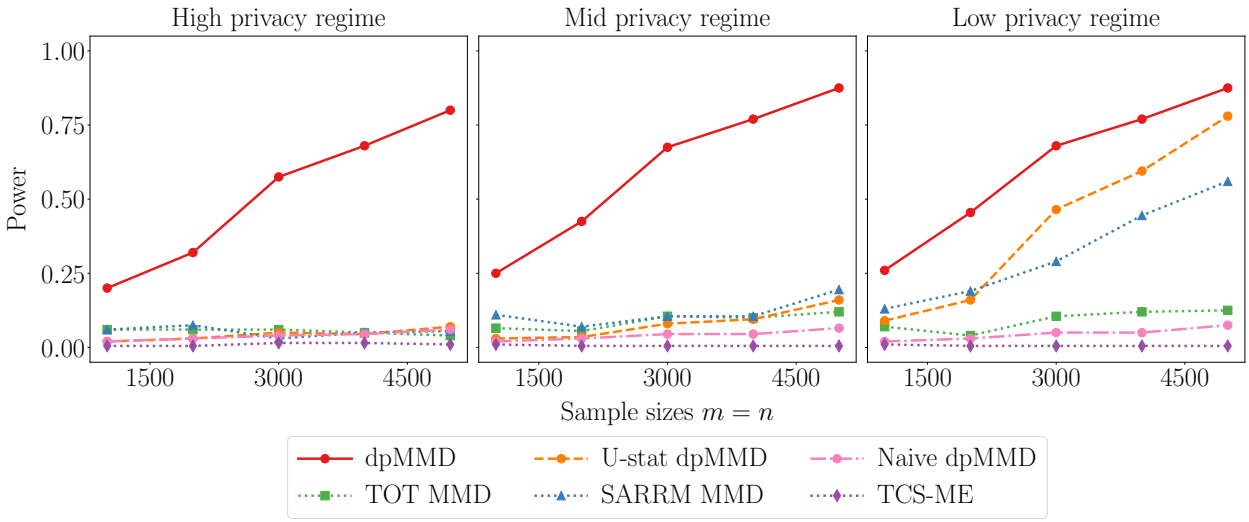

Figure 3: Comparing uniform vs. perturbed uniform while varying the sample sizes $m = n$. We set the dimension $d = 1$ and perturbation amplitude $a = 0.1$. We change the privacy level as follows: *(Left)* Privacy level $\varepsilon = 10/\sqrt{n}$. *(Middle)* Privacy level $\varepsilon = 1$. *(Right)* Privacy level $\varepsilon = \sqrt{n}/10$.

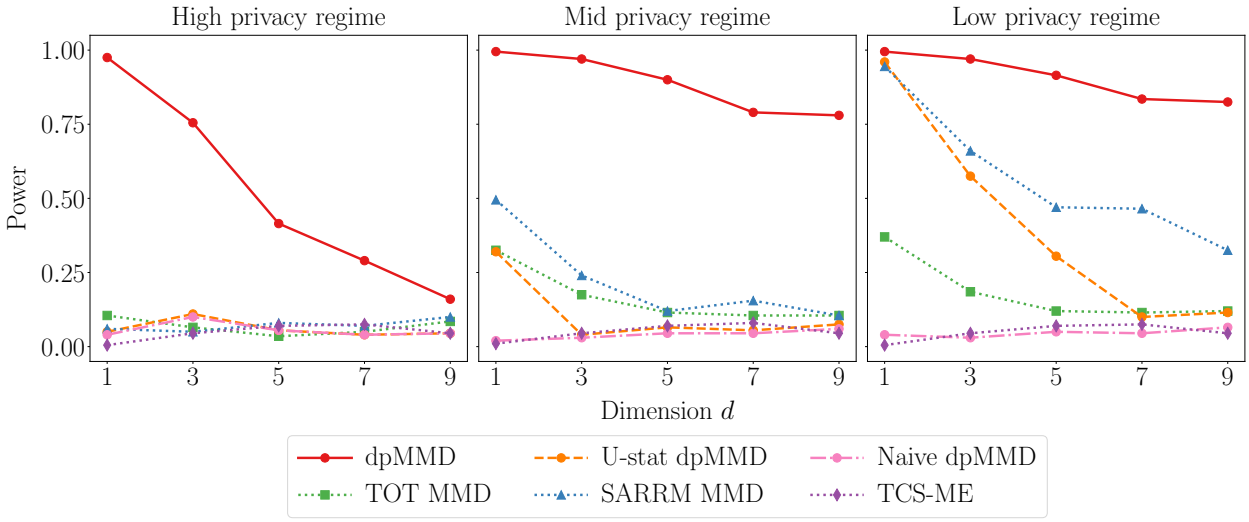

Figure 4: Comparing uniform vs. perturbed uniform while varying the dimension d. We set the sample sizes $m = n = 3000$ and perturbation amplitude $a = 0.2$. We change the privacy level as follows: *(Left)* Privacy level $\varepsilon = 10/\sqrt{n}$. *(Middle)* Privacy level $\varepsilon = 1$. *(Right)* Privacy level $\varepsilon = \sqrt{n}/10$.

We run our perturbed uniform experiments under three different settings where we vary the privacy level ε (Figure 2), the sample sizes $m = n$ (Figure 3), and the dimension d (Figure 4). Additionally, we provide experiments with 'strong-signal' alternatives in the low privacy regime in Figure 12 in Appendix C.2, as well as a level analysis in Figure 14 in Appendix C.3. We discuss all experimental results of Figures 2 to 4 in Section 6.4.

Figure 5: Selected CelebA images in dimension $3 \times 178 \times 218$.

6.3 CelebA

As some potential real-world applications of differentially private two-sample tests, we consider CelebA face images which in practice would be highly confidential, and hence the use of DP tests is thoroughly justified. The CelebA dataset (Liu et al., 2015) consists of $202,599$ face images of $10,177$ identities with a large diversity of face attributes, poses and backgrounds. For illustration purposes, we display a selection of CelebA images in Figure 5. It is worth highlighting that we run our tests on the original full-resolution images ($3 \times 178 \times 218$) without any modifications.

In our experiments, one sample consists of uniformly-sampled face images of women, while the other is 'corrupted' with corruption parameter $c \in [0, 1]$ in the following sense: we uniformly sample face images of women with probability $1 - c$, and of men with probability c.

We run several CelebA experiments while varying the privacy level ε (Figure 6), the sample sizes $m = n$ (Figure 7), and the corruption c (Figure 8). As in Section 6.2, we consider the high/mid/low privacy regimes for each of these. We also verify that all tests are well-calibrated in Figure 14 in Appendix C.3. TCS-ME is excluded from our power analysis on the CelebA data as we empirically observed that this method is not well-calibrated for this dataset.

Our results consistently demonstrate that kernel tests using a simple Gaussian kernel are able to capture complex image distribution shifts (see Figure 5) even in this extremely high dimensional setting with $d = 3 \times 178 \times 218 = 116,412$. This result is surprising given that the Gaussian kernel simply compares the distance between pixels at the same location without using information about the image structure. We discuss this aspect more in Section 6.4.

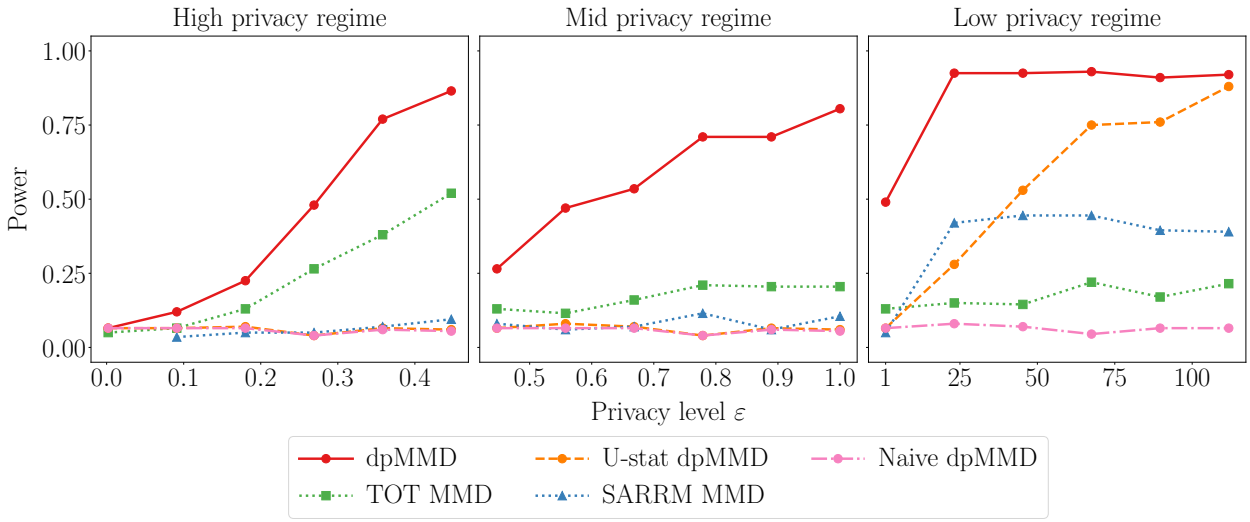

Figure 6: Comparing CelebA women/men images while varying the privacy level ε. We set the sample sizes $m = n = 500$, and change the parameters as follows: *(Left)* Privacy level ε from $1/n$ to $10/\sqrt{n}$, corruption $c = 1$. *(Middle)* Privacy level ε from $10/\sqrt{n}$ to 1, corruption $c = 0.6$. *(Right)* Privacy level ε from 1 to $\sqrt{n}$, corruption $c = 0.5$.

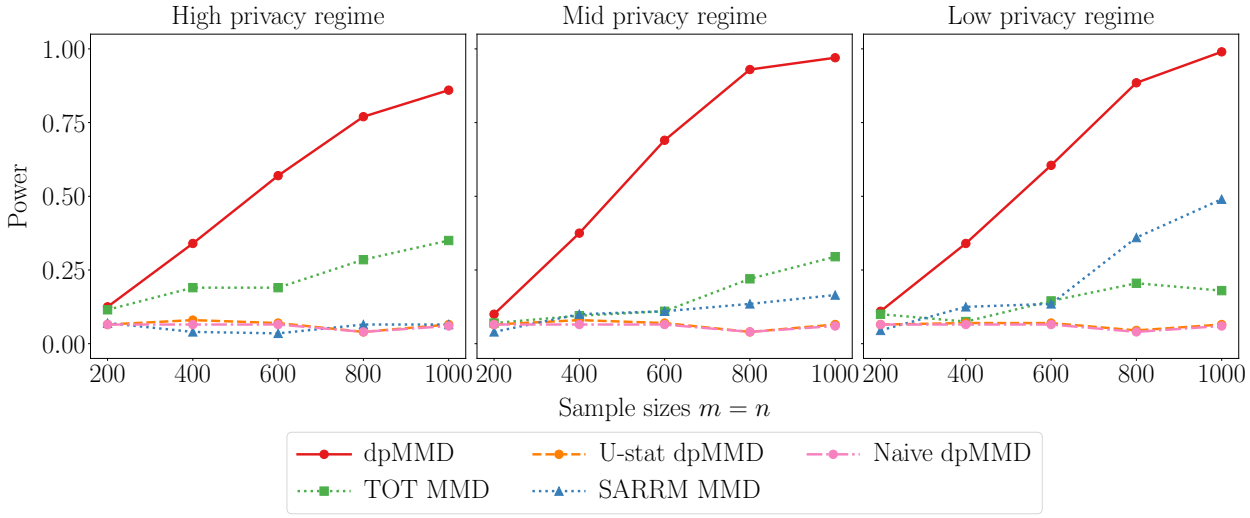

Figure 7: Comparing CelebA women/men images while varying the sample sizes $m = n$. We set the other parameters as follows: *(Left)* Privacy level $\varepsilon = 10/\sqrt{n}$, corruption $c = 0.7$. *(Middle)* Privacy level $\varepsilon = 1$, corruption $c = 0.5$. *(Right)* Privacy level $\varepsilon = \sqrt{n}/10$, corruption $c = 0.4$.

6.4 Analysis of Main Experimental Results

We now analyze the results of the perturbed uniform and CelebA experiments presented in Sections 6.2 and 6.3, respectively.

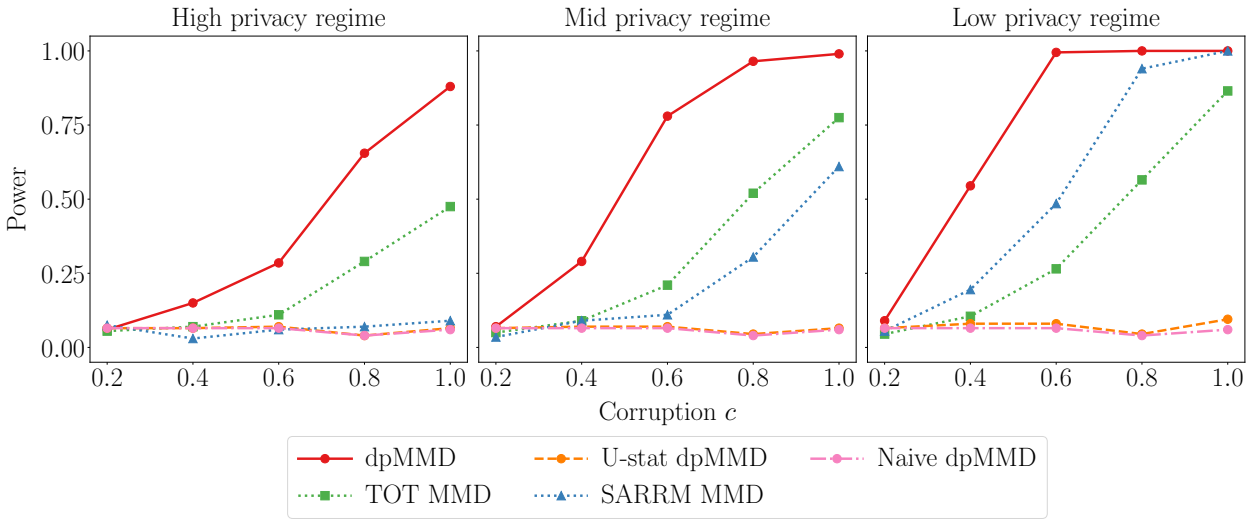

Figure 8: Comparing CelebA women/men images while varying the corruption parameter c. We set the sample sizes $m = n = 500$, and privacy parameter ε as follows: *(Left)* Privacy level $\varepsilon = 10/\sqrt{n}$. *(Middle)* Privacy level $\varepsilon = 1$. *(Right)* Privacy level $\varepsilon = \sqrt{n}/10$.

Overview. First and foremost, we observe that dpMMD achieves significantly higher power than all other tests across all privacy regimes. This trend remains consistent when varying the other parameters {privacy, sample sizes, dimension, corruption}. In the high and mid privacy regimes illustrated in Figures 2 to 4, only dpMMD is able to detect the perturbation on the uniform distribution. In the high and mid privacy regimes presented in Figures 6 to 8, dpMMD clearly outperforms all other tests but TOT MMD and SARRM MMD eventually manage to detect the CelebA image distributional shift. In the low privacy regime, dpMMD also achieves the highest power, which is eventually matched by U-stat dpMMD as the privacy parameter increases.

Varying privacy and free privacy. As can be seen in Figures 2 and 6, increasing ε (*i.e.*, lowering privacy) first leads to an increase in power. However, we observe that, after some threshold, further increasing ε does not increase the power of dpMMD. Essentially, in this low privacy regime, dpMMD already attains the power of the non-private MMD test and the (low) privacy guarantee then comes for free. This empirical observation is also supported by our theoretical findings that the non-private MMD and L_2 optimal uniform separation rates are attained by dpMMD in the low privacy regime (see Theorems 7 and 9).

Varying privacy for U-stat dpMMD. Furthermore, Figures 2 and 6 shows that the dpMMD test using the U-statistic is powerless in the high and mid privacy regimes, but it eventually reaches the power of dpMMD (using the V-statistic) in the low privacy regime, which is theoretically justified by our suboptimality result (Theorem 10) for dpMMD based on the U-statistic and by the fact that the non-private MMD U-statistic test is minimax rate optimal (Schrab et al., 2023).

Varying the sample size. When varying the sample sizes $m = n$ in Figures 3 and 7 with fixed high/mid/low privacy level $\varepsilon \in \{10/\sqrt{n}, 1, \sqrt{n}/10\}$, the power of all tests naturally increase, while the power of dpMMD increases faster than the others. In the low privacy regime of Figure 3, we see that the power of U-stat dpMMD approaches that of dpMMD as the same size increases. This can be explained by the aforementioned reasoning along with the observation that the privacy level $\varepsilon = \sqrt{n}/10$ also increases (*i.e.*, lower privacy) in this setting.

Varying the problem difficulty. In Figures 4 and 8, the sample sizes and privacy levels are fixed while the difficulty of the problem is varied. For the perturbed uniform experiment, as the dimension of the problem increases, the perturbation becomes more difficult to detect and hence the power decreases for each test. We observe nonetheless that the power of dpMMD deteriorates at a much slower rate than the one of the other tests. For the CelebA experiment, the test power increases with the corruption parameter of the image sampler, with dpMMD always achieving the highest power, followed by either TOT MMD or SARRM MMD.

The power of kernel methods. We end this discussion with some remarks regarding the CelebA experiments of Section 6.3. First, we emphasize again that the quadratic-time kernel tests run swiftly on the full-resolution CelebA image data, which has $116,412$ pixels per image. Second, given the large diversity of faces, poses and backgrounds (see Figure 5), it is remarkable that dpMMD is able to detect such complicated differences in high-dimensional image distributions while using an off-the-shelf Gaussian kernel which entirely ignores the image structure and simply averages distances between pixel values at the same locations. Third, the fact that such complex testing problems can now be solved while guaranteeing differential privacy is extremely important, especially when dealing with data as personal and sensitive as facial data. The DP constraint essentially guarantees privacy in the sense that confidential information about a single face image cannot be recovered. Fourth, we stress that the power results reported are meaningful and that dpMMD truly detects the difference between CelebA images of women and men, which is justified as type I error control is correctly retained (Appendix C.3). We believe these CelebA experiments strongly advocate the use of tests leveraging kernel methods for differential privacy.

6.5 Analysis of Additional Experimental Results

Before concluding the paper, we briefly summarize the results of the additional experiments in Appendix C.

Independence testing. In the HSIC testing experiments of Appendix C.1, we consider the problem of detecting the dependence of variables with the perturbed uniform joint density introduced in Section 6.2, and hence with uniform marginal densities. This setting is of particular interest as this corresponds to the joint density used in deriving the non-private L_2 minimax independence lower bound over Sobolev balls by Albert et al. (2022). The exact same power dynamics and aforementioned observations, which hold for the MMD-based tests, also apply to the HSIC variants of all tests, and indeed dpHSIC achieves substantially higher power than all the other tests.

High-signal & low-privacy alternatives. We remark that Naive dpMMD and TCS-ME have almost no power against the alternatives considered in the previous subsections. This is not because these tests are faulty but because the signal is too weak to be detected by these tests. In fact, the Naive dpMMD test with privacy ε is exactly equivalent to the dpMMD test with privacy $2/(B\varepsilon)$ (recall $B = 2000$ permutations), which justifies the poor performance of Naive dpMMD. For a sanity check, we consider 'high-signal' & 'low-privacy' alternatives in Appendix C.2 and show that dpMMD and TCS-ME are indeed able to detect the difference when the signal is large enough. The results can be found in Figure 12.

Type I error control. In Appendix C.3, we run experiments under the null hypothesis. This corresponds to no perturbation for the perturbed uniform two-sample and independence settings (amplitude $a = 0$), and to no corruption in the CelebA sampler (corruption $c = 0$). All tests are well-calibrated with empirical level around α under all settings considered, except TCS-ME. Indeed, we observe in Figure 14 that when testing two samples from a 100-dimensional uniform distribution, TCS-ME fails to control the type I error rate, which is estimated to be around $0.5 = 10\alpha$ instead of around $\alpha = 0.05$. This depicts a major limitation of TCS-ME especially in high-dimensional settings. However, we point out that this test is well-calibrated in the settings of Section 6.2, ensuring a fair comparison of power therein.

7 Discussion

In this work, we have proposed differentially private permutation tests and examined their theoretical and empirical performance. The prior work on differentially private testing has often been limited in its practical applicability, being restricted to discrete data or relying on asymptotic theory that does not offer confidence in finite sample scenarios. Our permutation framework addresses these challenges by introducing practical tools, which are applicable to diverse settings with finite sample guarantees for both type I error control and differential privacy. In addition to general power properties, we have provided a detailed power analysis of the proposed method in the context of kernel testing, and showed that the proposed private kernel tests achieve minimax optimal power in terms of kernel metrics in all privacy regimes. We have also analyzed the testing power against nonparametric L_2 alternatives, and established minimum separation rates in all privacy regimes. Finally, we have conducted an extensive simulation study to validate our theoretical findings as well as to highlight the practical value of our approach.

Our work raises several intriguing open questions that deserve further investigation, as outlined below.

- **Beyond Global DP.** Our work focuses on global (ε, δ)-differential privacy along with the Laplace mechanism. While global DP is a widely used as an effective concept of data protection, other privacy concepts may be more suitable depending on the context and requirements. Exploring the development of private permutation tests applicable to other privacy concepts or based on other privacy mechanisms would be an interesting direction for future investigation.

- **Other Applications.** We illustrated the proposed method in the context of two-sample and independence testing, with a focus on kernel-based tests. The permutation method has been employed successfully in other statistical problems such as testing for regression coefficients (DiCiccio and Romano, 2017) and conditional independence testing (Kim et al., 2022b). It is therefore compelling to broaden the application of our framework by tackling other statistical problems in privacy settings. Future work can also focus on conducting a detailed analysis of the private permutation test using non-kernel test statistics.

- **Variants of dpMMD and dpHSIC.** In our analysis, we utilized the plug-in estimators of MMD and HSIC based on single kernels. In recent years, significant progress has been made to reduce the computational complexity (Schrab et al., 2022; Domingo-Enrich et al., 2023) as well as to avoid the bandwidth selection issue (Schrab et al., 2023; Biggs et al., 2023) of this standard approach. Considering these developments, a promising avenue for future research would be to extend these recent advances to privacy-preserving settings.

- **Minimax Separation under DP.** Our results in Section 5.2 provide upper bounds for the minimax separation rates in terms of the L_2 metric, which match the lower bound in low privacy regimes. However, as mentioned earlier, it is unknown whether they are still tight in mid/high privacy regimes. Consequently, future work could focus on addressing this question by establishing matching lower bounds or sharper upper bounds. More broadly, it would be interesting to establish minimax separation rates under DP in terms of other metrics as well. We leave these important questions to future work.

References

Acharya, J., Canonne, C., Freitag, C., and Tyagi, H. (2019). Test without trust: Optimal locally private distribution testing. In *The 22nd International Conference on Artificial Intelligence and Statistics*.

Acharya, J., Sun, Z., and Zhang, H. (2018). Differentially private testing of identity and closeness of discrete distributions. *Advances in Neural Information Processing Systems*, 31.

Acharya, J., Sun, Z., and Zhang, H. (2021). Differentially Private Assouad, Fano, and Le Cam. In *Algorithmic Learning Theory*, pages 48–78. PMLR.

Alabi, D. and Vadhan, S. (2022). Hypothesis testing for differentially private linear regression. *Advances in Neural Information Processing Systems*, 35:14196–14209.

Albert, M., Laurent, B., Marrel, A., and Meynaoui, A. (2022). Adaptive test of independence based on HSIC measures. *The Annals of Statistics*, 50(2):858–879.

Aliakbarpour, M., Diakonikolas, I., Kane, D., and Rubinfeld, R. (2019). Private Testing of Distributions via Sample Permutations. *Advances in Neural Information Processing Systems*, 32.

Aliakbarpour, M., Diakonikolas, I., and Rubinfeld, R. (2018). Differentially Private Identity and Closeness Testing of Discrete Distributions. *In International Conference on Machine Learning*, pages 169–178.

Apple (2017). Learning with Privacy at Scale. https://machinelearning.apple.com/research/learning-with-privacy-at-scale.

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

Awan, J. and Slavković, A. (2018). Differentially Private Uniformly Most Powerful Tests for Binomial Data. *Advances in Neural Information Processing Systems*, 31.

Balle, B. and Wang, Y. (2018). Improving the gaussian mechanism for differential privacy: Analytical calibration and optimal denoising. In Dy, J. G. and Krause, A., editors, *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pages 403–412. PMLR.

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

Bartlett, P. L. and Mendelson, S. (2002). Rademacher and Gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research*, 3(Nov):463–482.

Berrett, T. and Butucea, C. (2020). Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms. *Advances in Neural Information Processing Systems*, 33:3164–3173.

Berrett, T. B., Kontoyiannis, I., and Samworth, R. J. (2021). Optimal rates for independence testing via U-statistic permutation tests. *The Annals of Statistics*, 49(5):2457–2490.

Biggs, F., Schrab, A., and Gretton, A. (2023). MMD-FUSE: Learning and Combining Kernels for Two-Sample Testing Without Data Splitting. *arXiv preprint arXiv:2306.08777*.

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press.

Bun, M. and Steinke, T. (2016). Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography Conference*, pages 635–658. Springer.

Cai, B., Daskalakis, C., and Kamath, G. (2017). Priv'it: Private and sample efficient identity testing. In *International Conference on Machine Learning*, pages 635–644. PMLR.

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

Campbell, Z., Bray, A., Ritz, A., and Groce, A. (2018). Differentially private ANOVA testing. In *2018 1st International Conference on Data Intelligence and Security (ICDIS)*, pages 281–285. IEEE.

Canonne, C. L. (2022). Topics and Techniques in Distribution Testing: A Biased but Representative Sample. *Foundations and Trends® in Communications and Information Theory*, 19(6):1032–1198.

Canonne, C. L., Kamath, G., McMillan, A., Smith, A., and Ullman, J. (2019). The structure of optimal private tests for simple hypotheses. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 310–321.

Canonne, C. L., Kamath, G., McMillan, A., Ullman, J., and Zakynthinou, L. (2020). Private identity testing for high-dimensional distributions. *Advances in Neural Information Processing Systems*, 33:10099–10111.

Couch, S., Kazan, Z., Shi, K., Bray, A., and Groce, A. (2019). Differentially private nonparametric hypothesis testing. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, pages 737–751.

Diakonikolas, I., Gouleakis, T., Kane, D. M., Peebles, J., and Price, E. (2021). Optimal testing of discrete distributions with high probability. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 542–555.

Diakonikolas, I., Hardt, M., and Schmidt, L. (2015). Differentially private learning of structured discrete distributions. *Advances in Neural Information Processing Systems*, 28.

Diakonikolas, I. and Kane, D. M. (2016). A new approach for testing properties of discrete distributions. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 685–694. IEEE.

DiCiccio, C. J. and Romano, J. P. (2017). Robust permutation tests for correlation and regression coefficients. *Journal of the American Statistical Association*, 112(519):1211–1220.

Ding, B., Kulkarni, J., and Yekhanin, S. (2017). Collecting Telemetry Data Privately. *Advances in Neural Information Processing Systems*, 30.

Dobriban, E. (2022). Consistency of invariance-based randomization tests. *The Annals of Statistics*, 50(4):2443–2466.

Domingo-Enrich, C., Dwivedi, R., and Mackey, L. (2023). Compress then test: Powerful kernel testing in near-linear time. In *International Conference on Artificial Intelligence and Statistics*.

Dong, J., Roth, A., and Su, W. J. (2022). Gaussian differential privacy. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(1):3–37.

Dubois, A., Berrett, T. B., and Butucea, C. (2023). Goodness-of-Fit Testing for Hölder Continuous Densities Under Local Differential Privacy. In *Foundations of Modern Statistics*, pages 53–119. Springer International Publishing.

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

Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., and Naor, M. (2006a). Our data, ourselves: Privacy via distributed noise generation. In *Advances in Cryptology-EUROCRYPT 2006: 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, St. Petersburg, Russia, May 28-June 1, 2006. Proceedings 25*, pages 486–503. Springer.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006b). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography Conference*, pages 265–284. Springer.

Dwork, C., Roth, A., et al. (2014). The Algorithmic Foundations of Differential Privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407.

Erlingsson, Ú., Pihur, V., and Korolova, A. (2014). Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pages 1054–1067.

Fernández, T. and Rivera, N. (2022). A general framework for the analysis of kernel-based tests. *arXiv preprint arXiv:2209.00124*.

Fienberg, S. E., Slavkovic, A., and Uhler, C. (2011). Privacy preserving GWAS data sharing. In *2011 IEEE 11th International Conference on Data Mining Workshops*, pages 628–635. IEEE.

Friedberg, R. and Rogers, R. (2023). Privacy aware experimentation over sensitive groups: A general chi square approach. In *Workshop on Algorithmic Fairness through the Lens of Causality and Privacy*, pages 23–66. PMLR.

Gaboardi, M., Lim, H., Rogers, R., and Vadhan, S. (2016). Differentially private chi-squared hypothesis testing: Goodness of fit and independence testing. In *International Conference on Machine Learning*, pages 2111–2120. PMLR.

Gaboardi, M. and Rogers, R. (2018). Local private hypothesis testing: Chi-square tests. In *International Conference on Machine Learning*, pages 1626–1635. PMLR.

Gretton, A. (2015). A simpler condition for consistency of a kernel independence test. *arXiv preprint arXiv:1501.06103*.

Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. (2012). A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773.

Gretton, A., Bousquet, O., Smola, A., and Schölkopf, B. (2005). Measuring statistical dependence with Hilbert-Schmidt norms. In *International Conference on Algorithmic Learning Theory*, pages 63–77. Springer.

Holohan, N., Leith, D. J., and Mason, O. (2015). Differential Privacy in Metric Spaces: Numerical, Categorical and Functional Data Under the One Roof. *Information Sciences*, 305:256–268.

Ingster, Y., Ingster, J. I., and Suslina, I. (2003). *Nonparametric Goodness-of-Fit Testing under Gaussian Models*. Springer.

Ingster, Y. I. (1994). Minimax detection of a signal in ℓ_p metrics. *Journal of Mathematical Sciences*, 68(4):503–515.

Ingster, Y. I. (2000). Adaptive chi-square tests. *Journal of Mathematical Sciences*, 99(2):1110–1119.

Jitkrittum, W., Szabó, Z., Chwialkowski, K. P., and Gretton, A. (2016). Interpretable distribution features with maximum testing power. In Lee, D. D., Sugiyama, M., von Luxburg, U., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 181–189.

Kakizaki, K., Fukuchi, K., and Sakuma, J. (2017). Differentially Private Chi-squared Test by Unit Circle Mechanism. In *International Conference on Machine Learning*, pages 1761–1770. PMLR.

Kalemaj, I., Kasiviswanathan, S. P., and Ramdas, A. (2023). Differentially private conditional independence testing. *arXiv preprint arXiv:2306.06721*.

Kamath, G., Li, J., Singhal, V., and Ullman, J. (2019). Privately learning high-dimensional distributions. In *Conference on Learning Theory*, pages 1853–1902. PMLR.

Kamath, G., Singhal, V., and Ullman, J. (2020). Private mean estimation of heavy-tailed distributions. In *Conference on Learning Theory*, pages 2204–2235. PMLR.

Kamath, G. and Ullman, J. (2020). A primer on private statistics. *arXiv preprint arXiv:2005.00010*.

Kazan, Z., Shi, K., Groce, A., and Bray, A. (2023). The Test of Tests: A Framework For Differentially Private Hypothesis Testing. *arXiv preprint arXiv:2302.04260*.

Kim, I. (2021). Comparing a large number of multivariate distributions. *Bernoulli*, 27(1):419–441.

Kim, I., Balakrishnan, S., and Wasserman, L. (2022a). Minimax optimality of permutation tests. *The Annals of Statistics (arXiv:2003.13208)*, 50(1):225–251.

Kim, I., Neykov, M., Balakrishnan, S., and Wasserman, L. (2022b). Local permutation tests for conditional independence. *The Annals of Statistics*, 50(6):3388–3414.

Kim, I., Neykov, M., Balakrishnan, S., and Wasserman, L. (2023). Conditional Independence Testing for Discrete Distributions: Beyond χ^2-and G-tests. *arXiv preprint arXiv:2308.05373*.

Kusner, M. J., Sun, Y., Sridharan, K., and Weinberger, K. Q. (2016). Private causal inference. In *Artificial Intelligence and Statistics*, pages 1308–1317. PMLR.

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

Le Cam, L. (1973). Convergence of Estimates Under Dimensionality Restrictions. *The Annals of Statistics*, 1(1):38–53.

Le Cam, L. (2012). *Asymptotic Methods in Statistical Decision Theory*. Springer Science & Business Media.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*, volume 3. Springer.

Li, T. and Yuan, M. (2019). On the optimality of Gaussian kernel based nonparametric tests against smooth alternatives. *arXiv preprint arXiv:1909.03302*.

Liao, J., Sankar, L., Tan, V. Y., and du Pin Calmon, F. (2017). Hypothesis testing under mutual information privacy constraints in the high privacy regime. *IEEE Transactions on Information Forensics and Security*, 13(4):1058–1071.

Liu, Z., Luo, P., Wang, X., and Tang, X. (2015). Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*.

Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability*, 18:1269–1283.

Mironov, I. (2017). Rényi differential privacy. In *2017 IEEE 30th computer security foundations symposium (CSF)*, pages 263–275. IEEE.

Narayanan, S. (2022). Private high-dimensional hypothesis testing. In *Conference on Learning Theory*, pages 3979–4027. PMLR.

Nissim, K., Raskhodnikova, S., and Smith, A. (2007). Smooth Sensitivity and Sampling in Private Data Analysis. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 75–84.

Peña, V. and Barrientos, A. F. (2022). Differentially Private Hypothesis Testing with the Subsampled and Aggregated Randomized Response Mechanism. *arXiv preprint arXiv:2208.06803*.

Raj, A., Law, H. C. L., Sejdinovic, D., and Park, M. (2020). A differentially private kernel two-sample test. *Lecture Notes in Computer Science*, 11906.

Rindt, D., Sejdinovic, D., and Steinsaltz, D. (2021). Consistency of permutation tests of independence using distance covariance, HSIC and dHSIC. *Stat*, 10(1):e364.

Rogers, R. and Kifer, D. (2017). A new class of private chi-square hypothesis tests. In *Artificial Intelligence and Statistics*, pages 991–1000. PMLR.

Romano, J. P. and Wolf, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108.

Schrab, A., Kim, I., Albert, M., Laurent, B., Guedj, B., and Gretton, A. (2023). MMD Aggregated Two-Sample Test. *Journal of Machine Learning Research*, 24(194):1–81.

Schrab, A., Kim, I., Guedj, B., and Gretton, A. (2022). Efficient Aggregated Kernel Tests using Incomplete U-statistics. *Advances in Neural Information Processing Systems*, 35:18793–18807.

Sheffet, O. (2017). Differentially Private Ordinary Least Squares. In *International Conference on Machine Learning*, pages 3105–3114. PMLR.

Sheffet, O. (2018). Locally Private Hypothesis Testing. In *International Conference on Machine Learning*, pages 4605–4614. PMLR.

Shekhar, S., Kim, I., and Ramdas, A. (2022a). A permutation-free kernel independence test. *arXiv preprint arXiv:2212.09108*.

Shekhar, S., Kim, I., and Ramdas, A. (2022b). A permutation-free kernel two-sample test. *Advances in Neural Information Processing Systems*, 35:18168–18180.

Song, L., Smola, A., Gretton, A., Bedo, J., and Borgwardt, K. (2012). Feature Selection via Dependence Maximization. *Journal of Machine Learning Research*, 13(5).

Sriperumbudur, B. K., Fukumizu, K., Gretton, A., Schölkopf, B., and Lanckriet, G. R. G. (2012). On the empirical estimation of integral probability metrics. *Electronic Journal of Statistics*, 6(none).

Sriperumbudur, B. K., Fukumizu, K., and Lanckriet, G. R. (2011). Universality, Characteristic Kernels and RKHS Embedding of Measures. *Journal of Machine Learning Research*, 12(7):2389–2410.

Steinke, T. (2022). Composition of Differential Privacy & Privacy Amplification by Subsampling. *arXiv preprint arXiv:2210.00597*.

Swanberg, M., Globus-Harris, I., Griffith, I., Ritz, A., Groce, A., and Bray, A. (2019). Improved differentially private analysis of variance. *Proceedings on Privacy Enhancing Technologies*, 2019(3):310–330.

Task, C. and Clifton, C. (2016). Differentially private significance testing on paired-sample data. In *Proceedings of the 2016 SIAM International Conference on Data Mining*, pages 153–161. SIAM.

Tolstikhin, I., Sriperumbudur, B. K., and Muandet, K. (2017). Minimax estimation of kernel mean embeddings. *The Journal of Machine Learning Research*, 18(1):3002–3048.

Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer New York, NY.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer New York.

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

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press.

Wang, Y., Lee, J., and Kifer, D. (2015). Revisiting Differentially Private Hypothesis Tests for Categorical Data. *arXiv preprint arXiv:1511.03376*.

Yang, Y., Adamczewski, K., Sutherland, D. J., Li, X., and Park, M. (2023). Differentially Private Neural Tangent Kernels for Privacy-Preserving Data Generation. *arXiv preprint arXiv:2303.01687*.

Zhang, Q., Filippi, S., Gretton, A., and Sejdinovic, D. (2018). Large-scale kernel methods for independence testing. *Statistics and Computing*, 28(1):113–130.

A Overview of Appendices

This supplementary material includes additional results, technical details and proofs, omitted in the main text. The remaining material is organized as follows.

- In Appendix B, we present additional results including
  - (i) Limiting null distributions of privatized kernel statistics (Appendix B.1),
  - (ii) General consistency results (Appendix B.2),
  - (iii) Detailed explanation of Example 1 (Appendix B.3),
  - (iv) Detailed explanation of Example 2 (Appendix B.4),
  - (v) Minimax separation rate in the HSIC metric (Appendix B.5),
  - (vi) Minimum separation of the dpHSIC test in the L_2 distance (Appendix B.6),
  - (vii) Negative results of HSIC U-statistic (Appendix B.7) and
  - (viii) Analyses of minimum separation rates (Appendix B.8).
- In Appendix C, we present additional simulations including
  - (i) Independence testing with dpHSIC (Appendix C.1),
  - (ii) Additional experiments on the power in the low privacy regime (Appendix C.2) and
  - (iii) Experiments on type I error rates (Appendix C.3),
- In Appendix D, we explain in detail other private tests considered in our simulation studies.
- In Appendix E, we provide the proofs of the results presented in the main text.
- In Appendix F, we collect the proofs of the additional results provided in Appendix B.
- In Appendix G, we collect technical lemmas used in the main proofs.

Throughout these appendices, we use an additional set of notation described below.

Additional Notation. For a sequence of random variables X_n, we say that $X_n = O_P(1)$ if for any $\epsilon > 0$, there exists $M_\epsilon, N_\epsilon > 0$ such that $\mathbb{P}(|X_n| > M_\epsilon) < \epsilon$ for all $n \geq N_\epsilon$. Similarly, we say that $X_n = o_P(1)$ if for any $\epsilon > 0$, there exists $N_\epsilon > 0$ such that $\mathbb{P}(|X_n| > \epsilon) < \epsilon$ for all $n \geq N_\epsilon$. For a sequence of positive numbers a_n, $X_n = O_P(a_n)$ (resp. $o_P(a_n)$) means $a_n^{-1} X_n = O_P(1)$ (resp. $a_n^{-1} X_n = o_P(1)$). For $x \in \mathbb{R}$, $\lceil x \rceil$ denotes the least integer greater than or equal to x. We use the notation $X_n \xrightarrow{d} X$ to denote the convergence in distribution and $X_n \xrightarrow{p} X$ to denote the convergence in probability. Consider two probability distributions P and Q on a measurable space $(\Omega, \mathcal{F})$. The total variation (TV) distance between P and Q is defined as

$$d_{\mathrm{TV}}(P, Q) = \sup_{A \in \mathcal{F}} \big|P(A) - Q(A)\big| = \frac{1}{2}\|P - Q\|_1 = \frac{1}{2}\int \left|\frac{\mathrm{d}P}{\mathrm{d}\nu} - \frac{\mathrm{d}Q}{\mathrm{d}\nu}\right| \mathrm{d}\nu,$$

where ν is a common dominating measure of P and Q, and $\frac{\mathrm{d}P}{\mathrm{d}\nu}$, $\frac{\mathrm{d}Q}{\mathrm{d}\nu}$ are their density functions with respect to ν. The Kullback–Leibler (KL) divergence of P from Q is given as

$$d_{\mathrm{KL}}(P,Q) = \int \frac{\mathrm{d}P}{\mathrm{d}\nu} \log\left(\frac{\mathrm{d}P/\mathrm{d}\nu}{\mathrm{d}Q/\mathrm{d}\nu}\right)\mathrm{d}\nu.$$

We let $\mathbf{i}_p^q$ denote the set of all p-tuples drawn without replacement from $[q]$. The n-fold product distribution of a distribution P is denoted as $P^{\otimes n}$.

B Additional Results

In this section, we provide additional technical results omitted in the main text. The proofs for these additional results are relegated to Appendix F.

B.1 Limiting Null Distributions

The following proposition derives the limiting distributions of privatized kernel test statistics. In particular, denote

$$M_{\mathrm{MMD}} := \widehat{\mathrm{MMD}}(\mathcal{X}_{n+m}) + \frac{2\sqrt{2K}}{n\xi_{\varepsilon,\delta}}\zeta \quad \text{and} \quad M_{\mathrm{HSIC}} := \widehat{\mathrm{HSIC}}(\mathcal{X}_n) + \frac{8(n-1)\sqrt{KL}}{n^2\xi_{\varepsilon,\delta}}\zeta,$$

where $\zeta \sim \mathsf{Laplace}(0,1)$ independent of everything else. The results developed in Section 4 guarantee that M_{MMD} and M_{HSIC} are (ε,δ)-DP for bounded kernels. These two private statistics have the following limiting behavior under the null. The proof of Proposition 1 can be found in Appendix F.1.

Proposition 1 (Asymptotic Null Distributions)**.**

- *(MMD) Assume that the kernel k is bounded as $0 \le k(x,y) \le K$ for all $x, y \in \mathbb{S}$, and $\frac{m}{n+m} \to \omega \in (0,1)$. Write $\sigma = 2\sqrt{2K}n^{-1}\xi_{\varepsilon,\delta}^{-1}$. Then there exists a deterministic sequence $\{\lambda_i\}_{i=1}^{\infty}$ such that*

$$\begin{cases} (n+m)^{1/2} M_{\mathrm{MMD}} \xrightarrow{d} \sqrt{\sum_{i=1}^{\infty} \lambda_i Z_i^2} + \eta\zeta & \text{if } \sqrt{n+m}\sigma \to \eta \in [0,\infty), \\ \sigma^{-1} M_{\mathrm{MMD}} \xrightarrow{d} \zeta & \text{if } \sqrt{n+m}\sigma \to \infty, \end{cases}$$

  where $\{Z_i\}_{i=1}^{\infty} \overset{\text{i.i.d.}}{\sim} N(0,1)$ and $\zeta \sim \mathsf{Laplace}(0,1)$ are independent.

- *(HSIC) Assume that the kernels k and ℓ are bounded as $0 \le k(y,y') \le K$ and $0 \le \ell(z,z') \le L$ for all $y, y' \in \mathbb{Y}$ and $z, z' \in \mathbb{Z}$. Write $\sigma = 8(n-1)n^{-2}\sqrt{KL}\xi_{\varepsilon,\delta}^{-1}$. Then there exist deterministic sequences $\{\lambda_i\}_{i=1}^{\infty}$ and $\{\eta_i\}_{i=1}^{\infty}$ such that*

$$\begin{cases} n^{1/2} M_{\mathrm{HSIC}} \xrightarrow{d} \sqrt{\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \lambda_i \eta_j Z_{i,j}^2} + \eta\zeta & \text{if } \sqrt{n}\sigma \to \eta \in [0,\infty), \\ \sigma^{-1} M_{\mathrm{HSIC}} \xrightarrow{d} \zeta & \text{if } \sqrt{n}\sigma \to \infty, \end{cases}$$

  where $\{Z_{i,j}\}_{i,j=1}^{\infty} \overset{\text{i.i.d.}}{\sim} N(0,1)$ and $\zeta \sim \mathsf{Laplace}(0,1)$ are independent.

Remark. Note that the sequences $\{\lambda_i\}_{i=1}^{\infty}$ and $\{\eta_i\}_{i=1}^{\infty}$ are associated with the eigenvalues of integral kernel operators. See Gretton et al. (2012) and Zhang et al. (2018) for details.

