# Coverage correlation: detecting singular dependencies between random variables

Xuzhi Yang, Mona Azadkia and Tengyao Wang
London School of Economics and Political Science

(August 11, 2025)

**Abstract**

We introduce the coverage correlation coefficient, a novel nonparametric measure of statistical association designed to quantifies the extent to which two random variables have a joint distribution concentrated on a singular subset with respect to the product of the marginals. Our correlation statistic consistently estimates an $f$-divergence between the joint distribution and the product of the marginals, which is 0 if and only if the variables are independent and 1 if and only if the copula is singular. Using Monge–Kantorovich ranks, the coverage correlation naturally extends to measure association between random vectors. It is distribution-free, admits an analytically tractable asymptotic null distribution, and can be computed efficiently, making it well-suited for detecting complex, potentially nonlinear associations in large-scale pairwise testing.

## 1 Introduction

Correlation is a measure of statistical association that quantifies how two variables tend to vary together. Classically, Pearson's correlation, $r^{X,Y}$, captures linear relationships between real-valued variables $X$ and $Y$ (Pearson, 1920). In contrast, Spearman's rank correlation, $\rho^{X,Y}$, and Kendall's rank correlation, $\tau^{X,Y}$, capture monotonic associations between $X$ and $Y$, using different definitions of rank concordance (Spearman, 1904; Kendall, 1938). A key limitation of these classical measures is their poor performance in detecting non-monotonic associations, even in noise-free data.

To overcome this limitation, numerous approaches have been proposed, including the maximal correlation coefficient (Hirschfeld, 1935; Gebelein, 1941; Rényi, 1959; Breiman and Friedman, 1985), various methods based on joint cumulative distribution functions and ranks (e.g. Hoeffding, 1948; Blum et al., 1961; Bergsma and Dassios, 2014; Drton et al., 2020; Deb and Sen, 2023) kernel-based methods (e.g. Gretton et al., 2005, 2008; Sen and Sen, 2014; Pfister et al., 2018; Zhang et al., 2018), information-theoretic coefficients (e.g. Linfoot, 1957; Kraskov et al., 2004; Reshef et al., 2011; Berrett and Samworth, 2019; Berrett et al., 2021), copula-based coefficients (e.g. Sklar, 1959; Schweizer and Wolff, 1981; Zhang, 2019) and coefficients based on pairwise distances (e.g. Friedman and Rafsky, 1983; Székely et al., 2007; Heller et al., 2013).

Although many of these coefficients are commonly applied, two significant drawbacks remain. Most are constructed with the primary goal of testing for independence, offering little direct information about the magnitude of the underlying dependence. Moreover, their null distributions are often analytically intractable, so p-values often must be obtained through computationally intensive permutation procedures.

Recently, there has been renewed interest in developing nonparametric measures of statistical association, driven in part by the need to identify relevant features and interactions in large datasets. Several new statistics have been proposed to capture the extent to which $Y$ can be expressed as a deterministic (measurable) function of $X$ (Dette et al., 2013; Chatterjee, 2021; Azadkia and Chatterjee, 2021; Deb et al., 2020; Wiesel, 2022; Azadkia and Roudaki, 2025).

Among these recent proposals, Chatterjee's correlation (Chatterjee, 2021) has seen remarkably rapid adoption in practice, particularly in fields such as bioinformatics, where uncovering complex and potentially nonlinear associations in high-dimensional data is a central challenge (e.g. Dong et al., 2023; Suo et al., 2024; Sansalone et al., 2024). Its popularity arises from several appealing properties: the statistic is distribution-free under the null hypothesis of independence, allowing precise characterisation of its asymptotic null distribution, and it is computationally efficient, scaling well to large datasets.

To illustrate, consider the task of detecting covariation in gene expression levels in a single-cell RNA sequencing experiment, where thousands of genes are measured across tens of thousands of cells. The scale of this problem creates a substantial multiple testing burden, often requiring raw p-values on the order of $10^{-8}$ or smaller to declare significance for any gene pair. In such settings, resampling-based tests such as permutation become computationally prohibitive, making access to an accurate asymptotic null distribution essential.

Chatterjee's correlation is specifically designed to capture the extent to which $Y$ can be expressed as a measurable function of $X$ and is therefore inherently asymmetric. While this asymmetry can be advantageous in certain contexts, such as when a clear predictor-response relationship is present, it may be less suitable in others. In the genetic association example above, the direction of dependence between gene expression levels is often not known *a priori*, and in fact, the relationship may not be directional at all. For instance, two genes, A and B, may exhibit strong statistical dependence simply because they are both downstream of a common regulator gene C, rather than one being a function of the other. Motivated by such considerations, we propose a new measure of statistical association, the *coverage correlation coefficient*, which is symmetric and designed to capture more general implicit functional relationships of the form $f(X, Y) = 0$. More precisely, the coverage correlation coefficient quantifies the extent to which the joint distribution $P^{(X,Y)}$ is singular with respect to the product of marginals $P^X \otimes P^Y$, thereby detecting dependencies that may lie on low-dimensional structures within the joint space.

### 1.1 Coverage correlation coefficient

To provide intuition for the coverage correlation coefficient, consider a simplified setting in which the random variables $X$ and $Y$ have marginal distributions $\mathrm{Unif}[0, 1]$. Given an i.i.d. sample $\{(X_i, Y_i)\}_{i=1}^n$ drawn from the joint distribution $P^{(X,Y)}$, we examine two extreme cases:

(i) $X$ and $Y$ are independent;
(ii) $P^{(X,Y)}$ is singular with respect to $\mathrm{Unif}([0, 1]^2)$.

In the first case, the sample points $(X_i, Y_i)$ are uniformly distributed over the unit square $[0, 1]^2$. In contrast, under the second scenario, the points $(X_i, Y_i)$ are concentrated on a subset of $[0, 1]^2$ with Lebesgue measure zero. These two cases yield qualitatively distinct scatter plots, reflecting the presence or absence of dependence.

To quantify this difference, consider the area of the region in $[0, 1]^2$ not covered by the union of $\ell_\infty$-balls (squares) centred at each sample point $(X_i, Y_i)$, where the area of each ball is fixed to be $1/n$. The behaviour of this uncovered area has been extensively studied in the context of *coverage processes* (see, e.g. Hall, 1988).

As $n \to \infty$, the limiting uncovered area exhibits fundamentally different behaviour in the two cases considered above: when $X$ and $Y$ are independent, it converges to $e^{-1}$, whereas when $P^{(X,Y)}$ is singular with respect to $\mathrm{Unif}([0,1]^2)$, it converges to 1 in probability.

Building on this geometric intuition, we propose the *coverage correlation coefficient* of random vectors $X \in \mathbb{R}^{d_X}$ and $Y \in \mathbb{R}^{d_Y}$ with $d_X, d_Y \in \mathbb{N}$, based on the (multivariate) ranks of $\boldsymbol{X} := (X_i)_{1\le i\le n}$ and $\boldsymbol{Y} := (Y_i)_{1\le i\le n}$. This statistic is powerful against the alternative where $X$ and $Y$ possess a singular dependence. Intuitively speaking, the proposed correlation coefficient measures the uncovered volume in $[0,1]^{d_X+d_Y}$ when small cubes of volume $1/n$, centred at the multivariate ranks of $(X_i, Y_i)$, are used to cover the space.

Given two sets of reference points $\boldsymbol{U} = (U_1, \dots, U_n)$ in $[0,1]^{d_X}$ and $\boldsymbol{V} = (V_1, \dots, V_n)$ in $[0,1]^{d_Y}$, let

$$\pi^X := \operatorname*{arg\,min}_{\pi \in \mathcal{S}_n} \frac{1}{n} \sum_{i=1}^n \|U_{\pi(i)} - X_i\|_2^2 \quad \text{and} \quad \pi^Y := \operatorname*{arg\,min}_{\pi \in \mathcal{S}_n} \frac{1}{n} \sum_{i=1}^n \|V_{\pi(i)} - Y_i\|_2^2$$

be the optimal transport maps from $\boldsymbol{X}$ to $\boldsymbol{U}$ and from $\boldsymbol{Y}$ to $\boldsymbol{V}$ respectively, where $\mathcal{S}_n$ is the set of all permutations on $[n] := \{1, \dots, n\}$. For each $i \in [n]$, the empirical multivariate ranks for $X_i$ and $Y_i$ are

$$R_i^X := U_{\pi^X(i)} \quad \text{and} \quad R_i^Y := V_{\pi^Y(i)}, \tag{1}$$

respectively, and we write

$$R_i := (R_i^X, R_i^Y) \in [0,1]^d, \tag{2}$$

for their joint rank, where $d := d_X + d_Y$. We remark that $R_i^X$ and $R_i^Y$ are known as the *Monge–Kantorovich ranks* in the literature (Chernozhukov et al., 2017; Hallin et al., 2021).