B.2 General Pointwise Consistency

One of the desiderata of nonparametric tests is their pointwise consistency: the power converges to one as the sample size increases against any fixed alternative of interest. The following lemma develops a general pointwise consistency result that can be applied broadly to resampling-based tests (*e.g.*, bootstrap and permutation tests). We then leverage this general result to derive conditions for pointwise consistency of the differentially private permutation test in Theorem 3.

Lemma 8 (General Conditions for Consistency)**.** *Let $\alpha \in (0,1)$ be a fixed constant. Suppose that $\{W_{1,n}, \ldots, W_{B_n,n}\}$ are i.i.d. random variables conditional on a sigma field $\mathcal{G}$, and $W_{0,n}$ is constant conditional on the same sigma field $\mathcal{G}$. Suppose further that $\lim_{n\to\infty} \mathbb{P}(W_{0,n} \leq W_{1,n}) = 0$. Then for any positive sequence of B_n such that $\min_{n\geq 1} B_n + 1 > \alpha^{-1}$, we have*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{B_n+1} \left\{ \sum_{i=1}^{B_n} \mathbb{1}(W_{0,n} \leq W_{i,n}) + 1 \right\} \leq \alpha \right) = 1.$$

The proof of this result can be found in Appendix F.2. When B_n is a fixed quantity, Lemma 8 can be proved using a union bound. As mentioned in the main text, proving this consistency result for a general sequence of B_n is non-trivial and thereby we highlight it as our contribution.

To illustrate Lemma 8 in a simple setting, let us consider a wild bootstrap test for normal mean testing described as follows.

Example 3 (Wild Bootstrap)**.** Suppose that we are given $\mathcal{X}_n = \{X_1, \ldots, X_n\} \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$, and our interest is in testing whether $H_0 : \mu = 0$ or $H_1 : \mu > 0$. Consider a test statistic $W_{0,n} = \frac{1}{n}\sum_{i=1}^n X_i$, and let $\mathcal{G}$ be the sigma field generated by $\mathcal{X}_n$. Then conditional on $\mathcal{G}$, the test statistic $W_{0,n}$ is constant. Now generate i.i.d. Rademacher random variables $\epsilon_1, \ldots, \epsilon_n$, and define $W_{1,n} = \frac{1}{n}\sum_{i=1} \epsilon_i X_i$. The other statistics $W_{2,n}, \ldots, W_{B_n,n}$ are defined similarly using independent sets of i.i.d. Rademacher random variables. Then conditional on $\mathcal{G}$, the sequence $\{W_{1,n}, \ldots, W_{B_n,n}\}$ consists of i.i.d. random variables. We reject the null when

$$\frac{1}{B_n+1} \left\{ \sum_{i=1}^{B_n} \mathbb{1}(W_{0,n} \leq W_{i,n}) + 1 \right\} \leq \alpha$$

and the resulting test is called a (wild) bootstrap test. By the law of large numbers, it can be seen that $W_{0,n} \xrightarrow{p} \mu$ and $W_{1,n} \xrightarrow{p} 0$ for fixed mean $\mu > 0$ and finite variance σ^2, implying that $\lim_{n\to\infty} \mathbb{P}(W_{0,n} \leq W_{1,n}) = 0$. Hence the bootstrap test is consistent according to Lemma 8. The type I error is also controlled since $\{W_{0,n}, W_{1,n}, \ldots, W_{B_n,n}\}$ are exchangeable under the null.

B.3 Details on Example 1

Let us denote the pooled sample as $\mathcal{X}_{n+m} = \mathcal{Y}_n \cup \mathcal{Z}_m = \{X_1, \ldots, X_{n+m}\}$, and the permutation counterpart permuted according to $\boldsymbol{\pi}$ as $\mathcal{X}^{\boldsymbol{\pi}}_{n+m}$. Then the plug-in estimator using $\mathcal{X}^{\boldsymbol{\pi}}_{n+m}$ is given as

$$T_{\text{IPM}}(\mathcal{X}^{\boldsymbol{\pi}}_{n+m}) = \sup_{f\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^n f(X_{\pi_i}) - \frac{1}{m}\sum_{i=1}^m f(X_{\pi_{n+i}}) \right|.$$

By the reverse triangle inequality, it holds that

$$\sup_{\substack{\mathcal{X}_{n+m},\tilde{\mathcal{X}}_{n+m}:\\ d_{\mathrm{ham}}(\mathcal{X}_{n+m},\tilde{\mathcal{X}}_{n+m})\le 1}} \left|T_{\mathrm{IPM}}(\mathcal{X}^{\boldsymbol{\pi}}_{n+m}) - T_{\mathrm{IPM}}(\tilde{\mathcal{X}}^{\boldsymbol{\pi}}_{n+m})\right| \le \widetilde{\Delta}_T \coloneqq \frac{1}{\min\{n,m\}} \sup_{X,X'\in\mathbb{S}} \sup_{f\in\mathcal{F}} |f(X) - f(X')|.$$

The above bound is independent of $\boldsymbol{\pi}$ and thus the global sensitivity is at most $\widetilde{\Delta}_T$. Indeed, this upper bound is tight. For $X, X' \in \mathbb{S}$, set $\mathcal{X}_{n+m} = \{X, \ldots, X\}$ and $\tilde{\mathcal{X}}_{n+m} = \{X', X, \ldots, X\}$. Then the global sensitivity Δ_T is lower bounded as

$$\frac{1}{\min\{n,m\}} \sup_{f\in\mathcal{F}} |f(X) - f(X')| \le \Delta_T,$$

which holds for all $X, X' \in \mathbb{S}$. Hence it holds $\widetilde{\Delta}_T \le \Delta_T$, and the global sensitivity becomes $\Delta_T = \widetilde{\Delta}_T$.

B.4 Details on Example 2

In view of condition (8), there are four terms that we need to investigate, namely $\mathbb{E}[T(\mathcal{X}_{n+m})]$, $\mathbb{E}[T(\mathcal{X}^{\boldsymbol{\pi}}_{n+m})]$, $\mathrm{Var}[T(\mathcal{X}_{n+m})]$ and $\mathrm{Var}[T(\mathcal{X}^{\boldsymbol{\pi}}_{n+m})]$.

Starting with the expected value $\mathbb{E}[T(\mathcal{X}_{n+m})]$, the triangle inequality yields

$$\begin{aligned}\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^{n} f(Y_i) - \frac{1}{m}\sum_{i=1}^{m} f(Z_i)\right| \;\ge\; \mathrm{IPM}_{\mathcal{F}}(P,Q) &- \sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^{n} f(Y_i) - \mathbb{E}_P[f(Y)]\right| \\ &- \sup_{f\in\mathcal{F}} \left|\mathbb{E}_Q[f(Z)] - \frac{1}{m}\sum_{i=1}^{m} f(Z_i)\right|\end{aligned}$$

and thus

$$\begin{aligned}\mathbb{E}[T(\mathcal{X}_{n+m})] \;&\ge\; \mathrm{IPM}_{\mathcal{F}}(P,Q) - \mathbb{E}\left[\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^{n} f(Y_i) - \mathbb{E}_P[f(Y)]\right|\right] \\ &\qquad - \mathbb{E}\left[\sup_{f\in\mathcal{F}} \left|\mathbb{E}_Q[f(Z)] - \frac{1}{m}\sum_{i=1}^{m} f(Z_i)\right|\right] \\ &\ge\; \mathrm{IPM}_{\mathcal{F}}(P,Q) - 4\mathcal{R}_n(\mathcal{F}),\end{aligned}$$

where the last inequality uses the standard symmetrization trick (*e.g.*, van der Vaart and Wellner, 1996, Chapter 2.3). In particular, introducing i.i.d. copies $\tilde{Y}_i$s of Y_is, Jensen's inequality along with the triangle inequality yields

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^{n} f(Y_i) - \mathbb{E}_P[f(Y)]\right|\right] \le \mathbb{E}\left[\sup_{f\in\mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^{n} \omega_i\{f(Y_i) - f(\tilde{Y}_i)\}\right|\right] \le 2\mathcal{R}_n(\mathcal{F}).$$

The other expectation can be handled exactly the same way, which allows us to obtain the lower bound for $\mathbb{E}[T(\mathcal{X}_{n+m})]$.

We next look at $\mathbb{E}[T(\mathcal{X}^{\boldsymbol{\pi}}_{n+m})]$, which is equal to

$$\mathbb{E}[T(\mathcal{X}^{\boldsymbol{\pi}}_{n+m})] = \mathbb{E}\bigg[\sup_{f\in\mathcal{F}}\bigg|\frac{1}{n}\sum_{i=1}^{n} f(X_{\pi_i}) - \frac{1}{m}\sum_{i=1}^{m} f(X_{\pi_{n+i}})\bigg|\bigg].$$

Let $I_1, \ldots, I_{\binom{m}{n}}$ denote all subsets $(i_1, \ldots, i_n)$ of $[m]$ satisfying $1 \le i_1 < \ldots < i_n \le m$. Observe that the sample mean can be expressed as the U-statistic with a kernel of order n as below:

$$\frac{1}{m}\sum_{i=1}^{m} f(X_{\pi_{n+i}}) = \frac{1}{\binom{m}{n}} \sum_{1\le j\le \binom{m}{n}} \bigg\{\frac{1}{n}\sum_{i\in I_j} f(X_{\pi_{n+i}})\bigg\}.$$

This observation along with Jensen's inequality yields

$$\mathbb{E}[T(\mathcal{X}^{\boldsymbol{\pi}}_{n+m})] \le \frac{1}{\binom{m}{n}} \sum_{1\le j\le \binom{m}{n}} \mathbb{E}\bigg[\sup_{f\in\mathcal{F}}\bigg|\frac{1}{n}\sum_{i=1}^{n} f(X_{\pi_i}) - \frac{1}{n}\sum_{i\in I_j} f(X_{\pi_{n+i}})\bigg|\bigg].$$

Without loss of generality, set $I_j = (1, \ldots, n)$ and then we have

$$\mathbb{E}\bigg[\sup_{f\in\mathcal{F}}\bigg|\frac{1}{n}\sum_{i=1}^{n} f(X_{\pi_i}) - \frac{1}{n}\sum_{i\in I_j} f(X_{\pi_{n+i}})\bigg|\bigg] = \mathbb{E}\bigg[\sup_{f\in\mathcal{F}}\bigg|\frac{1}{n}\sum_{i=1}^{n} \omega_i\{f(X_{\pi_i}) - f(X_{\pi_{n+i}})\}\bigg|\bigg],$$

as randomly switching the order between π_i and π_{n+i} for $i \in [n]$ does not change the distribution of

$$\sup_{f\in\mathcal{F}}\bigg|\frac{1}{n}\sum_{i=1}^{n} f(X_{\pi_i}) - f(X_{\pi_{n+i}})\bigg|.$$

Therefore we may upper bound $\mathbb{E}[T(\mathcal{X}^{\boldsymbol{\pi}}_{n+m})]$ as

$$\mathbb{E}[T(\mathcal{X}^{\boldsymbol{\pi}}_{n+m})] \le 2\mathcal{R}_n(\mathcal{F}).$$

Moving to the variance terms, since $T(\mathcal{X}_{n+m})$ is a function of independent random variables, we can apply the Efron–Stein inequality for the variance (Boucheron et al., 2013, Corollary 3.2) as

$$\begin{aligned}
\mathrm{Var}[T(\mathcal{X}_{n+m})] &\le \frac{1}{2}\sum_{i=1}^{n+m} \mathbb{E}\big[\big(T(\mathcal{X}_{n+m}) - T(\mathcal{X}^{(i)}_{n+m})\big)^2\big] \\
&\le \frac{1}{2n}\sup_{X,X'\in\mathbb{S}}\sup_{f\in\mathcal{F}}|f(X) - f(X')|^2 + \frac{1}{2m}\sup_{X,X'\in\mathbb{S}}\sup_{f\in\mathcal{F}}|f(X) - f(X')|^2 \\
&\le n\Delta_T^2,
\end{aligned}$$

where $\mathcal{X}^{(i)}_{n+m} = (X_1, \ldots, X'_i, \ldots, X_{n+m})$, and X'_i is an i.i.d. copy of X_i independent of everything else. This proves that $\mathrm{Var}[T(\mathcal{X}_{n+m})] \le n\Delta_T^2$.

For the last term, the law of total variance shows

$$\mathrm{Var}[T(\mathcal{X}^{\boldsymbol{\pi}}_{n+m})] = \mathrm{Var}[\mathbb{E}\{T(\mathcal{X}^{\boldsymbol{\pi}}_{n+m}) \,|\, \boldsymbol{\pi}\}] + \mathbb{E}[\mathrm{Var}\{T(\mathcal{X}^{\boldsymbol{\pi}}_{n+m}) \,|\, \boldsymbol{\pi}\}].$$

Conditional on $\boldsymbol{\pi}$, the same analysis as above yields $\mathrm{Var}\{T(\mathcal{X}^{\boldsymbol{\pi}}_{n+m}) \,|\, \boldsymbol{\pi}\} \le n\Delta_T^2$ based on the Efron–Stein inequality, which gives $\mathbb{E}[\mathrm{Var}\{T(\mathcal{X}^{\boldsymbol{\pi}}_{n+m}) \,|\, \boldsymbol{\pi}\}] \le n\Delta_T^2$. For the first term, using the same trick used in the analysis of $\mathbb{E}[T(\mathcal{X}^{\boldsymbol{\pi}}_{n+m})]$, we have

$$\begin{aligned}
\mathrm{Var}[\mathbb{E}\{T(\mathcal{X}^{\boldsymbol{\pi}}_{n+m}) \,|\, \boldsymbol{\pi}\}] &\le \mathbb{E}[(\mathbb{E}\{T(\mathcal{X}^{\boldsymbol{\pi}}_{n+m}) \,|\, \boldsymbol{\pi}\})^2] \\
&\le \mathbb{E}\left[\left(\mathbb{E}\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\omega_i\{f(X_{\pi_i}) - f(X_{\pi_{n+i}})\}\right| \,\middle|\, \boldsymbol{\pi}\right\}\right)^2\right] \\
&\le 4\mathcal{R}_n^2(\mathcal{F}).
\end{aligned}$$

Having these ingredients, one can directly check the condition (8) is fulfilled if

$$\mathrm{IPM}_{\mathcal{F}}(P,Q) \ge C_1\frac{\mathcal{R}_n(\mathcal{F})}{\sqrt{\alpha\beta}} + C_2\frac{\sqrt{n}\Delta_T}{\sqrt{\alpha\beta}} + C_3\frac{\Delta_T}{\xi_{\varepsilon,\delta}}\max\left\{\log\left(\frac{1}{\alpha}\right), \log\left(\frac{1}{\beta}\right)\right\},$$

where C_1, C_2, C_3 are some positive constants.

B.5 Separation in HSIC metric

In this subsection, we develop results similar to those in Section 5.1 in terms of the HSIC metric. We start by discussing the minimum separation for the dpHSIC test in Theorem 12 and then establish the matching lower bound in Theorem 13. Letting $\mathcal{P}_{\mathbb{Y}\times\mathbb{Z}}$ denote the class of distributions on $\mathbb{Y}\times\mathbb{Z}$ and $\rho > 0$, we define the set of alternative distributions as

$$\mathcal{P}_{\mathrm{HSIC}_{k\otimes\ell}}(\rho) \coloneqq \{P_{YZ} \in \mathcal{P}_{\mathbb{Y}\times\mathbb{Z}} : \mathrm{HSIC}_{k\otimes\ell}(P_{YZ}) \ge \rho\}.$$

For a given target type II error $\beta \in (0, 1-\alpha)$, the minimum separation for the dpHSIC test against $\mathcal{P}_{\mathrm{HSIC}_{k\otimes\ell}}(\rho)$ is given as

$$\rho_{\phi_{\mathrm{dpHSIC}}}(\alpha,\beta,\varepsilon,\delta,n) \coloneqq \inf\left\{\rho > 0 : \sup_{P_{YZ}\in\mathcal{P}_{\mathrm{HSIC}_{k\otimes\ell}}(\rho)} \mathbb{E}_{P_{YZ}}[1-\phi_{\mathrm{dpHSIC}}] \le \beta\right\}. \tag{19}$$

The next theorem, which is analogous to Theorem 7 for the dpMMD test, provides an upper bound for the minimum separation $\rho_{\phi_{\mathrm{dpHSIC}}}(\alpha,\beta,\varepsilon,\delta,n)$ as a function of α, β, ε, δ and n.

Theorem 12 (Minimum Separation of ϕ_{dpHSIC})**.** *Assume that the kernels k and ℓ are bounded as $0 \le k(y,y') \le K$ and $0 \le \ell(z,z') \le L$ for all $y, y' \in \mathbb{Y}$ and $z, z' \in \mathbb{Z}$. For all values of $\alpha \in (0,1)$, $\beta \in (0, 1-\alpha)$, $\varepsilon > 0$, $\delta \in [0,1)$ and $B \ge 16\alpha^{-2}\log(8/\beta)$, the minimum separation for ϕ_{dpHSIC} satisfies*

$$\rho_{\phi_{\mathrm{dpHSIC}}}(\alpha,\beta,\varepsilon,\delta,n) \le C_{K,L}\max\left\{\sqrt{\frac{\max\{\log(1/\alpha), \log(1/\beta)\}}{n}}, \frac{\max\{\log(1/\alpha), \log(1/\beta)\}}{n\xi_{\varepsilon,\delta}}\right\},$$

where $C_{K,L}$ is a positive constant that only depends on K and L, and $\xi_{\varepsilon,\delta}$ is given in (1).

The proof of Theorem 12, given in Appendix F.3, follows a similar approach to that of Theorem 8. A notable difference, however, is that we need to use exponential tail bounds for the HSIC statistic instead of the MMD statistic. To this end, we develop exponential concentration results for the empirical HSIC (Lemma 14) and the permuted HSIC (Lemma 12), by leveraging the recent result of Kim et al. (2022a) and McDiarmid's inequality. As clearly demonstrated in the proof, the use of these tools is crucial in obtaining the logarithmic factors in α and β.

We now examine minimax optimality of ϕ_{dpHSIC} with a focus on the cases where $\mathbb{Y} = \mathbb{R}^{d_Y}$ and $\mathbb{Z} = \mathbb{R}^{d_Z}$. For this direction, we establish a lower bound for the minimax separation $\rho^\star_{\text{HSIC}}$ defined below and compare it with $\rho_{\phi_{\text{dpHSIC}}}$. Let $\phi : \mathcal{X}_n \mapsto \{0, 1\}$ be a test function and $\mathcal{P}_0 := \{P_{YZ} \in \mathcal{P}_{\mathbb{Y}\times\mathbb{Z}} : P_{YZ} = P_Y P_Z\}$ be the set of null distributions on $\mathbb{R}^{d_Y} \times \mathbb{R}^{d_Z}$. We denote the set of (ε, δ)-DP level α tests as

$$\Phi_{\alpha,\varepsilon,\delta} := \Big\{\phi : \sup_{P_{YZ}\in\mathcal{P}_0} \mathbb{E}_{P_{YZ}}[\phi] \le \alpha \text{ and } \phi \text{ is } (\varepsilon, \delta)\text{-DP}\Big\},$$

and define the minimax separation in terms of the HSIC metric as

$$\rho^\star_{\text{HSIC}}(\alpha, \beta, \varepsilon, \delta, n) := \inf\Big\{\rho > 0 : \inf_{\phi\in\Phi_{\alpha,\varepsilon,\delta}} \sup_{P_{YZ}\in\mathcal{P}_{\text{HSIC}_{k\otimes\ell}}(\rho)} \mathbb{E}_{P_{YZ}}[1 - \phi] \le \beta\Big\}.$$

Similar to Theorem 8, the next result establishes a lower bound for the minimax separation $\rho^\star_{\text{HSIC}}$ under the DP constraint.

Theorem 13 (Minimax Separation in HSIC)**.** *Let α and β be real numbers in the interval $(0, 1/5)$, $\varepsilon > 0$ and $\delta \in [0, 1)$. Assume that the kernel functions k and ℓ are translation invariant on $\mathbb{R}^{d_Y}$ and $\mathbb{R}^{d_Z}$, respectively. In particular, there exist some functions κ_Y, κ_Z such that $k(y, y') = \kappa_Y(y - y')$ for all $y, y' \in \mathbb{R}^{d_Y}$ and $\ell(z, z') = \kappa_Z(z - z')$ for all $z, z' \in \mathbb{R}^{d_Z}$. Moreover, assume that the kernels are non-constant in the sense that there exist positive constants η_Y, η_Z such that $\kappa_Y(0) - \kappa_Y(y_0) \ge \eta_Y$ and $\kappa_Z(0) - \kappa_Z(z_0) \ge \eta_Z$ for some $y_0 \in \mathbb{R}^{d_Y}$ and $z_0 \in \mathbb{R}^{d_Z}$. Then the minimax separation over $\mathcal{P}_{\text{HSIC}_{k\otimes\ell}}(\rho)$ is lower bounded as*

$$\rho^\star_{\text{HSIC}}(\alpha, \beta, \varepsilon, \delta, n) \ge C_{\eta_Y,\eta_Z} \max\left\{\min\left\{\sqrt{\frac{\log(1/(\alpha+\beta))}{n}}, 1\right\}, \min\left\{\frac{\log(1/\beta)}{n\xi_{\varepsilon,\delta}}, 1\right\}\right\},$$

where C_{η_Y,η_Z} is a positive constant that only depends on η_Y and η_Z, and $\xi_{\varepsilon,\delta}$ is given in (1).

The proof of this result can be found in Appendix F.4. By comparing the result of Theorem 12 with the above result, it becomes evident that $\rho_{\phi_{\text{dpHSIC}}} \asymp \rho^\star_{\text{HSIC}}$ under the conditions $\alpha \asymp \beta$ and $\mathbb{Y} \times \mathbb{Z} = \mathbb{R}^{d_Y} \times \mathbb{R}^{d_Z}$. Consequently, ϕ_{dpHSIC} achieves minimax rate optimality across all parameters, namely $(\alpha, \beta, \varepsilon, n)$, provided that the type I error and the type II error are comparable. Other comments on Theorem 8 can similarly be applied to Theorem 13. For instance, numerous kernels frequently used in practice are translation invariant and the existence of η_Y and η_Z in the theorem statement is guaranteed as long as the kernels k and ℓ are characteristic (Tolstikhin et al., 2017).

The minimax results for both dpMMD and dpHSIC tests indicate the existence of an inherent trade-off between privacy and statistical power. More concretely, one cannot expect to achieve the

parametric $\sqrt{n}$-separation rate in both MMD and HSIC metrics when the privacy parameter $\xi_{\varepsilon,\delta}$ is much smaller than $n^{-1/2}$. This intrinsic trade-off has been observed in a variety of statistical problems (Diakonikolas et al., 2015; Acharya et al., 2018; Kamath et al., 2019, 2020; Acharya et al., 2021) and our work serves to reaffirm this trade-off in the context of kernel-based testing problems.

B.6 Separation in L_2 Metric

We next investigate the minimum separation of the dpHSIC test against a class of alternatives measured in terms of the L_2 metric. For $\rho > 0$, let $\tilde{\mathcal{P}}_{L_2}(\rho)$ be the collection of distributions P_{YZ} on $\mathbb{R}^{d_Y} \times \mathbb{R}^{d_Z}$, where P_{YZ} is equipped with the Lebesgue density function p_{YZ} and the product of the marginals $p_Y p_Z$ such that $\|p_{YZ} - p_Y p_Z\|_{L_2} \geq \rho$. The class of alternative distributions of interest is a subset of $\tilde{\mathcal{P}}_{L_2}(\rho)$ given as

$$\tilde{\mathcal{P}}^s_{L_2}(\rho) := \left\{ P_{YZ} \in \tilde{\mathcal{P}}_{L_2}(\rho) : p_{YZ} - p_Y p_Z \in \mathcal{S}^s_{d_Y+d_Z}(R),\ \max(\|p_{YZ}\|_{L_\infty}, \|p_Y p_Z\|_{L_\infty}) \leq M \right\},$$

where $\mathcal{S}^s_{d_Y+d_Z}(R)$ is the Sobolev ball defined in (14). As for the dpMMD test in Section 5.2, we focus on the use of the Gaussian kernels for simplicity. In particular, for $y, y' \in \mathbb{R}^{d_Y}$ and $z, z' \in \mathbb{R}^{d_Z}$, consider the Gaussian kernels with bandwidths $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{d_Y}) \in (0,\infty)^{d_Y}$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{d_Z}) \in (0,\infty)^{d_Z}$ given as

$$k_{\boldsymbol{\lambda}}(y, y') = \prod_{i=1}^{d_Y} \frac{1}{\sqrt{2\pi}\lambda_i} e^{-\frac{(y_i - y'_i)^2}{2\lambda_i^2}} \quad \text{and} \quad \ell_{\boldsymbol{\mu}}(z, z') = \prod_{i=1}^{d_Z} \frac{1}{\sqrt{2\pi}\mu_i} e^{-\frac{(z_i - z'_i)^2}{2\mu_i^2}}.$$

Let us denote the minimum separation of the dpHSIC test with the above Gaussian kernels against L_2 alternatives as

$$\rho_{\phi_{\text{dpHSIC}}, L_2}(\alpha, \beta, \varepsilon, \delta, n, d_Y, d_Z, s, R, M) := \inf \left\{ \rho > 0 : \sup_{P_{YZ} \in \tilde{\mathcal{P}}^s_{L_2}(\rho)} \mathbb{E}_{P_{YZ}}[1 - \phi_{\text{dpHSIC}}] \leq \beta \right\}.$$

The next theorem presents an upper bound for $\rho_{\phi_{\text{dpHSIC}}, L_2}$ in terms of a set of parameters including $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ and n. The proof of Theorem 14 is given in Appendix F.5.

Theorem 14 (Minimum Separation of dpHSIC over $\tilde{\mathcal{P}}^s_{L_2}$)**.** *Assume that $\alpha \in (0,1)$, $\beta \in (0, 1-\alpha)$, $\varepsilon > 0$, $\delta \in [0,1)$, $B \geq 16\alpha^{-2}\log(8/\beta)$, $\prod_{i=1}^{d_Y} \lambda_i \leq 1$ and $\prod_{i=1}^{d_Z} \mu_i \leq 1$. The minimum separation of the* dpHSIC *test with the Gaussian kernel over $\tilde{\mathcal{P}}^s_{L_2}$ is upper bounded as*

$$\rho^2_{\phi_{\text{dpHSIC}}, L_2} \leq C_{\alpha,\beta,s,R,M,d_Y,d_Z} \left\{ \sum_{i=1}^{d_Y} \lambda_i^{2s} + \sum_{i=1}^{d_Z} \mu_i^{2s} + \frac{1}{n\sqrt{\lambda_1 \cdots \lambda_{d_Y} \mu_1 \cdots \mu_{d_Z}}} \right.$$
$$\left. + \frac{1}{n^2 \lambda_1 \cdots \lambda_{d_Y} \mu_1 \cdots \mu_{d_Z} \xi^2_{\varepsilon,\delta}} + \frac{1}{n^{3/2} \lambda_1 \cdots \lambda_{d_Y} \mu_1 \cdots \mu_{d_Z} \xi_{\varepsilon,\delta}} \right\},$$

where $C_{\alpha,\beta,s,R,M,d_Y,d_Z}$ is a positive constant, depending only on $\alpha, \beta, s, R, M, d_Y, d_Z$, and $\xi_{\varepsilon,\delta}$ is given in (1).

We make a few comments on Theorem 14.

- In contrast to the prior work (Berrett et al., 2021; Albert et al., 2022; Kim et al., 2022a; Schrab et al., 2022) that considers U-statistic for independence testing against L_2 alternatives, our result is based on the V-statistic of the HSIC, which requires additional effort in dealing with bias terms. As emphasized before, the V-statistic is more favorable than the U-statistic in high privacy regimes as the former has smaller global sensitivity than the latter.

- The main idea behind the proof is similar to that of Theorem 9 for the dpMMD test where we first analyze the difference between the U- and V-statistics of the MMD, and then leverage the existing results of the U-statistic. Extending this strategy to the HSIC requires substantial effort as the U- and V-statistics of the HSIC are based on the fourth-order kernel (whereas, the corresponding MMD statistics are based on the second-order kernel).

- In the proof, we do not keep track of the dependence on the type I error rate α in deriving the minimum separation. We believe that the dependence on α can be improved in view of Schrab et al. (2022, Theorem 3) where they establish a logarithmic dependence on α using U-statistics. Nevertheless, extending this result to the V-statistic poses a non-trivial challenge, and we defer this investigation to future research.

- We remark that if $\xi_{\varepsilon,\delta}$ is sufficiently large so that the following inequality holds

$$\rho^2_{\phi_{\mathrm{dpHSIC}},L_2} \leq C_{\alpha,\beta,s,R,M,d_Y,d_Z}\left\{\sum_{i=1}^{d_Y}\lambda_i^{2s} + \sum_{i=1}^{d_Z}\mu_i^{2s} + \frac{1}{n\sqrt{\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}}}\right\},$$

  then we obtain the same separation rate as the non-private HSIC test studied in Albert et al. (2022). In particular, letting $\lambda_1 = \cdots \lambda_{d_Y} = \mu_1 = \cdots = \mu_{d_Z} = n^{-2/(4s+d_Y+d_Z)}$, we see that the dpHSIC test achieves minimax optimal rate $n^{-2s/(4s+d_Y+d_Z)}$ against the Sobolev ball.

- As for the result (16) of the dpMMD test, one can achieve different minimum separation rates by varying the bandwidth parameters of the Gaussian kernel. In particular, a similar analysis given in Appendix B.8 verifies that

$$\rho_{\phi_{\mathrm{dpHSIC}},L_2} \lesssim \begin{cases} n^{-\frac{2s}{4s+d}}, & \text{if } n^{-\frac{2s-d/2}{4s+d}} \lesssim \xi_{\varepsilon,\delta} \text{ (low privacy)}, \\ (n^{\frac{3}{2}}\xi_{\varepsilon,\delta})^{-\frac{s}{2s+d}}, & \text{if } n^{-\frac{1}{2}} \lesssim \xi_{\varepsilon,\delta} \lesssim n^{-\frac{2s-d/2}{4s+d}} \text{ (mid privacy)}, \\ (n\xi_{\varepsilon,\delta})^{-\frac{2s}{2s+d}}, & \text{if } \xi_{\varepsilon,\delta} \lesssim n^{-\frac{1}{2}} \text{ (high privacy)}. \end{cases} \tag{20}$$

- As in the dpMMD test, the optimal choice of the bandwidths assumes that the smoothness parameter s is known. In this regard, it would be interesting to develop an adaptive test that does not require the knowledge of s, while retaining (nearly) optimal power. We leave this direction for future work. We also note that establishing lower bounds in high privacy regimes is another interesting avenue for future work.

We next turn to the minimum separation of the private independence test based on a U-statistic, and compare it with the one established in Theorem 14.

B.7 Private Test based on the HSIC U-statistic

Mirroring Section 5.3, this subsection develops a similar negative result for the dpHSIC test based on a U-statistic. As an unbiased estimator of the square of $\mathrm{HSIC}_{k\otimes\ell}(P_{YZ})$, the U-statistic of HSIC is given as

$$\begin{aligned} U_{\mathrm{HSIC}} &= \frac{1}{n(n-1)} \sum_{(i,j)\in \mathbf{i}_2^n} k(Y_i, Y_j)\ell(Z_i, Z_j) + \frac{(n-4)!}{n!} \sum_{(i_1,i_2,j_1,j_2)\in \mathbf{i}_4^n} k(Y_{i_1}, Y_{j_1})\ell(Z_{i_2}, Z_{j_2}) \\ &\quad - \frac{2}{n(n-1)(n-2)} \sum_{(i,j_1,j_2)\in \mathbf{i}_3^n} k(Y_i, Y_{j_1})\ell(Z_i, Z_{j_2}), \end{aligned} \tag{21}$$

where we recall that $\mathbf{i}_p^q$ denotes the set of all p-tuples drawn without replacement from $[q]$. To privatize U_{HSIC} as well as the test function through the Laplace mechanism, the following lemma computes the global sensitivity of U_{HSIC}.

Lemma 9. *Assume that the kernels k and ℓ are bounded as $0 \le k(y, y') \le K$ and $0 \le \ell(z, z') \le L$ for all $y, y' \in \mathbb{Y}$ and $z, z' \in \mathbb{Z}$. In addition, assume that k and ℓ are translation invariant, and have non-empty level sets on $\mathbb{Y}$ and $\mathbb{Z}$, respectively. Then there exists a positive sequence $c_n \in [2, 24]$ such that for all $n \ge 4$,*

$$\sup_{\boldsymbol{\pi}\in\boldsymbol{\Pi}_n} \sup_{\substack{\mathcal{X}_n, \tilde{\mathcal{X}}_n: \\ d_{\mathrm{ham}}(\mathcal{X}_n, \tilde{\mathcal{X}}_n)\le 1}} \left| U_{\mathrm{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}}) - U_{\mathrm{HSIC}}(\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}}) \right| = \frac{c_n K L}{n}.$$

The proof of Lemma 9 is given in Appendix F.6. We would like to highlight that determining the precise global sensitivity of U_{HSIC} is more challenging than that of U_{MMD} as the former is a fourth-order U-statistic, whereas the latter is a second order one. Having established the global sensitivity of U_{HSIC}, consider the dpHSIC test in Algorithm 1 using the test statistic U_{HSIC} and the global sensitivity $\Delta_T = \frac{c_n K L}{n}$ where c_n can be an arbitrary sequence of constants between 2 and 24. We denote the resulting test as ϕ^u_{dpHSIC} and show its suboptimal power property. The proof of the result below can be found in Appendix F.7

Theorem 15 (Suboptimality of ϕ^u_{dpHSIC})**.** *Assume that the kernels k and ℓ fulfills the conditions specified in Lemma 9. Moreover, assume that if $P_{YZ} \in \mathcal{P}_{\mathbb{Y}\times\mathbb{Z}}$, then $wP_{YZ} + (1-w)P_Z P_Z \in \mathcal{P}_{\mathbb{Y}\times\mathbb{Z}}$ for all $w \in [0, 1]$, and there exists $P_{YZ} \in \mathcal{P}_{\mathbb{Y}\times\mathbb{Z}}$ such that $\mathrm{HSIC}_{k\otimes\ell}(P_{YZ}) = \varrho_0$ for some fixed $\varrho_0 > 0$. Let $\alpha > \frac{1}{B+1}$ and $\alpha \in (0, 1)$, $\beta \in (0, 1-\alpha)$ be fixed values. Consider the high privacy regime where $\xi_{\varepsilon,\delta} \asymp n^{-1/2-r}$ with fixed $r \in (0, 1/2)$, for $\xi_{\varepsilon,\delta}$ as in (1). Then the uniform power of ϕ^u_{dpHSIC} is asymptotically at most α over $\mathcal{P}_{\mathrm{HSIC}_{k\otimes\ell}}(\rho)$ where ρ is given in (17). In other words, it holds that*

$$\limsup_{n\to\infty} \inf_{P_{YZ}\in\mathcal{P}_{\mathrm{HSIC}_{k\otimes\ell}}(\rho)} \mathbb{E}_{P_{YZ}}[\phi^u_{\mathrm{dpHSIC}}] \le \alpha.$$

The comments made for Theorem 10 also apply to Theorem 15. Specifically, as in the MMD case, the U-statistic of the HSIC requires a higher noise level than the V-statistic to ensure differential privacy through the Laplace mechanism. To be more precise, ϕ^u_{dpHSIC} becomes powerful in the

privacy regime when the target parameter $\mathrm{HSIC}^2_{k\otimes\ell}$ is greater than the Laplace noise level $(n\xi_{\varepsilon,\delta})^{-1}$ or equivalently $\mathrm{HSIC}_{k\otimes\ell}$ is greater than $(n\xi_{\varepsilon,\delta})^{-1/2}$. However, the second term $\min\{(n\xi_{\varepsilon,\delta})^{-1}, 1\}$ in the minimax separation rate (Theorem 13) is smaller than $(n\xi_{\varepsilon,\delta})^{-1/2}$, and thus we may see that ϕ^u_{dpHSIC} becomes powerless when $\min\{(n\xi_{\varepsilon,\delta})^{-1}, 1\} \lesssim \mathrm{HSIC}_{k\otimes\ell} \lesssim (n\xi_{\varepsilon,\delta})^{-1/2}$. We make this intuition more precise in the proof of Theorem 15 and establish that the worse-case power is essentially bounded by significance level α.

Our next concern is the minimum separation of ϕ^u_{dpHSIC} test over $\tilde{\mathcal{P}}^s_{L_2}$, which mirrors Theorem 11 for the U-statistic of the MMD. As in Theorem 11, we focus on the Gaussian kernels for the HSIC, and establish the following result. The proof can be found in Appendix F.8.

Theorem 16 (Minimum Separation of ϕ^u_{dpHSIC} over $\tilde{\mathcal{P}}^s_{L_2}$)**.** *Assume that $\alpha \in (0,1)$, $\beta \in (0, 1-\alpha)$, $\varepsilon > 0$, $\delta \in [0,1)$, $B \geq 16\alpha^{-2}\log(8/\beta)$, $\prod_{i=1}^{d_Y} \lambda_i \leq 1$ and $\prod_{i=1}^{d_Z} \mu_i \leq 1$. Then the minimum separation of ϕ^u_{dpHSIC} with Gaussian kernels over $\tilde{\mathcal{P}}^s_{L_2}$ is upper bounded as*

$$\rho^2_{\phi^u_{\mathrm{dpHSIC}}, L_2} \leq C_{\alpha,\beta,s,R,M,d_Y,d_Z} \left\{ \sum_{i=1}^{d_Y} \lambda_i^{2s} + \sum_{i=1}^{d_Z} \mu_i^{2s} + \frac{1}{n\sqrt{\lambda_1 \cdots \lambda_{d_Y} \mu_1 \cdots \mu_{d_Z}}} + \frac{1}{n\lambda_1 \cdots \lambda_{d_Y} \mu_1 \cdots \mu_{d_Z} \xi_{\varepsilon,\delta}} \right\},$$

where $C_{\alpha,\beta,s,R,M,d_Y,d_Z}$ is a positive constant, depending only on $\alpha, \beta, R, M, d_Y, d_Z$, and $\xi_{\varepsilon,\delta}$ is given in (1).

By comparing Theorem 16 with Theorem 14, we see that the upper bound for $\rho^2_{\phi^u_{\mathrm{dpHSIC}}, L_2}$ is smaller than that of $\rho^2_{\phi_{\mathrm{dpHSIC}}, L_2}$ if $n\xi_{\varepsilon,\delta} \lesssim 1$. Since we assume $\prod_{i=1}^{d_Y} \lambda_i \leq 1$ and $\prod_{i=1}^{d_Z} \mu_i \leq 1$, the condition $n\xi_{\varepsilon,\delta} \lesssim 1$ implies that $\|p_{YZ} - p_Y p_Z\|_{L_2}$ needs to be sufficiently large to ensure significant power. However, the L_2 distance cannot be made sufficiently large as $\|p_{YZ}\|_{L_\infty}$ and $\|p_Y p_Z\|_{L_\infty}$ are assumed to be bounded. Therefore, the proposed test ϕ_{dpHSIC} is more favorable than the one based on the U-statistic in terms of obtaining a tight separation rate in the L_2 distance. Nevertheless, when $\xi_{\varepsilon,\delta}$ is sufficiently large, the minimum separation satisfies

$$\rho^2_{\phi^u_{\mathrm{dpHSIC}}, L_2} \leq C_{\alpha,\beta,s,R,M,d_Y,d_Z} \left\{ \sum_{i=1}^{d_Y} \lambda_i^{2s} + \sum_{i=1}^{d_Z} \mu_i^{2s} + \frac{1}{n\sqrt{\lambda_1 \cdots \lambda_{d_Y} \mu_1 \cdots \mu_{d_Z}}} \right\}.$$

Therefore, in low privacy regimes, both ϕ_{dpHSIC} and ϕ^u_{dpHSIC} tests achieve the same separation rate in terms of the L_2 distance, which can be optimal when $\lambda_1 = \cdots = \lambda_{d_Y} = \mu_1 = \cdots = \mu_{d_Z} \asymp n^{-2/(4s+d_Y+d_Z)}$.

Like the negative results of U_{MMD}, our negative results of U_{HSIC} are based solely on the Laplace mechanism. It is currently unknown whether there is an alternative privacy mechanism, enabling a test based on U_{HSIC} to be optimal in high privacy regimes. We leave this important direction for future work.

B.8 Analyses of Minimum Separation Rates

In this section, we provide details on the separation rate stated in (16). Throughout our discussion, we treat the level α as a fixed constant. Let us recall the three terms in the separation rate:

$$\begin{aligned}\text{(I)} &= \frac{\log(1/\alpha)}{n\sqrt{\lambda_1 \cdots \lambda_d}},\\ \text{(II)} &= \frac{\log(1/\alpha)}{n^{3/2}\lambda_1 \cdots \lambda_d \xi_{\varepsilon,\delta}},\\ \text{(III)} &= \frac{\log^2(1/\alpha)}{n^2 \lambda_1 \cdots \lambda_d \xi_{\varepsilon,\delta}^2}.\end{aligned}$$

Among these three terms, the dominant term becomes

$$\begin{cases} \text{(I)}, & \text{if } n^{-1/2}(\lambda_1 \dots \lambda_d)^{-1/2} \lesssim \xi_{\varepsilon,\delta} \text{ (low privacy regime)},\\ \text{(II)}, & \text{if } n^{-1/2} \lesssim \xi_{\varepsilon,\delta} \lesssim n^{-1/2}(\lambda_1 \dots \lambda_d)^{-1/2} \text{ (mid privacy regime)},\\ \text{(III)}, & \text{if } \xi_{\varepsilon,\delta} \lesssim n^{-1/2} \text{ (high privacy regime)}. \end{cases}$$

We choose the bandwidth parameters so that each dominating term matches the order of $\sum_{i=1}^d \lambda_i^{2s}$ and compute the resulting rate explained below.

- **Low privacy regime.** First, in the low privacy regime, we have the separation rate satisfying

$$\rho^2_{\phi_{\mathrm{dpMMD}},L_2} \le C_{\tau,\beta,s,R,M,d}\left\{\sum_{i=1}^d \lambda_i^{2s} + \frac{\log(1/\alpha)}{n\sqrt{\lambda_1 \cdots \lambda_d}}\right\}.$$

  Equating (I) with $\sum_{i=1}^d \lambda_i^{2s}$ yields optimal bandwidths $\lambda_i = n^{-\frac{2}{4s+d}}$ for $i = 1, \dots, d$, and the resulting rate is $n^{-\frac{4s}{4s+d}}$, *i.e.*, $\rho^2_{\phi_{\mathrm{dpMMD}},L_2} \lesssim n^{-\frac{4s}{4s+d}}$. This rate is known to be minimax optimal over the Sobolev ball under the non-DP setting (Li and Yuan, 2019; Schrab et al., 2023). Therefore, when $\lambda_i = n^{-\frac{2}{4s+d}}$ for $i \in [d]$, we can achieve an optimal separation rate whenever $\xi_{\varepsilon,\delta} \gtrsim n^{-\frac{2s-d/2}{4s+d}}$.

- **Mid privacy regime.** Second, in the mid privacy regime, we have the separation rate satisfying

$$\rho^2_{\phi_{\mathrm{dpMMD}},L_2} \le C_{\tau,\beta,s,R,M,d}\left\{\sum_{i=1}^d \lambda_i^{2s} + \frac{\log(1/\alpha)}{n^{3/2}\lambda_1 \cdots \lambda_d \xi_{\varepsilon,\delta}}\right\}.$$

  Equating (II) with $\sum_{i=1}^d \lambda_i^{2s}$ yields optimal bandwidths $\lambda_i = (n^{3/2}\xi_{\varepsilon,\delta})^{-\frac{1}{2s+d}}$ for $i = 1, \dots, d$, and the resulting rate becomes $(n^{3/2}\xi_{\varepsilon,\delta})^{-\frac{2s}{2s+d}}$. Therefore, we can achieve the separation rate $(n^{3/2}\xi_{\varepsilon,\delta})^{-\frac{2s}{2s+d}}$ using bandwidths $\lambda_i = (n^{3/2}\xi_{\varepsilon,\delta})^{-\frac{1}{2s+d}}$ for $i \in [d]$ whenever $n^{-1/2} \lesssim \xi_{\varepsilon,\delta} \lesssim n^{-1/2}(\lambda_1 \dots \lambda_d)^{-1/2}$, which is equivalent to

$$n^{-1/2} \lesssim \xi_{\varepsilon,\delta} \lesssim n^{-\frac{2s-d/2}{4s+d}}.$$

- **High privacy regime.** Lastly, in the high privacy regime, the separation rate satisfies

$$\rho^2_{\phi_{\mathrm{dpMMD}},L_2} \leq C_{\tau,\beta,s,R,M,d}\left\{\sum_{i=1}^{d}\lambda_i^{2s} + \frac{\log^2(1/\alpha)}{n^2\lambda_1\cdots\lambda_d\xi^2_{\varepsilon,\delta}}\right\}.$$

  Equating (III) with $\sum_{i=1}^{d}\lambda_i^{2s}$ yields optimal bandwidths $\lambda_i = (n\xi_{\varepsilon,\delta})^{-\frac{2}{2s+d}}$ for $i \in [d]$, and the resulting rate becomes $(n\xi_{\varepsilon,\delta})^{-\frac{4s}{2s+d}}$. Consequently, we have the separation rate $(n\xi_{\varepsilon,\delta})^{-\frac{4s}{2s+d}}$ whenever $\xi_{\varepsilon,\delta} \lesssim n^{-1/2}$.