We also define a $d$-dimensional $\ell_\infty$ neighbourhood of radius $r$ centred at $w \in [0,1]^d$ as

$$B(w, r) := \{z \in [0,1]^d : \inf_{\ell \in \{-1,0,1\}^d} \|z - w - \ell\|_\infty \le r\}.$$

Here, the infimum over $\ell$ defines the $\ell_\infty$ distance with a periodic boundary condition that identifies opposite sides of the unit cube $[0,1]^d$. This improves the empirical performance of the correlation coefficient statistic in small sample size. Writing $\mathrm{vol}(\cdot)$ for the $d$-dimensional Lebesgue measure, for $\gamma \in (0,1)$, we define

$$\mathcal{V}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{V}; \gamma) := 1 - \mathrm{vol}\Big(\bigcup_{i=1}^n B(R_i, \gamma)\Big) \tag{3}$$

to be the uncovered volume in the $d$-dimensional unit cube outside subcubes of radius $\gamma$ centred at the empirical ranks.

In what follows, we will mostly be working with the uniformly random reference points $(U_1, V_1), \dots, (U_n, V_n) \overset{\text{iid}}{\sim} \mathrm{Unif}[0,1]^d$ and $\gamma := \frac{1}{2n^{1/d}}$ so that each subcube has volume $1/n$. We write $\mathcal{V}_n = \mathcal{V}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{V}; \gamma)$ for this specific choice of reference points and $\gamma$.

**Definition 1.** *The* empirical coverage correlation coefficient *between samples $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ is defined as*

$$\kappa_n^{X,Y} := \frac{\mathcal{V}_n - e^{-1}}{1 - e^{-1}}.$$

We note that $\kappa_n^{X,Y}$ is random due to the uniformly random reference points, even when conditioning on $\boldsymbol{X}$ and $\boldsymbol{Y}$. If a non-random correlation statistic is preferred, it is possible to replace the randomly generated $\boldsymbol{U}$ and $\boldsymbol{V}$ by a set of fixed reference points that are sufficiently 'spread out' in $[0,1]^d$. For instance, when $d_X = d_Y = 1$, a natural choice could be $\boldsymbol{U} = \boldsymbol{V} = (1/n, \ldots, (n-1)/n, 1)$. We will discuss this further in Section 2.

We summarise several key features of the coverage correlation coefficient $\kappa_n^{X,Y}$ as follows:

1. When $d_X = d_Y = 1$, the coverage correlation statistic converges to 0 in probability if and only if $X$ is independent of $Y$, and to 1 in probability if and only if $P^{(X,Y)}$ is singular with respect to the product of the marginals.

2. More generally, $\kappa_n^{X,Y}$ converges to a population quantity that measures an $f$-divergence between the joint distribution and the product of the marginals with respect to the divergence generator function $f(x) = \frac{e^{-x}-e^{-1}}{1-e^{-1}}$. Notably, $\kappa_n^{X,Y}$ circumvents the need for density estimation, distinguishing it from numerous existing divergence estimators (e.g. Rubenstein et al., 2019).

3. For any $d_X, d_Y$, under the null hypothesis of independence between $X$ and $Y$, $\kappa_n^{X,Y}$ is asymptotically normally distributed, which allows us to construct asymptotically valid p-values.

4. The coverage correlation statistic is distribution-free, thanks to Monge–Kantorovich ranks used in (1). This is in contrast to other possible multivariate ranks such as depth-based ranks (Tukey, 1975; Liu and Singh, 1993; Zuo and Serfling, 2000), spatial ranks (Möttönen and Oja, 1995; Chaudhuri, 1996; Koltchinskii, 1997), componentwise ranks (Hodges, 1955; Bickel, 1965), and Mahalanobis ranks (Hallin and Paindaveine, 2002b,a). The distribution-free property of the coverage correlation statistic yields a pivotal null distribution, enabling easy computation of p-values.

5. For univariate marginal distributions, we develop an algorithm with $O(n \log n)$ time complexity (see Appendix C.1). The method is implemented as R and Python packages `covercorr`. Both the package and simulation code for reproducing figures and tables in the paper can be found at https://github.com/wangtengyao/covercorr.

### 1.2 Connection to Chatterjee's correlation

For random variables $X$ and $Y$ and given a sample $(X_i, Y_i)_{i\in[n]}$, let $X_{(1)} \le \cdots \le X_{(n)}$ denote the order statistics of the $X_i$'s and let $(Y_{(i)})_{i\in[n]}$ denote the corresponding concomitants. Assuming for simplicity that there are no ties, the empirical Chatterjee's correlation is defined as

$$\xi_n^{X,Y} := 1 - \frac{\sum_{i=1}^{n-1} |r_{i+1} - r_i|}{(n^2-1)/3}, \tag{4}$$

where $r_i := \#\{j : Y_{(j)} \le Y_{(i)}\}$ is the rank of $Y_{(i)}$. Chatterjee (2021, Theorem 1.1) shows that $\xi_n^{X,Y}$ converges stochastically to 0 when $X$ and $Y$ are independent, and to 1 when $Y$ is a function of $X$. This convergence is interpreted through the population statistic

$$\xi^{X,Y} := \frac{\int_{\mathbb{R}} \mathrm{Var}(\mathbb{E}[\mathbb{1}\{Y \ge t\} \mid X])\, dP^Y(t)}{\int_{\mathbb{R}} \mathrm{Var}(\mathbb{E}[\mathbb{1}\{Y \ge t\}])\, dP^Y(t)}.$$

We remark that (4) can also be interpreted as a measure of 'excess vacancy', similar to our coverage correlation coefficient, in the following sense. Writing $\tilde{\mathcal{V}} := 1 - \sum_{i=1}^{n-1} \left| \frac{r_{i+1}}{n} - \frac{r_i}{n} \right| \cdot \frac{1}{n}$ for the total area in $[0,1]^2$ not covered by the union of rectangles

$$\bigcup_{i=1}^{n-1} \Big( \Big[\frac{i-1}{n}, \frac{i}{n}\Big] \times \Big[\min\Big\{\frac{r_i}{n}, \frac{r_{i+1}}{n}\Big\}, \min\Big\{\frac{r_i}{n}, \frac{r_{i+1}}{n}\Big\}\Big]\Big),$$

we have

$$\xi_n^{X,Y} = \frac{\tilde{\mathcal{V}} - 2/3}{1/3} + O(n^{-2}).$$

In other words, if we draw a line plot of the ordered normalised $X_i$ ranks (which are $1/n, 2/n, \ldots, 1$) against the corresponding normalised $Y_i$ ranks $(r_1/n, \ldots, r_n/n)$, with a line 'thickness' of $1/n$, then $\tilde{\mathcal{V}}$ approximates the area in $[0,1]^2$ that remains uncovered by this thickened line plot.

Figure 1 illustrates this for several $(X, Y)$ distributions. In the case of independence (first column), the normalised $Y_i$ ranks resemble independent Unif$[0,1]$ values, and successive differences in rank are approximately $1/3$ on average, yielding $\tilde{\mathcal{V}} \approx 2/3$. On the other hand, when $Y$ is a function of $X$, the normalised $Y_i$ ranks become a deterministic function of the normalised $X_i$ ranks (modulo discretisation), which under mild conditions causes the uncovered area $\tilde{\mathcal{V}}$ to shrink towards 0. As seen in the second and third columns, when a fraction of the $Y_i$'s can be well-approximated by a function of the $X_i$'s, the area uncovered by the line plot decreases accordingly. Chatterjee's correlation captures this structure.

The last column of Figure 1 presents an interesting example in which both $X$ and $Y$ are generated as functions of a third latent variable $U$, with added noise. In this setting, even though a visually striking relationship exists between the $X_i$'s and $Y_i$'s, Chatterjee's correlation remains close to zero. This highlights a key limitation of the statistic: it is specifically designed to detect directional functional dependence, but it may fail to capture more symmetric or indirect relationships.

## 2 Theoretical guarantees

In this section, we present theoretical results for the coverage correlation coefficient $\kappa_n^{X,Y}$ defined in Definition 1. We first show that for univariate $X$ and $Y$, $\kappa_n^{X,Y}$ converges in probability to a population quantity that measures an $f$-divergence between $P^{(X,Y)}$ and $P^X \otimes P^Y$. We recall the definition of $f$-divergence (see, e.g. Samworth and Shah, 2025+, Definition 8.2).

**Definition 2.** *Let $f : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ be a convex function with $f(1) = 0$. For any two probability measures $\mu, \nu$ on a space $\mathcal{S}$, let $\mathrm{d}\mu = h\,\mathrm{d}\nu + \mathrm{d}\nu^{\perp}$ be the Lebesgue–Radon–Nikodym decomposition of $\mu$ with respect to $\nu$, where $h$ is $\nu$-integrable and $\nu^{\perp}$ is singular with respect to $\nu$. The $f$-*divergence *between $\mu$ and $\nu$ is defined as*

$$D_f(\mu \,\|\, \nu) = \int_{\mathcal{S}} f \circ h \, d\nu + f'(\infty)\nu^{\perp}(\mathcal{S}),$$

*where $f'(\infty) := \lim_{t\to\infty} t^{-1} f(t)$ is the asymptotic slope of $f$ at infinity.*

The following theorem shows that when $d_X = d_Y = 1$, the correlation coefficient $\kappa_n^{X,Y}$ converges stochastically to a population limit.

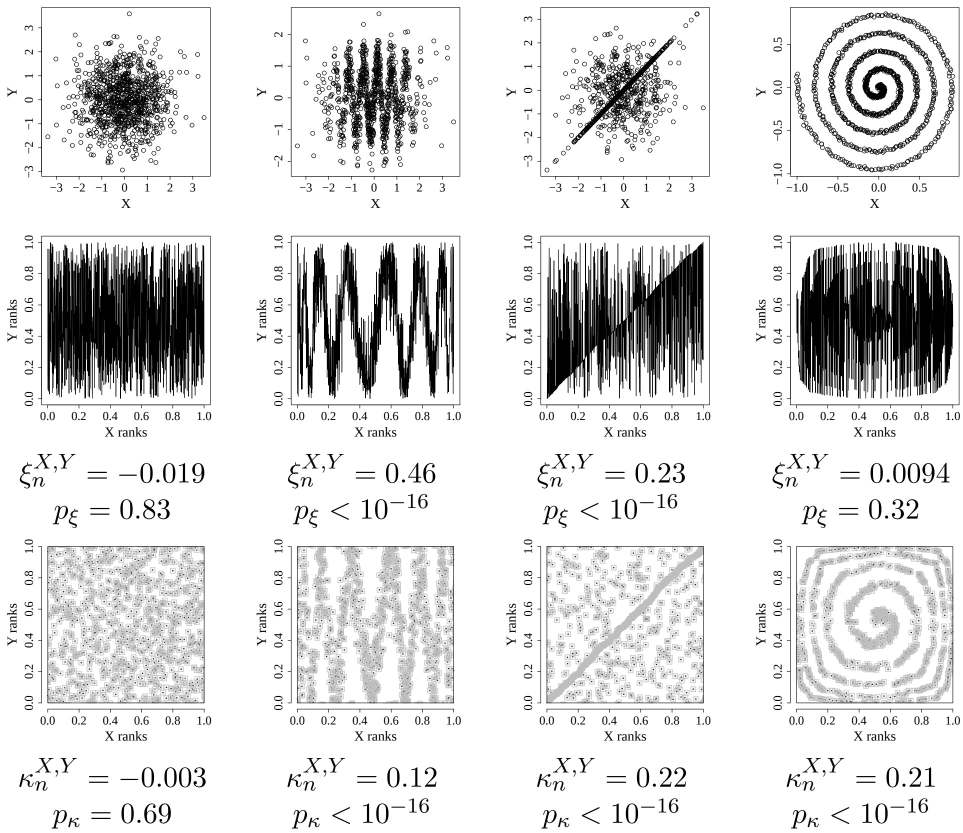


Figure 1: Chatterjee's correlation and coverage correlation for various joint distributions using a sample of $n = 1000$ observation pairs. Data generating mechanisms are as follows — first column: $X, Y \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$; second column: $X \sim \mathcal{N}(0,1)$ and $Y = \sin(10X) + 0.5\epsilon$ where $\epsilon \sim \mathcal{N}(0,1) \perp\!\!\!\perp (X,Y)$; third column: $X \sim \mathcal{N}(0,1)$ and $Y = XB + \epsilon(1-B)$, where $(B,\epsilon) \sim \text{Bernoulli}(1/2) \otimes \mathcal{N}(0,1) \perp\!\!\!\perp (X,Y)$; fourth column: $X = U\sin(10\pi U) + 0.01\epsilon_X$ and $Y = U\cos(10\pi U) + 0.01\epsilon_Y$, where $(U, \epsilon_X, \epsilon_Y) \sim \text{Unif}[0,1] \otimes \mathcal{N}(0,1) \otimes \mathcal{N}(0,1) \perp\!\!\!\perp (X,Y)$. For each column, the top panel shows the scatter plot, the middle panel shows the line plot of ordered normalised $X$ ranks against the corresponding $Y$ ranks, and the bottom panel shows the union of small squares of area $1/n$, centred at joint normalised ranks of $X$ and $Y$. Test statistics and p-values for both correlation measures are also displayed.

**Theorem 1.** *Let $P^{(X,Y)}$ be a Borel probability measure on $\mathbb{R}^2$ with marginals $P^X$ and $P^Y$. Define $f : \mathbb{R} \to \mathbb{R}$ as $f(x) = (e^{-x} - e^{-1})/(1 - e^{-1})$. Given $(X_1, Y_1), \ldots, (X_n, Y_n) \overset{\text{iid}}{\sim} P^{(X,Y)}$, we have*

$$\kappa_n^{X,Y} \xrightarrow{\text{p}} \kappa^{X,Y} := D_f(P^{(X,Y)} \,\|\, P^X \otimes P^Y), \quad \text{as } n \to \infty. \tag{5}$$

We do not impose any assumptions on the joint distribution $P^{(X,Y)}$, thus the variables $X$ and $Y$ may be continuous, discrete, or a mixture of both. By the data processing inequality (see Lemma 15), it suffices to consider distributions $P^{(X,Y)}$ with Unif$[0,1]$ marginals. Moreover, if $X$ and $Y$ are independent, then $\kappa_n^{X,Y} \xrightarrow{\text{p}} 0$, i.e. $\mathcal{V}_n \xrightarrow{\text{p}} e^{-1}$. This can be established relatively easily, without invoking Theorem 1, by observing that for an independent point $W \sim \text{Unif}([0,1]^2)$ we have

$$\begin{aligned} \mathbb{E}(\mathcal{V}_n) &= \mathbb{E}\Big[\mathbb{P}\Big\{W \notin \bigcup_{i=1}^n B\Big(R_i, \frac{1}{2\sqrt{n}}\Big) \;\Big|\; R_1, \ldots, R_n\Big\}\Big] = \mathbb{E}\Big[\mathbb{P}\Big\{R_i \notin B\Big(W, \frac{1}{2\sqrt{n}}\Big)\ \forall i \in [n] \;\Big|\; W\Big\}\Big] \\ &= (1 - 1/n)^n \to e^{-1}, \end{aligned} \tag{6}$$

and similarly (through a second moment calculation) $\text{Var}(\mathcal{V}_n) \to 0$. However, this argument does not extend to the general case, because the joint ranks $R_i$ are no longer independent, so the final equality above does not hold. This subtle dependence structure among the $R_i$'s is the main technical obstacle in proving Theorem 1.