To summarize, we can achieve the separation rate using different bandwidths in different privacy regimes as

$$\rho^2_{\phi_{\mathrm{dpMMD}},L_2} \lesssim \begin{cases} n^{-\frac{4s}{4s+d}}, & \text{in the low privacy regime with } n^{-\frac{2s-d/2}{4s+d}} \lesssim \xi_{\varepsilon,\delta}, \\ (n^{3/2}\xi_{\varepsilon,\delta})^{-\frac{2s}{2s+d}}, & \text{in the mid privacy regime with } n^{-1/2} \lesssim \xi_{\varepsilon,\delta} \lesssim n^{-\frac{2s-d/2}{4s+d}}, \\ (n\xi_{\varepsilon,\delta})^{-\frac{4s}{2s+d}}, & \text{in the high privacy regime with } \xi_{\varepsilon,\delta} \lesssim n^{-1/2}. \end{cases}$$

Each separation rate can be attained by using the bandwidths $\lambda_1 = \cdots = \lambda_d = n^{-\frac{2}{4s+d}}$ for the low privacy regime, $\lambda_1 = \cdots = \lambda_d = (n^{3/2}\xi_{\varepsilon,\delta})^{-\frac{1}{2s+d}}$ for the mid privacy regime and $\lambda_1 = \cdots = \lambda_d = (n\xi_{\varepsilon,\delta})^{-\frac{2}{2s+d}}$ for the high privacy regime.

C Additional Simulations

In this section, we present further experiments on:

(i) independence testing (Appendix C.1),

(ii) power in low privacy regimes (Appendix C.2) and

(iii) test significance levels (Appendix C.3).

These results are all analyzed in Section 6.4.

C.1 Perturbed Uniform Distributions for Independence Testing

We run independence testing simulations comparing the power of dpHSIC, U-Stat dpHSIC, Naive dpHSIC, TOT HSIC and SARRM HSIC, where the samples are drawn from the perturbed uniform joint density defined in Equation (18). The privacy level, sample size, and dimension are varied in Figures 9 to 11, respectively. As analyzed in Section 6.4, the same power trends are observed as in the two-sample MMD case of Section 6.2. Overall, dpHSIC achieves the highest power of all independence DP tests in all experimental settings and across all privacy regimes.

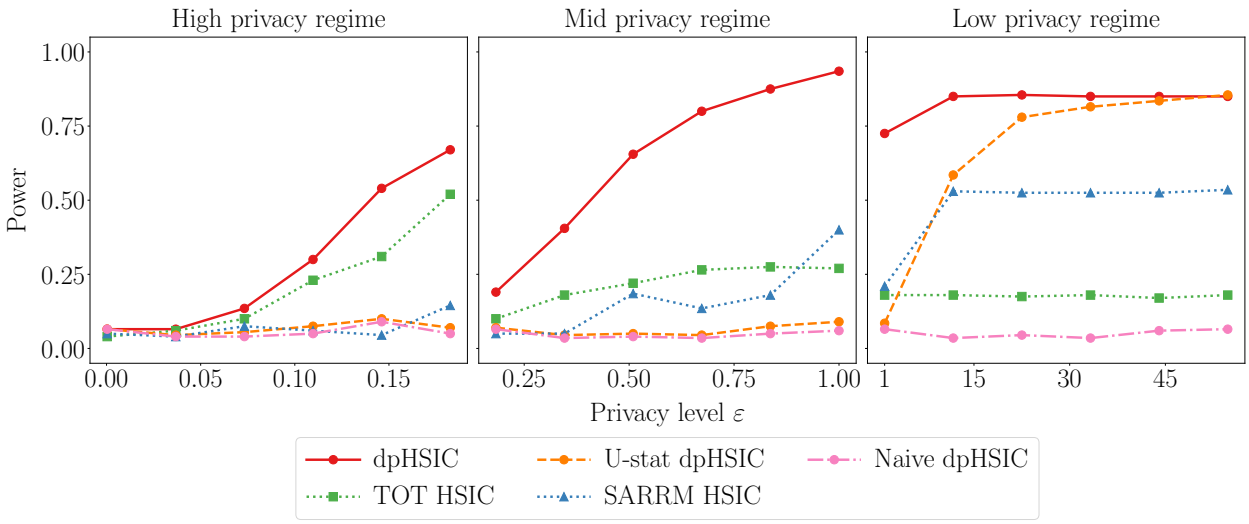

Figure 9: Measuring the dependence on joint perturbed uniform distributions while varying the privacy level ε. We set the sample size $n = 3000$ and dimensions $d_X = d_Y = 1$. We change the privacy level and perturbation amplitude as follows: *(Left)* Privacy level ε from $1/n$ to $10/\sqrt{n}$, perturbation amplitude $a = 0.4$. *(Middle)* Privacy level ε from $10/\sqrt{n}$ to 1, perturbation amplitude $a = 0.2$. *(Right)* Privacy level ε from 1 to $\sqrt{n}$, perturbation amplitude $a = 0.15$.

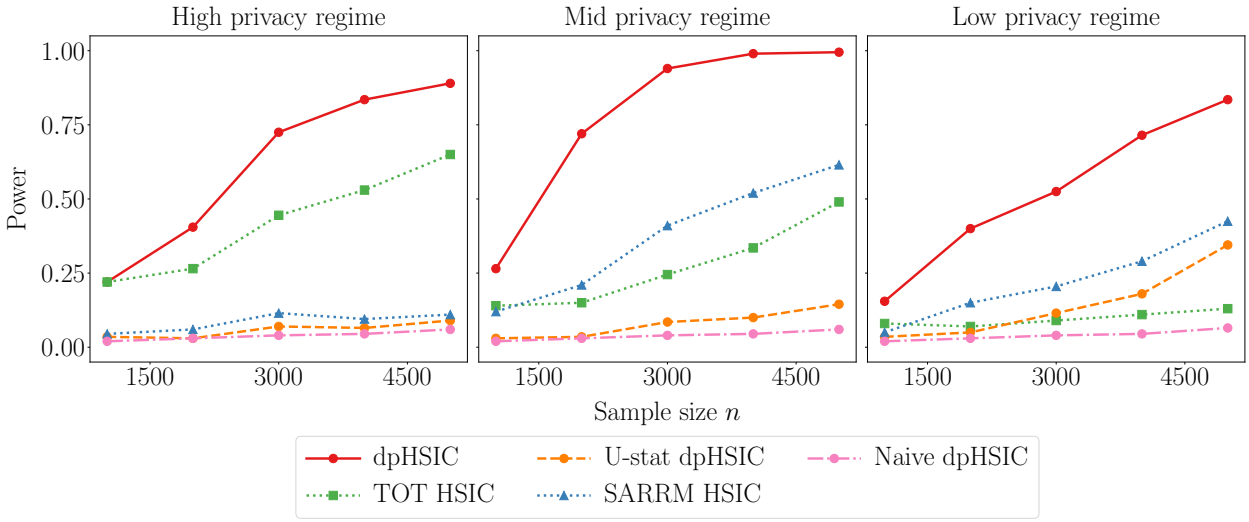

Figure 10: Measuring the dependence on joint perturbed uniform distributions while varying the sample size n. We set the dimensions $d_X = d_Y = 1$, and change other parameters as follows: *(Left)* Privacy level $\varepsilon = 10/\sqrt{n}$, perturbation amplitude $a = 0.4$. *(Middle)* Privacy level $\varepsilon = 1$, perturbation amplitude $a = 0.2$. *(Right)* Privacy level $\varepsilon = \sqrt{n}/10$, perturbation amplitude $a = 0.1$.

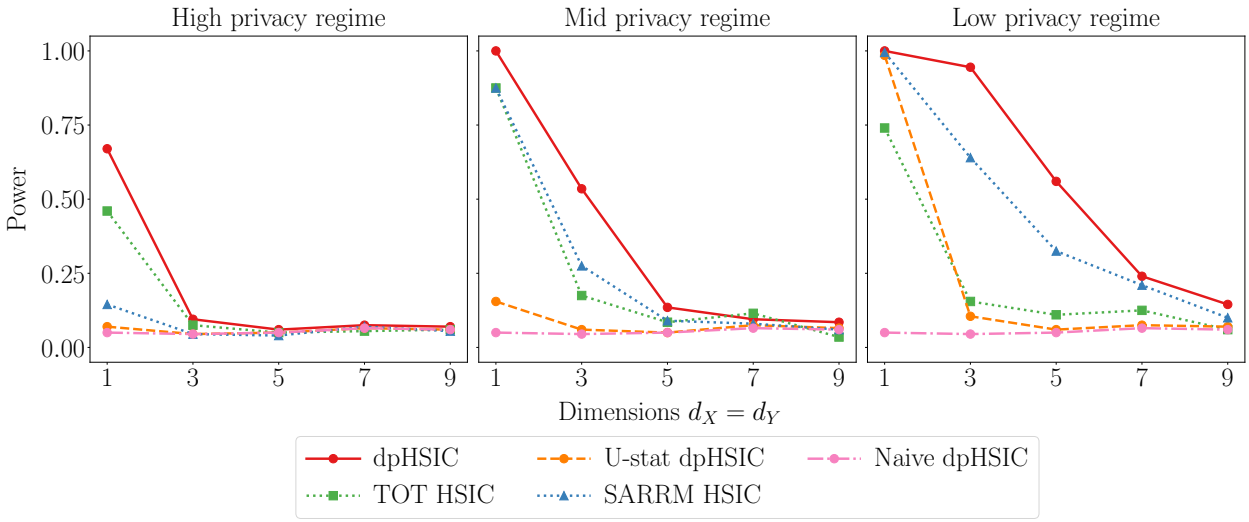

Figure 11: Measuring the dependence on joint perturbed uniform distributions while varying the dimensions $d_X = d_Y$. We set the sample size $n = 3000$, and change other parameters as follows: *(Left)* Privacy level $\varepsilon = 10/\sqrt{n}$, perturbation amplitude $a = 0.4$. *(Middle)* Privacy level $\varepsilon = 1$, perturbation amplitude $a = 0.35$. *(Right)* Privacy level $\varepsilon = \sqrt{n}/10$, perturbation amplitude $a = 0.3$.

C.2 High-Signal & Low-Privacy

We present power results on perturbed uniform two-sample and independence simulations in the low privacy regime with a high signal ($a = 1$) in Figures 12 and 13. This illustrates that TCS-ME, Naive dpMMD and Naive dpHSIC can indeed detect easier alternatives (while being almost powerless in other challenging alternatives) and have power eventually reaching one. We observe that TCS-ME and Naive dpMMD attain similar power, with Naive dpMMD actually being slightly more powerful.

C.3 Level

We verify in Figures 14 to 16 whether the tests are well-calibrated, *i.e.*, their type I error rates are well-controlled at the significance level $\alpha = 0.05$. To this end, we run experiments under the null where the underlying distribution is uniform without any perturbation (*i.e.*, the amplitude parameter $a = 0$ in the settings described in Section 6.2 and Appendix C.1) for both two-sample and independence testing. We also evaluate the type I error rates of private MMD tests by drawing CelebA face images of women (Section 6.3) for both samples (*i.e.*, the corruption parameter $c = 0$). Figures 14 to 16 indicate that all tests are well-calibrated in all experimental settings, except TCS-ME. Indeed, TCS-ME appears to be extremely miscalibrated, for example in dimension $d = 100$ where its level is around $10\alpha = 0.5$ instead of being around $\alpha = 0.05$. We note, nonetheless, that TCS-ME controls the type I error correctly at level α in the one-dimensional case considered in Section 6.2.

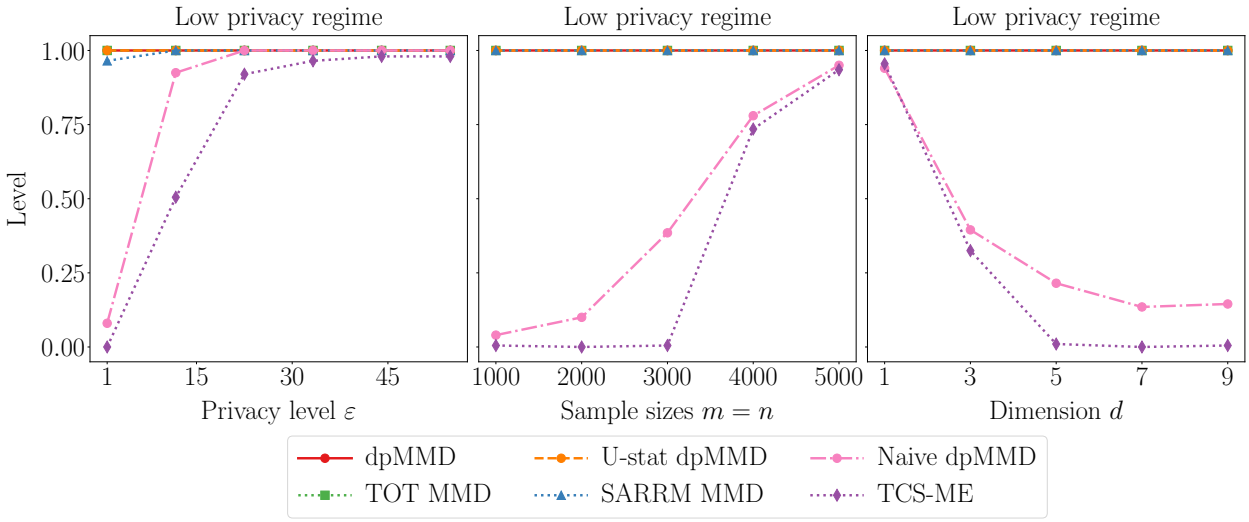

Figure 12: Comparing uniform vs. perturbed uniform in low privacy regimes. We set the perturbation amplitude $a = 1$ and change parameters as follows: *(Left)* Privacy level ε from 1 to $\sqrt{n}$, sample sizes $m = n = 3000$, dimension $d = 1$. *(Middle)* Privacy level $\varepsilon = \sqrt{n}/10$, dimension $d = 1$. *(Right)* Privacy level $\varepsilon = \sqrt{n}/10$, sample sizes $m = n = 5000$.

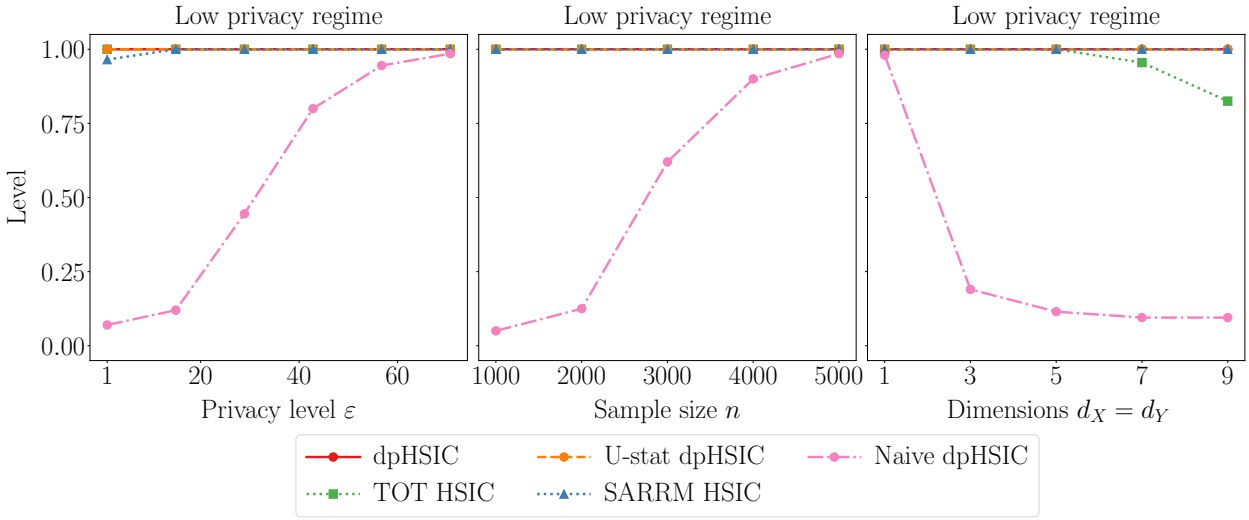

Figure 13: Measuring the dependence on joint perturbed uniform distributions in low privacy regimes. We set the perturbation amplitude $a = 1$ and change parameters as follows: *(Left)* Privacy level ε from 1 to $\sqrt{n}$, sample size $n = 5000$, dimensions $d_X = d_Y = 1$. *(Middle)* Privacy level $\varepsilon = \sqrt{n}$, dimensions $d_X = d_Y = 1$. *(Right)* Privacy level $\varepsilon = \sqrt{n}$, sample size $n = 5000$.

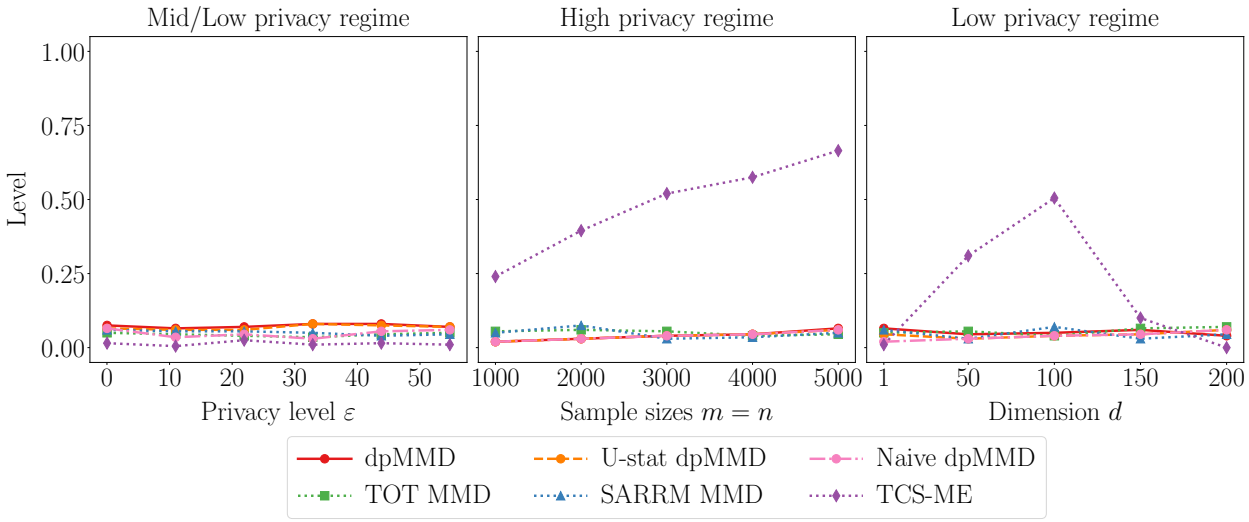

Figure 14: Type I error rates for two-sample testing under the uniform null distribution, *i.e.*, with perturbation amplitude $a = 0$. We vary parameters as follows: *(Left)* Privacy level ε from $1/\sqrt{n}$ to $\sqrt{n}$, sample sizes $m = n = 3000$, dimension $d = 1$. *(Middle)* Privacy level $\varepsilon = 10/\sqrt{n}$, dimension $d = 100$. *(Right)* Privacy level $\varepsilon = \sqrt{n}/10$, sample sizes $m = n = 3000$.

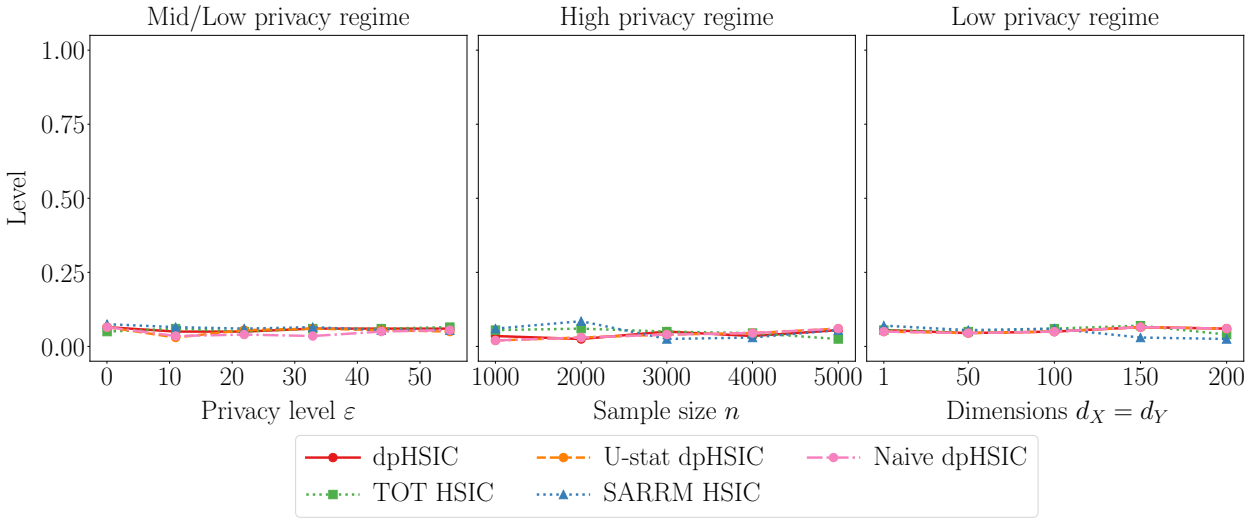

Figure 15: Type I error rates for independence testing under the uniform joint null distribution, *i.e.*, with perturbation amplitude $a = 0$. We vary parameters as follows: *(Left)* Privacy level ε from $1/\sqrt{n}$ to $\sqrt{n}$, sample size $n = 3000$, dimensions $d_X = d_Y = 1$. *(Middle)* Privacy level $\varepsilon = 10/\sqrt{n}$, dimensions $d_X = d_Y = 1$. *(Right)* Privacy level $\varepsilon = \sqrt{n}/10$, sample size $n = 3000$.

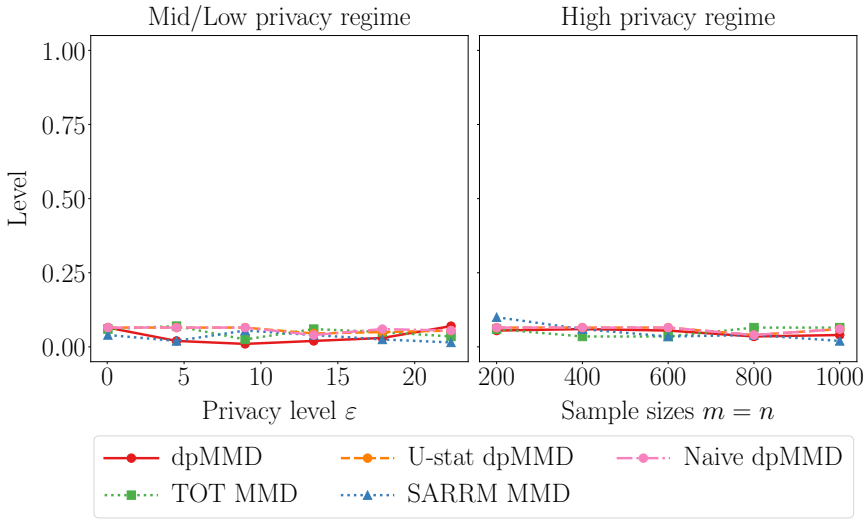

Figure 16: Type I error rates for two-sample testing for CelebA women images with zero-corruption ($c = 0$). We vary parameters as follows: *(Left)* Privacy level ε from $1/\sqrt{n}$ to $\sqrt{n}$, sample sizes $m = n = 500$. *(Right)* Privacy level $\varepsilon = 10/\sqrt{n}$.

D Alternative private tests

We first explain how standard MMD and HSIC permutation tests can be privatized using the methods of Kazan et al. (2023) in Appendix D.1 (TOT) and of Peña and Barrientos (2022) in Appendix D.2 (SARRM), both relying on the subsample-and-aggregate procedure of Canonne et al. (2019). We then present the approach of Raj et al. (2020) to constructing a private MMD test in Appendix D.3 (TCS-ME). We provide our own implementation of TOT and SARRM in JAX, and use the implemetation of Raj et al. (2020) directly for TCS-ME.

D.1 TOT: Test of Tests

Kazan et al. (2023) propose a procedure called 'Test of Tests' which allows us to privatize any existing test. The resulting test, which we refer to as TOT, is $(\varepsilon, 0)$-DP and has well-calibrated level (Kazan et al., 2023, Theorems 3.1 and 3.2). The test relies on the Truncated-Uniform-Laplace distribution Tulap(b) (Awan and Slavković, 2018, Definition 4.1) for positive real scale parameter b. It is defined as the distribution with the following cumulative distribution function

$$F_b^{\mathrm{Tulap}}(x) = \begin{cases} (1+b)^{-1}\, b^{-[x]_{\mathrm{near}}} \Big(b + \big(x - [x]_{\mathrm{near}} + \tfrac{1}{2}\big)(1-b) \Big) & \text{if } x \leq 0, \\ 1 - (1+b)^{-1}\, b^{[x]_{\mathrm{near}}} \Big(b + \big([x]_{\mathrm{near}} - x + \tfrac{1}{2}\big)(1-b) \Big) & \text{if } x > 0. \end{cases} \tag{22}$$

We summarize the procedure of the Test of Tests (Kazan et al., 2023, Algorithm 1) tailored to our permutation setting in Algorithm 2, and remark that their private test does not depend on the global sensitivity of the test statistic.

Algorithm 2 Permutation Test of Tests TOT (Kazan et al., 2023)

Input: Data $\mathcal{X}_n$, significance level $\alpha \in (0,1)$, privacy level $\varepsilon > 0$, test statistic function $T(\cdot)$, number of permutations $B \in \mathbb{N}$.

Procedure:
Choose α_0 (sub-test level) and S (number of subsets) according to Procedure 1.
Partition $\mathcal{X}_n$ into S disjoint subsets $\mathcal{X}_{n,1}, \ldots, \mathcal{X}_{n,S}$ (of equal sizes if possible)
for $s \in [S]$ **do**
  for $b \in [B]$ **do**
    Generate a random permutation $\boldsymbol{\pi}_{b,s}$ of $[|\mathcal{X}_{n,s}|]$.
    Compute $T(\mathcal{X}_{n,s}^{\boldsymbol{\pi}_{b,s}})$.
  end for
  Compute $T(\mathcal{X}_{n,s})$.
  Compute $\widehat{p}_s$ the permutation p-value for the subset $\mathcal{X}_{n,s}$ as in (3).
end for
Compute the number of rejects $a = |\{s : \widehat{p}_s < \alpha_0\}|$.
Privatize a using Truncated-Uniform-Laplace noise (Awan and Slavković, 2018, Algorithm 2):
  Generate random quantities $g_1, g_2 \sim \text{Geom}(1 - e^{-\varepsilon})$ and $u \sim \text{Unif}(-0.5, 0.5)$.
  Compute $z = a + g_1 - g_2 + u$.
Compute the p-value (Awan and Slavković, 2018, Algorithm 1):
  Compute $\widehat{p} = \sum_{s=0}^{S} \binom{S}{s} \alpha_0^s (1-\alpha_0)^{S-s} F_{e^{-\varepsilon}}^{\text{Tulap}}(s - z)$ with $F_{e^{-\varepsilon}}^{\text{Tulap}}$ as in (22).

Output: Reject H_0 if $\widehat{p} \leq \alpha$.

Procedure 1 (Kazan et al. 2023, Section 3.4)**.** *To the best of our understanding, in the non-parametric MMD and HSIC settings considered, the power of the permutation tests do not admit closed form expressions and, hence, it is not possible to perform the optimization procedure proposed by Kazan et al. (2023, Section 3.4) to select the optimal combination of the number of subsets S and of the sub-test significance threshold α_0. The authors point out that* 'the optimization tends to favor high values of S, with very small subsamples and high significance thresholds α_0 on the subtests.' *We verified these observations empirically and have found on separate data that setting the values to $S = \sqrt{n}$ and $\alpha_0 = 5\alpha$ consistently provide high power, we use these heuristics in our implementation.*

Note that the privatized statistic z in Algorithm 2 is the sum of two terms: "a" which under the null follows a Binomial(S, α_0) distribution, and a noise quantity which follows a Tulap$(e^{-\varepsilon})$ distribution. Hence, the p-value is $\mathbb{P}(B + N \geq z)$ where $B \sim \text{Binomial}(S, \alpha_0)$ and $N \sim \text{Tulap}(e^{-\varepsilon})$, and it can be estimated as in the last step of Algorithm 2 (Awan and Slavković, 2018, Algorithm 1).

While TOT is guaranteed to be $(\varepsilon, 0)$-DP and able to maintain a significance level α for any number of subsets S for the data partitioning, any sub-test significance level $\alpha_0 \in (0,1)$, and any sample size n (which is not the case of SARRM introduced below), this comes at the cost of having

to rely on heuristics for α_0 and S in the non-parametric testing setting (as explained in Procedure 1).

D.2 SARRM: Subsampled and Aggregated Randomized Response Mechanism

The privatization procedure of Peña and Barrientos (2022) is similar to the one of Kazan et al. (2023): both methods split the data into subsets on which the non-private test is run. Kazan et al. (2023) then runs the optimal binomial test of Awan and Slavković (2018), while Peña and Barrientos (2022) relies on a binomial test based on a randomized response mechanism.

For binary input $x \in \{0, 1\}$ and $p \in [0, 1]$, the randomized response mechanism is defined as

$$r_p(x) \coloneqq \begin{cases} x & \text{with probability } p, \\ 1 - x & \text{with probability } 1 - p. \end{cases} \tag{23}$$

We present the Subsampled and Aggregated Randomized Response Mechanism (SARRM) Test of Peña and Barrientos (2022) in Algorithm 3. Again, unlike our procedure, we point out that this test does not rely on the global sensitivity of the test statistic.

Algorithm 3 Permutation SARRM Test (Peña and Barrientos, 2022)

Input: Data $\mathcal{X}_n$, significance level $\alpha \in (0, 1)$, privacy level $\varepsilon > 0$, test statistic function $T(\cdot)$, number of permutations $B \in \mathbb{N}$.

Procedure:
Compute $p = e^\varepsilon/(1 + e^\varepsilon)$.
Choose α_0 (sub-test level) and k (determining the number of subsets) according to Procedure 2.
Partition $\mathcal{X}_n$ into $S = 2k + 1$ disjoint subsets $\mathcal{X}_{n,1}, \ldots, \mathcal{X}_{n,S}$ (of equal sizes if possible)
for $s \in [S]$ **do**
  for $b \in [B]$ **do**
    Generate a random permutation $\boldsymbol{\pi}_{b,s}$ of $[|\mathcal{X}_{n,s}|]$.
    Compute $T(\mathcal{X}_{n,s}^{\boldsymbol{\pi}_{b,s}})$.
  end for
  Compute $T(\mathcal{X}_{n,s})$.
  Compute $\widehat{p}_s$ the permutation p-value for the subset $\mathcal{X}_{n,s}$ as in (3).
  Compute sub-test output $t_s = \mathbb{1}(\widehat{p}_s \leq \alpha_0)$.
end for
Compute statistic $T = \sum_{s=1}^{S} r_p(t_s)$ with r_p as in (23).

Output: Reject H_0 if $T > k$.

Recall the meaning of the test parameters of SARRM: k determines the number of subsets $2k+1$, p is the randomized response mechanism probability, α_0 is the sub-test significance level, α is the significance level, and ε is the privacy level. Peña and Barrientos (2022, Proposition 2) show that SARRM is $(\varepsilon, 0)$-DP with

$$\varepsilon = \log\big(\mathbb{P}(B_1 > k)/\mathbb{P}(B_0 > k)\big) \tag{24}$$

where $B_0 \sim \text{Binomial}(2k+1, 1-p)$ and $B_1 \sim \text{Bernoulli}(p)+\text{Binomial}(2k, 1-p)$. Peña and Barrientos (2022, Proposition 7) use the fact that the test statistic $T = \sum_{s=1}^{2k+1} r_p(t_s)$ defined in Algorithm 3 follows a $\text{Binomial}(2k+1, q_{p,\alpha_0})$ distribution under the null, where $q_{p,\alpha_0} := p\alpha_0 + (1-p)(1-\alpha_0)$, from which it can be deduced that SARRM has significance level α provided that

$$\text{level}(k, p, \alpha_0) := \sum_{\ell=k+1}^{2k+1} \binom{2k+1}{\ell} q_{p,\alpha_0}^{\ell} (1 - q_{p,\alpha_0})^{2k+1-\ell} \leq \alpha. \tag{25}$$

Procedure 2 (Peña and Barrientos 2022, Section 3.3)**.** *The aim is to determine the parameters* $k \in \mathbb{N}\setminus\{0\}$, $p \in (0.5, 1)$ *and* $\alpha_0 \geq \alpha_{0,\min}$ *for* $\alpha_{0,\min} = 0.0025$ *(user-specified), such that SARRM achieves the given privacy level* ε *and significance level* α*. To start, set the value of* k *to 1. The parameter* p *can then be chosen according to (24) to guarantee the* $(\varepsilon, 0)$*-DP property. If* $\text{level}(k, p, \alpha_{0,\min}) \leq \alpha$*, then there exists some* $\alpha_0 \geq \alpha_{0,\min}$ *such that* $\text{level}(k, p, \alpha_0) = \alpha$ *which guarantees test level* α*, all required conditions are then satisfied, and the values* k*,* p *and* α_0 *are used for SARRM. If* $\text{level}(k, p, \alpha_{0,\min}) > \alpha$*, then level* α *cannot be attained, the value* $k+1$ *is then considered instead of* k*, and the procedure is repeated.*[2] *The procedure is guaranteed to terminate. If an upper bound on* k *is known (as in our experiments since the sample size is fixed), then the search of the parameter* k *can be performed using a bisection method which runs considerably faster.*

Note that, in order to achieve a given privacy level ε and significance level α, a certain number of subsets $2k+1$ (determined by Procedure 2) must be used. This is only possible if the sample size is greater than the number of subsets, that is, if $n \geq 2k+1$. If this is not the case, then SARRM simply cannot be run with that privacy level, significance level and sample size, which is a considerable limitation of SARRM.

D.3 TCS-ME: Trusted-Curator, perturbed Statistic, Mean Embedding

Raj et al. (2020) propose several variants of a DP kernel two-sample test:

- considering both the trusted-curator and the no-trusted-entity privacy settings,
- relying on the analytic Gaussian mechanism (Balle and Wang, 2018) to inject privacy noise either in the statistic or in the means and covariances of feature vectors,
- using the two kernel statistics of Jitkrittum et al. (2016) based on mean embeddings and on smooth characteristic functions,
- either optimizing the test locations and bandwidth, or sampling locations and using the median heuristic bandwidth,
- using the asymptotic χ^2 null distribution, or an approximation to the null distribution.

We compare against the most powerful of these tests (as can be seen in Raj et al. 2020, Figure 2), which they refer to as TCS-ME. This is the test based on the Trusted-Curator (TC) privacy

[2] Following this procedure, we are able to replicate the results of Peña and Barrientos (2022, Table 1).

setting (which is coherent with the privacy framework considered in for dpMMD), which directly perturbs the test Statistic (S) based on the kernel Mean Embeddings (ME), and which uses 20% of the data to optimize the test locations and kernel bandwidth for maximizing power, and which approximates the null.

TCS-ME is guaranteed to be (ε, δ)-DP for $\delta > 0$ as it relies on the (analytic) Gaussian mechanism. In our simulations, we set $\delta = 10^{-5}$ as in the experiments of Raj et al. (2020), and compare against $(\varepsilon, 0)$-DP tests (a slightly more restrictive constraint).

As can be seen in Appendix C.3, while TCS-ME controls the type I error at level α in the one-dimensional case (which is the setting considered Section 6.2 where we compared against TCS-ME), we observe that it can be extremely poorly calibrated (type I error up to 10α) when working in higher dimensions.

E Proofs for the Main Text

This section collects the proofs of the results provided in the main text.

E.1 Proof of Theorem 1

When $\mathcal{X}_n = (X_1, \ldots, X_n)$ are exchangeable, $M_0, \ldots, M_B$ are exchangeable by construction. Moreover, since $M_0, M_1, \ldots, M_B$ are distinct with probability one due to i.i.d. Laplace noises, the second result of Lemma 15 proves the equality.

E.2 Proof of Theorem 2

Let us denote the $1-\alpha$ quantile of $\{M_i\}_{i=0}^B$ and the $1-\alpha_\star$ quantile of $\{M_i\}_{i=1}^B$ by $q_{1-\alpha}$ and $r_{1-\alpha_\star}$, respectively, where

$$\alpha_\star = \max\left\{\left(\frac{B+1}{B}\alpha - \frac{1}{B}\right), 0\right\} \in [0, 1).$$

We then claim that

$$\mathbb{1}(\widehat{p}_{\mathrm{dp}} \leq \alpha) \overset{\text{(i)}}{=} \mathbb{1}(M_0 > q_{1-\alpha}) \overset{\text{(ii)}}{=} \mathbb{1}(M_0 > r_{1-\alpha_\star})\mathbb{1}\left(\alpha \geq \frac{1}{B+1}\right),$$

where identity (i) holds by the quantile representation of the permutation test in Lemma 17, and identity (ii) holds by Lemma 18. Moreover, Lemma 19 proves that $r_{1-\alpha_\star}$ has the global sensitivity at most Δ_T, and the triangle inequality yields that $T_0 - r_{1-\alpha_\star}$ has the global sensitivity at most $2\Delta_T$ under condition (4). Therefore the Laplace mechanism ensures that $M_0 - r_{1-\alpha_\star} = T_0 - r_{1-\alpha_\star} + 2\Delta_T \xi_{\varepsilon,\delta}^{-1} \zeta_0$ is (ε, δ)-DP. Finally, noting that the test function $\mathbb{1}(\widehat{p}_{\mathrm{dp}} \leq \alpha) = \mathbb{1}(M_0 - r_{1-\alpha_\star} > 0)\,\mathbb{1}(\alpha \geq 1/(B+1))$ is a function of (ε, δ)-DP statistic, the post processing property of differential privacy (Lemma 1) proves that the given permutation test is differentially private at privacy levels ε and δ. This completes the proof of Theorem 1.

E.3 Proof of Theorem 3

The result of Theorem 3 follows as a corollary of Lemma 8. In more detail, let $\mathcal{G}$ be the sigma field generated by $\{\mathcal{X}_n, \zeta_0\}$. Conditional on $\mathcal{G}$, observe that $\{M_1, \ldots, M_B\}$ are i.i.d. where the randomness arises from $\{\boldsymbol{\pi}_i, \zeta_i\}_{i=1}^B$, and M_0 is constant. Therefore, the proposed private test is pointwise consistent as the conditions in Lemma 8 are satisfied.

E.4 Proof of Theorem 4

By the quantile representation of the permutation test (Lemma 17), the type II error can be written as

$$\mathbb{P}_P(\widehat{p}_{\mathrm{dp}} > \alpha) = \mathbb{P}_P\big(M_0 \leq q_{1-\alpha,B}\big),$$

where $q_{1-\alpha,B}$ is the $1-\alpha$ quantile of $M_0, M_1, \ldots, M_B$. Let us split the proof into several steps:

- In the first step, we present an upper bound for $q_{1-\alpha,B}$ holding with high probability. In particular, the following holds

$$q_{1-\alpha,B} \leq \mathbb{E}_P[T(\mathcal{X}_n^{\boldsymbol{\pi}})] + 12\sqrt{\frac{\mathrm{Var}_P[T(\mathcal{X}_n^{\boldsymbol{\pi}})]}{\alpha\beta}} + 2\Delta_T \xi_{\varepsilon,\delta}^{-1} F_\zeta^{-1}(1-\alpha/4)$$

  with high probability (say $1-3\beta/4$).

- In the second step, we use the above quantile bound to show that control of the type II error is guaranteed under the condition in (8).

For simplicity, we omit the subscript P in the expectation, variance and probability operators throughout the proof.

Step 1 (Bounding the quantile). We first focus on the quantile $q_{1-\alpha,B}$ and present a high probability upper bound. Recall that we define M_i as $T_i + 2\Delta_T \xi_{\varepsilon,\delta}^{-1}\zeta_i$ and let $q^a_{1-\alpha/2,B}$ and $q^b_{1-\alpha/2,B}$ be the $1-\alpha/2$ quantiles of $\{T_0, T_1, \ldots, T_B\}$ and of $\{2\Delta_T\xi_{\varepsilon,\delta}^{-1}\zeta_0, \ldots, 2\Delta_T\xi_{\varepsilon,\delta}^{-1}\zeta_B\}$, respectively. Denote by $q^a_{1-\alpha/2,\infty}$ and $q^b_{1-\alpha/2,\infty}$ the corresponding $1-\alpha/2$ quantiles with $B=\infty$. In particular, we define $q^a_{1-\alpha,\infty}$ and $q^b_{1-\alpha,\infty}$ as

$$q^a_{1-\alpha,\infty} \coloneqq \inf\left\{x \in \mathbb{R} : \frac{1}{|\boldsymbol{\Pi}_n|}\sum_{\boldsymbol{\pi}\in\boldsymbol{\Pi}_n} \mathbb{1}\big(T(\mathcal{X}_n^{\boldsymbol{\pi}}) \leq x\big) \geq 1-\alpha\right\} \quad \text{and}$$

$$q^b_{1-\alpha,\infty} \coloneqq 2\Delta_T \xi_{\varepsilon,\delta}^{-1} F_\zeta^{-1}(1-\alpha),$$

where $|\boldsymbol{\Pi}_n|$ denotes the cardinality of the set $\boldsymbol{\Pi}_n$. Intuitively, when B is large, $q^a_{1-\alpha,B}$ (resp. $q^b_{1-\alpha,B}$) becomes close to $q^a_{1-\alpha,\infty}$ (resp. $q^b_{1-\alpha,\infty}$) with high probability. As noted in Kim et al. (2022a) and Schrab et al. (2023), their closeness can be precisely captured by the Dvoretzky–Kiefer–Wolfowitz

(DKW) inequality (Massart, 1990). More specifically, the DKW inequality ensures that the following event

$$E_1 \coloneqq \left\{ \sup_{x\in\mathbb{R}} \left| \frac{1}{B}\sum_{i=1}^{B} \mathbb{1}\big(T_i \le x\big) - \frac{1}{|\mathbf{\Pi}_n|}\sum_{\boldsymbol{\pi}\in\mathbf{\Pi}_n} \mathbb{1}\big(T(\mathcal{X}_n^{\boldsymbol{\pi}}) \le x\big) \right| \le \sqrt{\frac{1}{2B}\log\left(\frac{8}{\beta}\right)} \right\}$$

holds with probability at least $1-\beta/4$. Similarly, define the event E_2 as

$$E_2 \coloneqq \left\{ \sup_{x\in\mathbb{R}} \left| \frac{1}{B+1}\sum_{i=0}^{B} \mathbb{1}\big(\zeta_i \le x\big) - \mathbb{P}(\zeta \le x) \right| \le \sqrt{\frac{1}{2(B+1)}\log\left(\frac{8}{\beta}\right)} \right\},$$

which holds with probability at least $1-\beta/4$. Under the event E_1, we see that $q^a_{1-\alpha/2,B}$ is bounded as

$$\begin{aligned}
q^a_{1-\alpha/2,B} &= \inf\left\{ x\in\mathbb{R} : \frac{1}{B+1}\sum_{i=0}^{B} \mathbb{1}\big(T_i \le x\big) \ge 1-\frac{\alpha}{2} \right\} \\
&= \inf\left\{ x\in\mathbb{R} : \frac{1}{B}\sum_{i=1}^{B} \mathbb{1}\big(T_i \le x\big) \ge \frac{B+1}{B}\left(1-\frac{\alpha}{2}\right) \right\} \\
&\le \inf\left\{ x\in\mathbb{R} : \frac{1}{|\mathbf{\Pi}_n|}\sum_{\boldsymbol{\pi}\in\mathbf{\Pi}_n} \mathbb{1}\big(T(\mathcal{X}_n^{\boldsymbol{\pi}}) \le x\big) \ge 1-\frac{\alpha}{4} \right\} \\
&= q^a_{1-\alpha/4,\infty},
\end{aligned}$$

where the inequality holds when

$$\frac{B+1}{B}\left(1-\frac{\alpha}{2}\right) + \sqrt{\frac{1}{2B}\log\left(\frac{8}{\beta}\right)} \le 1-\frac{\alpha}{4}.$$

This inequality holds for $B \ge \frac{16}{\alpha^2}\times\big(\frac{1}{2}\log(\frac{8}{\beta}) + \frac{\alpha}{2}(1-\frac{\alpha}{2})\big)$, which is further implied by the condition $B \ge \frac{16}{\alpha^2}\log(\frac{8}{\beta})$. In more detail, we can obtain the last condition for B by noting that $\frac{\alpha}{2}(1-\frac{\alpha}{2}) \le \frac{1}{4} < \frac{\log(8)}{2} \le \frac{1}{2}\log(\frac{8}{\beta})$ for all $\alpha,\beta\in(0,1)$. Therefore, under the condition for B in the theorem statement, it holds that $q^a_{1-\alpha/2,B} \le q^a_{1-\alpha/4,\infty}$. Similarly, under the event E_2, the quantile $q^b_{1-\alpha/2,B}$ is bounded as

$$\begin{aligned}
q^b_{1-\alpha/2,B} &= \inf\left\{ x\in\mathbb{R} : \frac{1}{B+1}\sum_{i=0}^{B} \mathbb{1}\big(2\Delta_T\xi_{\varepsilon,\delta}^{-1}\zeta_i \le x\big) \ge 1-\alpha/2 \right\} \\
&\le \inf\left\{ x\in\mathbb{R} : \mathbb{P}\big(2\Delta_T\xi_{\varepsilon,\delta}^{-1}\zeta \le x\big) \ge 1-\alpha/2 + \sqrt{\frac{1}{2(B+1)}\log\left(\frac{8}{\beta}\right)} \right\} \\
&\le \inf\left\{ x\in\mathbb{R} : \mathbb{P}\big(2\Delta_T\xi_{\varepsilon,\delta}^{-1}\zeta \le x\big) \ge 1-\alpha/4 \right\}
\end{aligned}$$

$$= \; q^b_{1-\alpha/4,\infty},$$

where the last inequality holds when

$$1 - \frac{\alpha}{2} + \sqrt{\frac{1}{2(B+1)} \log\left(\frac{8}{\beta}\right)} \le 1 - \frac{\alpha}{4}.$$

Again, the above inequality is implied by our condition for B in the theorem statement. Note further that Lemma 20 gives the inequality that $q_{1-\alpha,B} \le q^a_{1-\alpha/2,B} + q^b_{1-\alpha/2,B}$. Therefore under the events E_1 and E_2 with the condition for B, it holds that $q_{1-\alpha,B} \le q^a_{1-\alpha/4,\infty} + q^b_{1-\alpha/4,\infty}$.