Interestingly, if we begin with a distribution $P^{(X,Y)}$ having Unif$([0,1]^{d_X})$ and Unif$([0,1]^{d_Y})$ marginals and compute the coverage correlation on the raw data $(X_i, Y_i)$ directly (i.e. setting $R_i = (X_i, Y_i)$ and skipping the rank transforms), then a similar argument to the one in (6) shows that $\kappa_n^{X,Y} \xrightarrow{\text{p}} \kappa^{X,Y}$ for any Borel probability measure $P^{(X,Y)}$ on $\mathbb{R}^{d_X+d_Y}$, instead of just for univariate $X$ and $Y$. Based on this heuristic and supporting numerical results, we conjecture that Theorem 1 holds for multivariate $X$ and $Y$. However, the current proof technique relies on the total ordering of $\mathbb{R}$, and thus does not immediately generalise to higher dimensions. Nonetheless, we are able to establish the following partial result in general dimensions.

**Proposition 2.** *Suppose $(X_1, Y_1), \ldots, (X_n, Y_n) \overset{\text{iid}}{\sim} P^{(X,Y)}$ for a Borel probability measure $P^{(X,Y)}$ on $\mathbb{R}^{d_X+d_Y}$ with marginals $P^X$ and $P^Y$ on $\mathbb{R}^{d_X}$ and $\mathbb{R}^{d_Y}$ respectively.*

*(i) If $P^{(X,Y)} = P^X \otimes P^Y$, then $\kappa_n^{X,Y} \xrightarrow{\text{p}} 0$.*

*(ii) If $P^{(X,Y)}$ is singular with respect to $P^X \otimes P^Y$, then $\kappa_n^{X,Y} \xrightarrow{\text{p}} 1$.*

We have defined the coverage correlation using uniformly distributed reference points $\boldsymbol{U}$ and $\boldsymbol{V}$. When $X$ and $Y$ are univariate, as considered in Theorem 1, it is also quite natural to use a grid reference $\boldsymbol{U} = \boldsymbol{V} = (1/n, 2/n, \ldots, 1)$. Such a grid reference derandomises the coverage correlation coefficient, which can be desirable in practice. As we show below, the empirical coverage correlation coefficient converges to the same population limit when defined through this grid reference.

**Theorem 3.** *Let $P^{(X,Y)}$ be a Borel probability measure on $\mathbb{R}^2$ with marginals $P^X$ and $P^Y$ and let $f$ be defined as in Theorem 1. Given $\boldsymbol{X} = (X_i)_{i\in[n]}$ and $\boldsymbol{Y} = (Y_i)_{i\in[n]}$ such that $(X_1, Y_1), \ldots, (X_n, Y_n) \overset{\text{iid}}{\sim} P^{(X,Y)}$, and $\boldsymbol{U} = \boldsymbol{V} = (1/n, 2/n, \ldots, 1)$, define*

$$\kappa_n^{X,Y;\text{grid}} := \frac{\mathcal{V}\big(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{V}; \frac{1}{2\sqrt{n}}\big) - e^{-1}}{1 - e^{-1}}.$$

*We have*

$$\kappa_n^{X,Y;\text{grid}} \xrightarrow{\text{p}} \kappa^{X,Y} := D_f(P^{(X,Y)} \,\|\, P^X \otimes P^Y), \quad \text{as } n \to \infty.$$

Theorem 1 allows us to characterise the joint distribution $P^{(X,Y)}$ when $\kappa_n^{X,Y} \xrightarrow{\text{p}} 0$ and $\kappa_n^{X,Y} \xrightarrow{\text{p}} 1$ respectively. The following proposition formalises this, as well as establishing several other properties of the population version of the coverage correlation coefficient.

**Proposition 4.** *Suppose $(X,Y) \sim \mathcal{P}^{(X,Y)} \in \mathcal{P}(\mathbb{R}^d)$ and $P^X$, $P^Y$ are the corresponding marginal probability measures. Let $\kappa^{X,Y}$ be defined as (5), then we have*

*(i)* $\kappa^{X,Y} = 0$ *if and only if $X$ is independent of $Y$;*
*(ii)* $\kappa^{X,Y} = 1$ *if and only if $P^{(X,Y)}$ is singular with respect to $P^X \otimes P^Y$.*
*(iii) For a random variable $Z$, if $X \perp\!\!\!\perp Y \mid Z$, then we have $\kappa^{X,Z} \geq \kappa^{X,Y}$;*
*(iv) For any sequence of random variables $X^{(n)}$ and $Y^{(n)}$ such that $P^{(X^{(n)},Y^{(n)})} \xrightarrow{\text{d}} P^{(X,Y)}$, we have $\liminf_{n\to\infty} \kappa^{X^{(n)},Y^{(n)}} \geq \kappa^{X,Y}$;*
*(v)* $\kappa^{X,Y} = \kappa^{Y,X}$.

We remark that parts (i), (iii), (iv) and (v) demonstrate that $\kappa^{X,Y}$, as a measure of statistical association, satisfies the 'zero-independence', 'information-monotonicity', 'lower semicontinuity' and 'symmetry axioms' considered in Borgonovo et al. (2025) (see also Móri and Székely, 2019; Rényi, 1959). Part (ii) is related to the 'max-functionality' axiom, though the coverage correlation measures a more general statistical association between $X$ and $Y$ than a purely directional functional relationship. Also, the following information gain inequality is an immediate consequence of part (iii).

(iii') for any random variables $X$, $X'$ and $Y$, we have $\kappa^{(X,X'),Y} \geq \kappa^{X,Y}$.

This inequality is not mentioned in Borgonovo et al. (2025), but it appears as an axiom for dependency measurement in Griessenberger et al. (2022).

The following result derives the asymptotic distribution of $\kappa_n^{X,Y}$ under the null. This allows us to use the coverage correlation to perform independence testing between $X$ and $Y$.

**Theorem 5.** *Suppose that $X$ and $Y$ are independent random vectors on $\mathbb{R}^{d_X}$ and $\mathbb{R}^{d_Y}$ respectively. Define*

$$\sigma_n^2 := \frac{1}{(1-e^{-1})^2} \sum_{k=2}^{n} \binom{n}{k} \Big(1 - \frac{2}{n}\Big)^{n-k} \Big\{ \Big(\frac{2}{k+1}\Big)^d n^{-k-1} - n^{-2k} \Big\}.$$

*Given an independent and identically distributed copies $(X_1,Y_1),\ldots,(X_n,Y_n)$ of $(X,Y)$, we have*

$$\frac{\sqrt{n}\kappa_n^{X,Y}}{\sigma_n} \xrightarrow{\text{d}} \mathcal{N}(0,1)$$

*as $n \to \infty$.*

Based on the above theorem, we can construct a test for

$$\mathcal{H}_0 : P^{(X,Y)} = P^X \otimes P^Y \quad \text{versus} \quad \mathcal{H}_1 : P^{(X,Y)} \neq P^X \otimes P^Y$$

by rejecting the null hypothesis if

$$\frac{\sqrt{n}\kappa_n^{X,Y}}{\sigma_n} \geq z_\alpha, \tag{7}$$

where $z_\alpha$ is the upper $\alpha$ quantile of the standard normal distribution. Since the limiting distribution is independent of the marginal distribution $P^X$ and $P^Y$, test (7) is distribution-free.

The normalising factor $\sigma_n/\sqrt{n}$ in Theorem 5 is the exact standard deviation of the empirical coverage correlation under the null. As can be seen in Lemma 24, we have

$$\lim_{n\to\infty} \sigma_n^2 = \frac{1}{(e-1)^2} \sum_{k=2}^{\infty} \frac{1}{k!} \left(\frac{2}{k+1}\right)^{d_X+d_Y} =: \sigma^2$$

One can equivalently construct the asymptotic test by replacing $\sigma_n$ with $\sigma$ in (7), though using the exact variance $\sigma_n^2$ improves the finite sample performance of the test when $n$ is relatively small. When $d_X = d_Y = 1$, the expression of the asymptotic variance $\sigma^2$ has an explicit value of $(e-1)^{-2}(4\mathrm{Ei}(1) - 4\gamma_0 - 5) \approx 0.091992$, where $\gamma_0$ is the Euler–Mascheroni constant and $\mathrm{Ei}(1)$ is the exponential integral evaluated at 1.

While the central limit theorem in Theorem 5 allows us to derive asymptotically valid p-values, the worst-case relative error of such p-values can still be large in the tails of the normal distribution (e.g. a bound of order $n^{-1/2}$ from the Berry–Esseen theorem (Berry, 1941; Esseen, 1942)). However, using the fact that the coverage correlation coefficient statistic exhibits weak dependence on individual data point $(X_i, Y_i)$, we are able to derive a finite-sample concentration inequality.

**Theorem 6.** *Let $P^{(X,Y)}$ be a Borel probability measure on $\mathbb{R}^2$. There exists a universal constant $C > 0$ such that for any $t > 0$, we have*

$$\mathbb{P}(|\mathcal{V}_n - \mathbb{E}\mathcal{V}_n| \geq t) \leq 2(n+1)e^{-C\min\{nt^2,(nt^2)^{1/3}\}}.$$

We remark that the above concentration is not sub-Gaussian for large deviations due to the fact that the spacings between consecutive order statistics of the uniformly random reference points $\boldsymbol{U}$ and $\boldsymbol{V}$ have sub-Gamma tails and deviate from the expected spacing of $1/(n+1)$. However, if we use the grid reference as in Theorem 3, which has constant spacing between consecutive grid points, a sub-Gaussian concentration is available.

**Proposition 7.** *Let $P^{(X,Y)}$ be a Borel probability measure on $\mathbb{R}^2$ and let $\kappa_n^{X,Y;\mathrm{grid}}$ be defined as in Theorem 3. There exists a universal constant $C > 0$ such that for any $t > 0$, we have*

$$\mathbb{P}\big(|\kappa_n^{X,Y;\mathrm{grid}} - \mathbb{E}(\kappa_n^{X,Y;\mathrm{grid}})| \geq t\big) \leq 2e^{-Cnt^2}.$$

# 3 Empirical studies

## 3.1 Numerical simulations

Table 1 summarises the finite-sample sizes of the independence test based on the coverage correlation coefficient $\kappa_n^{X,Y}$ at various nominal levels, for $n \in \{10, 100, 1000\}$ and $d_X = d_Y \in \{1, 2\}$. We see that the test is a bit conservative at relatively low sample sizes and well-calibrated for large sample sizes.

In Figure 2, we compare the power of the test based on $\xi_n^{X,Y}$ to those of Chatterjee's correlation (Chatterjee, 2021, implemented in the `XICOR` R package), distance correlation (dCor) (Székely et al., 2007, implemented in the `Rfast` R package), Hilbert–Schmidt Independence Criterion (HSIC) (Gretton et al., 2008, implemented in the `dHSIC` R package), the kernel measure of association (KMAc) (Deb et al., 2020, implemented in the `KPC` R package) and the U-statistics

| $n$ | $d_X$ | $\alpha = 1\%$ | $\alpha = 2.5\%$ | $\alpha = 5\%$ | $\alpha = 10\%$ |
|---|---|---|---|---|---|
| 10 | 1 | $0.69_{(0.03)}$ | $1.54_{(0.04)}$ | $3.03_{(0.05)}$ | $6.02_{(0.08)}$ |
| 100 | 1 | $0.93_{(0.03)}$ | $2.27_{(0.05)}$ | $4.34_{(0.06)}$ | $8.78_{(0.09)}$ |
| 1000 | 1 | $0.96_{(0.03)}$ | $2.34_{(0.05)}$ | $4.76_{(0.07)}$ | $9.50_{(0.09)}$ |
| 10 | 2 | $0.55_{(0.02)}$ | $1.18_{(0.03)}$ | $2.10_{(0.05)}$ | $4.08_{(0.06)}$ |
| 100 | 2 | $0.94_{(0.03)}$ | $2.11_{(0.05)}$ | $4.12_{(0.06)}$ | $8.03_{(0.09)}$ |
| 1000 | 2 | $0.97_{(0.03)}$ | $2.36_{(0.05)}$ | $4.66_{(0.07)}$ | $9.30_{(0.09)}$ |

Table 1: Empirical sizes (in percentage) of independent test based on coverage correlation coefficient at various nominal levels $\alpha$, estimated over 100000 Monte Carlo repetitions (with standard errors in brackets).

permutation test (USP) (Berrett et al., 2021, implemented in the `USP` R package). For dCor, HSIC, KMAc and USP, we run 100 permutations to obtain p-values. Also, for USP, we set $M = 3$ for the maximum frequency to use in the Fourier basis. We generate $n \in \{1000, 2000\}$ independent copies of $(X, Y)$ pair in $\mathbb{R}^{d_X} \times \mathbb{R}^{d_Y}$ for $d_X = d_Y \in \{1, 2\}$ from one of the six data generating mechanisms described below at different noise levels $\gamma \in \{0, 0.2, \ldots, 1.8, 2\}$ (all functions below are applied componentwise for vector inputs and $(\epsilon_X, \epsilon_Y) \sim \mathcal{N}(0, I_{d_X}) \otimes \mathcal{N}(0, I_{d_Y})$ is independent of all other randomness):

(i) sinusoidal: $X \sim \mathrm{Unif}([-1, 1]^{d_X})$, $Y = \cos(8\pi X) + \gamma\epsilon_Y$

(ii) zigzag: $X \sim \mathrm{Unif}([-1, 1]^{d_X})$, $Y = |X - 0.5\mathrm{sgn}(X)| + \gamma\epsilon_Y$

(iii) circle: $U \sim \mathrm{Unif}([0, 2\pi]^{d_X})$, $X = \cos(U) + 0.5\gamma\epsilon_X$, $Y = \sin(U) + 0.5\gamma\epsilon_Y$

(iv) spiral: $U \sim \mathrm{Unif}([0, 1]^{d_X})$, $X = U\sin(10\pi U) + 0.15\gamma\epsilon_X$, $Y = U\cos(10\pi U) + 0.15\gamma\epsilon_Y$

(v) Lissajous: $U \sim \mathrm{Unif}([0, 1]^{d_X})$, $X = \sin(3U + \pi/2) + 0.1\gamma\epsilon_X$, $Y = \sin(4U) + 0.1\gamma\epsilon_Y$

(vi) local: $Z \sim \mathcal{N}(0, I_{d_X})$, $W \sim \mathcal{N}(0, I_{d_Y})$, $X = Z + 0.8\epsilon_X$, $Y = \mathbb{1}_{\{Z>0, W>0\}} Z + (1 - \mathbb{1}_{\{Z>0, W>0\}})W + \epsilon_Y$.

Appendix C.2 includes representative scatter plots across various noise levels and reports the average runtime of the algorithms. We remark that in the 'sinusoidal' and 'zigzag' settings, $Y$ can be viewed as a function of $X$ with added noise. Here, Chatterjee's correlation performs best (though it is only applicable when $d_X = d_Y = 1$), and the coverage correlation, while testing against a wider range of alternatives, has similar power performance. In 'circle', 'spiral', 'Lissajous' settings, $X$ and $Y$ are, up to additive noise, functions of a latent variable $U$, and in the 'local' setting, the joint distribution of $(X, Y)$ has a lower-dimensional support in the first quadrant only. In these cases, coverage correlation shows better power than Chatterjee's correlation and KMAc, which are designed to test functional relationships. HSIC, dCor and USP perform well in a subset of these scenarios, though they all show poor performance in at least one setting. Moreover, since these methods rely on permutation-based p-value computation, they may not scale well, particularly when the multiple testing burden is high.

### 3.2 Real data

We use the coverage correlation coefficient to study two biological datasets. We first look at a single-cell RNA sequencing dataset from Suo et al. (2022). We use a subset of the data