Next we further upper bound $q^a_{1-\alpha/4,\infty}$ and $q^b_{1-\alpha/4,\infty}$ by more manageable terms. To begin with $q^a_{1-\alpha/4,\infty}$, Chebyshev's inequality conditional on $\mathcal{X}_n$ yields

$$\mathbb{P}_{\boldsymbol{\pi}}\big\{T(\mathcal{X}_n^{\boldsymbol{\pi}}) \ge \mathbb{E}_{\boldsymbol{\pi}}[T(\mathcal{X}_n^{\boldsymbol{\pi}}) \,|\, \mathcal{X}_n] + t \,|\, \mathcal{X}_n\big\} \le \frac{\mathrm{Var}_{\boldsymbol{\pi}}[T(\mathcal{X}_n^{\boldsymbol{\pi}}) \,|\, \mathcal{X}_n]}{t^2} \quad \text{for any } t > 0.$$

Consequently, $q^a_{1-\alpha/4,\infty}$ is bounded as

$$q^a_{1-\alpha/4,\infty} \le \sqrt{\frac{4\mathrm{Var}_{\boldsymbol{\pi}}[T(\mathcal{X}_n^{\boldsymbol{\pi}}) \,|\, \mathcal{X}_n]}{\alpha}} + \mathbb{E}_{\boldsymbol{\pi}}[T(\mathcal{X}_n^{\boldsymbol{\pi}}) \,|\, \mathcal{X}_n].$$

Now define other events E_3 and E_4,

$$E_3 := \Big\{\mathrm{Var}_{\boldsymbol{\pi}}[T(\mathcal{X}_n^{\boldsymbol{\pi}}) \,|\, \mathcal{X}_n] < 8\beta^{-1}\mathbb{E}\big(\mathrm{Var}_{\boldsymbol{\pi}}[T(\mathcal{X}_n^{\boldsymbol{\pi}}) \,|\, \mathcal{X}_n]\big)\Big\},$$

$$E_4 := \Big\{\big|\mathbb{E}_{\boldsymbol{\pi}}[T(\mathcal{X}_n^{\boldsymbol{\pi}}) \,|\, \mathcal{X}_n] - \mathbb{E}[T(\mathcal{X}_n^{\boldsymbol{\pi}})]\big| < \sqrt{8\beta^{-1}\mathrm{Var}\big(\mathbb{E}_{\boldsymbol{\pi}}[T(\mathcal{X}_n^{\boldsymbol{\pi}}) \,|\, \mathcal{X}_n]\big)}\Big\},$$

where each of the events holds with probability at least $1 - \beta/8$ by Markov's inequality and Chebyshev's inequality, respectively. We emphasize that the expectation $\mathbb{E}[T(\mathcal{X}_n^{\boldsymbol{\pi}})]$ is taken with respect to both $\mathcal{X}_n$ and $\boldsymbol{\pi}$. Under these events E_3 and E_4, it can be seen that

$$\begin{aligned} q^a_{1-\alpha/4,\infty} \;&\le\; \mathbb{E}[T(\mathcal{X}_n^{\boldsymbol{\pi}})] + \sqrt{\frac{36\mathbb{E}\big(\mathrm{Var}_{\boldsymbol{\pi}}[T(\mathcal{X}_n^{\boldsymbol{\pi}}) \,|\, \mathcal{X}_n]\big)}{\alpha\beta}} + \sqrt{\frac{8\mathrm{Var}\big(\mathbb{E}_{\boldsymbol{\pi}}[T(\mathcal{X}_n^{\boldsymbol{\pi}}) \,|\, \mathcal{X}_n]\big)}{\beta}} \\ &\le\; \mathbb{E}[T(\mathcal{X}_n^{\boldsymbol{\pi}})] + 12\sqrt{\frac{\mathrm{Var}[T(\mathcal{X}_n^{\boldsymbol{\pi}})]}{\alpha\beta}} \end{aligned}$$

where the second inequality uses the law of total variance. On the other hand, $q^b_{1-\alpha/4,\infty}$ is simply the $1 - \alpha/4$ quantile of $2\Delta_T\xi^{-1}_{\varepsilon,\delta}\zeta$, namely

$$q^b_{1-\alpha/4,\infty} = 2\Delta_T\xi^{-1}_{\varepsilon,\delta}F^{-1}_{\zeta}(1 - \alpha/4).$$

In summary, we have shown that with probability at least $1 - 3\beta/4$,

$$q_{1-\alpha,B} \le \mathbb{E}[T(\mathcal{X}_n^{\boldsymbol{\pi}})] + 12\sqrt{\frac{\mathrm{Var}[T(\mathcal{X}_n^{\boldsymbol{\pi}})]}{\alpha\beta}} + 2\Delta_T\xi^{-1}_{\varepsilon,\delta}F^{-1}_{\zeta}(1 - \alpha/4).$$

Step 2 (Bounding the type II error). Given the upper bound for $q_{1-\alpha,B}$ from the previous step, we now examine the type II error of the differentially private permutation test and show that it is bounded by β under the given condition. In particular, writing

$$R_1 := 12\sqrt{\frac{\mathrm{Var}[T(\mathcal{X}_n^{\boldsymbol{\pi}})]}{\alpha\beta}} + 2\Delta_T \xi_{\varepsilon,\delta}^{-1} F_\zeta^{-1}(1-\alpha/4),$$

$$R_2 := 2\Delta_T \xi_{\varepsilon,\delta}^{-1} F_\zeta^{-1}(1-\beta/8),$$

the result from Step 1 yields

$$\begin{aligned}
\mathbb{P}(M_0 \le q_{1-\alpha,B}) &\le \mathbb{P}\big(M_0 \le \mathbb{E}[T(\mathcal{X}_n^{\boldsymbol{\pi}})] + R_1\big) + \frac{3\beta}{4} \\
&= \mathbb{P}\big(T_0 + 2\Delta_T \xi_{\varepsilon,\delta}^{-1}\zeta_0 \le \mathbb{E}[T(\mathcal{X}_n^{\boldsymbol{\pi}})] + R_1\big) + \frac{3\beta}{4} \\
&\le \mathbb{P}\big(T_0 \le \mathbb{E}[T(\mathcal{X}_n^{\boldsymbol{\pi}})] + R_1 + R_2\big) + \frac{7\beta}{8}, \qquad (26)
\end{aligned}$$

where the last inequality holds since

$$\mathbb{P}\big(-\xi_{\varepsilon,\delta}^{-1}\zeta_0 > F_\zeta^{-1}(1-\beta/8)\big) \le \frac{\beta}{8}.$$

We also remark that for any $\alpha \in (0,1)$ and $\beta \in (0, 1-\alpha)$ (implying that $\min\{\alpha,\beta\} < 1/2$),

$$F_\zeta^{-1}(1-\alpha/4) + F_\zeta^{-1}(1-\beta/8) \le 5\max\{\log(1/\alpha), \log(1/\beta)\}. \qquad (27)$$

In more detail, the quantile function of ζ is given as

$$F_\zeta^{-1}(t) = \begin{cases} \log(2t) < 0, & \text{if } t < 0.5, \\ -\log\big(2(1-t)\big) \ge 0, & \text{if } t \ge 0.5. \end{cases}$$

Based on this expression along with the fact that $\min\{\alpha,\beta\} < 1/2$, we have

$$\begin{aligned}
F_\zeta^{-1}(1-\alpha/4) + F_\zeta^{-1}(1-\beta/8) &\le \log(2/\alpha) + \log(4/\beta) = 3\log(2) + \log(1/\alpha) + \log(1/\beta) \\
&\le 5\max\{\log(1/\alpha), \log(1/\beta)\}.
\end{aligned}$$

Hence, under the given condition of (8), we can ensure that

$$\mathbb{E}[T(\mathcal{X}_n^{\boldsymbol{\pi}})] + R_1 + R_2 \le \mathbb{E}[T(\mathcal{X}_n)] - \sqrt{8\beta^{-1}\mathrm{Var}[T(\mathcal{X}_n)]}.$$

For clarity, we emphasize the notational difference between $\mathbb{E}[T(\mathcal{X}_n)] = \mathbb{E}_P[T(\mathcal{X}_n)]$ and $\mathbb{E}[T(\mathcal{X}_n^{\boldsymbol{\pi}})] = \mathbb{E}_{P,\boldsymbol{\pi}}[T(\mathcal{X}_n^{\boldsymbol{\pi}})]$ where the former expectation is taken over $\mathcal{X}_n$, whereas the latter expectation is taken over both $\mathcal{X}_n$ and $\boldsymbol{\pi}$. This bound together with Chebyshev's inequality yields

$$\mathbb{P}\big(T_0 \le \mathbb{E}[T(\mathcal{X}_n^{\boldsymbol{\pi}})] + R_1 + R_2\big) \le \mathbb{P}\big(T_0 \le \mathbb{E}[T(\mathcal{X}_n)] - \sqrt{8\beta^{-1}\mathrm{Var}[T(\mathcal{X}_n)]}\big)$$

$$
\begin{aligned}
&= \mathbb{P}\big(\sqrt{8\beta^{-1}\mathrm{Var}[T(\mathcal{X}_n)]} \le \mathbb{E}[T(\mathcal{X}_n)] - T_0\big) \\
&\le \frac{\beta}{8}.
\end{aligned}
$$

Now combining the inequality (26) with the above yields that

$$\mathbb{P}(M_0 \le q_{1-\alpha,B}) \le \beta.$$

This bound holds for any $P \in \mathcal{P}_1$ and the upper bound is independent of P. Hence the uniform guarantee stated in Theorem 4 holds.

E.5 Proof of Lemma 5

We start with the proof of the upper bound, followed by the proof of the lower bound.

E.5.1 Upper Bound

To simplify the proof, we may exploit the permutation invariance property of $\widehat{\mathrm{MMD}}(\mathcal{X}^{\boldsymbol{\pi}}_{n+m})$ within $\{X_{\pi_1}, \ldots, X_{\pi_n}\}$ and $\{Y_{\pi_{n+1}}, \ldots, Y_{\pi_{n+m}}\}$, and focus on two specific cases of neighboring dataset:

$$\tilde{\mathcal{X}}^{\boldsymbol{\pi},a}_{n+m} = \{X'_{\pi_1}, \ldots, X_{\pi_n}, Y_{\pi_{n+1}}, \ldots, Y_{\pi_{n+m}}\} \quad \text{and} \quad \tilde{\mathcal{X}}^{\boldsymbol{\pi},b}_{n+m} = \{X_{\pi_1}, \ldots, X_{\pi_n}, Y'_{\pi_{n+1}}, \ldots, Y_{\pi_{n+m}}\},$$

where X'_{π_1} is an i.i.d. copy of X_{π_1} and $Y'_{\pi_{n+1}}$ is an i.i.d. copy of $Y_{\pi_{n+1}}$. The bound for the other neighboring datasets can be proven by following the same lines of the proof.

Starting with the neighboring dataset $\tilde{\mathcal{X}}^{\boldsymbol{\pi},a}_{n+m}$, note that

$$
\begin{aligned}
&\widehat{\mathrm{MMD}}(\mathcal{X}^{\boldsymbol{\pi}}_{n+m}) \\
&= \sup_{f\in\mathcal{F}_k}\left\{\frac{1}{n}\sum_{i=1}^{n} f(X_{\pi_i}) - \frac{1}{m}\sum_{j=1}^{m} f(Y_{\pi_{n+j}})\right\} \\
&= \sup_{f\in\mathcal{F}_k}\left\{\frac{1}{n}\sum_{i=1}^{n} f(X_{\pi_i}) - \frac{1}{m}\sum_{j=1}^{m} f(Y_{\pi_{n+j}}) + \frac{1}{n}f(X'_{\pi_1}) - \frac{1}{n}f(X'_{\pi_1})\right\} \\
&= \sup_{f\in\mathcal{F}_k}\left\{\frac{1}{n}\Big(f(X'_{\pi_1}) + \sum_{i=2}^{n} f(X_{\pi_i})\Big) - \frac{1}{m}\sum_{j=1}^{m} f(Y_{\pi_{n+j}}) + \frac{1}{n}f(X_{\pi_1}) - \frac{1}{n}f(X'_{\pi_1})\right\} \\
&\le \sup_{f\in\mathcal{F}_k}\left\{\frac{1}{n}\Big(f(X'_{\pi_1}) + \sum_{i=2}^{n} f(X_{\pi_i})\Big) - \frac{1}{m}\sum_{j=1}^{m} f(Y_{\pi_{n+j}})\right\} + \sup_{f\in\mathcal{F}_k}\left\{\frac{1}{n}f(X_{\pi_1}) - \frac{1}{n}f(X'_{\pi_1})\right\} \\
&= \widehat{\mathrm{MMD}}(\mathcal{X}^{\boldsymbol{\pi},a}_{n+m}) + \sup_{f\in\mathcal{F}_k}\left\{\frac{1}{n}f(X_{\pi_1}) - \frac{1}{n}f(X'_{\pi_1})\right\},
\end{aligned}
$$

where the inequality uses the triangle inequality. Since $|f(x) - f(y)| = |\langle f, k(x,\cdot) - k(y,\cdot)\rangle_{\mathcal{H}_k}| \le \|f\|_{\mathcal{H}_k}\sqrt{k(x,x) + k(y,y) - 2k(x,y)}$ by the Cauchy–Schwarz inequality as well as the reproducing

kernel property, we obtain

$$\sup_{f\in\mathcal{F}_k}\left\{\frac{1}{n}f(X_{\pi_1})-\frac{1}{n}f(X'_{\pi_1})\right\}\le\frac{1}{n}\sup_{f\in\mathcal{F}_k}|f(X_{\pi_1})-f(X'_{\pi_1})|\le\frac{\sqrt{2K}}{n},$$

where the second inequality uses the fact that $\|f\|_{\mathcal{H}_k}\le 1$ and $0\le k(x,y)\le K$ for all $x,y\in\mathbb{S}$. We therefore observe $\widehat{\mathrm{MMD}}(\mathcal{X}^{\boldsymbol{\pi}}_{n+m})-\widehat{\mathrm{MMD}}(\mathcal{X}^{\boldsymbol{\pi},a}_{n+m})\le\sqrt{2K}/n$. Since the same argument holds with the roles of $\mathcal{X}^{\boldsymbol{\pi}}_{n+m}$ and $\mathcal{X}^{\boldsymbol{\pi},a}_{n+m}$ reversed, we conclude that

$$\left|\widehat{\mathrm{MMD}}(\mathcal{X}^{\boldsymbol{\pi}}_{n+m})-\widehat{\mathrm{MMD}}(\mathcal{X}^{\boldsymbol{\pi},a}_{n+m})\right|\le\frac{\sqrt{2K}}{n}.$$

Next, for the dataset $\tilde{\mathcal{X}}^{\boldsymbol{\pi},b}_{n+m}$, we can use essentially the same argument to show that

$$\left|\widehat{\mathrm{MMD}}(\mathcal{X}^{\boldsymbol{\pi}}_{n+m})-\widehat{\mathrm{MMD}}(\mathcal{X}^{\boldsymbol{\pi},b}_{n+m})\right|\le\frac{\sqrt{2K}}{m}.$$

Since $n\le m$, the sensitivity is bounded by $\sqrt{2K}/n$ for any neighboring datasets and the upper bound is independent of $\boldsymbol{\pi}$. Therefore it holds that

$$\sup_{\boldsymbol{\pi}\in\boldsymbol{\Pi}_{n+m}}\sup_{\substack{\mathcal{X}_{n+m},\tilde{\mathcal{X}}_{n+m}:\\ d_{\mathrm{ham}}(\mathcal{X}_{n+m},\tilde{\mathcal{X}}_{n+m})\le 1}}\left|\widehat{\mathrm{MMD}}(\mathcal{X}^{\boldsymbol{\pi}}_{n+m})-\widehat{\mathrm{MMD}}(\tilde{\mathcal{X}}^{\boldsymbol{\pi}}_{n+m})\right|\le\frac{\sqrt{2K}}{n},$$

which in turn proves the first claim in Lemma 5.

E.5.2 Lower Bound

We now further assume that the kernel k is translation invariant, and has non-empty level sets in $\mathbb{S}$. In this case, we show that the inequality becomes an equality by considering appropriate sets $\mathcal{X}_{n+m}$ and $\tilde{\mathcal{X}}_{n+m}$. More specifically, for a translation invariant kernel k and a constant $\epsilon>0$, consider values x_a and x_b in $\mathbb{S}$ such that $k(x_a,x_b)=\epsilon_\star\le\epsilon$, which is guaranteed by our assumption that k has non-empty level sets. Let $Y_1=\ldots=Y_n=Z_1=\ldots=Z_m=x_a$. In this scenario, the empirical MMD becomes zero for any permutation $\boldsymbol{\pi}\in\boldsymbol{\Pi}_{n+m}$. Let $Y'_1=x_b$ so that $k(Y_1,Y'_1)=\epsilon_\star$. Then based on the closed form expression of the empirical MMD (10), we have

$$\begin{aligned}\widehat{\mathrm{MMD}}^2(\tilde{\mathcal{X}}^{\boldsymbol{\pi}}_{n+m}) &= \frac{(n-1)^2+1}{n^2}K+\frac{2(n-1)}{n^2}\epsilon_\star+K-\frac{2m(n-1)}{nm}K-\frac{2m}{nm}\epsilon_\star\\ &= \frac{2K}{n^2}-\frac{2\epsilon_\star}{n^2}.\end{aligned}$$

Therefore

$$\sup_{\boldsymbol{\pi}\in\boldsymbol{\Pi}_{n+m}}\sup_{\substack{\mathcal{X}_{n+m},\tilde{\mathcal{X}}_{n+m}:\\ d_{\mathrm{ham}}(\mathcal{X}_{n+m},\tilde{\mathcal{X}}_{n+m})\le 1}}\left|\widehat{\mathrm{MMD}}(\mathcal{X}^{\boldsymbol{\pi}}_{n+m})-\widehat{\mathrm{MMD}}(\tilde{\mathcal{X}}^{\boldsymbol{\pi}}_{n+m})\right|\ge\frac{\sqrt{2(K-\epsilon_\star)}}{n}.$$

Since ϵ is an arbitrary positive number and $\epsilon_\star\le\epsilon$, we may take $\epsilon\to 0$ and thus the lower bound becomes $\sqrt{2K}/n$, indicating that the upper bound $\sqrt{2K}/n$ is tight under the given conditions. This proves the second claim in Lemma 5.

E.6 Proof of Theorem 5

Let us prove the three claims made in Theorem 5 below.

- *Proof of Differential Privacy.* This result follows by Theorem 2 along with Lemma 5.
- *Proof of Validity.* This result follows by Theorem 1.
- *Proof of Consistency.* Having Lemma 8 in place, the only condition we need to verify for consistency is $\lim_{n\to\infty} \mathbb{P}(W_{0,n} \leq W_{1,n}) = 0$. Let $\mathcal{G}$ be the sigma field generated by $\{\zeta_0, \mathcal{X}_{n+m}\}$, and let

$$W_{i,n} = M_i = \widehat{\mathrm{MMD}}(\mathcal{X}_{n+m}^{\boldsymbol{\pi}_i}) + \frac{2\sqrt{2K}}{n\xi_{\varepsilon,\delta}}\zeta_i, \quad \text{for each } i \in \{0,1\},$$

  where we recall $\xi_{\varepsilon,\delta} = \varepsilon + \log(1/(1-\delta))$. Our goal is to show that $\lim_{n\to\infty} \mathbb{P}(M_0 \leq M_1) = 0$ under the conditions. Since the pair of distributions (P, Q) and kernel k do not vary with n, we consider the kernel bound K to be a fixed constant throughout. Since K is a constant and $(n\xi_{\varepsilon,\delta})^{-1} \to 0$, the scale of the Laplace noise $\frac{2\sqrt{2K}}{n\xi_{\varepsilon,\delta}}$ goes to zero as well. Thus Slutsky's theorem yields $\frac{2\sqrt{2K}}{n\xi_{\varepsilon,\delta}} \times \zeta_i = o_P(1)$ for each $i \in \{0,1\}$. Next, for $i = 0$, Gretton et al. (2012, Theorem 7) indicate that the unpermuted MMD statistic satisfies $\widehat{\mathrm{MMD}}(\mathcal{X}_{n+m}^{\boldsymbol{\pi}_0}) \xrightarrow{p} \mathrm{MMD}_k(P, Q)$ where the symbol $\xrightarrow{p}$ denotes convergence in probability. Hence another application of Slutsky's theorem yields $M_0 \xrightarrow{p} \mathrm{MMD}_k(P, Q)$. On the other hand, we have $\widehat{\mathrm{MMD}}(\mathcal{X}_{n+m}^{\boldsymbol{\pi}_1}) \xrightarrow{p} 0$ by Lemma 11, and again Slutsky's theorem gives $M_1 \xrightarrow{p} 0$. Combining these two results, we further have $M_0 - M_1 \xrightarrow{p} \mathrm{MMD}_k(P, Q)$. To complete the proof, note that $\mathrm{MMD}_k(P, Q) =: \epsilon > 0$ is assumed to be a fixed positive constant and therefore

$$\lim_{n\to\infty} \mathbb{P}(M_0 \leq M_1) \leq \lim_{n\to\infty} \mathbb{P}(|M_0 - M_1| \geq \epsilon) = 0,$$

  by the definition of convergence in probability. This completes the proof of Theorem 5.

E.7 Proof of Lemma 6

We prove the upper bound result and the lower bound result in order.

E.7.1 Upper Bound

We begin by considering the upper bound result of Lemma 6. Let us consider the dataset $\mathcal{X}_n = \{(Y_1, Z_1), \ldots, (Y_n, Z_n)\}$ and its neighboring dataset denoted by

$$\tilde{\mathcal{X}}_n = \{(Y_1', Z_1'), (Y_2, Z_2), \ldots, (Y_n, Z_n)\} \coloneqq \{(\tilde{Y}_1, \tilde{Z}_1), (\tilde{Y}_2, \tilde{Z}_2), \ldots, (\tilde{Y}_n, \tilde{Z}_n)\}$$

where (Y_1', Z_1') is an i.i.d. copy of (X_1, Y_1) independent of everything else. That is, $\mathcal{X}_n$ and $\tilde{\mathcal{X}}_n$ differ in their first component. We only prove the result of Lemma 6 focusing on $\tilde{\mathcal{X}}_n$, and the proof for the other neighboring datasets can be derived similarly due to the symmetric structure of the empirical HSIC.

For a given permutation $\boldsymbol{\pi}$, denote by $\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}}$ the neighboring dataset $\tilde{\mathcal{X}}_n$ whose Z values are permuted based on $\boldsymbol{\pi}$. Let us divide the cases into two: (i) $\pi_1 = 1$ and (ii) $\pi_1 \neq 1$, and provide the proofs separately. For the first case where $\pi_1 = 1$, we write $k(y, \cdot) = \psi_Y(y)$ and $\ell(z, \cdot) = \psi_Z(z)$. We also write the sample mean of $\psi_Y(Y_1), \ldots, \psi_Y(Y_n)$ as $\overline{\psi}_Y$ and the sample mean of $\psi_Z(Z_1), \ldots, \psi_Z(Z_n)$ as $\overline{\psi}_Z$. Similarly write the sample mean of $\psi_Y(\tilde{Y}_1), \ldots, \psi_Y(\tilde{Y}_n)$ as $\tilde{\psi}_Y$ and the sample mean of $\psi_Z(\tilde{Z}_1), \ldots, \psi_Z(\tilde{Z}_n)$ as $\tilde{\psi}_Z$. Then by adding and subtracting the same terms, we can connect $\widehat{\mathrm{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}})$ with $\widehat{\mathrm{HSIC}}(\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}})$ as follows:

$$
\begin{aligned}
&\widehat{\mathrm{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}}) \\
&= \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left\{ \frac{1}{n} \sum_{i=1}^{n} f(Y_i, Z_{\pi_i}) - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} f(Y_i, Z_{\pi_j}) \right\} \\
&= \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left\langle f, \frac{1}{n} \sum_{i=1}^{n} \big\{ \psi_Y(Y_i) - \overline{\psi}_Y \big\} \big\{ \psi_Z(Z_{\pi_i}) - \overline{\psi}_Z \big\} \right\rangle_{\mathcal{H}_{k\otimes\ell}} \\
&= \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left\langle f, \frac{1}{n} \sum_{i=1}^{n} \big\{ \psi_Y(Y_i) - \psi_Y(\tilde{Y}_i) + \psi_Y(\tilde{Y}_i) - \overline{\psi}_Y + \tilde{\psi}_Y - \tilde{\psi}_Y \big\} \big\{ \psi_Z(Z_{\pi_i}) - \overline{\psi}_Z \big\} \right\rangle_{\mathcal{H}_{k\otimes\ell}} \\
&\overset{\text{(i)}_a}{=} \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left\langle f, \frac{1}{n} \sum_{i=1}^{n} \big\{ \psi_Y(Y_i) - \psi_Y(\tilde{Y}_i) + \psi_Y(\tilde{Y}_i) - \tilde{\psi}_Y \big\} \big\{ \psi_Z(Z_{\pi_i}) - \overline{\psi}_Z \big\} \right\rangle_{\mathcal{H}_{k\otimes\ell}} \\
&\leq \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left\langle f, \frac{1}{n} \big\{ \psi_Y(Y_1) - \psi_Y(\tilde{Y}_1) \big\} \big\{ \psi_Z(Z_{\pi_1}) - \overline{\psi}_Z \big\} \right\rangle_{\mathcal{H}_{k\otimes\ell}} \\
&\qquad + \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left\langle f, \frac{1}{n} \sum_{i=1}^{n} \big\{ \psi_Y(\tilde{Y}_i) - \tilde{\psi}_Y \big\} \big\{ \psi_Z(Z_{\pi_i}) - \overline{\psi}_Z \big\} \right\rangle_{\mathcal{H}_{k\otimes\ell}} \\
&= \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left\langle f, \frac{1}{n} \big\{ \psi_Y(Y_1) - \psi_Y(\tilde{Y}_1) \big\} \big\{ \psi_Z(Z_{\pi_1}) - \overline{\psi}_Z \big\} \right\rangle_{\mathcal{H}_{k\otimes\ell}} \\
&\quad + \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left\langle f, \frac{1}{n} \sum_{i=1}^{n} \big\{ \psi_Y(\tilde{Y}_i) - \tilde{\psi}_Y \big\} \big\{ \psi_Z(Z_{\pi_i}) - \psi_Z(\tilde{Z}_{\pi_i}) + \psi_Z(\tilde{Z}_{\pi_i}) - \tilde{\psi}_Z + \tilde{\psi}_Z - \overline{\psi}_Z \big\} \right\rangle_{\mathcal{H}_{k\otimes\ell}} \\
&\overset{\text{(ii)}_a}{\leq} \widehat{\mathrm{HSIC}}(\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}}) + \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left\langle f, \frac{1}{n} \big\{ \psi_Y(Y_1) - \psi_Y(\tilde{Y}_1) \big\} \big\{ \psi_Z(Z_{\pi_1}) - \overline{\psi}_Z \big\} \right\rangle_{\mathcal{H}_{k\otimes\ell}} \\
&\qquad + \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left\langle f, \frac{1}{n} \big\{ \psi_Y(\tilde{Y}_1) - \tilde{\psi}_Y \big\} \big\{ \psi_Z(Z_{\pi_1}) - \psi_Z(\tilde{Z}_{\pi_1}) \big\} \right\rangle_{\mathcal{H}_{k\otimes\ell}},
\end{aligned}
$$

where step $(\mathrm{i})_a$ uses $\frac{1}{n}\sum_{i=1}^n(\overline{\psi}_Y-\tilde{\psi}_Y)\{\psi_Z(Z_{\pi_i})-\overline{\psi}_Z\}=0$, and step $(\mathrm{ii})_a$ follows similarly. Therefore

$$\begin{aligned}\left|\widehat{\mathrm{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}})-\widehat{\mathrm{HSIC}}(\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}})\right| &\leq \sup_{f\in\mathcal{F}_{k\otimes\ell}}\left|\left\langle f,\,\frac{1}{n}\{\psi_Y(Y_1)-\psi_Y(\tilde{Y}_1)\}\{\psi_Z(Z_{\pi_1})-\overline{\psi}_Z\}\right\rangle_{\mathcal{H}_{k\otimes\ell}}\right| \\ &\quad+\sup_{f\in\mathcal{F}_{k\otimes\ell}}\left|\left\langle f,\,\frac{1}{n}\{\psi_Y(\tilde{Y}_1)-\tilde{\psi}_Y\}\{\psi_Z(Z_{\pi_1})-\psi_Z(\tilde{Z}_{\pi_1})\}\right\rangle_{\mathcal{H}_{k\otimes\ell}}\right| \\ &= \mathrm{(I)}+\mathrm{(II)}.\end{aligned}$$

For the term (I), the Cauchy–Schwarz inequality along with the fact $\|f\|_{\mathcal{H}_{k\otimes\ell}}\leq 1$ yields

$$\begin{aligned}\mathrm{(I)} &\leq \left\|\frac{1}{n}\{\psi_Y(Y_1)-\psi_Y(\tilde{Y}_1)\}\{\psi_Z(Z_{\pi_1})-\overline{\psi}_Z\}\right\|_{\mathcal{H}_{k\otimes\ell}} \\ &= \frac{1}{n}\left\|\psi_Y(Y_1)-\psi_Y(\tilde{Y}_1)\right\|_{\mathcal{H}_k}\left\|\psi_Z(Z_{\pi_1})-\overline{\psi}_Z\right\|_{\mathcal{H}_\ell} \\ &\overset{(\mathrm{i})_b}{\leq} \frac{1}{n^2}\sum_{i=1}^n\left\|\psi_Y(Y_1)-\psi_Y(\tilde{Y}_1)\right\|_{\mathcal{H}_k}\left\|\psi_Z(Z_{\pi_1})-\psi_Z(Z_{\pi_i})\right\|_{\mathcal{H}_\ell} \\ &\overset{(\mathrm{ii})_b}{\leq} \frac{n-1}{n^2}\sqrt{2K}\sqrt{2L}.\end{aligned}$$

In the above, step $(\mathrm{i})_b$ uses Jensen's inequality, and step $(\mathrm{ii})_b$ follows since

$$\|\psi_Y(y)-\psi_Y(y')\|_{\mathcal{H}_k}=\sqrt{k(y,y)+k(y',y')-2k(y,y')}\leq\sqrt{2K}\quad\text{and}$$
$$\|\psi_Z(z)-\psi_Z(z')\|_{\mathcal{H}_\ell}=\sqrt{\ell(z,z)+\ell(z',z')-2\ell(z,z')}\leq\sqrt{2L},$$

for any $y,y'\in\mathbb{Y}$ and $z,z'\in\mathbb{Z}$. The second term (II) can be similarly handled and shown to be

$$\mathrm{(II)}\ \leq\ \frac{n-1}{n^2}\sqrt{2K}\sqrt{2L}.$$

Thus it holds that

$$\left|\widehat{\mathrm{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}})-\widehat{\mathrm{HSIC}}(\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}})\right|\leq\frac{4(n-1)\sqrt{KL}}{n^2}.$$

The other case where $\pi_1\neq 1$ can be proven similarly. Without loss of generality, we may assume that $\pi_1=2$. Then by following the same calculations as before, we have

$$\begin{aligned}\left|\widehat{\mathrm{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}})-\widehat{\mathrm{HSIC}}(\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}})\right| &\leq \sup_{f\in\mathcal{F}_{k\otimes\ell}}\left|\left\langle f,\,\frac{1}{n}\{\psi_Y(Y_1)-\psi_Y(\tilde{Y}_1)\}\{\psi_Z(Z_{\pi_1})-\overline{\psi}_Z\}\right\rangle_{\mathcal{H}_{k\otimes\ell}}\right| \\ &\quad+\sup_{f\in\mathcal{F}_{k\otimes\ell}}\left|\left\langle f,\,\frac{1}{n}\{\psi_Y(\tilde{Y}_2)-\tilde{\psi}_Y\}\{\psi_Z(Z_{\pi_2})-\psi_Z(\tilde{Z}_{\pi_2})\}\right\rangle_{\mathcal{H}_{k\otimes\ell}}\right| \\ &\leq \frac{4(n-1)\sqrt{KL}}{n^2},\end{aligned}$$

which completes the proof of the first claim of Lemma 6.

E.7.2 Lower Bound

We now construct an example where the sensitivity becomes $4(n-2.5)n^{-2}\sqrt{KL}$. For a given $\epsilon \in (0, \min\{K, L\})$, we may assume that there exist $y_a, y_b \in \mathbb{Y}$ and $z_a, z_b \in \mathbb{Z}$ such that $k(y_a, y_b) = \epsilon_\star \leq \epsilon$ and $\ell(z_a, z_b) = \epsilon'_\star \leq \epsilon$, since k and ℓ are assumed to have non-empty level sets on $\mathbb{Y}$ and $\mathbb{Z}$. Consider the two datasets

$$\mathcal{X}_n = \begin{bmatrix} y_a & z_a \\ y_a & z_a \\ y_b & z_b \\ \vdots & \vdots \\ y_b & z_b \end{bmatrix} \quad \text{and} \quad \tilde{\mathcal{X}}_n = \begin{bmatrix} y_a & z_a \\ y_b & z_b \\ y_b & z_b \\ \vdots & \vdots \\ y_b & z_b \end{bmatrix},$$

where $d_{\text{ham}}(\mathcal{X}_n, \tilde{\mathcal{X}}_n) = 1$. We then consider a permutation $\boldsymbol{\pi} = (2, 1, 3, 4, \ldots, n)$ and denote the corresponding permuted datasets as

$$\mathcal{X}_n^{\boldsymbol{\pi}} = \begin{bmatrix} y_a & z_a \\ y_a & z_a \\ y_b & z_b \\ \vdots & \vdots \\ y_b & z_b \end{bmatrix} \quad \text{and} \quad \tilde{\mathcal{X}}_n^{\boldsymbol{\pi}} = \begin{bmatrix} y_a & z_b \\ y_b & z_a \\ y_b & z_b \\ \vdots & \vdots \\ y_b & z_b \end{bmatrix}.$$

Using the closed form expression of the squared HSIC in (12) and the property of translation invariant kernels, in particular $k(y, y) = K$ and $\ell(z, z) = L$ for all y, z, it can be seen that

$$\begin{aligned}
\widehat{\text{HSIC}}^2(\mathcal{X}_n^{\boldsymbol{\pi}}) &= \frac{4 + (n-2)^2}{n^2} KL + \frac{4(n-2)\epsilon_\star \epsilon'_\star}{n^2} \\
&\quad + \left\{ \frac{4 + (n-2)^2}{n^2} K + \frac{4(n-2)}{n^2} \epsilon_\star \right\} \left\{ \frac{4 + (n-2)^2}{n^2} L + \frac{4(n-2)}{n^2} \epsilon'_\star \right\} \\
&\quad - 4 \left\{ \frac{4}{n^3} KL + \frac{2K\epsilon'_\star (n-2)}{n^3} + \frac{2L\epsilon_\star (n-2)}{n^3} + \frac{(n-2)^2 \epsilon_\star \epsilon'_\star}{n^3} \right\} \\
&\quad - 2(n-2) \left\{ \frac{4\epsilon_\star \epsilon'_\star}{n^3} + \frac{2\epsilon_\star L(n-2)}{n^3} + \frac{2K\epsilon'_\star (n-2)}{n^3} + \frac{KL(n-2)^2}{n^3} \right\} \\
&= \frac{16(n-2)^2}{n^4} KL + C_1 \epsilon_\star \epsilon'_\star + C_2 \epsilon_\star + C_3 \epsilon'_\star,
\end{aligned}$$

where C_1, C_2, C_3 are constants that only depend on K, L, n. A similar calculation shows

$$\widehat{\text{HSIC}}^2(\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}}) = \frac{4}{n^4} KL + C'_1 \epsilon_\star \epsilon'_\star + C'_2 \epsilon_\star + C'_3 \epsilon'_\star,$$

where C'_1, C'_2, C'_3 are constants that only depend on K, L, n. These results together with the reverse triangle inequality yields

$$\left| \widehat{\text{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}}) - \widehat{\text{HSIC}}(\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}}) \right| \geq \left| \sqrt{\frac{16(n-2)^2}{n^4} KL} - \sqrt{\frac{4}{n^4} KL} \right| - h(\epsilon_\star, \epsilon'_\star, K, L, n)$$

$$= \frac{4(n-2.5)}{n^2}\sqrt{KL} - h(\epsilon_\star, \epsilon'_\star, K, L, n),$$

where $h(\epsilon_\star, \epsilon'_\star, K, L, n)$ is some function of $\epsilon_\star, \epsilon'_\star, K, L, n$, which goes to zero as $\epsilon_\star \to 0$ and $\epsilon'_\star \to 0$ for each fixed K, L, n. Therefore it holds that

$$\sup_{\boldsymbol{\pi}\in\boldsymbol{\Pi}_n} \sup_{\substack{\mathcal{X}_n, \tilde{\mathcal{X}}_n: \\ d_{\mathrm{ham}}(\mathcal{X}_n, \tilde{\mathcal{X}}_n)\le 1}} \left|\widehat{\mathrm{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}}) - \widehat{\mathrm{HSIC}}(\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}})\right| \ge \frac{4(n-2.5)}{n^2}\sqrt{KL} - h(\epsilon_\star, \epsilon'_\star, K, L, n),$$

and the lower bound result follows by letting $\epsilon \to 0$ given that $\max\{\epsilon_\star, \epsilon'_\star\} \le \epsilon$. This completes the proof of Lemma 6.

E.8 Proof Theorem 6

Let us prove the three claims made in Theorem 6 below.

- *Proof of Differential Privacy.* This result follows by Theorem 2 along with Lemma 6.
- *Proof of Validity.* This result follows by Theorem 1.
- *Proof of Consistency.* The proof of consistency for ϕ_{dpHSIC} is essentially the same as that for ϕ_{dpMMD} in Theorem 5. We only need to verify that $\widehat{\mathrm{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}_0}) \xrightarrow{p} \mathrm{HSIC}_{k\otimes\ell}(P_{YZ})$ and $\widehat{\mathrm{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}_1}) \xrightarrow{p} 0$. Indeed, the first claim follows by Lemma 14 and the second claim follows by Lemma 12, given that α, K, L are fixed quantities. The remaining steps are the same as those for the proof of Theorem 5 and thus we omit the details.

E.9 Proof of Theorem 7

The proof of Theorem 7 follows the same structure as that of Theorem 4. The main difference is that we make use of exponential concentration inequalities of the MMD statistic in order to obtain the logarithmic dependence on both α and β. Throughout the proof, we denote positive constants that only depend on K and τ by $C_1, C_2, \ldots$ whose values may vary in different places. As in the proof of Theorem 4, we build on the quantile representation of the permutation test (Lemma 17) from which we see that the type II error can be written as

$$\mathbb{E}(1 - \phi_{\mathrm{dpMMD}}) = \mathbb{P}(M_0 \le q_{1-\alpha,B}),$$

where $q_{1-\alpha,B}$ denotes the $1-\alpha$ quantile of $\{M_0, M_1, \ldots, M_B\}$, and M_i is given as $\widehat{\mathrm{MMD}}(\mathcal{X}_{n+m}^{\boldsymbol{\pi}_i}) + 2\Delta_T \xi_{\varepsilon,\delta}^{-1}\zeta_i$ for $i \in \{0\} \cup [B]$.

Step 1 (Bounding the quantile). As in the proof of Theorem 4, let $q^a_{1-\alpha/2,B}$ and $q^b_{1-\alpha/2,B}$ denote the $1-\alpha/2$ quantiles of $\{\widehat{\mathrm{MMD}}(\mathcal{X}_{n+m}^{\boldsymbol{\pi}_0}), \ldots, \widehat{\mathrm{MMD}}(\mathcal{X}_{n+m}^{\boldsymbol{\pi}_B})\}$ and of $\{2\Delta_T\xi_{\varepsilon,\delta}^{-1}\zeta_0, \ldots, 2\Delta_T\xi_{\varepsilon,\delta}^{-1}\zeta_B\}$, respectively. Denote by $q^a_{1-\alpha/4,\infty}$ and $q^b_{1-\alpha/4,\infty}$ the corresponding $1-\alpha/4$ quantiles with $B=\infty$. Then following the argument given in Appendix E.4 and under the condition for B, we see that the inequality $q_{1-\alpha,B} \le q^a_{1-\alpha/4,\infty} + q^b_{1-\alpha/4,\infty}$ holds with probability at least $1-\beta/2$.

Next we further upper bound $q^a_{1-\alpha/4,\infty}$ and $q^b_{1-\alpha/4,\infty}$ by more manageable terms. Applying Lemma 10 with the condition $n \le m \le \tau n$ yields the following inequality:

$$q^a_{1-\alpha/4,\infty} \le C_1 \sqrt{\frac{1}{n} \log\left(\frac{4}{\alpha}\right)}.$$

On the other hand, as noted in Appendix E.4, $q^b_{1-\alpha/4,\infty}$ is simply the $1-\alpha/4$ quantile of $2\Delta_T \xi^{-1}_{\varepsilon,\delta} \zeta$, namely

$$q^b_{1-\alpha/4,\infty} = 2\Delta_T \xi^{-1}_{\varepsilon,\delta} F^{-1}_{\zeta}(1-\alpha/4).$$

Therefore, the following inequality holds with probability at least $1-\beta/2$:

$$q_{1-\alpha,B} \le C_1 \sqrt{\frac{1}{n} \log\left(\frac{4}{\alpha}\right)} + 2\Delta_T \xi^{-1}_{\varepsilon,\delta} F^{-1}_{\zeta}(1-\alpha/4). \tag{28}$$

Remark 3 (Condition for the sample-size ratio)**.** The condition $n \le m \le \tau n$ is required to apply Lemma 10 under which we can obtain the logarithmic factor of α in the upper bound for $q^a_{1-\alpha/4,\infty}$. This condition can be eliminated by using Lemma 11, which shows that

$$q^a_{1-\alpha/4,\infty} \le C_2 \sqrt{\frac{1}{n\alpha}},$$

without the condition for the sample-size ratio constraint. It remains an open question whether we can achieve the logarithmic factor of α in the upper bound without the constraint $n \le m \le \tau n$, which we leave for future research.