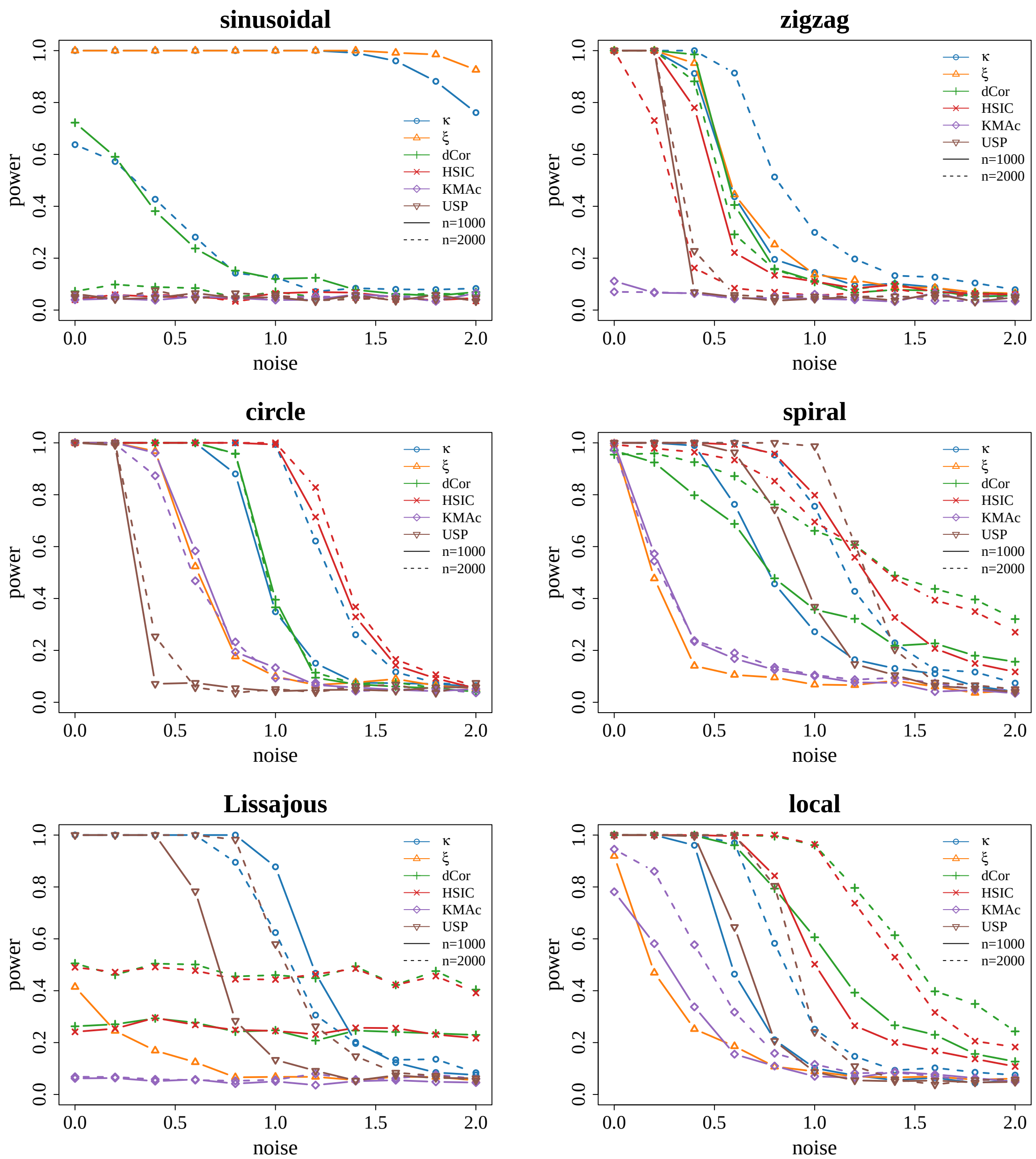


Figure 2: Power curves, estimated over 500 Monte Carlo repetitions, are presented for the coverage correlation, Chatterjee's correlation, distance correlation, HSIC, KMAc and USP in six data-generating scenarios described in Section 3.1 for $(n, d_X, d_Y) \in \{(1000, 1, 1), (2000, 2, 2)\}$ and noise level $\gamma \in \{0, 0.2, 0.4, \ldots, 2\}$. Power is evaluated at the nominal level 0.05. The solid lines correspond to the setting $(n, d_X, d_Y) = (1000, 1, 1)$ while the dashed lines represent the setting $(n, d_X, d_Y) = (2000, 2, 2)$.

used in the paper, consisting of the gene expression levels of top $p = 1000$ highly variable genes measured in $n = 9369$ $CD8^+$ T cells (the processed data is available in the `covercorr` R package). We compute all $\binom{p}{2}$ pairwise correlations using Pearson's correlation, Spearman's correlation, Chatterjee's correlation and the coverage correlation and adjust the corresponding p-values via Bonferroni correction. We identified 54 gene pairs as significant by coverage correlation but not by any of the other methods. The two most significant pairs are plotted in Figure 3. As can be seen, for both pairs, the scatter plots of pairwise gene expression levels exhibit a clear L-shaped relationship suggestive of an implicit functional dependence.

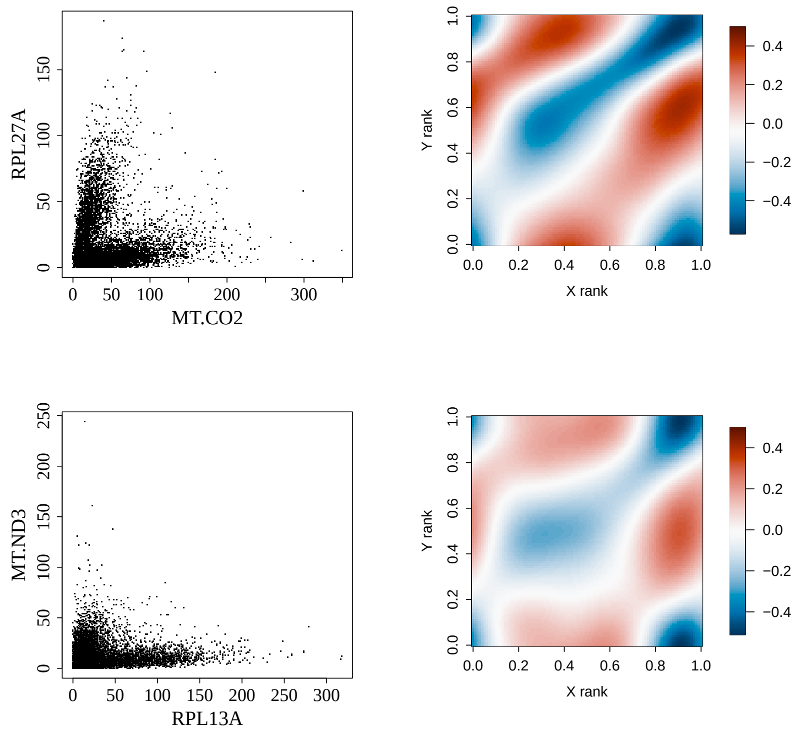


Figure 3: The top two gene pairs with significant coverage correlation after Bonferroni correction, but are not significant under Pearson, Spearman, or Chatterjee's correlation tests. The right column shows heatmaps of excess density of covered area in the coverage correlation calculation of the corresponding gene pairs on the left column.

We next look at the yeast gene expression data from Spellman et al. (1998) (available from R package `minerva`). The dataset measures the expression level of $p = 4381$ genes over $n = 23$ time points. We compute all $\binom{p}{2}$ pairwise correlations using Pearson's, Spearman's, Chatterjee's, and coverage correlation and adjust p-values via Bonferroni correction. There are 85 pairs of genes whose association are statistically significant at 0.05 level after Bonferroni correction using coverage correlation, but no other methods. The top four such pairs with the largest coverage correlation coefficients are shown in Figure 4.

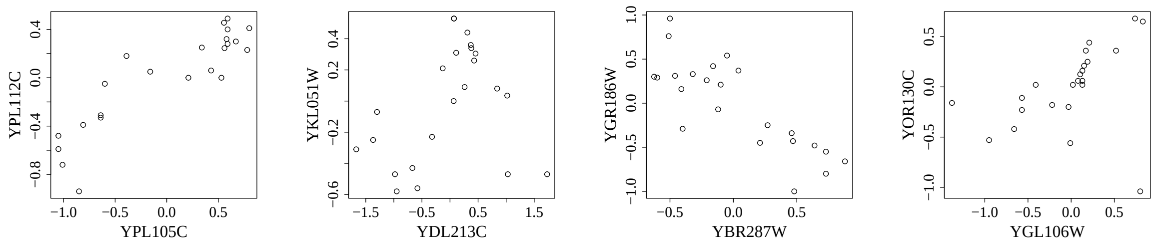


Figure 4: Top four gene pairs with significant coverage correlation after Bonferroni correction, but not significant under Pearson's, Spearman's, or Chatterjee's correlation tests.

# A Proof of main results

## A.1 Proof of Theorem 1

The proof strategy for Theorem 1 is as follows. We first reduce the problem to the case where the marginal distributions $P^X$ and $P^Y$ are both $\mathrm{Unif}[0,1]$, so that the joint distribution $P^{(X,Y)}$ is simply the copula. Next, we approximate the absolutely continuous part of $P^{(X,Y)}$ by a piecewise constant density on rectangular pieces and the singular part of $P^{(X,Y)}$ by a singular measure supported on a Lebesgue null compact set. Finally, we show that the contribution to the coverage from each rectangular piece with constant density $q$ is proportional to $e^{-q}$ and the contribution from the singular measure with Lebesgue null support is 0 to complete the argument. Following this strategy, we will lay down some preliminary results before presenting the overall proof.

The first preliminary result controls the contribution to the coverage from a rectangular region in the domain where the density is constant.

**Proposition 8.** *Suppose $P^{(X,Y)}$ is a probability measure on $[0,1]^2$ with $\mathrm{Unif}[0,1]$ marginals. Suppose that for some $0 \le a_1 < a_2 \le 1$ and $0 \le b_1 < b_2 \le 1$, $P^{(X,Y)}$ is equal to $q \cdot \mathrm{vol}$ when restricted to $[a_1,a_2]\times[b_1,b_2]$, where $\mathrm{vol}$ denotes the Lebesgue measure. Let $(X_1,Y_1),\ldots,(X_n,Y_n) \overset{\mathrm{iid}}{\sim} P^{(X,Y)}$ and let $R_i$ be defined as in (2) with respect to reference points $(U_i,V_i)_{i\in[n]} \overset{\mathrm{iid}}{\sim} \mathrm{Unif}[0,1]^2$. Then*

$$\mathrm{vol}\biggl(\bigcup_{i:(X_i,Y_i)\in(a_1,a_2]\times(b_1,b_2]} B\Bigl(R_i, \frac{1}{2\sqrt{n}}\Bigr)\biggr) = (1-e^{-q})(a_2-a_1)(b_2-b_1)(1+o_p(1)).$$

*Proof.* Define $\mathcal{I} := \{i \in [n] : a_1 < X_i \le a_2\}$, $\mathcal{I}_- := \{i \in [n] : X_i \le a_1\}$, $\mathcal{J} := \{j \in [n] : b_1 < Y_j \le b_2\}$ and $\mathcal{J}_- := \{j \in [n] : Y_j \le b_1\}$. We write $S_0 := |\mathcal{I}_-|$, $S_1 := |\mathcal{I}_- \cup \mathcal{I}|$, $T_0 := |\mathcal{J}_-|$ and $T_1 := |\mathcal{J}_- \cup \mathcal{J}|$. Also, let $M := |\mathcal{I} \cap \mathcal{J}|$.

Given $(U_i,V_i)_{i\in[n]} \overset{\mathrm{iid}}{\sim} \mathrm{Unif}[0,1]^2$, observe that for $R_i^X$ and $R_i^Y$ be defined in (1), we have

$$\{R_i^X : i \in \mathcal{I}\} = \{U_{(S_0+1)},\ldots,U_{(S_1)}\} \quad \text{and} \quad \{R_j^Y : j \in \mathcal{J}\} = \{V_{(T_0+1)},\ldots,V_{(T_1)}\}.$$

Let $\mathcal{F}$ be the $\sigma$-algebra generated by $U_{(S_0)}$, $U_{(S_1+1)}$, and $V_{(T_0)}$, $V_{(T_1+1)}$ and $M$. By David and Nagaraja (2004, Theorem 2.5), we have

$$(R_i^X : i \in \mathcal{I}) \mid \mathcal{F} \overset{\mathrm{iid}}{\sim} \mathrm{Unif}[U_{(S_0)}, U_{(S_1+1)}] \quad \text{and} \quad (R_j^Y : j \in \mathcal{J}) \mid \mathcal{F} \overset{\mathrm{iid}}{\sim} \mathrm{Unif}[V_{(T_0)}, V_{(T_1+1)}].$$

Furthermore, since

$$\big((X_i, Y_i) : i \in \mathcal{I} \cap \mathcal{J}\big) \bigm| M \overset{\text{iid}}{\sim} \mathrm{Unif}(a_1, a_2] \otimes \mathrm{Unif}(b_1, b_2], \tag{8}$$

we have

$$(R_i : i \in \mathcal{I} \cap \mathcal{J}) \mid \mathcal{F} \overset{\text{iid}}{\sim} \mathrm{Unif}[U_{(S_0)}, U_{(S_1+1)}] \otimes \mathrm{Unif}[V_{(T_0)}, V_{(T_1+1)}].$$

By law of large numbers, there is an event $\Omega$ with probability 1 on which we have $M/n \to q(a_2 - a_1)(b_2 - b_1)$, $S_0/n \to a_1$, $S_1/n \to a_2$, $T_0/n \to b_1$, $T_1/n \to b_2$, $U_{(S_0)} \to a_1$, $U_{(S_1+1)} \to a_2$, $V_{(T_0)} \to b_1$ and $V_{(T_1+1)} \to b_2$. We will work on this event henceforth.

As $n \to \infty$, the contribution of the covered area by points near the boundary of any rectangle is negligible (so we may ignore the periodic boundary condition), hence Lemma 24 and a linear rescaling, conditional on $\mathcal{F}$, we have

$$\mathbb{E}\bigg\{\mathrm{vol}\Big(\bigcup_{i \in \mathcal{I} \cap \mathcal{J}} B\Big(R_i, \frac{1}{2\sqrt{n}}\Big)\Big) \biggm| \mathcal{F}\bigg\} \to (1 - e^{-q})(a_2 - a_1)(b_2 - b_1),$$
$$\mathrm{Var}\bigg\{\mathrm{vol}\Big(\bigcup_{i \in \mathcal{I} \cap \mathcal{J}} B\Big(R_i, \frac{1}{2\sqrt{n}}\Big)\Big) \biggm| \mathcal{F}\bigg\} \to 0.$$

By the Dominated Convergence Theorem, the same result holds unconditionally, which implies the desired result by an application of Chebyshev's inequality. □

The next preliminary result shows that the coverage correlation coefficients of samples generated from two probability measures close in total variation distance are (stochastically) close to each other.

**Proposition 9.** *Let $\mu$ and $\nu$ be two probability measures on $\mathbb{R}^2$ with $d_{\mathrm{TV}}(\mu, \nu) \le \epsilon$ and suppose $(X_i, Y_i)_{i \in [n]} \overset{\text{iid}}{\sim} \mu$, $(\tilde{X}_i, \tilde{Y}_i) \overset{\text{iid}}{\sim} \nu$. Also suppose $(U_i, V_i)_{i \in [n]} \overset{\text{iid}}{\sim} \mathrm{Unif}[0,1]^2$ and $(\tilde{U}_i, \tilde{V}_i)_{i \in [n]} \overset{\text{iid}}{\sim} \mathrm{Unif}[0,1]^2$ are independent of $(X_i, Y_i)_{i \in [n]}$ and $(\tilde{X}_i, \tilde{Y}_i)_{i \in [n]}$ respectively. There exists a coupling between $(X_i, Y_i, U_i, V_i)_{i \in [n]}$ and $(\tilde{X}_i, \tilde{Y}_i, \tilde{U}_i, \tilde{V}_i)_{i \in [n]}$ such that*

$$\bigg|\mathcal{V}\Big((X_i)_{i \in [n]}, (Y_i)_{i \in [n]}, (U_i)_{i \in [n]}, (V_i)_{i \in [n]}; \frac{1}{2\sqrt{n}}\Big) - \mathcal{V}\Big((\tilde{X}_i)_{i \in [n]}, (\tilde{Y}_i)_{i \in [n]}, (\tilde{U}_i)_{i \in [n]}, (\tilde{V}_i)_{i \in [n]}; \frac{1}{2\sqrt{n}}\Big)\bigg| \le \epsilon + O_p(n^{-1/2}).$$

*Proof.* For notational simplicity, write $\boldsymbol{X} := (X_i)_{i \in [n]}$, $\boldsymbol{Y} := (Y_i)_{i \in [n]}$, $\tilde{\boldsymbol{X}} := (\tilde{X}_i)_{i \in [n]}$, $\tilde{\boldsymbol{Y}} := (\tilde{Y}_i)_{i \in [n]}$, $\boldsymbol{U} := (U_i)_{i \in [n]}$, $\boldsymbol{V} := (V_i)_{i \in [n]}$, $\tilde{\boldsymbol{U}} := (\tilde{U}_i)_{i \in [n]}$, $\tilde{\boldsymbol{V}} := (\tilde{V}_i)_{i \in [n]}$. Also, let $(\boldsymbol{X}^{\mathrm{Pois}}, \boldsymbol{Y}^{\mathrm{Pois}}) := (X_i^{\mathrm{Pois}}, Y_i^{\mathrm{Pois}})_{i \in [N]} \sim \mathrm{PP}(n\mu)$ be a Poisson point process with intensity $n\mu$ with $N \sim \mathrm{Poi}(n)$ points. Similarly, define mutually independent Poisson point processes $(\tilde{\boldsymbol{X}}^{\mathrm{Pois}}, \tilde{\boldsymbol{Y}}^{\mathrm{Pois}}) \sim \mathrm{PP}(n\nu)$, $(\boldsymbol{U}^{\mathrm{Pois}}, \boldsymbol{V}^{\mathrm{Pois}}) \sim \mathrm{PP}(n \cdot \mathrm{vol})$ and $(\tilde{\boldsymbol{U}}^{\mathrm{Pois}}, \tilde{\boldsymbol{V}}^{\mathrm{Pois}}) \sim \mathrm{PP}(n \cdot \mathrm{vol})$.