Step 2 (Bounding the type II error). Let (P, Q) be a pair of distributions in $\mathcal{P}_{\mathrm{MMD}_k}(\rho)$. Given (P, Q), Lemma 13 (concentration inequality for $\widehat{\mathrm{MMD}}$) ensures that the following event E_1:

$$E_1 := \left\{ \left| \mathrm{MMD}_k(P,Q) - \widehat{\mathrm{MMD}}(\mathcal{X}_{n+m}) \right| \le C_3 \sqrt{\frac{\log(8/\beta)}{n}} \right\}$$

holds with probability at least $1-\beta/4$. Notice that the inverse cumulative distribution function of $\zeta \sim \mathsf{Laplace}(0,1)$ is given by

$$F^{-1}_{\zeta}(p) := -\mathrm{sign}(p-0.5) \times \log(1-2|p-0.5|) \quad \text{for } p \in (0,1).$$

This yields that the probability of the event

$$E_2 := \left\{ 2\Delta_T \xi^{-1}_{\varepsilon,\delta} \zeta_0 > 2\Delta_T \xi^{-1}_{\varepsilon,\delta} F^{-1}_{\zeta}(\beta/4) \right\}$$

is equal to $1-\beta/4$. Using these preliminary results, it can be seen that the type II error of ϕ_{dpMMD} satisfies

$$\mathbb{E}(1-\phi_{\mathrm{dpMMD}}) \overset{\text{(i)}}{=} \mathbb{P}(M_0 \le q_{1-\alpha,B}) \overset{\text{(ii)}}{=} \mathbb{P}\big(\widehat{\mathrm{MMD}}(\mathcal{X}_{n+m}) + 2\Delta_T \xi^{-1}_{\varepsilon,\delta} \zeta_0 \le q_{1-\alpha,B}\big)$$

$$
\begin{aligned}
&\overset{\text{(iii)}}{\leq} \mathbb{P}\bigg(\mathrm{MMD}_k(P,Q) + 2\Delta_T \xi_{\varepsilon,\delta}^{-1} F_\zeta^{-1}(\beta/4) \leq q_{1-\alpha,B} + C_3\sqrt{\frac{\log(8/\beta)}{n}}\bigg) + \mathbb{P}(E_1^c) + \mathbb{P}(E_2^c) \\
&\overset{\text{(iv)}}{\leq} \mathbb{P}\bigg(\mathrm{MMD}_k(P,Q) \leq q_{1-\alpha,B} + C_3\sqrt{\frac{\log(8/\beta)}{n}} - 2\Delta_T \xi_{\varepsilon,\delta}^{-1} F_\zeta^{-1}(\beta/4)\bigg) + \frac{\beta}{2} \\
&\overset{\text{(v)}}{\leq} \mathbb{P}\bigg(\mathrm{MMD}_k(P,Q) \leq C\sqrt{\frac{\log(4/\alpha)}{n}} + 2\Delta_T \xi_{\varepsilon,\delta}^{-1} F_\zeta^{-1}(1-\alpha/4) \\
&\qquad\qquad + C_3\sqrt{\frac{\log(8/\beta)}{n}} - 2\Delta_T \xi_{\varepsilon,\delta}^{-1} F_\zeta^{-1}(\beta/4)\bigg) + \beta,
\end{aligned}
$$

where step (i) uses Lemma 17, step (ii) uses the definition of M_0, step (iii) follows by the union bound, step (iv) uses the properties of the events E_1 and E_2, and step (v) uses the inequality given in (28) and the union bound.

Since $\beta/4 < 1/2$ and $1 - \alpha/4 > 1/2$, it follows that

$$
F_\zeta^{-1}(\beta/4) = \log\bigg(\frac{\beta}{2}\bigg) \quad \text{and} \quad F_\zeta^{-1}(1-\alpha/4) = \log\bigg(\frac{2}{\alpha}\bigg),
$$

which gives

$$
2\Delta_T \xi_{\varepsilon,\delta}^{-1} F_\zeta^{-1}(1-\alpha/4) - 2\Delta_T \xi_{\varepsilon,\delta}^{-1} F_\zeta^{-1}(\beta/4) = 2\Delta_T \xi_{\varepsilon,\delta}^{-1} \bigg\{\log\bigg(\frac{2}{\alpha}\bigg) + \log\bigg(\frac{2}{\beta}\bigg)\bigg\}.
$$

Consequently, the type II error of the dpMMD test is bounded by

$$
\begin{aligned}
&\mathbb{E}(1 - \phi_{\mathrm{dpMMD}} \\
&\leq \mathbb{P}\bigg(\mathrm{MMD}_k(P,Q) \leq C_4\sqrt{\frac{\max\{\log(1/\alpha), \log(1/\beta)\}}{n}} + C_5\frac{\max\{\log(1/\alpha), \log(1/\beta)\}}{n\xi_{\varepsilon,\delta}}\bigg) + \beta \\
&\leq \beta,
\end{aligned}
$$

where the last inequality holds by taking $C_{K,\tau}$ to be larger than, for instance, $2\max\{C_4, C_5\} + 1$ in the theorem statement. Finally, we note that the above upper bound is independent of (P, Q), and thus complete the proof by taking the supremum on both sides over $\mathcal{P}_{\mathrm{MMD}_k}(\rho)$.

E.10 Proof of Theorem 8

Under the conditions of Theorem 8, we analyze that the minimax separation $\rho^\star_{\mathrm{MMD}}(\alpha, \beta, \varepsilon, \delta, m, n) = \rho^\star_{\mathrm{MMD}}$ in the non-privacy regime and in the privacy regime, separately. In both regimes, we use a common trick that reduces the two-sample problem to the one-sample problem (also known as goodness-of-fit testing) stated below.

Reducing the two-sample problem to the one-sample problem. We can think of the one-sample problem as a special case of the two-sample problem by assuming that we have access to

an infinite number of observations from one of the two distributions (say Q). That is, we know the entire information of the distribution Q and $m = \infty$. Arias-Castro et al. (2018, Lemma 1) builds on this simple observation and formalizes that the minimax risk of the two-sample problem is greater than or equal to that of the one-sample problem. We follow this strategy and work with the one-sample problem.

More formally, pick one distribution Q_0 (specified later in the proof) from $\mathcal{P}$ and fix it throughout. Let $\phi : \mathcal{X}_n \mapsto \{0, 1\}$ be a test function and define the set of (ε, δ)-DP level α one-sample tests as

$$\Phi_{\alpha,\varepsilon,\delta,Q_0} \coloneqq \Big\{\phi : \mathbb{E}_{Q_0}[\phi] \le \alpha \text{ and } \phi \text{ is } (\varepsilon, \delta)\text{-DP}\Big\}.$$

Then, letting $\mathcal{P}_{1,\mathrm{MMD}_k}(\rho; Q_0) \coloneqq \{P \in \mathcal{P} : \mathrm{MMD}_k(P, Q_0) \ge \rho\}$, Arias-Castro et al. (2018, Lemma 1) yields

$$\rho^\star_{\mathrm{MMD}} \ge \inf\Big\{\rho > 0 : \inf_{\phi \in \Phi_{\alpha,\varepsilon,\delta,Q_0}} \sup_{P \in \mathcal{P}_{1,\mathrm{MMD}_k}(\rho;Q_0)} \mathbb{E}_P[1 - \phi] \le \beta\Big\}.$$

Therefore, once we prove

$$\begin{aligned}
&\inf\Big\{\delta > 0 : \inf_{\phi \in \Phi_{\alpha,\varepsilon,Q_0}} \sup_{P \in \mathcal{P}_{1,\mathrm{MMD}_k}(\rho;Q_0)} \mathbb{E}_P[1 - \phi] \le \beta\Big\} \\
&\ge\ C_\eta \max\left\{\min\left\{\sqrt{\frac{\log(1/(\alpha+\beta))}{n}}, 1\right\}, \min\left\{\frac{\log(1/\beta)}{n(\varepsilon+\delta)}, 1\right\}\right\} \qquad (29a) \\
&\ge\ C_\eta \max\left\{\min\left\{\sqrt{\frac{\log(1/(\alpha+\beta))}{n}}, 1\right\}, \min\left\{\frac{\log(1/\beta)}{n\big(\varepsilon + \log(1/(1-\delta))\big)}, 1\right\}\right\}, \qquad (29b)
\end{aligned}$$

for some $Q_0 \in \mathcal{P}$, then the desired result of Theorem 8 follows. The two lower bounds in (29b) are proved in Appendices E.10.1 and E.10.2. The inequality (29a) $\ge$ (29b) holds as $\log(1/(1-\delta)) \ge \delta$ for all $\delta \in [0, 1)$, and is sufficient to prove Theorem 8. Nonetheless, to provide intuition on the tightness of the lower bound, it is interesting to understand that when $\alpha \asymp \beta$ the two rates in Equations (29a) and (29b) are the same, which we prove in Appendix E.10.3.

E.10.1 Non-privacy regime

We first show that $\rho^\star_{\mathrm{MMD}} \ge C_\eta \min\big\{\sqrt{\log(1/(\alpha+\beta))/n}, 1\big\}$ in the non-privacy regime. Writing the set of non-private level α tests (*i.e.*, $\varepsilon \to \infty$) as

$$\Phi_{\alpha,\infty,Q_0} \coloneqq \Big\{\phi : \mathbb{E}_{Q_0}[\phi] \le \alpha\Big\},$$

we clearly have $\Phi_{\alpha,\varepsilon,\delta,Q_0} \subset \Phi_{\alpha,\infty,Q_0}$. Hence, in this non-private regime, it suffices to show

$$\inf\Big\{\rho > 0 : \inf_{\phi \in \Phi_{\alpha,\infty,Q_0}} \sup_{P \in \mathcal{P}_{1,\mathrm{MMD}_k}(\rho;Q_0)} \mathbb{E}_P[1 - \phi] \le \beta\Big\} \ge C_\eta \min\left\{\sqrt{\frac{\log(1/(\alpha+\beta))}{n}}, 1\right\},$$

as the infimum term above is smaller than or equal to $\rho^\star_{\mathrm{MMD}}$. To this end, we use the standard Le Cam's two point method (Le Cam, 1973, 2012). In particular, for any $P_0 \in \mathcal{P}_{1,\mathrm{MMD}_k}(\widetilde{\rho}; Q_0)$ with

$$\widetilde{\rho} = C_\eta \min\left\{\sqrt{\frac{\log(1/(\alpha+\beta))}{n}}, 1\right\},$$

we have

$$\begin{aligned}
\inf_{\phi \in \Phi_{\alpha,\infty,Q_0}} \sup_{P \in \mathcal{P}_{1,\mathrm{MMD}_k}(\widetilde{\rho};Q_0)} \mathbb{E}_P[1-\phi] &\geq \inf_{\phi \in \Phi_{\alpha,\infty,Q_0}} \mathbb{E}_{P_0}[1-\phi] = 1 - \sup_{\phi \in \Phi_{\alpha,\infty,Q_0}} \mathbb{E}_{P_0}[\phi] \\
&\geq 1 - \alpha - d_{\mathrm{TV}}(P_0^{\otimes n}, Q_0^{\otimes n}) \\
&\overset{\text{(i)}}{\geq} 1 - \alpha - 1 + \frac{1}{2} e^{-d_{\mathrm{KL}}(P_0^{\otimes n} \| Q_0^{\otimes n})} \overset{\text{(ii)}}{=} \frac{1}{2} e^{-n \times d_{\mathrm{KL}}(P_0 \| Q_0)} - \alpha,
\end{aligned}$$

where step (i) holds by Bretagnolle–Huber inequality (Canonne, 2022, Lemma B.4) and step (ii) uses the chain rule of the KL divergence. Therefore the minimax type II error is bounded by β, that is

$$\inf_{\phi \in \Phi_{\alpha,\infty,Q_0}} \sup_{P \in \mathcal{P}_{1,\mathrm{MMD}_k}(\widetilde{\rho};Q_0)} \mathbb{E}_P[1-\phi] \geq \beta,$$

provided that $\alpha + \beta < 0.4$ and

$$d_{\mathrm{KL}}(P_0 \| Q_0) \leq \frac{1}{n} \log\left(\frac{1}{2(\alpha+\beta)}\right).$$

Hence it is enough to find an instance of (P_0, Q_0) in $\mathbb{R}^d$ such that

$$\mathrm{MMD}_k(P_0, Q_0) \geq C_\eta \min\left\{\sqrt{\frac{\log(1/(\alpha+\beta))}{n}}, 1\right\} \quad \text{and} \tag{30a}$$

$$d_{\mathrm{KL}}(P_0 \| Q_0) \leq \frac{1}{n} \log\left(\frac{1}{2(\alpha+\beta)}\right). \tag{30b}$$

Motivated by the proof of Tolstikhin et al. (2017, Theorem 1), we pick two discrete distributions $P_0 = p_0\delta_x + (1-p_0)\delta_v$ and $Q_0 = q_0\delta_x + (1-q_0)\delta_v$, where $x, v \in \mathbb{R}^d$, $0 < p_0, q_0 < 1$ and δ_x is a Dirac measure at x. In this setting, the calculations in Tolstikhin et al. (2017) show that

$$\mathrm{MMD}_k(P_0, Q_0) = \sqrt{2(p_0 - q_0)^2\big(\kappa(0) - \kappa(x-v)\big)}. \tag{31}$$

In addition, under the same setting, the KL divergence of P_0 from Q_0 is bounded by

$$d_{\mathrm{KL}}(P_0 \| Q_0) \leq \frac{(p_0 - q_0)^2}{q_0(1-q_0)}.$$

Now set

$$p_0 = \frac{1}{2} + \min\left\{\sqrt{\frac{1}{4n}\log\left(\frac{1}{2(\alpha+\beta)}\right)}, \frac{1}{2}\right\} \quad \text{and} \quad q_0 = \frac{1}{2},$$

so that both P_0 and Q_0 are valid probability measures, which leads to

$$d_{\mathrm{KL}}(P_0\|Q_0) \le \min\left\{\frac{1}{n}\log\left(\frac{1}{2(\alpha+\beta)}\right), 1\right\} \le \frac{1}{n}\log\left(\frac{1}{2(\alpha+\beta)}\right).$$

We also choose x and v such that $\kappa(0)-\kappa(z)\ge\eta$ for $z=x-v$ where η is given in the theorem assumption. In this setting, the explicit expression of the MMD metric (31) can be used to show

$$\begin{aligned}\mathrm{MMD}_k(P_0,Q_0) &\ge \sqrt{2\eta}\min\left\{\sqrt{\frac{1}{4n}\log\left(\frac{1}{2(\alpha+\beta)}\right)}, \frac{1}{2}\right\}\\ &\ge C_\eta \min\left\{\sqrt{\frac{1}{n}\log\left(\frac{1}{\alpha+\beta}\right)}, 1\right\}\end{aligned}$$

where the second inequality holds by taking C_η sufficiently small under the condition $\alpha+\beta<0.4$. Hence we have shown that the sufficient conditions in (30a) and (30b) are fulfilled, and thus the minimax separation satisfies

$$\rho^\star_{\mathrm{MMD}} \ge C_\eta \min\left\{\sqrt{\frac{\log(1/(\alpha+\beta))}{n}}, 1\right\}.$$

Next we focus on the privacy regime and prove the second term in the lower bound.

E.10.2 Privacy regime

The second term in the lower bound can be proved based on the coupling idea described in Acharya et al. (2018, Theorem 11) and Acharya et al. (2021, Theorem 1). In detail, we pick P_0 from $\mathcal{P}_{1,\mathrm{MMD}_k}(\delta;Q_0)$, and let $(\mathcal{X}_n,\mathcal{Y}_n)$ be a coupling between $P_0^{\otimes n}$ and $Q_0^{\otimes n}$ with $D=\mathbb{E}[d_{\mathrm{ham}}(\mathcal{X}_n,\mathcal{Y}_n)]$. Then for any $\phi\in\Phi_{\alpha,\varepsilon,Q_0}$, the proof of Acharya et al. (2021, Theorem 1) shows

$$\begin{aligned}1-\alpha \le \mathbb{E}_{Q_0}[1-\phi] &\le \mathbb{E}_{P_0}[1-\phi]\cdot e^{10D\varepsilon} + 0.1 + 10D\delta e^{10D\varepsilon}\\ \mathbb{E}_{P_0}[1-\phi] &\ge e^{-10D\varepsilon}(0.9-\alpha) - 10D\delta.\end{aligned}$$

For completeness, we give details here. By Markov's inequality, the event $\mathcal{E}^c=\{d_{\mathrm{ham}}(\mathcal{X}_n,\mathcal{Y}_n)>10D\}$ holds with probability at most $1/10$ (here one can replace the constant 10 with some other positive number). Moreover,

$$\begin{aligned}\mathbb{E}_{Q_0}[1-\phi]=\mathbb{P}_{Q_0}(\phi=0) &\le \mathbb{P}_{Q_0}(\phi=0\,|\,\mathcal{E})\mathbb{P}(\mathcal{E})+\mathbb{P}(\mathcal{E}^c)\\ &\le \mathbb{P}_{Q_0}(\phi=0\,|\,\mathcal{E})\mathbb{P}(\mathcal{E})+0.1\\ &\overset{(\dagger)}{\le} \mathbb{P}_{P_0}(\phi=0\,|\,\mathcal{E})\mathbb{P}(\mathcal{E})\cdot e^{10D\varepsilon}+0.1+10D\delta e^{(10D-1)\varepsilon}\\ &\le \mathbb{E}_{P_0}[1-\phi]\cdot e^{10D\varepsilon}+0.1+10D\delta e^{10D\varepsilon},\end{aligned}$$

where step (†) holds due to the fact that ϕ is ε-differentially private and inequality (2). Letting $E := 10D$ for notation purposes, and using the fact that $\alpha \in (0, 1/5)$, we obtain

$$\begin{aligned}\inf_{\phi \in \Phi_{\alpha,\varepsilon,Q_0}} \sup_{P \in \mathcal{P}_{1,\mathrm{MMD}_k}(\rho;Q_0)} \mathbb{E}_P[1-\phi] &\geq \inf_{\phi \in \Phi_{\alpha,\varepsilon,Q_0}} \mathbb{E}_{P_0}[1-\phi] \\ &\geq e^{-E\varepsilon}(0.9-\alpha) - E\delta \\ &\geq 0.5e^{-E\varepsilon} - E\delta \\ &\geq 0.25e^{-E\varepsilon},\end{aligned}$$

provided that the condition $E\delta \leq 0.25e^{-E\varepsilon}$ holds. Hence, we require

$$\inf_{\phi \in \Phi_{\alpha,\varepsilon,Q_0}} \sup_{P \in \mathcal{P}_{1,\mathrm{MMD}_k}(\rho;Q_0)} \mathbb{E}_P[1-\phi] \;\geq\; 0.25e^{-E\varepsilon} \;\geq\; \beta,$$

which is fulfilled when

$$E \;\leq\; \frac{1}{\varepsilon}\log\left(\frac{1}{4\beta}\right) \qquad \text{and} \qquad E\delta \;\leq\; \frac{1}{4}e^{-E\varepsilon}. \tag{32}$$

The second condition, equivalently $E\delta e^{E\varepsilon} \leq 1/4$, is trivial when $\delta = 0$. Hence by assuming $\delta > 0$, we verify that

$$E \leq \frac{1}{4(\varepsilon+\delta)} \quad \text{implies} \quad E\delta e^{E\varepsilon} \leq \frac{1}{4}.$$

To see this, under the condition $E \leq \frac{1}{4(\varepsilon+\delta)}$ and letting $x = \varepsilon/\delta$,

$$E\delta e^{E\varepsilon} \leq \frac{\delta}{4(\varepsilon+\delta)} e^{\frac{\varepsilon}{4(\varepsilon+\delta)}} = \frac{1}{4(x+1)} e^{\frac{x}{4(x+1)}}$$

Hence it suffices to show that for all $x > 0$ we have

$$\frac{1}{4(x+1)} e^{\frac{x}{4(x+1)}} \leq \frac{1}{4} \qquad \Longleftrightarrow \qquad f(x) := 4(x+1)\log(x+1) - x \geq 0.$$

The first derivative of f is given as $f'(x) = 4\log(x+1) + 3 > 0$, which implies that f is monotone increasing and $f(x) \geq 0$ for all $x > 0$. Hence, for all $\varepsilon > 0$ and $\delta \in [0, 1)$, the conditions in Equation (32) are satisfied when

$$E \;\leq\; \frac{1}{\varepsilon}\log\left(\frac{1}{4\beta}\right) \qquad \text{and} \qquad E \;\leq\; \frac{1}{4(\varepsilon+\delta)},$$

which are in particular satisfied when

$$E \;\leq\; \frac{1}{4(\varepsilon+\delta)} \min\left\{\log\left(\frac{1}{4\beta}\right), 1\right\}. \tag{33}$$

Furthermore, given P_0 and Q_0, Acharya et al. (2021, Lemma 20) ensures the existence of a coupling between $P_0^{\otimes n}$ and $Q_0^{\otimes n}$ such that

$$D = \mathbb{E}[d_{\text{ham}}(\mathcal{X}_1^n, \mathcal{Y}_1^n)] = \frac{n}{2}\|P_0 - Q_0\|_1,$$

so $E = 10D = 5n\|P_0 - Q_0\|_1$. The condition in Equation (33) becomes

$$\|P_0 - Q_0\|_1 \;\leq\; \frac{1}{20n(\varepsilon+\delta)} \min\left\{\log\left(\frac{1}{4\beta}\right), 1\right\}. \tag{34}$$

As in the non-private case, we can pick two discrete distributions $P_0 = p_0\delta_x + (1-p_0)\delta_v$ and $Q_0 = q_0\delta_x + (1-q_0)\delta_v$, where $q_0 = 1/2$ and

$$p_0 = \frac{1}{2} + \min\left\{\frac{1}{40n(\varepsilon+\delta)}\log\left(\frac{1}{4\beta}\right), \frac{1}{2}\right\}.$$

This choice ensures that

$$\|P_0 - Q_0\|_1 = 2|p_0 - q_0| = 2\min\left\{\frac{1}{40n(\varepsilon+\delta)}\log\left(\frac{1}{4\beta}\right), \frac{1}{2}\right\}.$$

We also choose x and v such that $\kappa(0) - \kappa(z) \geq \eta$ for $z = x - v$ under which

$$\begin{aligned}\text{MMD}_k(P_0, Q_0) &\geq \sqrt{2\eta}\min\left\{\frac{1}{40n(\varepsilon+\delta)}\log\left(\frac{1}{4\beta}\right), \frac{1}{2}\right\} \\ &\geq C_\eta \min\left\{\frac{1}{n(\varepsilon+\delta)}\log\left(\frac{1}{\beta}\right), 1\right\},\end{aligned}$$

where the second inequality holds under the condition $\beta \in (0, 1/5)$. Therefore, under privacy regime, for all $\varepsilon > 0$ and all $\delta \in [0, 1)$, it holds that

$$\rho^\star_{\text{MMD}} \geq C_\eta \min\left\{\frac{\log(1/\beta)}{n(\varepsilon+\delta)}, 1\right\},$$

which proves the inequality in (29a).

E.10.3 Equivalence of rates

We now prove the equivalence in (29b) when $\alpha \asymp \beta$. As previously noted, we have (29a) $\geq$ (29b) as

$$\log(1/(1-\delta)) \geq \delta \quad \text{for all } \delta \in [0, 1).$$

Hence, it remains to show that (29a) $\leq$ (29b), so that, up to constants, we have

$$\max\left\{\min\left\{\sqrt{\frac{\log(1/\beta)}{n}}, 1/2\right\}, \min\left\{\frac{\log(1/\beta)}{n(\varepsilon+\delta)}, 1/2\right\}\right\}$$

$$\overset{(\star)}{\asymp} \max\left\{\min\left\{\sqrt{\frac{\log(1/\beta)}{n}}, 1/2\right\}, \min\left\{\frac{\log(1/\beta)}{n\big(\varepsilon+\log(1/(1-\delta))\big)}, 1/2\right\}\right\}.$$

Here, we have used the fact that $\alpha \asymp \beta$ and have absorbed some terms in the constants to replace the 1's with 1/2's in the minima. For the result to be non-trivial we need to assume

$$\sqrt{\frac{\log(1/\beta)}{n}} < 1/2 \qquad \text{and} \qquad \frac{\log(1/\beta)}{n\big(\varepsilon+\log(1/(1-\delta))\big)} < 1/2. \tag{35}$$

- If $\delta \in [0, 1/2)$, then $2\log(2)\delta \geq \log(1/(1-\delta)) \geq \delta$, and so we get

  $$\frac{\log(1/\beta)}{n\big(\varepsilon+\log(1/(1-\delta))\big)} \leq \frac{\log(1/\beta)}{n(\varepsilon+\delta)} \leq 2\log(2)\frac{\log(1/\beta)}{n\big(\varepsilon+\log(1/(1-\delta))\big)}$$

  which proves $(\star)$ for $\delta \in [0, 1/2)$.

- If $\delta \in [1/2, 1)$, then $\delta \geq 1/2 > \sqrt{\log(1/\beta)/n}$, hence we get

  $$\frac{\log(1/\beta)}{n\big(\varepsilon+\log(1/(1-\delta))\big)} \leq \frac{\log(1/\beta)}{n\delta} < \sqrt{\frac{\log(1/\beta)}{n}}$$

  and

  $$\frac{\log(1/\beta)}{n(\varepsilon+\delta)} < \sqrt{\frac{\log(1/\beta)}{n}},$$

  so both sides of $(\star)$ are equal to $\sqrt{\log(1/\beta)/n}$ when $\delta \in [1/2, 1)$. So, when $\delta \geq \sqrt{\log(1/\beta)/n}$, the non-DP rate dominates.

We have shown that $(\star)$ holds for all $\delta \in [0, 1)$, which completes the proof.

E.11 Proof of Theorem 9

To simplify the notation, we denote the square of the empirical MMD based on $\mathcal{X}_{n+m}$ as V_{MMD} (or equivalently the V-statistic of MMD), and that based on permuted data $\mathcal{X}^{\boldsymbol{\pi}}_{n+m}$ as $V_{\boldsymbol{\pi},\text{MMD}}$. We also denote the U-statistic of the MMD as U_{MMD}, *i.e.*,

$$U_{\text{MMD}} = \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} k_{\boldsymbol{\lambda}}(Y_i, Y_j) + \frac{1}{m(m-1)} \sum_{1\leq i\neq j\leq m} k_{\boldsymbol{\lambda}}(Z_i, Z_j) - \frac{2}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} k_{\boldsymbol{\lambda}}(Y_i, Z_j),$$

and $U_{\boldsymbol{\pi},\text{MMD}}$ is similarly defined based on $\mathcal{X}^{\boldsymbol{\pi}}_{n+m}$. Remark that the difference between the V-statistic and the U-statistic of MMD is

$$\begin{aligned} V_{\text{MMD}} - U_{\text{MMD}} \;=\; & \frac{1}{n}\prod_{i=1}^{d}\frac{1}{\sqrt{2\pi}\lambda_i} + \frac{1}{m}\prod_{i=1}^{d}\frac{1}{\sqrt{2\pi}\lambda_i} \\ & - \frac{1}{n^2(n-1)} \sum_{1\leq i\neq j\leq n} k_{\boldsymbol{\lambda}}(Y_i, Y_j) - \frac{1}{m^2(m-1)} \sum_{1\leq i\neq j\leq m} k_{\boldsymbol{\lambda}}(Z_i, Z_j). \end{aligned} \tag{36}$$

Given this connection, our proof strategy is to leverage known results of the U-statistic along with a careful analysis of the difference between V_{MMD} and U_{MMD}. Writing the sensitivity of $\sqrt{V_{\mathrm{MMD}}}$ as $\Delta_{V^{1/2}}$, Lemma 17 ensures that the dpMMD test rejects the null if and only if

$$\sqrt{V_{\mathrm{MMD}}} + \frac{2\Delta_{V^{1/2}}}{\xi_{\varepsilon,\delta}}\zeta_0 > q_{1-\alpha,B},$$

where $q_{1-\alpha,B}$ is the $1-\alpha$ quantile of $\{\sqrt{V_{\boldsymbol{\pi}_i,\mathrm{MMD}}} + 2\Delta_{V^{1/2}}\xi_{\varepsilon,\delta}^{-1}\zeta_i\}_{i=0}^{B}$. We then closely follow the proof steps given in Appendix E.4 to prove the claim.

Bounding the type II error. As in the proof of Theorem 4 in Appendix E.4, we let $q^a_{1-\alpha,B}$ and $q^b_{1-\alpha,B}$ denote the $1-\alpha$ quantiles of $\{V^{1/2}_{\boldsymbol{\pi}_0,\mathrm{MMD}}, V^{1/2}_{\boldsymbol{\pi}_1,\mathrm{MMD}}, \ldots, V^{1/2}_{\boldsymbol{\pi}_B,\mathrm{MMD}}\}$ and of $\{2\Delta_{V^{1/2}}\xi^{-1}_{\varepsilon,\delta}\zeta_0, \ldots, 2\Delta_{V^{1/2}}\xi^{-1}_{\varepsilon,\delta}\zeta_B\}$, respectively. Similarly, denote by $q^a_{1-\alpha,\infty}$ and $q^b_{1-\alpha,\infty}$, the corresponding $1-\alpha$ quantiles with $B=\infty$. Then for $B \geq 16\alpha^{-2}\log(8/\beta)$, the analysis given in the proof of Theorem 4 shows that the type II error of the dpMMD test can be bounded as

$$\begin{aligned}
\mathbb{P}\big(\sqrt{V_{\mathrm{MMD}}} + 2\Delta_{V^{1/2}}\xi^{-1}_{\varepsilon,\delta}\zeta_0 \leq q_{1-\alpha,B}\big) &\leq \mathbb{P}\big(\sqrt{V_{\mathrm{MMD}}} + 2\Delta_{V^{1/2}}\xi^{-1}_{\varepsilon,\delta}\zeta_0 \leq q^a_{1-\alpha/4,\infty} + q^b_{1-\alpha/4,\infty}\big) + \beta/2 \\
&\leq \mathbb{P}\big(\sqrt{V_{\mathrm{MMD}}} \leq q^a_{1-\alpha/4,\infty} + q^b_{1-\alpha/4,\infty} + R_2\big) + 5\beta/8,
\end{aligned}$$

where we recall $R_2 = 2\Delta_{V^{1/2}}\xi^{-1}_{\varepsilon,\delta}F^{-1}_{\zeta}(1-\beta/8)$. We also note that the quantile $q^b_{1-\alpha/4,\infty}$ can be explicitly written as

$$q^b_{1-\alpha/4,\infty} = 2\Delta_{V^{1/2}}\xi^{-1}_{\varepsilon,\delta}F^{-1}_{\zeta}(1-\alpha/4),$$

and using the inequality (27), the type II error is upper bounded as

$$\begin{aligned}
&\mathbb{P}\big(\sqrt{V_{\mathrm{MMD}}} + 2\Delta_{V^{1/2}}\xi^{-1}_{\varepsilon,\delta}\zeta_0 \leq q_{1-\alpha,B}\big) \\
\leq\ &\mathbb{P}\big(\sqrt{V_{\mathrm{MMD}}} \leq q^a_{1-\alpha/4,\infty} + 10\Delta_{V^{1/2}}\xi^{-1}_{\varepsilon,\delta}\max\{\log(1/\alpha), \log(1/\beta)\}\big) + 5\beta/8. \qquad (37)
\end{aligned}$$

Given this bound for the type II error, our next effort lies in bounding $q^a_{1-\alpha/4,\infty}$.

Bounding $q^a_{1-\alpha/4,\infty}$. To obtain an upper bound for $q^a_{1-\alpha/4,\infty}$, we leverage the exponential inequality for permuted U-statistics studied in Kim et al. (2022a) and Schrab et al. (2023). To this end, denote the probability function, which is taken over $\boldsymbol{\pi}$ conditional on $\mathcal{X}_{m+n}$, as $\mathbb{P}_{\boldsymbol{\pi}}$ and note that for any $t>0$,

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\pi}}\left(\sqrt{V_{\boldsymbol{\pi},\mathrm{MMD}}} \geq \sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)\prod_{i=1}^{d}\frac{1}{\sqrt{2\pi}\lambda_i} + t}\right) \\
=\ &\mathbb{P}_{\boldsymbol{\pi}}\left(V_{\boldsymbol{\pi},\mathrm{MMD}} \geq \left(\frac{1}{n}+\frac{1}{m}\right)\prod_{i=1}^{d}\frac{1}{\sqrt{2\pi}\lambda_i} + t\right) \\
=\ &\mathbb{P}_{\boldsymbol{\pi}}\left(V_{\boldsymbol{\pi},\mathrm{MMD}} - U_{\boldsymbol{\pi},\mathrm{MMD}} + U_{\boldsymbol{\pi},\mathrm{MMD}} \geq \left(\frac{1}{n}+\frac{1}{m}\right)\prod_{i=1}^{d}\frac{1}{\sqrt{2\pi}\lambda_i} + t\right)
\end{aligned}$$

$$= \mathbb{P}_{\boldsymbol{\pi}}\left(\left\{V_{\boldsymbol{\pi},\mathrm{MMD}} - U_{\boldsymbol{\pi},\mathrm{MMD}} - \left(\frac{1}{n}+\frac{1}{m}\right)\prod_{i=1}^{d}\frac{1}{\sqrt{2\pi}\lambda_i}\right\} + U_{\boldsymbol{\pi},\mathrm{MMD}} \geq t\right)$$

$$\leq \mathbb{P}_{\boldsymbol{\pi}}\big(U_{\boldsymbol{\pi},\mathrm{MMD}} \geq t\big),$$

where the last equality holds since for any permutation $\boldsymbol{\pi}$

$$V_{\boldsymbol{\pi},\mathrm{MMD}} - U_{\boldsymbol{\pi},\mathrm{MMD}} - \left(\frac{1}{n}+\frac{1}{m}\right)\prod_{i=1}^{d}\frac{1}{\sqrt{2\pi}\lambda_i} \leq 0,$$

which can be seen from the previous expression of $V_{\mathrm{MMD}} - U_{\mathrm{MMD}}$ in (36) and non-negativity of $k_{\boldsymbol{\lambda}}$. Hence $q^a_{1-\alpha/4,\infty}$ is bounded by

$$q^a_{1-\alpha/4,\infty} \leq \sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)\prod_{i=1}^{d}\frac{1}{\sqrt{2\pi}\lambda_i} + q^U_{1-\alpha/4,\infty}},$$

where $q^U_{1-\alpha/4,\infty}$ denotes the $1-\alpha/4$ quantile of the permutation distribution of $U_{\boldsymbol{\pi},\mathrm{MMD}}$. Now using the exponential bound for $U_{\boldsymbol{\pi},\mathrm{MMD}}$ in Kim et al. (2022a, Equation 59), and following the proof of Schrab et al. (2023, Proposition 4), we see that there exists a constant $C_{\tau,\beta,M,d}$ such that

$$\mathbb{P}\left(q^U_{1-\alpha/4,\infty} \leq C_{\tau,\beta,M,d}\frac{\log(1/\alpha)}{n\sqrt{\lambda_1\cdots\lambda_d}}\right) \geq 1-\beta/8,$$

under the conditions of Theorem 9. For simplicity, let us write

$$b_{n,m,\boldsymbol{\lambda}} = \left(\frac{1}{n}+\frac{1}{m}\right)\prod_{i=1}^{d}\frac{1}{\sqrt{2\pi}\lambda_i} \leq \frac{C_{\tau,d}}{n\lambda_1\ldots\lambda_d}. \tag{38}$$

Observe that we have $\max\{\log(1/\alpha),\log(1/\beta)\} \leq C_\beta\log(1/\alpha)$ for $\alpha \in (0,e^{-1})$. Putting these pieces together and continuing from (37), the type II error is further bounded by

$$\begin{aligned}
&\mathbb{P}\big(\sqrt{V_{\mathrm{MMD}}} + 2\Delta_{V^{1/2}}\xi^{-1}_{\varepsilon,\delta}\zeta_0 \leq q_{1-\alpha,B}\big)\\
\leq\ &\mathbb{P}\left(\sqrt{V_{\mathrm{MMD}}} \leq \sqrt{b_{n,m,\boldsymbol{\lambda}} + C_{\tau,\beta,M,d}\frac{\log(1/\alpha)}{n\sqrt{\lambda_1\cdots\lambda_d}}} + C_\beta\Delta_{V^{1/2}}\xi^{-1}_{\varepsilon,\delta}\log(1/\alpha)\right) + 6\beta/8\\
=\ &\mathbb{P}\left(V_{\mathrm{MMD}} \leq b_{n,m,\boldsymbol{\lambda}} + C_{\tau,\beta,M,d}\frac{\log(1/\alpha)}{n\sqrt{\lambda_1\cdots\lambda_d}} + C^2_\beta\Delta^2_{V^{1/2}}\xi^{-2}_{\varepsilon,\delta}\log^2(1/\alpha)\right.\\
&\qquad\left. + 2\sqrt{b_{n,m,\boldsymbol{\lambda}} + C_{\tau,\beta,M,d}\frac{\log(1/\alpha)}{n\sqrt{\lambda_1\cdots\lambda_d}}}C_\beta\Delta_{V^{1/2}}\xi^{-1}_{\varepsilon,\delta}\log(1/\alpha)\right) + 6\beta/8.
\end{aligned} \tag{39}$$

Connecting V_{MMD} with U_{MMD}. In order to express the condition for type II error control in terms of the L_2 distance, we make use of the existing results for the U-statistic in Schrab et al.

(2023). For simplicity, we use the notation $C_1, C_2, \ldots$ to represent constants that may depend on τ, β, s, R, M, d. The specific values of these constants may vary in different places. To proceed, let us make an observation that

$$\mathbb{E}\left[\frac{1}{n^2(n-1)}\sum_{(i,j)\in \mathbf{i}_2^n} k_{\boldsymbol{\lambda}}(Y_i, Y_j)\right] = \frac{1}{n}\mathbb{E}\big[k_{\boldsymbol{\lambda}}(Y_1, Y_2)\big] = \frac{1}{n}\int\int p(y_1)p(y_2)k_{\boldsymbol{\lambda}}(y_1, y_2)dy_1 dy_2$$
$$\leq \frac{\|p\|_{L_\infty}}{n}$$

and

$$\mathrm{Var}\left[\frac{1}{n^2(n-1)}\sum_{(i,j)\in \mathbf{i}_2^n} k_{\boldsymbol{\lambda}}(Y_i, Y_j)\right] \leq \frac{C_1}{n^3}\mathbb{E}[k_{\boldsymbol{\lambda}}^2(Y_1, Y_2)] \leq \frac{C_2}{n^3\lambda_1 \ldots \lambda_d}. \tag{40}$$

To explain the last display, note that the U-statistic

$$\frac{1}{n(n-1)}\sum_{(i,j)\in \mathbf{i}_2^n} k_{\boldsymbol{\lambda}}(Y_i, Y_j)$$

achieves the minimum variance among all unbiased estimators of $k_{\boldsymbol{\lambda}}(Y_1, Y_2)$, including a linear estimator given as

$$\frac{1}{\lfloor n/2 \rfloor}\sum_{i=1}^{\lfloor n/2 \rfloor} k_{\boldsymbol{\lambda}}(Y_{2i-1}, Y_{2i}).$$

The variance of the above linear estimator is bounded by $n^{-1}\mathbb{E}[k_{\boldsymbol{\lambda}}^2(Y_1, Y_2)]$, up to a constant factor, thereby the inequality (40) holds. A similar calculation shows that

$$\mathbb{E}\left[\frac{1}{m^2(m-1)}\sum_{(i,j)\in \mathbf{i}_2^m} k_{\boldsymbol{\lambda}}(Z_i, Z_j)\right] \leq \frac{\|p\|_{L_\infty}}{n} \quad \text{and}$$
$$\mathrm{Var}\left[\frac{1}{m^2(m-1)}\sum_{(i,j)\in \mathbf{i}_2^m} k_{\boldsymbol{\lambda}}(Z_i, Z_j)\right] \leq \frac{C_3}{n^3\lambda_1 \ldots \lambda_d}.$$

Based on these results combined with Chebyshev's inequality, we have the following statement, which holds with probability at least $1 - \beta/8$,

$$\begin{aligned}
& V_{\mathrm{MMD}} - U_{\mathrm{MMD}} + U_{\mathrm{MMD}} \\
= \; & b_{n,m,\boldsymbol{\lambda}} - \frac{1}{n^2(n-1)}\sum_{(i,j)\in \mathbf{i}_2^n} k_{\boldsymbol{\lambda}}(Y_i, Y_j) - \frac{1}{m^2(m-1)}\sum_{(i,j)\in \mathbf{i}_2^m} k_{\boldsymbol{\lambda}}(Z_i, Z_j) + U_{\mathrm{MMD}} \\
\geq \; & b_{n,m,\boldsymbol{\lambda}} - \frac{C_4}{n} - \frac{C_5}{n^{3/2}\sqrt{\lambda_1 \ldots \lambda_d}} + U_{\mathrm{MMD}}.
\end{aligned}$$

Hence, continuing from (39), the type II error is upper bounded by

$$\mathbb{P}\Bigg(U_{\mathrm{MMD}} \le \frac{C_1}{n} + \frac{C_2}{n^{3/2}\sqrt{\lambda_1\cdots\lambda_d}} + \frac{C_3\log(1/\alpha)}{n\sqrt{\lambda_1\cdots\lambda_d}} + \frac{C_4\Delta_{V^{1/2}}^2\log^2(1/\alpha)}{\xi_{\varepsilon,\delta}^{-2}} \\ + 2C_5\sqrt{b_{n,m,\boldsymbol{\lambda}} + \frac{C_6\log(1/\alpha)}{n\sqrt{\lambda_1\cdots\lambda_d}}}\frac{\Delta_{V^{1/2}}\log(1/\alpha)}{\xi_{\varepsilon,\delta}}\Bigg) + 7\beta/8. \tag{41}$$

Recall that we assume $\alpha \in (0, e^{-1})$ and $\lambda_1\cdots\lambda_d \le 1$. Moreover, for the Gaussian kernel, Lemma 5 shows that the sensitivity of $\sqrt{V_{\mathrm{MMD}}}$ satisfies

$$\Delta_{V^{1/2}} \le \frac{C_6}{n\sqrt{\lambda_1\ldots\lambda_d}}.$$

Using these conditions along with the inequality (38) for $b_{n,m\boldsymbol{\lambda}}$, the previous bound (41) can be upper bounded as

$$\mathbb{P}\Bigg(U_{\mathrm{MMD}} \le \frac{C_1\log(1/\alpha)}{n\sqrt{\lambda_1\cdots\lambda_d}} + \frac{C_2\log^2(1/\alpha)}{n^2\lambda_1\cdots\lambda_d\xi_{\varepsilon,\delta}^2} \\ + \frac{C_3\log(1/\alpha)}{n^{3/2}\lambda_1\cdots\lambda_d\xi_{\varepsilon,\delta}} + \frac{C_4\log^{3/2}(1/\alpha)}{n^{3/2}(\lambda_1\cdots\lambda_d)^{3/4}\xi_{\varepsilon,\delta}}\Bigg) + 7\beta/8.$$

Condition in terms of L_2 distance. As shown in Schrab et al. (2023, Lemma 2), a sufficient condition for the first probability term in the above display to be less than $\beta/8$ is

$$\mathrm{MMD}_{k_{\boldsymbol{\lambda}}}^2 \ge C_5\sqrt{\mathrm{Var}[U_{\mathrm{MMD}}]} + \frac{C_1\log(1/\alpha)}{n\sqrt{\lambda_1\cdots\lambda_d}} + \frac{C_2\log^2(1/\alpha)}{n^2\lambda_1\cdots\lambda_d\xi_{\varepsilon,\delta}^2} \\ + \frac{C_3\log(1/\alpha)}{n^{3/2}\lambda_1\cdots\lambda_d\xi_{\varepsilon,\delta}} + \frac{C_4\log^{3/2}(1/\alpha)}{n^{3/2}(\lambda_1\cdots\lambda_d)^{3/4}\xi_{\varepsilon,\delta}}.$$

Moreover, writing the difference of two densities as $\psi = p - q$ and the convolution of ψ and $k_{\boldsymbol{\lambda}}$ as $\psi * k_{\boldsymbol{\lambda}}$, the proof of Schrab et al. (2023, Theorem 5) yields that the previous condition is implied by

$$\|\psi\|_{L_2}^2 \ge \|\psi - \psi*\varphi_\lambda\|_{L_2}^2 + C_6\Bigg\{\frac{\log(1/\alpha)}{n\sqrt{\lambda_1\cdots\lambda_d}} + \frac{\log^2(1/\alpha)}{n^2\lambda_1\cdots\lambda_d\xi_{\varepsilon,\delta}^2} \\ + \frac{\log(1/\alpha)}{n^{3/2}\lambda_1\cdots\lambda_d\xi_{\varepsilon,\delta}} + \frac{\log^{3/2}(1/\alpha)}{n^{3/2}(\lambda_1\cdots\lambda_d)^{3/4}\xi_{\varepsilon,\delta}}\Bigg\}.$$

Lastly, the proof of Schrab et al. (2023, Theorem 6) yields that over the Sobolev ball, a sufficient condition for the previous inequality is

$$\|\psi\|_{L_2}^2 \ge C_7\Bigg\{\sum_{i=1}^d \lambda_i^{2s} + \frac{\log(1/\alpha)}{n\sqrt{\lambda_1\cdots\lambda_d}} + \frac{\log^2(1/\alpha)}{n^2\lambda_1\cdots\lambda_d\xi_{\varepsilon,\delta}^2}$$

$$+\frac{\log(1/\alpha)}{n^{3/2}\lambda_1\cdots\lambda_d\xi_{\varepsilon,\delta}}+\frac{\log^{3/2}(1/\alpha)}{n^{3/2}(\lambda_1\cdots\lambda_d)^{3/4}\xi_{\varepsilon,\delta}}\Bigg\}$$

as claimed.

E.12 Proof of Lemma 7

Let U'_{MMD} be the U-statistic similarly defined as U_{MMD} by replacing Y_1 with Y'_1. Then it can be seen that

$$\begin{aligned}|U_{\mathrm{MMD}}-U'_{\mathrm{MMD}}| &= \left|\frac{2}{n(n-1)}\sum_{i=2}^{n}\left\{k(Y_1,Y_j)-k(Y'_1,Y_j)\right\}-\frac{2}{nm}\sum_{j=1}^{m}\left\{k(Y_1,Z_j)-k(Y'_1,Z_j)\right\}\right| \\ &\leq \frac{8K}{n}.\end{aligned}$$

Similarly, the difference is bounded by $8K/n$ for other neighboring datasets, and thus the global sensitivity of U_{MMD} is bounded by $8K/n$. Let ϵ be a number between $(0,K)$. For a translation invariant kernel with non-empty level sets, we can ensure the existence of an instance where $Y_1=\ldots=Y_n$ and $Z_1=\ldots=Z_m$ such that $k(Y_1,Z_1)=\epsilon_\star$ for some $\epsilon_\star\in[0,\epsilon]$. Moreover, by taking $Y'_1=Z_1$, we have $k(Y_1,Y'_1)=\epsilon_\star$ and $k(Y'_1,Z_1)=K$. Under this setting, the difference between U_{MMD} and U'_{MMD} becomes

$$|U_{\mathrm{MMD}}-U'_{\mathrm{MMD}}|=\frac{4(K-\epsilon_\star)}{n}.$$

Since ϵ (and so $\epsilon_\star$) can be arbitrarily small, we see that the global sensitivity defined with the supremum is lower bounded by $4K/n$. This concludes the claim of Lemma 7. This concludes the claim of Lemma 7.

E.13 Proof of Theorem 10

For notational convenience, let us write $U_{\mathrm{MMD}}(\mathcal{X}^{\boldsymbol{\pi}_0}_{n+m})=U_{\mathrm{MMD}}$ and $U_{\mathrm{MMD}}(\mathcal{X}^{\boldsymbol{\pi}_i}_{n+m})=U_{\boldsymbol{\pi}_i,\mathrm{MMD}}$ for $i\in[B]$. For $\alpha>1/(B+1)$, Lemma 18 yields that

$$\phi^{u}_{\mathrm{dpMMD}}=\mathbb{1}\left(U_{\mathrm{MMD}}+\frac{2c_{m,n}K}{n\xi_{\varepsilon,\delta}}\zeta_0>r_{1-\alpha_\star}\right),$$

where $\alpha_\star=\frac{B+1}{B}\alpha-\frac{1}{B}$ and $r_{1-\alpha_\star}$ is the $1-\alpha_\star$ quantile of $\{U_{\boldsymbol{\pi}_i,\mathrm{MMD}}+\frac{2c_{m,n}K}{n\xi_{\varepsilon,\delta}}\zeta_i\}_{i=1}^{B}$. Let us denote the V-statistic of MMD as

$$V_{\mathrm{MMD}}=\frac{1}{n^2}\sum_{i,j=1}^{n}k(Y_i,Y_j)+\frac{1}{m^2}\sum_{i,j=1}^{m}k(Z_i,Z_j)-\frac{2}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}k(Y_i,Z_j),$$

which is greater than or equal to U_{MMD} since

$$\frac{1}{n^2}\sum_{i,j=1}^{n}k(Y_i,Y_j)\geq\frac{1}{n(n-1)}\sum_{(i,j)\in\mathbf{i}_2^n}k(Y_i,Y_j)$$

$$\iff \frac{1}{n}\sum_{i=1}^{n} k(Y_i, Y_i) = K \geq \frac{1}{n(n-1)} \sum_{(i,j)\in \mathbf{i}_2^n} k(Y_i, Y_j)$$

and similarly

$$\frac{1}{m^2}\sum_{i,j=1}^{m} k(Z_i, Z_j) \geq \frac{1}{m(m-1)} \sum_{(i,j)\in \mathbf{i}_2^m} k(Z_i, Z_j)$$

$$\iff \frac{1}{m}\sum_{i=1}^{m} k(Z_i, Z_i) = K \geq \frac{1}{m(m-1)} \sum_{(i,j)\in \mathbf{i}_2^m} k(Z_i, Z_j).$$

Moreover, since $V_{\mathrm{MMD}} \geq 0$, the U-statistic is lower bounded by $U_{\mathrm{MMD}} \geq U_{\mathrm{MMD}} - V_{\mathrm{MMD}}$. Consider a lower bound for $U_{\mathrm{MMD}} - V_{\mathrm{MMD}}$ as

$$\begin{aligned}
U_{\mathrm{MMD}} - V_{\mathrm{MMD}} &= \frac{1}{n(n-1)} \sum_{(i,j)\in \mathbf{i}_2^n} k(Y_i, Y_j) - \frac{1}{n^2}\sum_{i,j=1}^{n} k(Y_i, Y_j) \\
&\quad + \frac{1}{m(m-1)} \sum_{(i,j)\in \mathbf{i}_2^m} k(Z_i, Z_j) - \frac{1}{m^2}\sum_{i,j=1}^{m} k(Z_i, Z_j) \\
&= \frac{1}{n}\left[\frac{1}{n(n-1)} \sum_{(i,j)\in \mathbf{i}_2^n} k(Y_i, Y_j) - \frac{1}{n}\sum_{i=1}^{n} k(Y_i, Y_i)\right] \\
&\quad + \frac{1}{m}\left[\frac{1}{m(m-1)} \sum_{(i,j)\in \mathbf{i}_2^m} k(Z_i, Z_j) - \frac{1}{m}\sum_{i=1}^{m} k(Z_i, Z_i)\right] \\
&\geq -\frac{K}{n} - \frac{K}{m}.
\end{aligned}$$

Thereby, $U_{\mathrm{MMD}} \geq -\frac{K}{n} - \frac{K}{m}$. We use this observation to lower bound the critical value $r_{1-\alpha_\star}$ as

$$r_{1-\alpha_\star} \geq \frac{2c_{m,n}K}{n\xi_{\varepsilon,\delta}} \mathrm{Quantile}_{1-\alpha_\star}\big\{\zeta_1, \ldots, \zeta_B\big\} - \frac{K}{n} - \frac{K}{m},$$

where $\mathrm{Quantile}_{1-\alpha_\star}\{\zeta_1, \ldots, \zeta_B\}$ denotes the $1-\alpha_\star$ quantile of $\zeta_1, \ldots, \zeta_B$. Putting pieces together yields

$$\begin{aligned}
&\phi^u_{\mathrm{dpMMD}} \\
&\leq \mathbb{1}\left(V_{\mathrm{MMD}} + \frac{2c_{m,n}K}{n\xi_{\varepsilon,\delta}}\zeta_0 > \frac{2c_{m,n}K}{n\xi_{\varepsilon,\delta}} \mathrm{Quantile}_{1-\alpha_\star}\big\{\zeta_1, \ldots, \zeta_B\big\} - \frac{K}{n} - \frac{K}{m}\right) \\
&\leq \mathbb{1}\left(\frac{n\xi_{\varepsilon,\delta}}{2c_{m,n}K} V_{\mathrm{MMD}} + \zeta_0 > \mathrm{Quantile}_{1-\alpha_\star}\big\{\zeta_1, \ldots, \zeta_B\big\} - \frac{\xi_{\varepsilon,\delta}}{c_{m,n}}\right).
\end{aligned}$$

Further note that the statistic V_{MMD} is the square of the empirical MMD, which satisfies

$$V_{\mathrm{MMD}} = \big\{\mathrm{MMD}_k(P, Q) + R_{m,n}\big\}^2,$$

where $R_{m,n} = O_P(n^{-1/2})$ due to Gretton et al. (2012, Theorem 7), recalled in Lemma 13.

Let us consider $P_0, Q_0 \in \mathcal{P}_{\mathbb{S}}$ such that $\mathrm{MMD}_k(P_0, Q_0) = \varrho_0$ in the theorem statement. Let $P = P_0$, and let Q be a mixture distribution $Q = wP_0 + (1-w)Q_0$. Since $\mathcal{P}_{\mathbb{S}}$ is a convex set, Q belongs to $\mathcal{P}_{\mathbb{S}}$, and it can be seen that

$$\mathrm{MMD}_k(P, Q) = (1-w)\varrho_0.$$

Now for $\gamma \in (1/2, 1)$, we set $w = 1 - (n\xi_{\varepsilon,\delta})^{-\gamma}$. Since $\xi_{\varepsilon,\delta} \asymp n^{-1/2-r}$ with $r \in (0, 1/2)$, we can ensure that $w \in (0,1)$ and $\mathrm{MMD}_k(P,Q) = (n\xi_{\varepsilon,\delta})^{-\gamma}\varrho_0$ for sufficiently large n. Moreover, this pair of distributions (P, Q) belongs to $\mathcal{P}_{\mathrm{MMD}_k}(\rho)$ for ρ as in (17) since

$$\begin{aligned}\mathrm{MMD}_k(P,Q) = \frac{\varrho_0}{\big(n\xi_{\varepsilon,\delta}\big)^{\gamma}} \;\geq\; \rho \;\asymp\; \frac{\log(n)}{n\xi_{\varepsilon,\delta}} \quad &\Longleftrightarrow \quad \big(n\xi_{\varepsilon,\delta}\big)^{1-\gamma} \;\gtrsim\; \log(n) \\ &\Longleftrightarrow \quad n^{\frac{(1-2r)(1-\gamma)}{2}} \;\gtrsim\; \log(n).\end{aligned}$$

Moreover, as $n\xi_{\varepsilon,\delta} \to \infty$ and $\xi_{\varepsilon,\delta} \to 0$,

$$\frac{n\xi_{\varepsilon,\delta}}{2c_{m,n}K} V_{\mathrm{MMD}} = \frac{1}{2c_{m,n}K}\big\{\sqrt{n\xi_{\varepsilon,\delta}}\mathrm{MMD}_k(P,Q) + \sqrt{n\xi_{\varepsilon,\delta}}R_{m,n}\big\}^2 = o_P(1),$$

where we use the fact that $c_{m,n} \in [4, 8]$. Having this observation in place, for any fixed $t > 0$, we can take $N_t > 0$ such that for all $n \geq N_t$, it holds that

$$\begin{aligned}\mathbb{E}[\phi^u_{\mathrm{dpMMD}}] \;&\leq\; \mathbb{P}\Big(\zeta_0 > \mathrm{Quantile}_{1-\alpha_\star}\big\{\zeta_1, \ldots, \zeta_B\big\} - \frac{\xi_{\varepsilon,\delta}}{c_{m,n}} - \frac{n\xi_{\varepsilon,\delta}}{2c_{m,n}K} V_{\mathrm{MMD}}, \\ &\qquad\qquad \Big|\frac{\xi_{\varepsilon,\delta}}{c_{m,n}} + \frac{n\xi_{\varepsilon,\delta}}{2c_{m,n}K} V_{\mathrm{MMD}}\Big| < t\Big) + \mathbb{P}\Big(\Big|\frac{\xi_{\varepsilon,\delta}}{c_{m,n}} + \frac{n\xi_{\varepsilon,\delta}}{2c_{m,n}K} V_{\mathrm{MMD}}\Big| \geq t\Big) \\ &\leq\; \mathbb{P}\Big(\zeta_0 > \mathrm{Quantile}_{1-\alpha_\star}\big\{\zeta_1, \ldots, \zeta_B\big\} - t\Big) + t \\ &=\; \mathbb{E}\Big[\mathbb{P}\Big(\zeta_0 > \mathrm{Quantile}_{1-\alpha_\star}\big\{\zeta_1, \ldots, \zeta_B\big\} - t \,\Big|\, \zeta_1, \ldots, \zeta_B\Big)\Big] + t \\ &\leq\; \mathbb{P}\Big(\zeta_0 > \mathrm{Quantile}_{1-\alpha_\star}\big\{\zeta_1, \ldots, \zeta_B\big\}\Big) + t\|f_\zeta\|_{L_\infty} + t,\end{aligned}$$

where f_ζ denotes the density function of $\mathsf{Laplace}(0,1)$ and the last inequality uses the fact that

$$\big|\mathbb{P}(\zeta > a + b) - \mathbb{P}(\zeta > b)\big| \leq \|f_\zeta\|_{L_\infty}|a| \quad \text{for all } a, b \in \mathbb{R}.$$

Since $\|f_\zeta\|_{L_\infty} \leq \frac{1}{2}$ and

$$\mathbb{P}\Big(\zeta_0 > \mathrm{Quantile}_{1-\alpha_\star}\big\{\zeta_1, \ldots, \zeta_B\big\}\Big) = \mathbb{P}\Big(\zeta_0 > \mathrm{Quantile}_{1-\alpha}\big\{\zeta_0, \zeta_1, \ldots, \zeta_B\big\}\Big) \leq \alpha,$$

we have

$$\limsup_{n\to\infty} \inf_{(P,Q)\in\mathcal{P}_{\mathrm{MMD}_k}(\rho)} \mathbb{E}_{P,Q}[\phi^u_{\mathrm{dpMMD}}] \leq \alpha + \frac{3t}{2}.$$

The result follows as t can be made arbitrarily small.

E.14 Proof of Theorem 11

Let us denote by $C_1, C_2, \ldots$ constants that may depend on τ, β, s, R, M, d. Following the proofs of Theorem 4 and Theorem 9 along with the sensitivity result for U_{MMD} in Lemma 7, we may arrive at the point where the type II error of ϕ^u_{dpMMD} is upper bounded as

$$\mathbb{E}[1 - \phi^u_{\mathrm{dpMMD}}] \leq \mathbb{P}\left(U_{\mathrm{MMD}} \leq \frac{C_1 \log(1/\alpha)}{n\sqrt{\lambda_1 \cdots \lambda_d}} + \frac{C_2 \log(1/\alpha)}{n\lambda_1 \cdots \lambda_d \xi_{\varepsilon,\delta}} \right) + 7\beta/8.$$

We then use the proofs of Schrab et al. (2023, Theorem 5) and Schrab et al. (2023, Theorem 6), and show that the probability term in the above display is less than or equal to $\beta/8$ once

$$\|p - q\|_{L_2}^2 \;\geq\; C_3 \left\{ \sum_{i=1}^{d} \lambda_i^{2s} + \frac{\log(1/\alpha)}{n\sqrt{\lambda_1 \cdots \lambda_d}} + \frac{\log(1/\alpha)}{n\lambda_1 \cdots \lambda_d \xi_{\varepsilon,\delta}} \right\}.$$

This proves Theorem 11.

F Proofs for Appendix B

This section collects the proofs of the results in Appendix B.

F.1 Proof of Proposition 1

Focusing on MMD, the square of the empirical MMD satisfies

$$(n+m)\widehat{\mathrm{MMD}}^2(\mathcal{X}_{n+m}) \xrightarrow{d} \sum_{i=1}^{\infty} \lambda_i Z_i^2,$$

which can be seen by the standard asymptotic theory of V-statistics. See Fernández and Rivera (2022, Proposition 9). Moreover, note that the empirical MMD and the Laplace noise are independent. Therefore, the continuous mapping theorem along with Slutsky's theorem proves the result when $\sqrt{n+m}\sigma \to \eta \in [0, \infty)$. On the other hand, when $\sqrt{n+m}\sigma \to \infty$, it holds that

$$\sigma^{-1} M_{\mathrm{MMD}} = \underbrace{\frac{1}{\sqrt{n+m}\,\sigma}}_{=o_P(1)} \times \underbrace{\sqrt{n+m}\,\widehat{\mathrm{MMD}}(\mathcal{X}_{n+m})}_{=O_P(1)} + \zeta = o_P(1) + \zeta.$$

This proves the results for MMD. The proof for HSIC is completely analogous and thereby omitted, which can be derived by using Zhang et al. (2018, Theorem 1) in place of Fernández and Rivera (2022, Proposition 9).

F.2 Proof of Lemma 8

We first prove the claim by considering two scenarios: (i) B_n is fixed in n and (ii) B_n increases with n, and prove the consistency result. We then turn to a general case of B_n and complete the proof building on the prior result. Throughout this proof, we often omit the dependence of n on B_n and write it as B for simplicity.

F.2.1 Simple cases

Fixed B. Assume that B is fixed and $B+1 > \alpha^{-1}$ or equivalently $\alpha(B+1) \geq 1$. As mentioned in the main text, we simply exploit the union bound to prove the claim for fixed B. Concretely,

$$\begin{aligned}
\mathbb{P}\Big[\frac{1}{B+1}\Big(\sum_{i=1}^{B}\mathbb{1}(W_{0,n} \leq W_{i,n}) + 1\Big) \leq \alpha\Big] &= \mathbb{P}\Big[\sum_{i=1}^{B}\mathbb{1}(W_{0,n} \leq W_{i,n}) \leq \alpha(B+1) - 1\Big] \\
&\overset{\text{(i)}}{\geq} \mathbb{P}\Big[\sum_{i=1}^{B}\mathbb{1}(W_{0,n} \leq W_{i,n}) \leq 0\Big] \\
&= \mathbb{P}\Big[W_{0,n} - \max_{i\in[B]} W_{i,n} > 0\Big] \\
&\overset{\text{(ii)}}{\geq} 1 - \sum_{i=1}^{B}\mathbb{P}(W_{0,n} \leq W_{i,n}),
\end{aligned}$$

where step (i) uses condition $\alpha(B+1) > 1$ and step (ii) uses de Morgan's law and the union bound. By the condition that $\lim_{n\to\infty}\mathbb{P}(W_{0,n} \leq W_{1,n}) = 0$ and since $\{W_{0,n} - W_{i,n}\}_{i=2}^{B}$ are identically distributed as $W_{0,n} - W_{1,n}$, the lower bound converges to one for fixed B and thus the consistency result follows.

Increasing B. Next we consider the case where B increases to infinity with n. For some $\eta > 0$ (specified later), define an event $\mathcal{A}$ as

$$\mathcal{A} = \Big\{\Big|\frac{1}{B}\sum_{i=1}^{B}\mathbb{1}(W_{0,n} \leq W_{i,n}) - \mathbb{P}(W_{0,n} \leq W_{1,n} \,|\, \mathcal{G})\Big| \leq \sqrt{\frac{1}{2B}\log\Big(\frac{2}{\eta}\Big)}\Big\}.$$

Since $\mathbb{1}(W_{0,n} \leq W_{1,n}), \ldots, \mathbb{1}(W_0 \leq W_{B,n})$ are i.i.d. random variables conditional on the sigma field $\mathcal{G}$ under the condition of Lemma 8, Hoeffding's inequality yields that

$$\mathbb{P}(\mathcal{A}^c \,|\, \mathcal{G}) \leq \eta.$$

By taking the expectation over $\mathcal{G}$ on both sides, it also holds marginally that $\mathbb{P}(\mathcal{A}^c) \leq \eta$. Now

$$\begin{aligned}
&\mathbb{P}\Big[\frac{1}{B+1}\Big(\sum_{i=1}^{B}\mathbb{1}(W_{0,n} \leq W_{i,n}) + 1\Big) > \alpha\Big] \\
&= \mathbb{P}\Big[\frac{1}{B}\sum_{i=1}^{B}\mathbb{1}(W_{0,n} \leq W_{i,n}) > \frac{B+1}{B}\Big(\alpha - \frac{1}{B+1}\Big)\Big] \\
&\leq \mathbb{P}\Big[\frac{1}{B}\sum_{i=1}^{B}\mathbb{1}(W_{0,n} \leq W_{i,n}) > \frac{B+1}{B}\Big(\alpha - \frac{1}{B+1}\Big), \mathcal{A}\Big] + \mathbb{P}(\mathcal{A}^c) \\
&\leq \mathbb{P}\Big[\mathbb{P}(W_{0,n} \leq W_{1,n} \,|\, \mathcal{G}) + \sqrt{\frac{1}{2B}\log\Big(\frac{2}{\eta}\Big)} > \frac{B+1}{B}\Big(\alpha - \frac{1}{B+1}\Big)\Big] + \eta.
\end{aligned} \tag{42}$$

By taking $\eta = B^{-1}$, we know $\eta \to 0$ (since $B \to \infty$) and

$$\frac{B+1}{B}\left(\alpha - \frac{1}{B+1}\right) - \sqrt{\frac{1}{2B}\log\left(\frac{2}{\eta}\right)} \to \alpha \quad \text{as } n \to \infty.$$

Moreover our condition guarantees that $\lim_{n\to\infty} \mathbb{P}(W_{0,n} \le W_{1,n}) = 0$. Hence for any $\epsilon > 0$, we can take N such that statement (i) $\eta < \epsilon/2$ and statement (ii)

$$\begin{aligned}
&\mathbb{P}\left[\mathbb{P}(W_{0,n} \le W_{1,n} \,|\, \mathcal{G}) + \sqrt{\frac{1}{2B}\log\left(\frac{2}{\eta}\right)} > \frac{B+1}{B}\left(\alpha - \frac{1}{B+1}\right)\right] \\
\le\ &\mathbb{P}\left[\mathbb{P}(W_{0,n} \le W_{1,n} \,|\, \mathcal{G}) > \frac{\alpha}{2}\right] \qquad (43) \\
\overset{(\dagger)}{\le}\ &\frac{2}{\alpha}\mathbb{P}(W_{0,n} \le W_{1,n}) \le \frac{\epsilon}{2},
\end{aligned}$$

hold for all $n \ge N$. Here step (†) holds by Markov's inequality. Thus for all $n \ge N$ we have

$$\mathbb{P}\left[\frac{1}{B_n+1}\left(\sum_{i=1}^{B_n} \mathbb{1}(W_{0,n} \le W_{i,n}) + 1\right) > \alpha\right] < \epsilon.$$

Since ϵ is arbitrary, the above argument proves the claim that

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{1}{B_n+1}\left(\sum_{i=1}^{B_n} \mathbb{1}(W_{0,n} \le W_{i,n}) + 1\right) \le \alpha\right] = 1.$$

F.2.2 Main proof

We now deal with an arbitrary sequence of B_n such that $\inf_{n\ge1} B_n + 1 > \alpha^{-1}$ and prove the claim. First note that

$$\begin{aligned}
&\mathbb{P}\left[\frac{1}{B+1}\left(\sum_{i=1}^{B} \mathbb{1}(W_{0,n} \le W_{i,n}) + 1\right) > \alpha\right] = \mathbb{P}\left[\frac{1}{B+1}\sum_{i=1}^{B} \mathbb{1}(W_{0,n} \le W_{i,n}) > \left(\alpha - \frac{1}{B+1}\right)\right] \\
\overset{(\mathrm{i})}{\le}\ &\mathbb{P}\left[\sum_{i=1}^{B} \mathbb{1}(W_{0,n} \le W_{i,n}) > 0\right] = 1 - \mathbb{P}\left[\sum_{i=1}^{B} \mathbb{1}(W_{0,n} \le W_{i,n}) = 0\right] \\
=\ &1 - \mathbb{E}\left\{\mathbb{P}\left[\sum_{i=1}^{B} \mathbb{1}(W_{0,n} \le W_{i,n}) = 0 \,\middle|\, \mathcal{G}\right]\right\} = 1 - \mathbb{E}\left[\{\mathbb{P}(W_{0,n} > W_{1,n} \,|\, \mathcal{G})\}^B\right] \\
\overset{(\mathrm{ii})}{\le}\ &1 - \{\mathbb{P}(W_{0,n} > W_{1,n})\}^B,
\end{aligned}$$

where step (i) uses the condition $\alpha > \frac{1}{B+1}$ and step (ii) uses Jensen's inequality. On the other hand, from the previous results (42) and Markov's inequality with $\eta = B^{-1}$, we have

$$\mathbb{P}\left[\frac{1}{B+1}\left(\sum_{i=1}^{B} \mathbb{1}(W_{0,n} \le W_{i,n}) + 1\right) > \alpha\right] \le \frac{1}{B} + \frac{\mathbb{P}(W_{0,n} \le W_{1,n}) + \sqrt{\frac{1}{2B}\log(2B)}}{\frac{B+1}{B}(\alpha - \frac{1}{B+1})}.$$

Thus combining the two bounds yields

$$\mathbb{P}\left[\frac{1}{B+1}\left(\sum_{i=1}^{B}\mathbb{1}(W_{0,n}\le W_{i,n})+1\right)>\alpha\right]$$

$$\le \min\left\{1-\{\mathbb{P}(W_{0,n}>W_{1,n})\}^B,\ \frac{1}{B}+\frac{\mathbb{P}(W_{0,n}\le W_{1,n})+\sqrt{\frac{1}{2B}\log(2B)}}{\frac{B+1}{B}(\alpha-\frac{1}{B+1})}\right\}.$$

Interestingly, a careful argument shows that the above upper bound can be made independent of B. Concretely, write $p_n \coloneqq \mathbb{P}(W_{0,n} > W_{1,n})$ for simplicity, which by assumption tends to 1 as n tends to infinity, and let b_n be a positive sequence that goes to infinity and satisfies $p_n^{b_n} \to 1$. For instance, one can take $b_n = \sqrt{-\log(1-p_n)}$. Using this notation in place, observe

$$\min\left\{1-\{\mathbb{P}(W_{0,n}>W_{1,n})\}^B,\ \frac{1}{B}+\frac{\mathbb{P}(W_{0,n}\le W_{1,n})+\sqrt{\frac{1}{2B}\log(2B)}}{\frac{B+1}{B}(\alpha-\frac{1}{B+1})}\right\}$$

$$\le \mathbb{1}(B<b_n)\{1-p_n^B\}+\mathbb{1}(B\ge b_n)\left\{\frac{1}{B}+\frac{1-p_n+\sqrt{\frac{1}{2B}\log(2B)}}{\frac{B+1}{B}(\alpha-\frac{1}{B+1})}\right\}$$

$$\le 1-p_n^{b_n}+\left\{\frac{1}{b_n}+\frac{1-p_n+\sqrt{\frac{1}{2b_n}\log(2b_n)}}{\frac{b_n+1}{b_n}(\alpha-\frac{1}{b_n+1})}\right\}.$$

Now the upper bound does not depend on B and converges to zero, which is ensured by the condition $\lim_{n\to\infty} p_n = 1$. As a result, we have

$$\lim_{n\to\infty}\mathbb{P}\left[\frac{1}{B_n+1}\left(\sum_{i=1}^{B_n}\mathbb{1}(W_{0,n}\le W_{i,n})+1\right)\le\alpha\right]=1,$$

and complete the proof of Lemma 8.

F.3 Proof of Theorem 12

The proof of Theorem 12 can follow a similar approach to that of Theorem 7 by replacing concentration inequalities for MMD statistics (Lemma 10 and Lemma 13) with the corresponding ones for HSIC statistics (Lemma 12 and Lemma 14). Throughout this proof, we denote positive constants that only depend on K and L by $C_1, C_2, C_3, \ldots$

Step 1 (Bounding the quantile). Let us remind that M_i is defined as $\widehat{\mathrm{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}_i}) + 2\Delta_T\xi_{\varepsilon,\delta}^{-1}\zeta_i$ for $i \in \{0\}\cup[B]$. We first examine the $1-\alpha$ quantile of $M_0, M_1, \ldots, M_B$ denoted as $q_{1-\alpha,B}$ and establish an upper bound for $q_{1-\alpha,B}$ with high probability. Again, the proof strategy is similar to that of Theorem 7. With an abuse of notation, let $q^a_{1-\alpha/2,B}$ and $q^b_{1-\alpha/2,B}$ be the $1-\alpha/2$ quantiles of $\{\widehat{\mathrm{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}_0}),\ldots,\widehat{\mathrm{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}_B})\}$ and of $\{2\Delta_T\xi_{\varepsilon,\delta}^{-1}\zeta_0,\ldots,2\Delta_T\xi_{\varepsilon,\delta}^{-1}\zeta_B\}$, respectively. Denote by $q^a_{1-\alpha/2,\infty}$ and $q^b_{1-\alpha/2,\infty}$ the corresponding $1-\alpha/2$ quantiles with $B=\infty$. Following the same lines of

the proof of Theorem 7, we see that $q_{1-\alpha,B} \leq q^a_{1-\alpha/4,\infty} + q^b_{1-\alpha/4,\infty}$ with probability at least $1-\beta/2$ under our condition for B given as $B \geq 16\alpha^{-2}\log\bigl(8/\beta\bigr)$. Moreover applying Lemma 12 yields

$$q^a_{1-\alpha/4,\infty} \leq C_1\sqrt{\frac{1}{n}\max\left\{\log\left(\frac{4}{\alpha}\right), \log^{1/2}\left(\frac{4}{\alpha}\right), 1\right\}},$$

whereas $q^b_{1-\alpha/4,\infty} = 2\Delta_T\xi^{-1}_{\varepsilon,\delta}F^{-1}_{\zeta}(1-\alpha/4)$. Therefore, with probability at least $1-\beta/2$, it holds that

$$q_{1-\alpha,B} \leq C_1\sqrt{\frac{1}{n}\max\left\{\log\left(\frac{4}{\alpha}\right), \log^{1/2}\left(\frac{4}{\alpha}\right), 1\right\}} + 2\Delta_T\xi^{-1}_{\varepsilon,\delta}F^{-1}_{\zeta}(1-\alpha/4). \tag{44}$$

Step 2 (Bounding the type II error). For the empirical HSIC, Lemma 14 shows that the following event

$$E'_1 := \left\{\left|\mathrm{HSIC}_{k\otimes\ell}(P_{YZ}) - \widehat{\mathrm{HSIC}}(\mathcal{X}_n)\right| \leq C_2\sqrt{\frac{\log(8/\beta)}{n}}\right\}$$

holds with probability at least $1-\beta/4$. On the other hand, the condition $\zeta_0 \sim \mathsf{Laplace}(0,1)$ ensures that the probability of the event

$$E'_2 := \left\{2\Delta_T\xi^{-1}_{\varepsilon,\delta}\zeta_0 > 2\Delta_T\xi^{-1}_{\varepsilon,\delta}F^{-1}_{\zeta}(\beta/4)\right\}$$

is equal to $1-\beta/4$ where F^{-1}_{ζ} is the inverse cumulative distribution function of ζ. Using these results as well as inequality (44) from Step 1, it can be seen similarly to the proof of Theorem 7 that

$$\begin{aligned}
\mathbb{E}[1-\phi_{\mathrm{dpHSIC}}] \;&\leq\; \mathbb{P}\left(\mathrm{HSIC}_{k\otimes\ell}(P_{YZ}) \leq q_{1-\alpha,B} + C_2\sqrt{\frac{\log(8/\beta)}{n}} - 2\Delta_T\xi^{-1}_{\varepsilon,\delta}F^{-1}_{\zeta}(\beta/4)\right) + \frac{\beta}{2} \\
&\leq\; \mathbb{P}\left(\mathrm{HSIC}_{k\otimes\ell}(P_{YZ}) \leq C_1\sqrt{\frac{1}{n}\max\left\{\log\left(\frac{4}{\alpha}\right), \log^{1/2}\left(\frac{4}{\alpha}\right), 1\right\}}\right. \\
&\qquad \left. + 2\Delta_T\xi^{-1}_{\varepsilon,\delta}F^{-1}_{\zeta}(1-\alpha/4) + C_2\sqrt{\frac{\log(8/\beta)}{n}} - 2\Delta_T\xi^{-1}_{\varepsilon,\delta}F^{-1}_{\zeta}(\beta/4)\right) + \beta.
\end{aligned}$$

Next, recall that we assume $\alpha \in (0,1)$ and $\beta \in (0, 1-\alpha)$ under which it holds

$$\max\left\{\log\left(\frac{4}{\alpha}\right), \log^{1/2}\left(\frac{4}{\alpha}\right), 1, \log\left(\frac{8}{\beta}\right)\right\} \;\leq\; C_3\max\left\{\log\left(\frac{1}{\alpha}\right), \log\left(\frac{1}{\beta}\right)\right\}. \tag{45}$$

To see this inequality, first assume that $\alpha \in (0, 1/2)$. Then

$$\begin{aligned}
\max\left\{\log\left(\frac{4}{\alpha}\right), \log^{1/2}\left(\frac{4}{\alpha}\right), 1\right\} \;&\leq\; C_4\log\left(\frac{1}{\alpha}\right) \leq C_4\max\left\{\log\left(\frac{1}{\alpha}\right), \log\left(\frac{8}{\beta}\right)\right\} \\
&\leq\; C_5\max\left\{\log\left(\frac{1}{\alpha}\right), \log\left(\frac{1}{\beta}\right)\right\}
\end{aligned}$$

as $\beta \in (0, 1-\alpha)$. On the other hand, if $\alpha \in [1/2, 1)$, β should be less than $1/2$, which implies that

$$\begin{aligned}\max\left\{\log\left(\frac{4}{\alpha}\right), \log^{1/2}\left(\frac{4}{\alpha}\right), 1\right\} &\leq C_6 \log\left(\frac{8}{\beta}\right) \leq C_7 \log\left(\frac{1}{\beta}\right) \\ &\leq C_7 \max\left\{\log\left(\frac{1}{\alpha}\right), \log\left(\frac{1}{\beta}\right)\right\}.\end{aligned}$$

Therefore inequality (45) follows.

Having these ingredients in place, we follow the same lines of the proof of Theorem 7 and see that

$$\begin{aligned}&\mathbb{E}[1 - \phi_{\mathrm{dpHSIC}}] \\ &\leq \mathbb{P}\left(\mathrm{HSIC}_{k\otimes\ell}(P_{YZ}) \leq C_8 \sqrt{\frac{\max\{\log(1/\alpha), \log(1/\beta)\}}{n}} + C_9 \frac{\max\{\log(1/\alpha), \log(1/\beta)\}}{n\xi_{\varepsilon,\delta}}\right) + \beta \\ &\leq \beta,\end{aligned}$$

where the last inequality holds by taking $C_{K,L}$ to be larger than, for instance, $2\max\{C_8, C_9\} + 1$ in the theorem statement. We conclude the proof by noting that the upper bound is independent of P_{YZ} and taking the supremum on both sides over $\mathcal{P}_{\mathrm{HSIC}_{k\otimes\ell}}(\rho)$.

F.4 Proof of Theorem 13

The proof of Theorem 13 relies on the same idea as that of Theorem 8, which uses Le Cam's two point method and coupling method. Hence we omit the details explained in the proof of Theorem 8, and instead focus on the key differences. As in the proof of Theorem 8, we simply let $\rho^\star_{\mathrm{HSIC}} = \rho^\star_{\mathrm{HSIC}}(\alpha, \beta, \varepsilon, \delta, n)$ and establish the minimax separation by examining the non-privacy regime and privacy regime in order.

F.4.1 Non-privacy regime

We begin by proving that $\rho^\star_{\mathrm{HSIC}} \geq C_{\eta_Y,\eta_Z} \min\{\sqrt{\log(1/(\alpha+\beta))/n}, 1\}$ in the non-privacy regime. We pick one distribution $P_{YZ,0}$ from $\mathcal{P}_{\mathrm{HSIC}_{k\otimes\ell}}(\widetilde{\rho})$ with

$$\widetilde{\rho} = C_{\eta_Y,\eta_Z} \min\left\{\sqrt{\frac{\log(1/(\alpha+\beta))}{n}}, 1\right\}$$

and denote the product of its marginals as $P_{Y,0}P_{Z,0}$. Then, as in Appendix E.10.1, an application of Le Cam's two point method (Le Cam, 1973, 2012) and Bretagnolle–Huber inequality (Canonne, 2022, Lemma B.4) yields

$$\begin{aligned}\inf_{\phi\in\Phi_{\alpha,\varepsilon,\delta}} \sup_{P_{YZ}\in\mathcal{P}_{\mathrm{HSIC}_{k\otimes\ell}}(\widetilde{\rho})} \mathbb{E}_{P_{YZ}^n}[1-\phi] &\geq \inf_{\phi\in\Phi_{\alpha,\infty}} \sup_{P_{YZ}\in\mathcal{P}_{\mathrm{HSIC}_{k\otimes\ell}}(\widetilde{\rho})} \mathbb{E}_{P_{YZ}}[1-\phi] \\ &\geq \inf_{\phi\in\Phi_{\alpha,\infty}} \mathbb{E}_{P_{YZ,0}}[1-\phi] = 1 - \sup_{\phi\in\Phi_{\alpha,\infty}} \mathbb{E}_{P_{YZ,0}}[\phi]\end{aligned}$$

$$\geq 1 - \alpha - d_{\mathrm{TV}}(P_{YZ,0}^{\otimes n}, P_{Y,0}^{\otimes n} P_{Z,0}^{\otimes n})$$
$$\geq \frac{1}{2} e^{-n \times d_{\mathrm{KL}}(P_{YZ,0} \| P_{Y,0} P_{Z,0})} - \alpha.$$

Therefore the minimax type II error is at least β if $\alpha + \beta < 0.4$ as well as

$$d_{\mathrm{KL}}(P_{YZ,0} \| P_{Y,0} P_{Z,0}) \leq \frac{1}{n} \log\left(\frac{1}{2(\alpha+\beta)}\right).$$

Hence $\rho^\star_{\mathrm{HSIC}} \geq C_{\eta_Y, \eta_Z} \min\{\sqrt{\log(1/(\alpha+\beta))/n}, 1\}$ follows if we find $P_{YZ,0}$ such that

$$\mathrm{HSIC}_{k \otimes \ell}(P_{YZ,0}, P_{Y,0} P_{Z,0}) \geq C_{\eta_Y, \eta_Z} \min\left\{\sqrt{\frac{\log(1/(\alpha+\beta))}{n}}, 1\right\} \quad \text{and} \tag{46a}$$

$$d_{\mathrm{KL}}(P_{YZ,0} \| P_{Y,0} P_{Z,0}) \leq \frac{1}{n} \log\left(\frac{1}{2(\alpha+\beta)}\right). \tag{46b}$$

To this end, consider (discrete) random vectors $Y \in \{y_1, y_2\}$ and $Z \in \{z_1, z_2\}$ where $y_1, y_2 \in \mathbb{R}^{d_Y}$ and $z_1, z_2 \in \mathbb{R}^{d_Z}$. Further assume that $\mathbb{P}(Y = y_1) = \mathbb{P}(Y = y_2) = 1/2$ and $\mathbb{P}(Z = z_1) = \mathbb{P}(Z = z_2) = 1/2$. Suppose that the joint probabilities of (Y, Z) are given as

$$\mathbb{P}(Y = y_1, Z = z_1) = \mathbb{P}(Y = y_2, Z = z_2) = 1/4 + \nu \quad \text{and}$$
$$\mathbb{P}(Y = y_1, Z = z_2) = \mathbb{P}(Y = y_2, Z = z_1) = 1/4 - \nu,$$

where $\nu \in (0, 1/4]$. In this setting, it can be seen that $P_{YZ,0} \neq P_{Y,0} P_{Z,0}$ and thus Y and Z are trivially dependent. For such $P_{YZ,0}$ and translation invariant kernels, we have

$$\begin{aligned}
\mathrm{HSIC}_{k \otimes \ell}(P_{YZ,0}) &\overset{\text{(i)}}{=} \sqrt{\mathbb{E}[k(Y_1, Y_2)\ell(Z_1, Z_2)] + \mathbb{E}[k(Y_1, Y_2)\ell(Z_3, Z_4)] - 2\mathbb{E}[k(Y_1, Y_2)\ell(Z_1, Z_3)]} \\
&\overset{\text{(ii)}}{=} \sqrt{4\nu^2 \{\kappa_Y(0) - \kappa_Y(y_1 - y_2)\}\{\kappa_Z(0) - \kappa_Z(z_1 - z_2)\}} \\
&\overset{\text{(iii)}}{\geq} \sqrt{4\nu^2 \eta_Y \eta_Z} = 2\nu\sqrt{\eta_Y \eta_Z},
\end{aligned}$$

where step (i) follows by Gretton et al. (2005, Lemma 1) with $\{(Y_i, Z_i)\}_{i=1}^4 \overset{\text{i.i.d.}}{\sim} P_{YZ,0}$, step (ii) can be verified through algebra and the last inequality (iii) holds by choosing (y_1, y_2) and (z_1, z_2) such that $y_1 - y_2 = y_0$ and $z_1 - z_2 = z_0$.

On the other hand, we verify condition (46b) based on the well-known fact (*e.g.*, Tsybakov, 2009, Lemma 2.7) that the Kullback–Leibler divergence is upper bounded by χ^2 divergence denoted by $d_{\chi^2}(P_{YZ,0} \| P_{Y,0} P_{Z,0})$. This gives

$$\begin{aligned}
d_{\mathrm{KL}}(P_{YZ,0} \| P_{Y,0} P_{Z,0}) &\leq d_{\chi^2}(P_{YZ,0} \| P_{Y,0} P_{Z,0}) = 8\left(\frac{1}{4} + \nu\right)^2 + 8\left(\frac{1}{4} - \nu\right)^2 - 1 \\
&= 16\nu^2.
\end{aligned}$$

Using the above inequality, condition (46b) is fulfilled by taking

$$\nu = \min\left\{\sqrt{\frac{1}{16n}\log\left(\frac{1}{2(\alpha+\beta)}\right)}, \frac{1}{4}\right\}$$

for which condition (46a) is also satisfied as

$$\begin{aligned}\mathrm{HSIC}_{k\otimes\ell}(P_{YZ,0}) &\geq 2\nu\sqrt{\eta_Y\eta_Z} = \frac{\sqrt{\eta_Y\eta_Z}}{2}\min\left\{\sqrt{\frac{1}{n}\log\left(\frac{1}{2(\alpha+\beta)}\right)}, 1\right\} \\ &\geq C_{\eta_Y,\eta_Z}\min\left\{\sqrt{\frac{\log(1/(\alpha+\beta))}{n}}, 1\right\}\end{aligned}$$

as $\alpha+\beta<0.4$. Therefore it holds that $\rho^\star_{\mathrm{HSIC}} \geq C_{\eta_Y,\eta_Z}\min\{\sqrt{\log(1/(\alpha+\beta))/n}, 1\}$ as desired.

F.4.2 Privacy regime

The proof of the separation rate under the privacy regime is essentially the same as that for the dpMMD test in Appendix E.10.2. All we need is to find an instance of $P_{YZ,0}$ such that

$$\mathrm{HSIC}_{k\otimes\ell}(P_{YZ,0}) \geq C_{\eta_Y,\eta_Z}\min\left\{\frac{\log(1/\beta)}{n\big(\varepsilon+\delta\big)}, 1\right\} \quad \text{and} \tag{47a}$$

$$\|P_{YZ,0} - P_{Y,0}P_{Z,0}\|_1 \leq \frac{1}{20n\big(\varepsilon+\delta\big)}\log\left(\frac{1}{4\beta}\right). \tag{47b}$$

To this end, choose the distribution of (Y,Z) as in the case of the non-privacy regime. For such $P_{YZ,0}$, a direct calculation gives $\|P_{YZ,0} - P_{Y,0}P_{Z,0}\|_1 = 4\nu$ and thus condition (47b) is satisfied if we take

$$\nu = \min\left\{\frac{1}{80n\big(\varepsilon+\delta\big)}\log\left(\frac{1}{4\beta}\right), \frac{1}{4}\right\}.$$

On the other hand, the previous calculation of the lower bound for $\mathrm{HSIC}_{k\otimes\ell}(P_{YZ,0})$ yields

$$\begin{aligned}\mathrm{HSIC}_{k\otimes\ell}(P_{YZ,0}) &\geq 2\nu\sqrt{\eta_Y\eta_Z} = 2\min\left\{\frac{1}{80n\big(\varepsilon+\delta\big)}\log\left(\frac{1}{4\beta}\right), \frac{1}{4}\right\}\sqrt{\eta_Y\eta_Z} \\ &\geq C_{\eta_Y,\eta_Z}\min\left\{\frac{\log(1/\beta)}{n\big(\varepsilon+\delta\big)}, 1\right\}\end{aligned}$$

which in turn verifies condition (47a) under $\beta \in (0, 1/5)$. Therefore it holds that the minimax separation under privacy regime $\rho^\star_{\mathrm{HSIC}} \geq C_{\eta_Y,\eta_Z}\min\big\{\log(1/\beta)/(n(\varepsilon+\delta)), 1\big\}$.

F.4.3 Combining bounds

Combining the lower bounds for the non-privacy and privacy regimes, for all $\varepsilon > 0$ and $\delta \in [0, 1)$, we obtain

$$\begin{aligned}\rho^{\star}_{\mathrm{HSIC}} &\geq C_{\eta_Y,\eta_Z} \max\left\{\min\left\{\sqrt{\frac{\log(1/(\alpha+\beta))}{n}}, 1\right\}, \min\left\{\frac{\log(1/\beta)}{n(\varepsilon+\delta)}, 1\right\}\right\} \\ &\geq C_{\eta_Y,\eta_Z} \max\left\{\min\left\{\sqrt{\frac{\log(1/(\alpha+\beta))}{n}}, 1\right\}, \min\left\{\frac{\log(1/\beta)}{n\big(\varepsilon+\log(1/(1-\delta))\big)}, 1\right\}\right\}.\end{aligned}$$

The last inequality holds since $\log(1/(1-\delta)) \geq \delta$ for all $\delta \in [0, 1)$, and is actually tight when $\alpha \asymp \beta$ as explained in Appendix E.10.3. This concludes the proof of Theorem 12.

F.5 Proof of Theorem 14

The proof of Theorem 14 follows the same structure as the proof of Theorem 9 in Appendix E.11. For simplicity, write $\widehat{\mathrm{HSIC}}^2$ given in (12) with the Gaussian kernels as V_{HSIC} and similarly the U-statistic given in (21) with the Gaussian kernel as U_{HSIC}. We also denote the corresponding statistics based on permuted data $\mathcal{X}_n^{\boldsymbol{\pi}}$ as $V_{\boldsymbol{\pi},\mathrm{HSIC}}$ and $U_{\boldsymbol{\pi},\mathrm{HSIC}}$, respectively. Letting

$$K = \prod_{i=1}^{d_Y} \frac{1}{\sqrt{2\pi}\lambda_i} \quad \text{and} \quad L = \prod_{i=1}^{d_Z} \frac{1}{\sqrt{2\pi}\mu_i},$$

Lemma 21 ensures that the difference between V_{HSIC} and U_{HSIC} can be written as

$$V_{\mathrm{HSIC}} - U_{\mathrm{HSIC}} = D_1 + D_2,$$

where

$$D_1 = \frac{n-1}{n^2} KL - \frac{L}{n^3} \sum_{(i,j)\in \mathbf{i}_2^n} k_{\boldsymbol{\lambda}}(Y_i, Y_j) - \frac{K}{n^3} \sum_{(i,j)\in \mathbf{i}_2^n} \ell_{\boldsymbol{\mu}}(Z_i, Z_j),$$

$$\begin{aligned}D_2 = &- \frac{3n^2-4n+2}{(n-1)n^4} \sum_{(i,j)\in \mathbf{i}_2^n} k_{\boldsymbol{\lambda}}(Y_i, Y_j)\ell_{\boldsymbol{\mu}}(Z_i, Z_j) + \frac{2(5n^2-8n+4)}{n^4(n-1)(n-2)} \sum_{(i,j_1,j_2)\in \mathbf{i}_3^n} k_{\boldsymbol{\lambda}}(Y_i, Y_{j_1})\ell_{\boldsymbol{\mu}}(Z_i, Z_{j_2}) \\ &- \frac{6n^2-11n+6}{n^4(n-1)(n-2)(n-3)} \sum_{(i_1,i_2,j_1,j_2)\in \mathbf{i}_4^n} k_{\boldsymbol{\lambda}}(Y_{i_1}, Y_{i_2})\ell_{\boldsymbol{\mu}}(Z_{j_1}, Z_{j_2}).\end{aligned}$$

The term D_1 is invariant to any permutation of Z values, and thus the permutation distribution of $V_{\boldsymbol{\pi},\mathrm{HSIC}}$ is equivalent to the permutation distribution of $U_{\boldsymbol{\pi},\mathrm{HSIC}} + D_1 + D_{2,\boldsymbol{\pi}}$ where $D_{2,\boldsymbol{\pi}}$ has the same form of D_2 but computed based on permuted data $\mathcal{X}_n^{\boldsymbol{\pi}}$. Writing the sensitivity of $\sqrt{V_{\mathrm{HSIC}}}$ as $\Delta_{V^{1/2}}$, Lemma 17 yields that the dpHSIC test rejects the null if and only if

$$\sqrt{V_{\mathrm{HSIC}}} + \frac{2\Delta_{V^{1/2}}}{\xi_{\varepsilon,\delta}}\zeta_0 > q_{1-\alpha,B},$$

where $q_{1-\alpha,B}$ is the $1-\alpha$ quantile of $\{\sqrt{V_{\boldsymbol{\pi}_i,\mathrm{HSIC}}}+2\Delta_{V^{1/2}}\xi_{\varepsilon,\delta}^{-1}\zeta_i\}_{i=0}^{B}$. As in the proof of Theorem 9 in Appendix E.11, we let $q^a_{1-\alpha,B}$ denote the $1-\alpha$ quantile of $\{V^{1/2}_{\boldsymbol{\pi}_0,\mathrm{HSIC}}, V^{1/2}_{\boldsymbol{\pi}_1,\mathrm{HSIC}},\dots,V^{1/2}_{\boldsymbol{\pi}_B,\mathrm{HSIC}}\}$. Similarly, denote by $q^a_{1-\alpha,\infty}$ the corresponding $1-\alpha$ quantile with $B=\infty$. Then for $B\geq 16\alpha^{-2}\log(8/\beta)$, the analysis given in the proof of Theorem 9 shows that the type II error of the dpHSIC test can be bounded as

$$\begin{aligned}
&\mathbb{P}\big(\sqrt{V_{\mathrm{HSIC}}}+2\Delta_{V^{1/2}}\xi_{\varepsilon,\delta}^{-1}\zeta_0\leq q_{1-\alpha,B}\big)\\
\leq\ &\mathbb{P}\big(\sqrt{V_{\mathrm{HSIC}}}\leq q^a_{1-\alpha/4,\infty}+10\Delta_{V^{1/2}}\xi_{\varepsilon,\delta}^{-1}\max\{\log(1/\alpha),\log(1/\beta)\}\big)+5\beta/8. \qquad (48)
\end{aligned}$$

We next delve into $q^a_{1-\alpha/4,\infty}$, and then continue to upper bound the type II error. In what follows, we use the notation $C_1, C_2, C_3, \dots$ to represent constants, which could be dependent on $\alpha, \beta, s, R, M, d_Y, d_Z$. The values of these constants may vary in different places.

Bounding $q^a_{1-\alpha/4,\infty}$. Using the relationship between V_{HSIC} and U_{HSIC} in Lemma 21 and the quantile inequality in Lemma 20, we have

$$\begin{aligned}
q^a_{1-\alpha/4,\infty} &= \sqrt{\mathrm{Quantile}_{1-\alpha/4}\big(\{U_{\boldsymbol{\pi}_i,\mathrm{HSIC}}+D_{2,\boldsymbol{\pi}_i}\}_{i=0}^{\infty}+D_1\big)}\\
&\leq \sqrt{\mathrm{Quantile}_{1-\alpha/8}\big(\{U_{\boldsymbol{\pi}_i,\mathrm{HSIC}}\}_{i=0}^{\infty}\big)+\mathrm{Quantile}_{1-\alpha/8}\big(\{D_{2,\boldsymbol{\pi}_i}\}_{i=0}^{\infty}\big)+D_1}.
\end{aligned}$$

The quantile of the permuted U-statistic $U_{\boldsymbol{\pi},\mathrm{HSIC}}$ has been studied by Kim et al. (2022a). In particular, the proof of Kim et al. (2022a, Theorem 5.1) along with the proof of Albert et al. (2022, Proposition 2) shows that

$$\mathbb{E}_{\boldsymbol{\pi}}[U_{\boldsymbol{\pi},\mathrm{HSIC}}\,|\,\mathcal{X}_n]=0 \quad\text{and}\quad \mathbb{E}\big[\mathrm{Var}_{\boldsymbol{\pi}}\big(U_{\boldsymbol{\pi},\mathrm{HSIC}}\,|\,\mathcal{X}_n\big)\big]\leq\frac{C_1}{n^2\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}}.$$

Thus by Chebyshev's inequality, it can be seen that the following inequality holds with probability at least $1-\beta/8$:

$$\mathrm{Quantile}_{1-\alpha/8}\big(\{U_{\boldsymbol{\pi}_i,\mathrm{HSIC}}\}_{i=0}^{\infty}\big)\leq\frac{C_2}{n\sqrt{\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}}}, \qquad (49)$$

For the $1-\alpha/8$ quantile of $\{D_{2,\boldsymbol{\pi}_i}\}_{i=0}^{\infty}$, note that

$$D_{2,\boldsymbol{\pi}}\leq\frac{C_3}{n^4}\sum_{(i,j_1,j_2)\in\mathbf{i}_3^n}k_{\boldsymbol{\lambda}}(Y_i,Y_{j_1})\ell_{\boldsymbol{\mu}}(Z_{\pi_i},Z_{\pi_{j_2}})$$

due to the non-negativity of $k_{\boldsymbol{\lambda}}$ and $\ell_{\boldsymbol{\mu}}$, and also note that

$$\mathbb{E}_{\boldsymbol{\pi}}\left[\sum_{(i,j_1,j_2)\in\mathbf{i}_3^n}k_{\boldsymbol{\lambda}}(Y_i,Y_{j_1})\ell_{\boldsymbol{\mu}}(Z_{\pi_i},Z_{\pi_{j_2}})\,\Big|\,\mathcal{X}_n\right]\leq\frac{C_4}{n}\left[\sum_{(i_1,i_2)\in\mathbf{i}_2^n}k_{\boldsymbol{\lambda}}(Y_{i_1},Y_{i_2})\right]\left[\sum_{(j_1,j_2)\in\mathbf{i}_2^n}\ell_{\boldsymbol{\mu}}(Z_{j_1},Z_{j_2})\right].$$

Then Markov's inequality yields that the $1-\alpha/8$ quantile of $\{D_{2,\boldsymbol{\pi}_i}\}_{i=0}^{\infty}$ is bounded as

$$\mathrm{Quantile}_{1-\alpha/8}\big(\{D_{2,\boldsymbol{\pi}_i}\}_{i=0}^{\infty}\big)\leq\frac{C_5}{n^5}\left[\sum_{(i_1,i_2)\in\mathbf{i}_2^n}k_{\boldsymbol{\lambda}}(Y_{i_1},Y_{i_2})\right]\left[\sum_{(j_1,j_2)\in\mathbf{i}_2^n}\ell_{\boldsymbol{\mu}}(Z_{j_1},Z_{j_2})\right].$$

Moreover, since we assume $\|p_{YZ}\|_{L_\infty} \leq M$ and $\|p_Y p_Z\|_{L_\infty} \leq M$,

$$\max\Big\{\mathbb{E}[k_{\boldsymbol{\lambda}}(Y_1,Y_2)\ell_{\boldsymbol{\mu}}(Z_1,Z_2)],\ \mathbb{E}[k_{\boldsymbol{\lambda}}(Y_1,Y_2)\ell_{\boldsymbol{\mu}}(Z_1,Z_3)],\ \mathbb{E}[k_{\boldsymbol{\lambda}}(Y_1,Y_2)\ell_{\boldsymbol{\mu}}(Z_3,Z_4)]\Big\} \leq C_6.$$

For instance, we see that

$$\begin{aligned}
\mathbb{E}[k_{\boldsymbol{\lambda}}(Y_1,Y_2)\ell_{\boldsymbol{\mu}}(Z_1,Z_2)] &= \int\cdots\int k_{\boldsymbol{\lambda}}(y_1,y_2)\ell_{\boldsymbol{\mu}}(z_1,z_2)p_{YZ}(y_1,z_1)p_{YZ}(y_2,z_2)\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}z_1\mathrm{d}z_2 \\
&\leq \|p_{YZ}\|_{L_\infty}\int\int \underbrace{\int k_{\boldsymbol{\lambda}}(y_1,y_2)\mathrm{d}y_1}_{=1}\underbrace{\int \ell_{\boldsymbol{\mu}}(z_1,z_2)\mathrm{d}z_1}_{=1} p_{YZ}(y_2,z_2)\mathrm{d}y_2\mathrm{d}z_2 \\
&\leq M,
\end{aligned}$$

and the other terms can be similarly analyzed. Given this ingredient, we have

$$\mathbb{E}\Bigg\{\Bigg[\sum_{(i_1,i_2)\in\mathbf{i}_2^n} k_{\boldsymbol{\lambda}}(Y_{i_1},Y_{i_2})\Bigg]\Bigg[\sum_{(j_1,j_2)\in\mathbf{i}_2^n} \ell_{\boldsymbol{\mu}}(Z_{j_1},Z_{j_2})\Bigg]\Bigg\} \leq C_7 n^4.$$

Therefore, another application of Markov's inequality yields

$$\mathrm{Quantile}_{1-\alpha/8}\big(\{D_{2,\boldsymbol{\pi}_i}\}_{i=0}^\infty\big) \leq \frac{C_8}{n}, \tag{50}$$

with probability at least $1-\beta/8$, and observe that $D_1 \leq \frac{n-1}{n^2}KL \leq C_9/\{n\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}\}$.

In summary, with probability at least $1-\beta/4$,

$$q^a_{1-\alpha/4,\infty} \leq \frac{C_{10}}{\sqrt{n\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}}}. \tag{51}$$

given that $D_1 \leq KL/n$ and $\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z} \leq 1$.

Bounding the type II error. We continue from (48), and note that

$$\begin{aligned}
&\mathbb{P}\Big(\sqrt{V_{\mathrm{HSIC}}} \leq q^a_{1-\alpha/4,\infty} + 10\Delta_{V^{1/2}}\xi^{-1}_{\varepsilon,\delta}\max\{\log(1/\alpha),\log(1/\beta)\}\Big) \\
=\ &\mathbb{P}\Big(V_{\mathrm{HSIC}} \leq (q^a_{1-\alpha/4,\infty})^2 + 10^2\Delta^2_{V^{1/2}}\xi^{-2}_{\varepsilon,\delta}\max\{\log^2(1/\alpha),\log^2(1/\beta)\} \\
&\quad + 20q^a_{1-\alpha/4,\infty}\Delta_{V^{1/2}}\xi^{-1}_{\varepsilon,\delta}\max\{\log(1/\alpha),\log(1/\beta)\}\Big) \\
=\ &\mathbb{P}\Big(U_{\mathrm{HSIC}} \leq \mathrm{Quantile}_{1-\alpha/8}\big(\{U_{\boldsymbol{\pi}_i,\mathrm{HSIC}}\}_{i=0}^\infty\big) + \mathrm{Quantile}_{1-\alpha/8}\big(\{D_{2,\boldsymbol{\pi}_i}\}_{i=0}^\infty\big) - D_2 \\
&\quad + 10^2\Delta^2_{V^{1/2}}\xi^{-2}_{\varepsilon,\delta}\max\{\log^2(1/\alpha),\log^2(1/\beta)\} + 20q^a_{1-\alpha/4,\infty}\Delta_{V^{1/2}}\xi^{-1}_{\varepsilon,\delta}\max\{\log(1/\alpha),\log(1/\beta)\}\Big),
\end{aligned}$$

where the first equality simply follows by taking the square on both sides, and the second equality uses the identity $V_{\mathrm{HSIC}} = U_{\mathrm{HSIC}} + D_1 + D_2$ from Lemma 21. Next, by Lemma 6, the global sensitivity of $\sqrt{V_{\mathrm{HSIC}}}$ is bounded as

$$\Delta_{V^{1/2}} \leq \frac{C_1}{n\sqrt{\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}}}.$$

In addition, observe that $|D_2| \leq C_2 n^{-1}$ with probability at least $1 - \beta/16$, which can be proven by Markov's inequality. Therefore, equipped with the previous ingredients (48), (49), (50) and (51), the type II error bound can be bounded as

$$\begin{aligned}
&\mathbb{P}\big(\sqrt{V_{\mathrm{HSIC}}} + 2\Delta_{V^{1/2}}\xi_{\varepsilon,\delta}^{-1}\zeta_0 \leq q_{1-\alpha,B}\big) \\
\leq\ &\mathbb{P}\bigg(U_{\mathrm{HSIC}} \leq \frac{C_3}{n\sqrt{\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}}} + \frac{C_4}{n^2\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}\xi_{\varepsilon,\delta}^2} \\
&\qquad + \frac{C_5}{n^{3/2}\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}\xi_{\varepsilon,\delta}}\bigg) + \frac{15}{16}\beta.
\end{aligned}$$

Condition in terms of L_2 distance. A slight modification of Albert et al. (2022, Lemma 1) yields that a sufficient condition for the first probability in the above display to be less than $\beta/16$ is

$$\begin{aligned}
\mathrm{HSIC}^2_{k_{\boldsymbol{\lambda}}\otimes\ell_{\boldsymbol{\mu}}} \ \geq\ &\sqrt{\mathrm{Var}[U_{\mathrm{HSIC}}]} + \frac{C_3}{n\sqrt{\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}}} + \frac{C_4}{n^2\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}\xi_{\varepsilon,\delta}^2} \\
&+ \frac{C_5}{n^{3/2}\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}\xi_{\varepsilon,\delta}}.
\end{aligned}$$

In addition, by writing $\psi = p_{YZ} - p_Y p_Z$ and the convolution of ψ and $k_{\boldsymbol{\lambda}} \otimes \ell_{\boldsymbol{\mu}}$ as $\psi * (k_{\boldsymbol{\lambda}} \otimes \ell_{\boldsymbol{\mu}})$, Albert et al. (2022, Theorem 1 and Proposition 4) show that the previous condition is implied by

$$\begin{aligned}
\|\psi\|^2_{L_2} \ \geq\ &\|\psi - \psi * (k_{\boldsymbol{\lambda}} \otimes \ell_{\boldsymbol{\mu}})\|^2_{L_2} + C_6\bigg\{\frac{1}{n\sqrt{\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}}} + \frac{1}{n^2\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}\xi_{\varepsilon,\delta}^2} \\
&+ \frac{1}{n^{3/2}\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}\xi_{\varepsilon,\delta}}\bigg\}.
\end{aligned}$$

Finally the proof of (Albert et al., 2022, Theorem 2) yields that over the Sobolev ball, a sufficient condition for the previous inequality is

$$\begin{aligned}
\|\psi\|^2_{L_2} \ \geq\ &C_7\bigg\{\sum_{i=1}^{d_Y}\lambda_i^{2s} + \sum_{i=1}^{d_Z}\mu_i^{2s} + \frac{1}{n\sqrt{\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}}} \\
&+ \frac{1}{n^2\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}\xi_{\varepsilon,\delta}^2} + \frac{1}{n^{3/2}\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}\xi_{\varepsilon,\delta}}\bigg\}
\end{aligned}$$

as claimed.

F.6 Proof of Lemma 9

Sensitivity upper bound. For simplicity, write $U_{\mathrm{HSIC}} = U_{\mathrm{HSIC},a} + U_{\mathrm{HSIC},b} - 2U_{\mathrm{HSIC},c}$ where

$$U_{\mathrm{HSIC},a} \ := \ \frac{1}{n(n-1)}\sum_{(i,j)\in\mathbf{i}_2^n} k(Y_i, Y_j)\ell(Z_i, Z_j),$$

$$U_{\mathrm{HSIC},b} := \frac{(n-4)!}{n!} \sum_{(i_1,i_2,j_1,j_2)\in \mathbf{i}_4^n} k(Y_{i_1}, Y_{j_1})\ell(Z_{i_2}, Z_{j_2}),$$

$$U_{\mathrm{HSIC},c} := \frac{1}{n(n-1)(n-2)} \sum_{(i,j_1,j_2)\in \mathbf{i}_3^n} k(Y_i, Y_{j_1})\ell(Z_i, Z_{j_2}).$$

To establish an upper bound for the global sensitivity, we first consider a neighboring dataset $\tilde{\mathcal{X}}_n = \{(Y_1', Z_1'), (Y_2, Z_2), \ldots, (Y_n, Z_n)\}$ and denote the U-statistic of HSIC based on $\tilde{\mathcal{X}}_n$ as $U'_{\mathrm{HSIC}} := U'_{\mathrm{HSIC},a} + U'_{\mathrm{HSIC},b} - 2U'_{\mathrm{HSIC},c}$. By the triangle inequality, the absolute deviation between U_{HSIC} and U'_{HSIC} is then bounded as

$$\begin{aligned} |U_{\mathrm{HSIC}} - U'_{\mathrm{HSIC}}| &\le |U_{\mathrm{HSIC},a} - U'_{\mathrm{HSIC},a}| + |U_{\mathrm{HSIC},b} - U'_{\mathrm{HSIC},b}| + 2|U_{\mathrm{HSIC},c} - U'_{\mathrm{HSIC},c}| \\ &\le \frac{C_1 KL}{n}, \end{aligned}$$

where $C_1, C_2, \ldots$ denote some universal positive constants throughout.

For another neighboring dataset $\tilde{\mathcal{X}}_n = \{(Y_1', Z_1), (Y_2, Z_2'), \ldots, (Y_n, Z_n)\}$, a similar analysis shows that $|U_{\mathrm{HSIC}} - U'_{\mathrm{HSIC}}| \le \frac{C_2 KL}{n}$. Since U_{HSIC} is invariant to the permutation of the paired indices, we may conclude that

$$\sup_{\boldsymbol{\pi}\in\boldsymbol{\Pi}_n} \sup_{\substack{\mathcal{X}_n, \tilde{\mathcal{X}}_n: \\ d_{\mathrm{ham}}(\mathcal{X}_n, \tilde{\mathcal{X}}_n)\le 1}} \left|U_{\mathrm{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}}) - U_{\mathrm{HSIC}}(\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}})\right| \le \frac{C_3 KL}{n}. \tag{52}$$

While it is loose, the inequality holds with $C_3 = 24$ for $n \ge 4$.

Sensitivity lower bound. We now show that this upper bound is tight up to a constant factor. We treat the cases of n being even or odd separately, and follow a similar reasoning to the one in Appendix E.7. Recall that kernels k and ℓ are assumed to have non-empty level sets on $\mathbb{Y}$ and $\mathbb{Z}$. Hence, for a given $\epsilon \in (0, \min\{K, L\})$, we may assume that there exist $y_a, y_b \in \mathbb{Y}$ and $z_a, z_b \in \mathbb{Z}$ such that $k(y_a, y_b) = \epsilon_1$ and $\ell(z_a, z_b) = \epsilon_2$ where $0 \le \epsilon_1, \epsilon_2 \le \epsilon$.

Sensitivity lower bound for $n = 2k$ even. We begin by considering the case where n is even, and compute the difference between U-statistics of the HSIC based on two specific neighboring datasets given below:

$$\mathcal{X}_n = \begin{bmatrix} y_a & z_a \\ \vdots & \vdots \\ y_a & z_a \\ y_a & z_a \\ y_a & z_a \\ y_b & z_b \\ \vdots & \vdots \\ y_b & z_b \end{bmatrix} \quad \text{and} \quad \tilde{\mathcal{X}}_n = \begin{bmatrix} y_a & z_a \\ \vdots & \vdots \\ y_a & z_a \\ y_a & z_a \\ y_b & z_b \\ y_b & z_b \\ \vdots & \vdots \\ y_b & z_b \end{bmatrix}.$$

The indices displayed above are $1, \ldots, k-2, k-1, k, k+1, \ldots 2k$, and it is clear that $d_{\text{ham}}(\mathcal{X}_n, \tilde{\mathcal{X}}_n) = 1$. We further consider the permutation $\boldsymbol{\pi}$ which permutes only the $k-1$ and k entries, *i.e.*, $\boldsymbol{\pi} = (1, 2, \ldots, k-2, k, k-1, k+1, \ldots, n)$, so that

$$\mathcal{X}_n^{\boldsymbol{\pi}} = \begin{bmatrix} y_a & z_a \\ \vdots & \vdots \\ y_a & z_a \\ y_a & z_a \\ y_a & z_a \\ y_b & z_b \\ \vdots & \vdots \\ y_b & z_b \end{bmatrix} \quad \text{and} \quad \tilde{\mathcal{X}}_n^{\boldsymbol{\pi}} = \begin{bmatrix} y_a & z_a \\ \vdots & \vdots \\ y_a & z_a \\ y_a & z_b \\ y_b & z_a \\ y_b & z_b \\ \vdots & \vdots \\ y_b & z_b \end{bmatrix}.$$

We now compute the U-statistic of the HSIC based on these two permuted datasets. Unfortunately, a direct computation of U_{HSIC} requires a complicated case-by-case analysis. Instead, we consider the following trick to simplify calculations. First of all, we pretend that (y_a, z_a) and (y_b, z_b) take some specific values, say $(y_a, z_a) = (1, 1)$ and $(y_b, z_b) = (0, 0)$, and consider indicator kernels $k(y, y') = K \times \mathbb{1}(y = y')$ and $\ell(z, z') = L \times \mathbb{1}(z = z')$. Under this simplified setting, U_{HSIC} is essentially the statistic considered in Kim et al. (2023, Proposition 1) for multinomial data, which can be computed in a straightforward manner. In addition, we observe that U_{HSIC} based on the indicator kernels remains the same as U_{HSIC} based on generic kernels k and ℓ up to some quantities tending to zero as $\epsilon_1, \epsilon_2 \to 0$. Using this trick along with the observation that $k(y, y) = K$ and $\ell(z, z) = L$ for all $y \in \mathbb{Y}$ and $z \in \mathbb{Z}$, it can be seen that for all $k \geq 2$

$$U_{\text{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}}) = \frac{k(k-1)}{(2k-1)(2k-3)} KL + C_1\epsilon_1 + C_2\epsilon_2 + C_3\epsilon_1\epsilon_2,$$

where C_1, C_2, C_3 are constants that only depend on K, L, n. A similar calculation yields

$$U_{\text{HSIC}}(\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}}) = \frac{(k-2)(k^2-4k+1)}{(k-1)(2k-1)(2k-3)} KL + C_1'\epsilon_1 + C_2'\epsilon_2 + C_3'\epsilon_1\epsilon_2,$$

where C_1', C_2', C_3' are constants that only depend on K, L, n. From these results, we deduce that

$$\begin{aligned} & U_{\text{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}}) - U_{\text{HSIC}}(\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}}) \\ = \; & \left(\frac{2}{k-1} - \frac{1}{2k-1} - \frac{1}{2k-3} \right) KL + (C_1 - C_1')\epsilon_1 + (C_2 - C_2')\epsilon_2 + (C_3 - C_3')\epsilon_1\epsilon_2 \\ \overset{(\star)}{\to} \; & \left(\frac{2}{k-1} - \frac{1}{2k-1} - \frac{1}{2k-3} \right) KL \geq \frac{KL}{k} \end{aligned}$$

where convergence $(\star)$ holds as $\epsilon_1, \epsilon_2 \to 0$ for each fixed K, L, n, and the last inequality holds for all $k \geq 2$. Therefore, by letting $k = n/2 \geq 2$, it holds that

$$\sup_{\boldsymbol{\pi} \in \boldsymbol{\Pi}_n} \sup_{\substack{\mathcal{X}_n, \tilde{\mathcal{X}}_n: \\ d_{\text{ham}}(\mathcal{X}_n, \tilde{\mathcal{X}}_n) \leq 1}} \left| U_{\text{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}}) - U_{\text{HSIC}}(\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}}) \right| \geq \frac{2KL}{n}.$$

Sensitivity lower bound for $n = 2k+1$ odd. We now consider the same two datasets as in the even case but with an additional row consisting of (y_b, z_b):

$$\mathcal{X}_n = \begin{bmatrix} y_a & z_a \\ \vdots & \vdots \\ y_a & z_a \\ y_a & z_a \\ y_a & z_a \\ y_b & z_b \\ \vdots & \vdots \\ y_b & z_b \\ y_b & z_b \end{bmatrix} \quad \text{and} \quad \tilde{\mathcal{X}}_n = \begin{bmatrix} y_a & z_a \\ \vdots & \vdots \\ y_a & z_a \\ y_a & z_a \\ y_b & z_b \\ y_b & z_b \\ \vdots & \vdots \\ y_b & z_b \\ y_b & z_b \end{bmatrix},$$

where the indices displayed are $1, \ldots, k-2, k-1, k, k+1, \ldots 2k, 2k+1$. As before, we indeed have $d_{\text{ham}}(\mathcal{X}_n, \tilde{\mathcal{X}}_n) = 1$. Once again, we consider the permutation $\boldsymbol{\pi}$ which permutes only the $k-1$ and k entries, which leads to

$$\mathcal{X}_n^{\boldsymbol{\pi}} = \begin{bmatrix} y_a & z_a \\ \vdots & \vdots \\ y_a & z_a \\ y_a & z_a \\ y_a & z_a \\ y_b & z_b \\ \vdots & \vdots \\ y_b & z_b \\ y_b & z_b \end{bmatrix} \quad \text{and} \quad \tilde{\mathcal{X}}_n^{\boldsymbol{\pi}} = \begin{bmatrix} y_a & z_a \\ \vdots & \vdots \\ y_a & z_a \\ y_a & z_a \\ y_b & z_a \\ y_b & z_b \\ \vdots & \vdots \\ y_b & z_b \\ y_b & z_b \end{bmatrix}.$$

To compute $U_{\text{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}})$ and $U_{\text{HSIC}}(\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}})$, we use the trick considered in the even case, and find that

$$U_{\text{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}}) \;=\; \frac{k(k+1)}{4k^2-1} KL + C_1\epsilon_1 + C_2\epsilon_2 + C_3\epsilon_1\epsilon_2$$

and

$$U_{\text{HSIC}}(\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}}) \;=\; \frac{(k+1)(k-2)(k^2-3k-2)}{k(k-1)(2k-1)(2k+1)} KL + C_1'\epsilon_1 + C_2'\epsilon_2 + C_3'\epsilon_1\epsilon_2,$$

where $C_1, \ldots, C_3'$ are constants that only depend on K, L, n. As a result, we have

$$\begin{aligned} U_{\text{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}}) - U_{\text{HSIC}}(\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}}) \;&=\; \frac{4(k^3-2k+1)}{k(k-1)(4k^2-1)} KL + (C_1 - C_1')\epsilon_1 + (C_2 - C_2')\epsilon_2 + (C_3 - C_3')\epsilon_1\epsilon_2 \\ &\overset{(\star)}{\to}\; \frac{4(k^3-2k+1)}{k(k-1)(4k^2-1)} KL \geq \frac{2}{2k+1} KL, \end{aligned}$$

where convergence $(\star)$ holds as $\epsilon_1, \epsilon_2 \to 0$ for each fixed K, L, n, and the last inequality holds for all $k \geq 2$. Therefore we conclude that

$$\sup_{\boldsymbol{\pi} \in \boldsymbol{\Pi}_n} \sup_{\substack{\mathcal{X}_n, \tilde{\mathcal{X}}_n: \\ d_{\mathrm{ham}}(\mathcal{X}_n, \tilde{\mathcal{X}}_n) \leq 1}} \left| U_{\mathrm{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}}) - U_{\mathrm{HSIC}}(\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}}) \right| \geq \frac{2KL}{n},$$

for any $n \geq 4$ (either even or odd). This completes the derivation of the lower bound on the sensitivity of the HSIC U-statistic. Together with the upper bound (52), this verifies the claim of Lemma 9.

F.7 Proof of Theorem 15

Let us write the V-statistic of HSIC as

$$\begin{aligned} V_{\mathrm{HSIC}} &= \frac{1}{n^2} \sum_{i,j=1}^{n} k(Y_i, Y_j) \ell(Z_i, Z_j) + \frac{1}{n^4} \sum_{i_1, i_2, j_1, j_2 = 1}^{n} k(Y_{i_1}, Y_{j_1}) \ell(Z_{i_2}, Z_{j_2}) \\ &\quad - \frac{2}{n^3} \sum_{i, j_1, j_2 = 1}^{n} k(Y_i, Y_{j_1}) \ell(Z_i, Z_{j_2}) \\ &= V_{\mathrm{HSIC},a} + V_{\mathrm{HSIC},b} - 2V_{\mathrm{HSIC},c}, \end{aligned}$$

which is equivalent to the square of $\widehat{\mathrm{HSIC}}$. Using the bounded kernel property, it can be seen that

$$\max \left\{ |U_{\mathrm{HSIC},a} - V_{\mathrm{HSIC},a}|, \ |U_{\mathrm{HSIC},b} - V_{\mathrm{HSIC},b}|, \ |U_{\mathrm{HSIC},c} - V_{\mathrm{HSIC},c}| \right\} \leq C_1 \frac{KL}{n}$$

for some constant $C_1 > 0$, which leads to

$$|U_{\mathrm{HSIC}} - V_{\mathrm{HSIC}}| \leq C_2 \frac{KL}{n}.$$

Again, C_2 is some positive number. By treating K and L as some fixed numbers, we also note that Lemma 14 yields

$$V_{\mathrm{HSIC}} = \left\{ \mathrm{HSIC}_{k \otimes \ell}(P_{YZ}) + R_n \right\}^2 \quad \text{where } R_n = O_P(n^{-1/2}).$$

Having these ingredients in place, we can essentially follow the same proof strategy as for Theorem 10 and conclude that

$$\limsup_{n \to \infty} \inf_{P \in \mathcal{P}_{\mathrm{HSIC}_{k \otimes \ell}}(\rho)} \mathbb{E}_{P_{YZ}}[\phi^u_{\mathrm{dpHSIC}}] \leq \alpha.$$

This completes the proof of Theorem 15.

F.8 Proof of Theorem 16

Let us denote by $C_1, C_2, \ldots$ constants that may depend on $\alpha, \beta, R, M, d_Y, d_Z$. Following the proofs of Theorem 4 and Theorem 14 along with the sensitivity result for U_{HSIC} in Lemma 9, we may arrive at the point where the type II error of ϕ^u_{dpHSIC} is upper bounded as

$$\mathbb{E}[1-\phi^u_{\mathrm{dpHSIC}}] \le \mathbb{P}\left(U_{\mathrm{HSIC}} \le \frac{C_1}{n\sqrt{\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}}} + \frac{C_2}{n\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}\xi_{\varepsilon,\delta}}\right) + \frac{15}{16}\beta$$

We then use the proofs of Albert et al. (2022, Theorem 1) and Albert et al. (2022, Theorem 2), and show that the probability term in the above display is less than or equal to $\beta/16$ once

$$\|\psi\|^2_{L_2} \;\ge\; C_3\left\{\sum_{i=1}^{d_Y}\lambda_i^{2s} + \sum_{i=1}^{d_Z}\mu_i^{2s} + \frac{1}{n\sqrt{\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}}} + \frac{1}{n\lambda_1\cdots\lambda_{d_Y}\mu_1\cdots\mu_{d_Z}\xi_{\varepsilon,\delta}}\right\}.$$

This proves Theorem 11.

G Technical Lemmas

This section collects some technical lemmas used in the main proofs. The first lemma from Kim (2021) provides an exponential concentration inequality for the permuted MMD statistic with kernel k.

Lemma 10 (Theorem 5.1 of Kim 2021)**.** *Consider two samples $Y_1,\ldots,Y_n$ and $Z_1,\ldots,Z_m$ and write $\{X_1,\ldots,X_N\} = \{Y_1,\ldots,Y_n,Z_1,\ldots,Z_m\}$. Let us further write $\widetilde{k}(x,y) = k(x,x)+k(y,y)-2k(x,y)$ and $\gamma = nm/(n+m)^2$. Define*

$$\widehat{\sigma}^2 = \frac{1}{N(N-1)}\sum_{(i,j)\in \boldsymbol{i}_2^N}\widetilde{k}(X_i,X_j).$$

Then for all $t>0$,

$$\mathbb{P}_{\boldsymbol{\pi}}\left[\sup_{f\in\mathcal{F}_k}\left(\frac{1}{n}\sum_{i=1}^{n}f(X_{\pi_i}) - \frac{1}{m}\sum_{i=1}^{m}f(X_{\pi_{i+n}})\right) \ge t + \sqrt{\frac{\widehat{\sigma}^2}{2N\gamma}} \;\middle|\; X_1,\ldots,X_N\right] \le \exp\left(-\frac{N\gamma^2 t^2}{2\widehat{\sigma}^2}\right).$$

An exponential tail decay in the above comes at a price — it requires that n and m are well-balanced in the sense that $N\gamma^2$ is large. Thus it can be used to show that $\widehat{\mathrm{MMD}}(\mathcal{X}^{\boldsymbol{\pi}}_{n+m}) = o_P(1)$ under the limited regime where $N\gamma^2 \to \infty$. The lemma below removes this unnecessary constraint at the expense of having a polynomial tail decay.

Lemma 11 (Markov for the Permuted MMD)**.** *Under the setting of Lemma 10 with $N = n+m$, we have*

$$\mathbb{P}\left[\sup_{f\in\mathcal{F}_k}\left(\frac{1}{n}\sum_{i=1}^{n}f(X_{\pi_i}) - \frac{1}{m}\sum_{i=1}^{m}f(X_{\pi_{i+n}})\right) \ge t\right] \le \frac{2K}{nt^2} \quad \text{for all } t>0.$$

Hence $\widehat{\mathrm{MMD}}(\mathcal{X}^{\boldsymbol{\pi}}_{n+m}) = o_P(1)$, provided that $Kn^{-1}\to 0$.

Proof. Markov's inequality together with the V-statistic representation of the empirical MMD yields that for any $t > 0$

$$\begin{aligned}
&\mathbb{P}\left[\sup_{f\in\mathcal{F}_k}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_{\pi_i}) - \frac{1}{m}\sum_{i=1}^{m} f(X_{\pi_{i+n}})\right) \geq t\right] \\
&\leq \frac{1}{t^2}\Bigg\{\underbrace{\frac{1}{n^2}\sum_{i,j=1}^{n}\mathbb{E}[k(X_{\pi_i}, X_{\pi_j})]}_{\text{(I)}} + \underbrace{\frac{1}{m^2}\sum_{i,j=1}^{m}\mathbb{E}[k(X_{\pi_{i+n}}, X_{\pi_{j+n}})]}_{\text{(II)}} - \underbrace{\frac{2}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\mathbb{E}[k(X_{\pi_i}, X_{\pi_{j+n}})]}_{\text{(III)}}\Bigg\}.
\end{aligned}$$

Exploiting the uniformity of permutation $\boldsymbol{\pi}$, we analyze the terms (I), (II) and (III), separately. The analyses of the first two terms (I) and (II) are similar, and we note that

$$\begin{aligned}
\mathbb{E}[k(X_{\pi_i}, X_{\pi_i})] = \mathbb{E}[k(X_{\pi_{i+n}}, X_{\pi_{i+n}})] &= \frac{1}{N}\sum_{i=1}^{n}\mathbb{E}[k(Y_i, Y_i)] + \frac{1}{N}\sum_{j=1}^{m}\mathbb{E}[k(Z_j, Z_j)] \\
&= \frac{n}{N}\mathbb{E}[k(Y_1, Y_1)] + \frac{m}{N}\mathbb{E}[k(Z_1, Z_1)]
\end{aligned}$$

and for $i \neq j$,

$$\begin{aligned}
&\mathbb{E}[k(X_{\pi_i}, X_{\pi_j})] = \mathbb{E}[k(X_{\pi_{i+n}}, X_{\pi_{j+n}})] \\
&= \frac{1}{N(N-1)}\Bigg\{\sum_{(i,j)\in \boldsymbol{i}_2^n}\mathbb{E}[k(Y_i, Y_j)] + \sum_{(i,j)\in \boldsymbol{i}_2^m}\mathbb{E}[k(Z_i, Z_j)] + 2\sum_{i=1}^{n}\sum_{j=1}^{m}\mathbb{E}[k(Y_i, Z_j)]\Bigg\} \\
&= \frac{n(n-1)}{N(N-1)}\mathbb{E}[k(Y_1, Y_2)] + \frac{m(m-1)}{N(N-1)}\mathbb{E}[k(Z_1, Z_2)] + \frac{2nm}{N(N-1)}\mathbb{E}[k(Y_1, Z_1)].
\end{aligned}$$

Therefore

$$\begin{aligned}
\text{(I)} &= \frac{n}{nN}\mathbb{E}[k(Y_1, Y_1)] + \frac{m}{nN}\mathbb{E}[k(Z_1, Z_1)] \\
&\quad + \frac{n-1}{n}\Bigg\{\frac{n(n-1)}{N(N-1)}\mathbb{E}[k(Y_1, Y_2)] + \frac{m(m-1)}{N(N-1)}\mathbb{E}[k(Z_1, Z_2)] + \frac{2nm}{N(N-1)}\mathbb{E}[k(Y_1, Z_1)]\Bigg\}
\end{aligned}$$

and

$$\begin{aligned}
\text{(II)} &= \frac{n}{mN}\mathbb{E}[k(Y_1, Y_1)] + \frac{m}{mN}\mathbb{E}[k(Z_1, Z_1)] \\
&\quad + \frac{m-1}{m}\Bigg\{\frac{n(n-1)}{N(N-1)}\mathbb{E}[k(Y_1, Y_2)] + \frac{m(m-1)}{N(N-1)}\mathbb{E}[k(Z_1, Z_2)] + \frac{2nm}{N(N-1)}\mathbb{E}[k(Y_1, Z_1)]\Bigg\}.
\end{aligned}$$

On the other hand, the last term (III) is

$$\text{(III)} = \frac{2n(n-1)}{N(N-1)}\mathbb{E}[k(Y_1, Y_2)] + \frac{2m(m-1)}{N(N-1)}\mathbb{E}[k(Z_1, Z_2)] + \frac{4nm}{N(N-1)}\mathbb{E}[k(Y_1, Z_1)].$$

Given that kernel k is non-negative and bounded by K, we can upper bound (I) + (II) − (III) as

$$\begin{aligned}(\mathrm{I})+(\mathrm{II})-(\mathrm{III}) &\leq \left(\frac{n}{nN}+\frac{m}{nN}+\frac{n}{mN}+\frac{m}{mN}\right)K = \frac{n+m}{nm}K \\ &\leq \frac{2K}{n}.\end{aligned}$$

Hence the result follows. □

The lemma below presents a concentration bound for the permuted HSIC statistic. It is worth noting that obtaining the logarithmic dependence on α is non-trivial, and thus we highlight it as our contribution.

Lemma 12 (Concentration Inequality for Permuted HSIC)**.** *Assume that the kernels k and ℓ are bounded as $0 \leq k(y, y') \leq K$ and $0 \leq \ell(z, z') \leq L$ for all $y, y' \in \mathbb{Y}$ and $z, z' \in \mathbb{Z}$. Then for any $\alpha \in (0, 1)$,*

$$\mathbb{P}_{\boldsymbol{\pi}}\left(\widehat{\mathrm{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}}) \geq C\sqrt{\frac{KL}{n}}\max\left\{\log^{1/4}\left(\frac{1}{\alpha}\right), \log^{1/2}\left(\frac{1}{\alpha}\right), 1\right\} \,\middle|\, \mathcal{X}_n\right) \leq \alpha,$$

where C is some positive constant.

Proof. The proof is based on Kim et al. (2022a, Theorem 6.2) that establishes an exponential tail bound for a permuted U-statistic of HSIC. Given $\mathcal{X}_n^{\boldsymbol{\pi}}$, let us denote the permuted U-statistic by

$$\begin{aligned}U_{\boldsymbol{\pi},\mathrm{HSIC}} &= \frac{(n-2)!}{n!}\sum_{(i,j)\in \boldsymbol{i}_2^n} k(Y_i, Y_j)\ell(Z_{\pi_i}, Z_{\pi_j}) + \frac{(n-4)!}{n!}\sum_{(i_1,i_2,j_1,j_2)\in \boldsymbol{i}_4^n} k(Y_{i_1}, Y_{j_1})\ell(Z_{\pi_{i_2}}, Z_{\pi_{j_2}}) \\ &\quad - \frac{2(n-3)!}{n!}\sum_{(i,j_1,j_2)\in \boldsymbol{i}_3^n} k(Y_i, Y_{j_1})\ell(Z_{\pi_i}, Z_{\pi_{j_2}}),\end{aligned}$$

where $\boldsymbol{i}_m^n$ stands for the set of all m-tuples drawn from $[n]$ without replacement. Kim et al. (2022a, Theorem 6.2) shows that $U_{\boldsymbol{\pi},\mathrm{HSIC}}$ satisfies

$$\mathbb{P}_{\boldsymbol{\pi}}\big(U_{\boldsymbol{\pi},\mathrm{HSIC}} \geq t \,|\, \mathcal{X}_n\big) \leq \exp\left\{-C_1 \min\left(\frac{t^2}{\Sigma_n^2}, \frac{t}{\Sigma_n}\right)\right\} \quad \text{for all } t > 0,$$

where

$$\Sigma_n^2 = \frac{1}{n^2(n-1)^2}\sup_{\boldsymbol{\pi}\in\boldsymbol{\Pi}_n}\left\{\sum_{(i,j)\in \boldsymbol{i}_2^n} k^2(Y_i, Y_j)\ell^2(Z_{\pi_i}, Z_{\pi_j})\right\} \leq \frac{1}{n(n-1)}K^2L^2.$$

The inequality above holds under the assumption that k and ℓ are uniformly bounded by K and L. Therefore, under the assumption of Lemma 12,

$$\mathbb{P}_{\boldsymbol{\pi}}\big(U_{\boldsymbol{\pi},\mathrm{HSIC}} \geq t \,|\, \mathcal{X}_n\big) \leq \exp\left\{-C_2 \min\left(\frac{n^2t^2}{K^2L^2}, \frac{nt}{KL}\right)\right\} \quad \text{for all } t > 0. \tag{53}$$

On the other hand, the squared empirical HSIC can be written in the form of a V-statistic given as

$$\begin{aligned}\widehat{\mathrm{HSIC}}^2(\mathcal{X}_n^{\boldsymbol{\pi}}) &= \frac{1}{n^2}\sum_{i,j=1}^{n} k(Y_i,Y_j)\ell(Z_{\pi_i},Z_{\pi_j}) + \frac{1}{n^4}\sum_{i_1,i_2,j_1,j_2=1}^{n} k(Y_{i_1},Y_{j_1})\ell(Z_{\pi_{i_2}},Z_{\pi_{j_2}}) \\ &\quad -\frac{2}{n^3}\sum_{i,j_1,j_2=1}^{n} k(Y_i,Y_{j_1})\ell(Z_{\pi_i},Z_{\pi_{j_2}}),\end{aligned}$$

and we note that the difference between the U-statistic and the V-statistic is bounded as

$$\begin{aligned}&\left|U_{\boldsymbol{\pi},\mathrm{HSIC}} - \widehat{\mathrm{HSIC}}^2(\mathcal{X}_n^{\boldsymbol{\pi}})\right| \\ &\le \left|\frac{(n-2)!}{n!}\sum_{(i,j)\in \boldsymbol{i}_2^n} k(Y_i,Y_j)\ell(Z_{\pi_i},Z_{\pi_j}) - \frac{1}{n^2}\sum_{i,j=1}^{n} k(Y_i,Y_j)\ell(Z_{\pi_i},Z_{\pi_j})\right| \\ &+ \left|\frac{(n-4)!}{n!}\sum_{(i_1,i_2,j_1,j_2)\in \boldsymbol{i}_4^n} k(Y_{i_1},Y_{j_1})\ell(Z_{\pi_{i_2}},Z_{\pi_{j_2}}) - \frac{1}{n^4}\sum_{i_1,i_2,j_1,j_2=1}^{n} k(Y_{i_1},Y_{j_1})\ell(Z_{\pi_{i_2}},Z_{\pi_{j_2}})\right| \\ &+ \left|\frac{2(n-3)!}{n!}\sum_{(i,j_1,j_2)\in \boldsymbol{i}_3^n} k(Y_i,Y_{j_1})\ell(Z_{\pi_i},Z_{\pi_{j_2}}) - \frac{2}{n^3}\sum_{i,j_1,j_2=1}^{n} k(Y_i,Y_{j_1})\ell(Z_{\pi_i},Z_{\pi_{j_2}})\right| \\ &\le C_3\frac{KL}{n}, \end{aligned} \tag{54}$$

for any $\boldsymbol{\pi}$ and $\mathcal{X}_n$. Using the above bound (54) and letting $t > 0$, we convert the concentration inequality for the permuted empirical HSIC into that for the permuted U-statistic as

$$\begin{aligned}\mathbb{P}_{\boldsymbol{\pi}}\big(\widehat{\mathrm{HSIC}}(\mathcal{X}_n^{\boldsymbol{\pi}}) \ge t \,|\, \mathcal{X}_n\big) &= \mathbb{P}_{\boldsymbol{\pi}}\big(\widehat{\mathrm{HSIC}}^2(\mathcal{X}_n^{\boldsymbol{\pi}}) \ge t^2 \,|\, \mathcal{X}_n\big) \\ &= \mathbb{P}_{\boldsymbol{\pi}}\left(\widehat{\mathrm{HSIC}}^2(\mathcal{X}_n^{\boldsymbol{\pi}}) - U_{\boldsymbol{\pi},\mathrm{HSIC}} + U_{\boldsymbol{\pi},\mathrm{HSIC}} \ge t^2 \,|\, \mathcal{X}_n\right) \\ &\le \mathbb{P}_{\boldsymbol{\pi}}\big(\big|\widehat{\mathrm{HSIC}}^2(\mathcal{X}_n^{\boldsymbol{\pi}}) - U_{\boldsymbol{\pi},\mathrm{HSIC}}\big| + U_{\boldsymbol{\pi},\mathrm{HSIC}} \ge t^2 \,|\, \mathcal{X}_n\big) \\ &\overset{\text{(i)}}{\le} \mathbb{P}_{\boldsymbol{\pi}}\left(U_{\boldsymbol{\pi},\mathrm{HSIC}} \ge t^2 - C_3\frac{KL}{n} \,|\, \mathcal{X}_n\right) \\ &\overset{\text{(ii)}}{\le} \mathbb{P}_{\boldsymbol{\pi}}\left(U_{\boldsymbol{\pi},\mathrm{HSIC}} \ge \frac{t^2}{2} \,|\, \mathcal{X}_n\right) \\ &\overset{\text{(iii)}}{\le} \exp\left\{-C_4\min\left(\frac{n^2t^4}{K^2L^2}, \frac{nt^2}{KL}\right)\right\}\end{aligned}$$

where step (i) uses the bound (54), step (ii) assumes that $t^2 \ge 2C_3\frac{KL}{n}$ and step (iii) follows by concentration inequality (53). Setting the last exponential bound to α and solving for t yield the desired result. □

We recall the concentration inequality of the empirical MMD presented by Gretton et al. (2012), which has been used in various places throughout the paper.

Lemma 13 (Theorem 7 of Gretton et al. 2012)**.** *Assume that the kernel k is bounded as $0 \le k(x,y) \le K$ for all $x, y \in \mathbb{S}$. Then for any $t > 0$*

$$\mathbb{P}\left\{\left|\widehat{\mathrm{MMD}}(\mathcal{X}_{n+m}) - \mathrm{MMD}_k(P,Q)\right| > 2\left(\sqrt{\frac{K}{m}} + \sqrt{\frac{K}{n}}\right) + t\right\} \le 2\exp\left(-\frac{t^2 mn}{2K(m+n)}\right).$$

The following is a counterpart to Lemma 13 for the empirical HSIC. While similar results exist in the literature (e.g., Gretton et al., 2005, Theorem 3), none precisely align with our specific requirements. Hence we opt to present a detailed proof of the following result.

Lemma 14 (Exponential Inequality for the Empirical HSIC)**.** *Assume that the kernels k and ℓ are bounded as $0 \le k(y,y') \le K$ and $0 \le \ell(z,z') \le L$ for all $y, y' \in \mathbb{Y}$ and $z, z' \in \mathbb{Z}$. Then for any $t \ge 0$,*

$$\mathbb{P}\left\{\left|\widehat{\mathrm{HSIC}}(\mathcal{X}_n) - \mathrm{HSIC}_{k\otimes\ell}(P_{YZ})\right| \ge C_1\sqrt{\frac{KL}{n}} + t\right\} \le 2\exp\left(-\frac{C_2 t^2 n}{KL}\right),$$

where C_1, C_2 are positive constants.

Proof. Throughout the proof, we denote by $C_1, C_2, \ldots$ some generic positive constants. By Lemma 6, the empirical HSIC has global sensitivity at most $C_1\sqrt{KL}/n$. Thus an application of McDiarmid's inequality (Gretton et al., 2012, Theorem 29) yields

$$\mathbb{P}\left\{\left|\widehat{\mathrm{HSIC}}(\mathcal{X}_n) - \mathbb{E}\left[\widehat{\mathrm{HSIC}}(\mathcal{X}_n)\right]\right| \ge t\right\} \le 2\exp\left(-\frac{C_2 t^2 n}{KL}\right) \quad \text{for all } t \ge 0. \tag{55}$$

Now we bound the difference between the expectation of $\widehat{\mathrm{HSIC}}(\mathcal{X}_n)$ and the population HSIC as

$$\begin{aligned}
&\left|\mathbb{E}\left[\widehat{\mathrm{HSIC}}(\mathcal{X}_n)\right] - \mathrm{HSIC}_{k\otimes\ell}(P_{YZ})\right| \\
&= \left|\mathbb{E}\left[\sup_{f\in\mathcal{F}_{k\otimes\ell}}\left\{\frac{1}{n}\sum_{i=1}^n f(Y_i,Z_i) - \frac{1}{n^2}\sum_{i,j=1}^n f(Y_i,Z_j)\right\}\right] - \sup_{f\in\mathcal{F}_{k\otimes\ell}}\left\{\mathbb{E}_{P_{YZ}}[f(Y,Z)] - \mathbb{E}_{P_Y P_Z}[f(Y,Z)]\right\}\right| \\
&\overset{\text{(i)}}{\le} \mathbb{E}\left|\sup_{f\in\mathcal{F}_{k\otimes\ell}}\left\{\frac{1}{n}\sum_{i=1}^n f(Y_i,Z_i) - \frac{1}{n^2}\sum_{i,j=1}^n f(Y_i,Z_j)\right\} - \sup_{f\in\mathcal{F}_{k\otimes\ell}}\left\{\mathbb{E}_{P_{YZ}}[f(Y,Z)] - \mathbb{E}_{P_Y P_Z}[f(Y,Z)]\right\}\right| \\
&\overset{\text{(ii)}}{\le} \mathbb{E}\left|\sup_{f\in\mathcal{F}_{k\otimes\ell}}\left\{\frac{1}{n}\sum_{i=1}^n f(Y_i,Z_i) - \mathbb{E}_{P_{YZ}}[f(Y,Z)] - \frac{1}{n^2}\sum_{i,j=1}^n f(Y_i,Z_j) + \mathbb{E}_{P_Y P_Z}[f(Y,Z)]\right\}\right| \\
&\overset{\text{(iii)}}{\le} \underbrace{\mathbb{E}\sup_{f\in\mathcal{F}_{k\otimes\ell}}\left|\frac{1}{n}\sum_{i=1}^n f(Y_i,Z_i) - \mathbb{E}_{P_{YZ}}[f(Y,Z)]\right|}_{:=\text{(I)}} + \underbrace{\mathbb{E}\sup_{f\in\mathcal{F}_{k\otimes\ell}}\left|\frac{1}{n^2}\sum_{i,j=1}^n f(Y_i,Z_j) - \mathbb{E}_{P_Y P_Z}[f(Y,Z)]\right|}_{:=\text{(II)}},
\end{aligned}$$

where step (i) uses Jensen's inequality, step (ii) uses the reverse triangle inequality and step (iii) follows by the triangle inequality. Let $\{(\tilde{Y}_i, \tilde{Z}_i)\}_{i=1}^n$ be i.i.d. copies of (Y_1, Z_1), and $\{\epsilon_i\}_{i=1}^n$ be i.i.d. Rademacher random variables. Then Jensen's inequality in conjunction with symmetrization yields

$$\begin{aligned}
\text{(I)} &\leq \mathbb{E} \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left| \frac{1}{n} \sum_{i=1}^n \epsilon_i \big\{ f(Y_i, Z_i) - f(\tilde{Y}_i, \tilde{Z}_i) \big\} \right| \\
&\leq 2\mathbb{E} \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left| \frac{1}{n} \sum_{i=1}^n \epsilon_i f(Y_i, Z_i) \right| \leq 2\sqrt{\frac{KL}{n}}, \qquad (56)
\end{aligned}$$

where the last inequality is due to Bartlett and Mendelson (2002, Lemma 22). For the second term,

$$\begin{aligned}
\text{(II)} &\overset{\text{(i)}}{\leq} \frac{1}{n} \sum_{j=1}^n \mathbb{E} \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left| \frac{1}{n} \sum_{i=1}^n f(Y_i, Z_j) - \mathbb{E}_{P_Y P_Z}[f(Y, Z)] \right| \\
&= \mathbb{E} \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left| \frac{1}{n} \sum_{i=2}^n f(Y_i, Z_1) + \frac{f(Y_1, Z_1)}{n} - \mathbb{E}_{P_Y P_Z}[f(Y, Z)] \right| \\
&\overset{\text{(ii)}}{\leq} \mathbb{E} \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left| \frac{1}{n} \sum_{i=2}^n f(Y_i, Z_1) - \mathbb{E}_{P_Y P_Z}[f(Y, Z)] \right| + \frac{1}{n} \mathbb{E} \sup_{f \in \mathcal{F}_{k\otimes\ell}} |f(Y_1, Z_1)| \\
&\overset{\text{(iii)}}{\leq} \mathbb{E} \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left| \frac{1}{n-1} \sum_{i=2}^n f(Y_i, Z_1) - \mathbb{E}_{P_Y P_Z}[f(Y, Z)] \right| \\
&\quad + \frac{1}{n(n-1)} \sum_{i=2}^n \mathbb{E} \sup_{f \in \mathcal{F}_{k\otimes\ell}} |f(Y_i, Z_1)| + \frac{1}{n} \mathbb{E} \sup_{f \in \mathcal{F}_{k\otimes\ell}} |f(Y_1, Z_1)|,
\end{aligned}$$

where step (i) follows by Jensen's inequality and step (ii) and step (iii) use the triangle inequality. Let $\mathcal{H}_{k\otimes\ell}$ be the reproducing kernel Hilbert space with the product kernel $k \otimes \ell$. Then by the reproducing property and the Cauchy–Schwarz inequality, we have

$$\sup_{f \in \mathcal{F}_{k\otimes\ell}} |f(y, z)| = \sup_{f \in \mathcal{F}_{k\otimes\ell}} |\langle k(\cdot, y)\ell(\cdot, z), f \rangle_{\mathcal{H}_{k\otimes\ell}}| \leq \sqrt{k(y, y)\ell(z, z)} \sup_{f \in \mathcal{F}_{k\otimes\ell}} \|f\|_{\mathcal{H}_{k\otimes\ell}} \leq \sqrt{KL}$$

for all $y \in \mathbb{Y}, z \in \mathbb{Z}$. Using this inequality and letting $\{(\tilde{Y}_i, \tilde{Z}_1)\}_{i=2}^n$ be i.i.d. copies of $\{(Y_i, Z_1)\}_{i=2}^n$ and recalling that $\{\epsilon_i\}_{i=1}^n$ are i.i.d. Rademacher random variables, the second term can be further bounded by

$$\begin{aligned}
\text{(II)} &\leq \mathbb{E} \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left| \frac{1}{n-1} \sum_{i=2}^n f(Y_i, Z_1) - \mathbb{E}_{P_Y P_Z}[f(Y, Z)] \right| + \frac{2\sqrt{KL}}{n} \\
&= \mathbb{E} \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left| \frac{1}{n-1} \sum_{i=2}^n f(Y_i, Z_1) - \frac{1}{n-1} \sum_{i=2}^n \mathbb{E}[f(\tilde{Y}_i, \tilde{Z}_1)] \right| + \frac{2\sqrt{KL}}{n} \qquad (57) \\
&\leq \mathbb{E} \sup_{f \in \mathcal{F}_{k\otimes\ell}} \left| \frac{1}{n-1} \sum_{i=2}^n f(Y_i, Z_1) - f(\tilde{Y}_i, \tilde{Z}_1) \right| + \frac{2\sqrt{KL}}{n}
\end{aligned}$$

$$= \mathbb{E}\left\{\mathbb{E}\left[\sup_{f\in\mathcal{F}_{k\otimes\ell}}\left|\frac{1}{n-1}\sum_{i=2}^{n}\epsilon_i\{f(Y_i,Z_1)-f(\tilde{Y}_i,\tilde{Z}_1)\}\right|\,\middle|\,Z_1,\tilde{Z}_1\right]\right\}+\frac{2\sqrt{KL}}{n} \tag{58}$$

$$\leq 2\mathbb{E}\left\{\mathbb{E}\left[\sup_{f\in\mathcal{F}_{k\otimes\ell}}\left|\frac{1}{n-1}\sum_{i=2}^{n}\epsilon_i f(Y_i,Z_1)\right|\,\middle|\,Z_1\right]\right\}+\frac{2\sqrt{KL}}{n}$$

$$\leq 2\sqrt{\frac{KL}{n-1}}+\frac{2\sqrt{KL}}{n}, \tag{59}$$

where the last inequality is due to Bartlett and Mendelson (2002, Lemma 22). Putting the inequalities (56) and (59) together, we have

$$\left|\mathbb{E}\left[\widehat{\mathrm{HSIC}}(\mathcal{X}_n)\right]-\mathrm{HSIC}_{k\otimes\ell}(P_{YZ})\right|\leq C_3\sqrt{\frac{KL}{n}}.$$

The above inequality along with McDiarmid's inequality (55) yields

$$\mathbb{P}\left\{\left|\widehat{\mathrm{HSIC}}(\mathcal{X}_n)-\mathrm{HSIC}_{k\otimes\ell}(P_{YZ})\right|\geq t+C_3\sqrt{\frac{KL}{n}}\right\}\leq 2\exp\left(-\frac{C_2t^2n}{KL}\right)\quad\text{for all } t\geq 0.$$

Hence we complete the proof of Lemma 14. □

The below lemma is folklore in the literature (*e.g.*, Romano and Wolf, 2005, Lemma 1) where we recall $\lfloor x\rfloor$ denotes the largest integer smaller than or equal to x. We provide a proof for completeness.

Lemma 15 (Permutation p-value)**.** *Suppose that $X_1,\ldots,X_n,X_{n+1}$ are exchangeable random variables. Then for any $\alpha\in[0,1]$, it holds that*

$$\mathbb{P}\left(\frac{1}{n+1}\left\{\sum_{i=1}^{n}\mathbb{1}(X_{n+1}\leq X_i)+1\right\}\leq\alpha\right)\leq\frac{\lfloor(n+1)\alpha\rfloor}{n+1}\leq\alpha.$$

Suppose further that $X_1,\ldots,X_{n+1}$ are all distinct with probability one. Then

$$\mathbb{P}\left(\frac{1}{n+1}\left\{\sum_{i=1}^{n}\mathbb{1}(X_{n+1}\leq X_i)+1\right\}\leq\alpha\right)=\frac{\lfloor(n+1)\alpha\rfloor}{n+1}.$$

Proof. Let $X_{(1)}\leq X_{(2)}\leq\ldots\leq X_{(n+1)}$ be the order statistics of $X_1,\ldots,X_{n+1}$. Then we have a series of the identities:

$$\frac{1}{n+1}\left\{\sum_{i=1}^{n}\mathbb{1}(X_{n+1}\leq X_i)+1\right\}=\frac{1}{n+1}\sum_{i=1}^{n+1}\mathbb{1}(X_{n+1}\leq X_i)\leq\alpha$$

$$\iff\sum_{i=1}^{n+1}\mathbb{1}(X_{n+1}\leq X_i)\leq\lfloor(n+1)\alpha\rfloor$$

$$\iff\sum_{i=1}^{n+1}\mathbb{1}(X_i<X_{n+1})\geq(n+1)-\lfloor(n+1)\alpha\rfloor:=k$$

$$\iff X_{n+1} > X_{(k)}.$$

Now by the exchangeability condition, we have

$$\mathbb{P}(X_{n+1} > X_{(k)}) \;=\; \mathbb{E}\left[\frac{1}{n+1}\sum_{i=1}^{n+1}\mathbb{1}(X_i > X_{(k)})\right].$$

On the other hand, by the definition of $X_{(k)}$,

$$\frac{1}{n+1}\sum_{i=1}^{n+1}\mathbb{1}(X_i > X_{(k)}) \le \frac{n+1-k}{n+1} = \frac{\lfloor (n+1)\alpha \rfloor}{n+1}.$$

Hence the first result follows. When $X_1, \ldots, X_{n+1}$ are all distinct, observe

$$\sum_{i=1}^{n+1}\mathbb{1}(X_i > X_{(k)}) = n+1-k.$$

Thus

$$\mathbb{E}\left[\frac{1}{n+1}\sum_{i=1}^{n+1}\mathbb{1}(X_i > X_{(k)})\right] = \frac{n+1-k}{n+1} = \frac{\lfloor (n+1)\alpha \rfloor}{n+1}.$$

□

The permutation test is often slightly conservative due to its discrete nature. The following randomization trick is well-known in the literature (e.g., Lehmann and Romano, 2005, Chapter 15), which modifies the permutation test to have type I error exactly equal to α.

Lemma 16 (Randomized Tests)**.** *For given $\alpha \in (0,1)$ and $\gamma \in (0, \alpha]$, consider a test function ϕ such that $\mathbb{P}_{H_0}(\phi = 1) = \gamma \le \alpha$. Letting U be a uniform random variable on $[0,1]$ independent of ϕ, define a randomized test ϕ_{rand} as*

$$\phi_{\text{rand}} = \phi + (1-\phi) \times \mathbb{1}\left(U \le \frac{\alpha - \gamma}{1-\gamma}\right).$$

The randomized test ϕ_{rand} has type I error exactly equal to α, and power never worse than that of ϕ, i.e., *$\mathbb{P}_{H_0}(\phi_{\text{rand}} = 1) = \alpha$ and $\mathbb{P}_{H_1}(\phi_{\text{rand}} = 1) \ge \mathbb{P}_{H_1}(\phi = 1)$.*

Proof. Type I error control is immediate given that

$$\begin{aligned}\mathbb{E}_{H_0}[\phi_{\text{rand}}] \;&=\; \mathbb{E}_{H_0}[\phi] + \mathbb{E}_{H_0}[1-\phi] \times \mathbb{P}_{H_0}\left(U \le \frac{\alpha-\gamma}{1-\gamma}\right) \\ &=\; \gamma + (1-\gamma) \times \frac{\alpha-\gamma}{1-\gamma} = \alpha.\end{aligned}$$

The second claim about the power is also immediate given that $\phi_{\text{rand}} \ge \phi$. □

The permutation test can be expressed in terms of the permutation p-value as well as the quantile of the permutation distribution, which can be justified by the following lemma.

Lemma 17 (Quantile Representation)**.** *For any $\alpha \in [0,1]$ and dataset $\{X_1, \ldots, X_{n+1}\}$, we have the identity*

$$\mathbb{1}\left(\frac{1}{n+1}\left\{\sum_{i=1}^{n} \mathbb{1}(X_{n+1} \leq X_i) + 1\right\} \leq \alpha\right) = \mathbb{1}\big(X_{n+1} > q_{1-\alpha}\big),$$

where $q_{1-\alpha}$ is the $1-\alpha$ quantile of $\{X_1, \ldots, X_{n+1}\}$ given as

$$q_{1-\alpha} = \inf\left\{t \in \mathbb{R} : \frac{1}{n+1}\sum_{i=1}^{n+1} \mathbb{1}(X_i \leq t) \geq 1-\alpha\right\}.$$

Moreover, by letting $X_{(\lceil (1-\alpha)(n+1) \rceil)}$ denote the $\lceil (1-\alpha)(n+1) \rceil$th order statistic of $X_1, \ldots, X_{n+1}$ and $X_{(0)} = -\infty$, we have $q_{1-\alpha} = X_{(\lceil (1-\alpha)(n+1) \rceil)}$.

Proof. The second claim follows by noting that

$$\begin{aligned}
q_{1-\alpha} &= \inf\left\{t \in \mathbb{R} : \frac{1}{n+1}\sum_{i=1}^{n+1} \mathbb{1}(X_i \leq t) \geq 1-\alpha\right\} \\
&= \inf\left\{t \in \mathbb{R} : \frac{1}{n+1}\sum_{i=1}^{n+1} \mathbb{1}(X_i < t) \geq 1-\alpha\right\} \qquad (60) \\
&= \inf\left\{t \in \mathbb{R} : \sum_{i=1}^{n+1} \mathbb{1}(X_i < t) \geq (1-\alpha)(n+1)\right\} \\
&= X_{(\lceil (1-\alpha)(n+1) \rceil)}.
\end{aligned}$$

For the first claim, denote $G(x) = \sum_{i=1}^{n+1} \mathbb{1}(X_i < x)$, which is a left-continuous step function. We then have

$$\begin{aligned}
\mathbb{1}\left(\frac{1}{n+1}\left\{\sum_{i=1}^{n} \mathbb{1}(X_{n+1} \leq X_i) + 1\right\} \leq \alpha\right) &= \mathbb{1}\big(G(X_{n+1}) \geq (1-\alpha)(n+1)\big) \\
&\leq \mathbb{1}(X_{n+1} > q_{1-\alpha})
\end{aligned}$$

where the last step follows since G is a left-continuous step function. In more detail, suppose that $G(X_{n+1}) \geq (1-\alpha)(n+1)$. Since G is a left-continuous step function, there exists a small constant $\epsilon > 0$ such that $G(X_{n+1} - \epsilon) = G(X_{n+1}) \geq (1-\alpha)(n+1)$. Therefore, X_{n+1} cannot be the $1-\alpha$ quantile and should be greater than $q_{1-\alpha}$.

Moreover, the event $X_{n+1} > q_{1-\alpha}$ implies that $G(X_{n+1}) \geq (1-\alpha)(n+1)$ by the definition of $q_{1-\alpha}$ as in expression (60). Hence we conclude that

$$\mathbb{1}\big(G(X_{n+1}) \geq (1-\alpha)(n+1)\big) = \mathbb{1}(X_{n+1} > q_{1-\alpha}),$$

and the first claim follows. □

The next lemma plays a crucial role in proving the validity of the differentially private permutation test in Algorithm 1.

Lemma 18 (Alternative Expression)**.** *Given $\alpha \in (0,1)$ and $n \geq 1$, set*

$$\alpha_\star = \max\left\{\left(\frac{n+1}{n}\alpha - \frac{1}{n}\right), 0\right\}.$$

Then for dataset $\{X_1, \ldots, X_{n+1}\}$, we have the identity

$$\mathbb{1}(X_{n+1} > q_{1-\alpha}) = \mathbb{1}(X_{n+1} > r_{1-\alpha_\star})\mathbb{1}\left(\alpha \geq \frac{1}{n+1}\right),$$

where $q_{1-\alpha}$ and $r_{1-\alpha_\star}$ are the $1-\alpha$ quantile of $\{X_i\}_{i=1}^{n+1}$ and the $1-\alpha_\star$ quantile of $\{X_i\}_{i=1}^{n}$, respectively.

Proof. For $\alpha \in (0, \frac{1}{n+1})$, $q_{1-\alpha}$ becomes the maximum of $X_1, \ldots, X_{n+1}$ for which $\mathbb{1}(X_{n+1} > q_{1-\alpha}) = 0$. Similarly, it becomes

$$\mathbb{1}(X_{n+1} > r_{1-\alpha_\star})\mathbb{1}\left(\alpha \geq \frac{1}{n+1}\right) = 0.$$

Hence we only need to verify the identity under $\alpha \geq \frac{1}{n+1}$. In what follows, we assume $\alpha \geq \frac{1}{n+1}$ and show that $\mathbb{1}(X_{n+1} > q_{1-\alpha}) = \mathbb{1}(X_{n+1} > r_{1-\alpha_\star})$.

Remark that the $1-\alpha$ quantile of $\{X_1 + c, \ldots, X_{n+1} + c\}$ is the same as the $1-\alpha$ quantile of $\{X_1, \ldots, X_{n+1}\}$ plus c for any $c \in \mathbb{R}$. Using this location-shift property of quantiles, observe that

$$\begin{aligned}
\mathbb{1}(X_{n+1} > q_{1-\alpha}) &= \mathbb{1}\left(0 > \inf\left\{x \in \mathbb{R} : \frac{1}{n+1}\sum_{i=1}^{n+1}\mathbb{1}(X_i - X_{n+1} \leq x) \geq 1-\alpha\right\}\right) \\
&= \mathbb{1}\left(0 > \inf\left\{x \in \mathbb{R} : \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(X_i - X_{n+1} \leq x) \geq \frac{n+1}{n}(1-\alpha) - \frac{\mathbb{1}(0 \leq x)}{n}\right\}\right) \\
&\geq \mathbb{1}\left(0 > \inf\left\{x \in \mathbb{R} : \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(X_i - X_{n+1} \leq x) \geq \frac{n+1}{n}(1-\alpha)\right\}\right) \\
&\overset{(\dagger)}{=} \mathbb{1}\left(0 > \inf\left\{x \in \mathbb{R} : \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(X_i - X_{n+1} \leq x) \geq \min\left[\frac{n+1}{n}(1-\alpha), 1\right]\right\}\right) \\
&= \mathbb{1}(X_{n+1} > r_{1-\alpha_\star}).
\end{aligned}$$

where the equality (†) holds under $\alpha \geq \frac{1}{n+1}$. Therefore it holds that $\mathbb{1}(X_{n+1} > q_{1-\alpha}) \geq \mathbb{1}(X_{n+1} > r_{1-\alpha_\star})$.

Next we prove the other direction $\mathbb{1}(X_{n+1} > q_{1-\alpha}) \leq \mathbb{1}(X_{n+1} > r_{1-\alpha_\star})$. Note that the infimum in the definition of a quantile can be replaced by the minimum so that

$$\frac{1}{n+1}\sum_{i=1}^{n+1}\mathbb{1}(X_i - X_{n+1} \leq \underbrace{q_{1-\alpha} - X_{n+1}}_{:=\tilde{q}_{1-\alpha}}) \geq 1-\alpha.$$

Having this in mind, for $\alpha \geq \frac{1}{n+1}$, assume $X_{n+1} > q_{1-\alpha}$ (equivalently $0 > \tilde{q}_{1-\alpha}$) under which it holds that

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i - X_{n+1} \leq \tilde{q}_{1-\alpha}) \geq \frac{n+1}{n}(1-\alpha) - \frac{\mathbb{1}(0 \leq \tilde{q}_{1-\alpha})}{n} = \frac{n+1}{n}(1-\alpha)$$

$$\iff \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i - X_{n+1} \leq \tilde{q}_{1-\alpha}) \geq \min\left\{\frac{n+1}{n}(1-\alpha),\, 1\right\}.$$

This implies that $\tilde{q}_{1-\alpha} \geq r_{1-\alpha_\star} - X_{n+1} \coloneqq \tilde{r}_{1-\alpha_\star}$ by the definition of $r_{1-\alpha_\star}$. Consequently, $\mathbb{1}(X_{n+1} > q_{1-\alpha}) \leq \mathbb{1}(\tilde{q}_{1-\alpha} \geq \tilde{r}_{1-\alpha_\star})$, which further implies that

$$\mathbb{1}(X_{n+1} > q_{1-\alpha}) \leq \mathbb{1}(\tilde{q}_{1-\alpha} \geq \tilde{r}_{1-\alpha_\star})\mathbb{1}(X_{n+1} > q_{1-\alpha}) \leq \mathbb{1}(0 > \tilde{r}_{1-\alpha_\star}) = \mathbb{1}(X_{n+1} > r_{1-\alpha_\star}).$$

Thus we conclude that $\mathbb{1}(X_{n+1} > q_{1-\alpha}) = \mathbb{1}(X_{n+1} > r_{1-\alpha_\star})$ for $\alpha \geq \frac{1}{n+1}$ as well. This completes the proof of Lemma 18. □

The next result is concerned with the global sensitivity of the quantiles.

Lemma 19 (Sensitivity of Quantiles)**.** *Suppose that the test statistic T has the global sensitivity at most Δ_T as in (4). Let us denote by $r_{1-\alpha}\big(\mathcal{X}_n; \{\boldsymbol{\pi}_i, \zeta_i\}_{i=1}^B\big)$ the $1-\alpha$ quantile of $\{M_i\}_{i=1}^B$ where $M_i = T(\mathcal{X}_n^{\boldsymbol{\pi}_i}) + 2\Delta_T \xi_{\varepsilon,\delta}^{-1}\zeta_i$ with $\xi_{\varepsilon,\delta} = \varepsilon + \log(1/(1-\delta))$ and $\zeta_i \overset{\text{i.i.d.}}{\sim} \mathsf{Laplace}(0,1)$ for $i \in [B]$. Then for any $\alpha \in [0,1)$, the global sensitivity of the $1-\alpha$ quantile satisfies*

$$\sup_{\substack{\mathcal{X}_n, \tilde{\mathcal{X}}_n: \\ d_{\text{ham}}(\mathcal{X}_n, \tilde{\mathcal{X}}_n) \leq 1}} \left| r_{1-\alpha}\big(\mathcal{X}_n; \{\boldsymbol{\pi}_i, \zeta_i\}_{i=1}^B\big) - r_{1-\alpha}\big(\tilde{\mathcal{X}}_n; \{\boldsymbol{\pi}_i, \zeta_i\}_{i=1}^B\big) \right| \;\leq\; \Delta_T,$$

for any permutations $\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_B$ and any $\zeta_1, \ldots, \zeta_B \overset{\text{i.i.d.}}{\sim} \mathsf{Laplace}(0,1)$.

Proof. Let $\tilde{\mathcal{X}}_n$ be a neighboring dataset of $\mathcal{X}_n$ where $\tilde{\mathcal{X}}_n$ and $\mathcal{X}_n$ differ only in their kth component for some $k \in [n]$. Denote the permuted test statistics computed on $\tilde{\mathcal{X}}_n^{\boldsymbol{\pi}_1}, \ldots, \tilde{\mathcal{X}}_n^{\boldsymbol{\pi}_B}$ by $\tilde{T}_1, \ldots, \tilde{T}_B$. For simplicity, we write $r_{1-\alpha} = r_{1-\alpha}\big(\mathcal{X}_n; \{\boldsymbol{\pi}_i, \zeta_i\}_{i=1}^B\big)$ and $\tilde{r}_{1-\alpha} = r_{1-\alpha}\big(\tilde{\mathcal{X}}_n; \{\boldsymbol{\pi}_i, \zeta_i\}_{i=1}^B\big)$. Having this notation, first note that $T_i \geq \tilde{T}_i - \Delta_T$ for all $i \in [B]$ under the assumption of (4) and thus

$$1-\alpha \leq \frac{1}{B}\sum_{i=1}^{B} \mathbb{1}(T_i + 2\Delta_T \xi_{\varepsilon,\delta}^{-1}\zeta_i \leq r_{1-\alpha}) \leq \frac{1}{B}\sum_{i=1}^{B} \mathbb{1}(\tilde{T}_i + 2\Delta_T \xi_{\varepsilon,\delta}^{-1}\zeta_i \leq r_{1-\alpha} + \Delta_T),$$

which implies that $\tilde{r}_{1-\alpha} \leq r_{1-\alpha} + \Delta_T$.

Next we argue that $\tilde{r}_{1-\alpha} \geq r_{1-\alpha} - \Delta_T$. For this direction, let $\epsilon > 0$ be an arbitrary constant. Then by the definition of $r_{1-\alpha}$ and $T_i \leq \tilde{T}_i + \Delta_T$ for all $i \in [B]$,

$$1-\alpha > \frac{1}{B}\sum_{i=1}^{B} \mathbb{1}(T_i + 2\Delta_T \xi_{\varepsilon,\delta}^{-1}\zeta_i \leq r_{1-\alpha} - \epsilon) \geq \frac{1}{B}\sum_{i=1}^{B} \mathbb{1}(\tilde{T}_i + 2\Delta_T \xi_{\varepsilon,\delta}^{-1}\zeta_i \leq r_{1-\alpha} - \epsilon - \Delta_T).$$

Hence $\tilde{r}_{1-\alpha} > r_{1-\alpha} - \epsilon - \Delta_T$. Since ϵ is arbitrary, we conclude $\tilde{r}_{1-\alpha} \geq r_{1-\alpha} - \Delta_T$. In summary, we have established that $|r_{1-\alpha} - \tilde{r}_{1-\alpha}| \leq \Delta_T$, which holds for any $k \in [n]$, any permutations $\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_B$ and any $\zeta_1, \ldots, \zeta_B \overset{\text{i.i.d.}}{\sim} \mathsf{Laplace}(0,1)$. Therefore, the desired claim follows. □

The Laplace mechanism introduces a perturbed statistic by adding Laplace noise. Instead of directly studying quantiles of this perturbed statistic, it is often easier to analyze the quantiles of the original statistic and the quantiles of the added Laplace noise separately. These results can then be combined using the following quantile inequality.

Lemma 20 (Quantile Inequality). *Let $q_{1-\alpha}^{X+Y}$ be the $1-\alpha$ quantile of $X+Y$. Similarly, let $q_{1-\alpha/2}^{X}$ and $q_{1-\alpha/2}^{Y}$ be the $1-\alpha/2$ quantile of X and Y, respectively. Then $q_{1-\alpha}^{X+Y} \leq q_{1-\alpha/2}^{X} + q_{1-\alpha/2}^{Y}$.*

Proof. Note that if $\{X+Y > q_{1-\alpha/2}^{X} + q_{1-\alpha/2}^{Y}\}$, then at least one of the events $\{X > q_{1-\alpha/2}^{X}\}$ and $\{Y > q_{1-\alpha/2}^{Y}\}$ should occur (otherwise contradiction). This together with the union bound yields

$$\begin{aligned}\mathbb{P}\big(X+Y > q_{1-\alpha/2}^{X} + q_{1-\alpha/2}^{Y}\big) &\leq \mathbb{P}\big(X > q_{1-\alpha/2}^{X}\big) + \mathbb{P}\big(Y > q_{1-\alpha/2}^{Y}\big) \\ &\leq \alpha,\end{aligned}$$

where the last inequality follows by the definition of quantile. The above implies that the $1-\alpha$ quantile of $X+Y$ is less than or equal to $q_{1-\alpha/2}^{X} + q_{1-\alpha/2}^{Y}$ and therefore the claim follows. □

The following lemma computes the difference between the V-statistic and U-statistic of the squared HSIC, which is useful in the proof of Theorem 14.

Lemma 21 (Difference between V_{HSIC} and U_{HSIC}). *Recall $\widehat{\text{HSIC}}^2$ given in (12) and U_{HSIC} given in (21). Suppose that kernels satisfy $k(y,y) = K$ and $\ell(z,z) = L$ for all $y \in \mathbb{Y}$ and $z \in \mathbb{Z}$. Then the difference between $\widehat{\text{HSIC}}^2$ and U_{HSIC} is computed as follows:*

$$\widehat{\text{HSIC}}^2 - U_{\text{HSIC}} = D_1 + D_2,$$

where

$$\begin{aligned}D_1 &= \frac{n-1}{n^2}KL - \frac{L}{n^3}\sum_{(i,j)\in \mathbf{i}_2^n} k(Y_i,Y_j) - \frac{K}{n^3}\sum_{(i,j)\in \mathbf{i}_2^n} \ell(Z_i,Z_j) \\ D_2 &= -\frac{3n^2-4n+2}{(n-1)n^4}\sum_{(i,j)\in \mathbf{i}_2^n} k(Y_i,Y_j)\ell(Z_i,Z_j) + \frac{2(5n^2-8n+4)}{n^4(n-1)(n-2)}\sum_{(i,j_1,j_2)\in \mathbf{i}_3^n} k(Y_i,Y_{j_1})\ell(Z_i,Z_{j_2}) \\ &\quad - \frac{6n^2-11n+6}{n^4(n-1)(n-2)(n-3)}\sum_{(i_1,i_2,j_1,j_2)\in \mathbf{i}_4^n} k(Y_{i_1},Y_{i_2})\ell(Z_{j_1},Z_{j_2}).\end{aligned}$$

Remark. The term D_1 is invariant to any permutation of either Y values or Z values.

Proof. Write the squared empirical HSIC in (12) as V_{HSIC} and the U-statistic of the squared HSIC as U_{HSIC}. Note that $V_{\text{HSIC}} = V_a + V_b - 2V_c$ where

$$V_a = \frac{1}{n^2}\sum_{i,j=1}^{n} k(Y_i,Y_j)\ell(Z_i,Z_j), \quad V_b = \frac{1}{n^4}\sum_{i_1,i_2,j_1,j_2=1}^{n} k(Y_{i_1},Y_{i_2})\ell(Z_{j_1},Z_{j_2}),$$

$$V_c \;=\; \frac{1}{n^3}\sum_{i,j_1,j_2=1}^{n} k(Y_i,Y_{j_1})\ell(Z_i,Z_{j_2}),$$

and each term can be expressed as

$$\begin{aligned}
V_a \;&=\; \frac{KL}{n} + \frac{1}{n^2}\sum_{(i,j)\in \mathbf{i}_2^n} k(Y_i,Y_j)\ell(Z_i,Z_j),\\
V_b \;&=\; \frac{KL}{n^3} + \frac{2K}{n^4}\sum_{(i,j)\in \mathbf{i}_2^n} \ell(Z_i,Z_j) + \frac{2L}{n^4}\sum_{(i,j)\in \mathbf{i}_2^n} k(Y_i,Y_j) + \frac{KLn(n-1)}{n^4}\\
&\quad + \frac{K(n-2)}{n^4}\sum_{(i,j)\in \mathbf{i}_2^n} \ell(Z_i,Z_j) + \frac{L(n-2)}{n^4}\sum_{(i,j)\in \mathbf{i}_2^n} k(Y_i,Y_j)\\
&\quad + \frac{2}{n^4}\sum_{(i,j)\in \mathbf{i}_2^n} k(Y_i,Y_j)\ell(Z_i,Z_j) + \frac{4}{n^4}\sum_{(i,j_1,j_2)\in \mathbf{i}_3^n} k(Y_i,Y_{j_1})\ell(Z_i,Z_{j_2})\\
&\quad + \frac{1}{n^4}\sum_{(i_1,i_2,j_1,j_2)\in \mathbf{i}_3^n} k(Y_{i_1},Y_{i_2})\ell(Z_{j_1},Z_{j_2}) \quad \text{and}\\
V_c \;&=\; \frac{KL}{n^2} + \frac{K}{n^3}\sum_{(i,j)\in \mathbf{i}_2^n} \ell(Z_i,Z_j) + \frac{L}{n^3}\sum_{(i,j)\in \mathbf{i}_2^n} k(Y_i,Y_j)\\
&\quad + \frac{1}{n^3}\sum_{(i,j)\in \mathbf{i}_2^n} k(Y_i,Y_j)\ell(Z_i,Z_j) + \frac{1}{n^3}\sum_{(i,j_1,j_2)\in \mathbf{i}_3^n} k(Y_i,Y_{j_1})\ell(Z_i,Z_{j_2}).
\end{aligned}$$

Note also that U_{HSIC} can be written as $U_{\text{HSIC}} = U_a + U_b - 2U_c$ where

$$\begin{aligned}
U_a \;&=\; \frac{1}{n(n-1)}\sum_{(i,j)\in \mathbf{i}_2^n} k(Y_i,Y_j)\ell(Z_i,Z_j),\\
U_b \;&=\; \frac{(n-4)!}{n!}\sum_{(i_1,i_2,j_1,j_2)\in \mathbf{i}_4^n} k(Y_{i_1},Y_{j_1})\ell(Z_{i_2},Z_{j_2}) \quad \text{and}\\
U_c \;&=\; \frac{1}{n(n-1)(n-2)}\sum_{(i,j_1,j_2)\in \mathbf{i}_3^n} k(Y_i,Y_{j_1})\ell(Z_i,Z_{j_2}).
\end{aligned}$$

The result follows by calculating $V_a - U_a$, $V_b - U_b$ and $-2V_c + 2U_c$, and simplifying their sum via algebra. □