We define a new measure $\lambda$ that is the maximum of $\mu$ and $\nu$ as follows. Let $f = d\mu/d(\mu + \nu)$ and $g = d\nu/d(\mu + \nu)$ be densities with respect to the common dominating measure $\mu + \nu$, we set $d\lambda := \max(f, g)\, d(\mu + \nu)$. It is clear that $\lambda \ge \mu$ and $\lambda \ge \nu$, and from the total variation bound between $\mu$ and $\nu$ we have $\lambda(\mathbb{R}^2) - \max\{\mu(\mathbb{R}^2), \nu(\mathbb{R}^2)\} \le \epsilon$. We define another two independent Poisson point process $(\boldsymbol{X}^{\max}, \boldsymbol{Y}^{\max}) \sim \mathrm{PP}(n\lambda)$ and $(\boldsymbol{U}^{\max}, \boldsymbol{V}^{\max}) \sim \mathrm{PP}(n \cdot \mathrm{vol})$.

We now define a chain of couplings by Lemmas 22 and 23 as follows:

$$\begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \\ \boldsymbol{U} \\ \boldsymbol{V} \end{pmatrix} \xleftrightarrow{\textit{Lemma } 22} \begin{pmatrix} \boldsymbol{X}^{\mathrm{Pois}} \\ \boldsymbol{Y}^{\mathrm{Pois}} \\ \boldsymbol{U}^{\mathrm{Pois}} \\ \boldsymbol{V}^{\mathrm{Pois}} \end{pmatrix} \xleftrightarrow{\textit{Lemma } 23} \begin{pmatrix} \boldsymbol{X}^{\max} \\ \boldsymbol{Y}^{\max} \\ \boldsymbol{U}^{\max} \\ \boldsymbol{V}^{\max} \end{pmatrix} \xleftrightarrow{\textit{Lemma } 23} \begin{pmatrix} \tilde{\boldsymbol{X}}^{\mathrm{Pois}} \\ \tilde{\boldsymbol{Y}}^{\mathrm{Pois}} \\ \tilde{\boldsymbol{U}}^{\mathrm{Pois}} \\ \tilde{\boldsymbol{V}}^{\mathrm{Pois}} \end{pmatrix} \xleftrightarrow{\textit{Lemma } 22} \begin{pmatrix} \tilde{\boldsymbol{X}} \\ \tilde{\boldsymbol{Y}} \\ \tilde{\boldsymbol{U}} \\ \tilde{\boldsymbol{V}} \end{pmatrix}$$

In particular, under this coupling, the cardinality of $\boldsymbol{X}^{\mathrm{Pois}}$, $\boldsymbol{Y}^{\mathrm{Pois}}$, $\boldsymbol{U}^{\mathrm{Pois}}$, $\boldsymbol{V}^{\mathrm{Pois}}$ and their tilde-ed counterpart are all equal to some $N \sim \mathrm{Poi}(n)$ and the cardinality of $\boldsymbol{X}^{\max}$, $\boldsymbol{Y}^{\max}$, $\boldsymbol{U}^{\max}$, $\boldsymbol{V}^{\max}$ are equal to $M \sim \mathrm{Poi}(n\lambda(\mathbb{R}^2))$. By Lemma 22, we have

$$\left|\mathcal{V}\Big(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{U},\boldsymbol{V};\frac{1}{2\sqrt{n}}\Big) - \mathcal{V}\Big(\boldsymbol{X}^{\mathrm{Pois}},\boldsymbol{Y}^{\mathrm{Pois}},\boldsymbol{U}^{\mathrm{Pois}}\boldsymbol{V}^{\mathrm{Pois}};\frac{1}{2\sqrt{N}}\Big)\right| = O_p(n^{-1/2})$$
$$\left|\mathcal{V}\Big(\tilde{\boldsymbol{X}},\tilde{\boldsymbol{Y}},\tilde{\boldsymbol{U}},\tilde{\boldsymbol{V}};\frac{1}{2\sqrt{n}}\Big) - \mathcal{V}\Big(\tilde{\boldsymbol{X}}^{\mathrm{Pois}},\tilde{\boldsymbol{Y}}^{\mathrm{Pois}},\tilde{\boldsymbol{U}}^{\mathrm{Pois}}\tilde{\boldsymbol{V}}^{\mathrm{Pois}};\frac{1}{2\sqrt{N}}\Big)\right| = O_p(n^{-1/2})$$

By Lemma 23, we have

$$\left|\mathcal{V}\Big(\boldsymbol{X}^{\max},\boldsymbol{Y}^{\max},\boldsymbol{U}^{\max},\boldsymbol{V}^{\max};\frac{1}{2\sqrt{M}}\Big) - \mathcal{V}\Big(\boldsymbol{X}^{\mathrm{Pois}},\boldsymbol{Y}^{\mathrm{Pois}},\boldsymbol{U}^{\mathrm{Pois}}\boldsymbol{V}^{\mathrm{Pois}};\frac{1}{2\sqrt{N}}\Big)\right| \le \epsilon + O_p(n^{-1/2})$$
$$\left|\mathcal{V}\Big(\boldsymbol{X}^{\max},\boldsymbol{Y}^{\max},\boldsymbol{U}^{\max},\boldsymbol{V}^{\max};\frac{1}{2\sqrt{M}}\Big) - \mathcal{V}\Big(\tilde{\boldsymbol{X}}^{\mathrm{Pois}},\tilde{\boldsymbol{Y}}^{\mathrm{Pois}},\tilde{\boldsymbol{U}}^{\mathrm{Pois}}\tilde{\boldsymbol{V}}^{\mathrm{Pois}};\frac{1}{2\sqrt{N}}\Big)\right| \le \epsilon + O_p(n^{-1/2})$$

The desired result follows by combining the above four vacancy difference bounds under the coupling. □

We are now in a position to prove Theorem 1.

*Proof of Theorem 1.* Let $F_X$ and $F_Y$ be the distribution functions of random variables $X$ and $Y$, respectively. By Lemma 15 and the fact that the coverage correlation is preserved under monotonic transformation of $X$ and $Y$, we may replace $(X_i, Y_i)_{i\in[n]}$ by $(F_X(X_i), F_Y(Y_i))_{i\in[n]}$ and assume without loss of generality that $P^X = P^Y = \mathrm{Unif}[0,1]$, so that $P^{(X,Y)}$ is simply a copula.

By Lebesgue–Radon–Nikodym decomposition, we can write $dP^{(X,Y)} = h\,d\mathrm{vol} + d\nu$ for some measurable function $h$ and a measure $\nu$ singular with respect to the Lebesgue measure. Fix $\epsilon > 0$. For all sufficiently large integer $N$, we can find a function $\tilde{h}$ piecewise constant on each $Q_{j,k} := [(j-1)/N, j/N) \times [(k-1)/N, k/N)$ such that $\int |h - \tilde{h}|\,d\mathrm{vol} \le \epsilon$ (for instance, $\tilde{h}$ can be defined to be equal to the mean value of $h$ in each $Q_{j,k}$ and the claim follows from the Lebesgue Differentiation Theorem and Dominated Convergence Theorem). Since $P^{(X,Y)}$, and hence also $\nu$, is Radon, and in particular inner regular, we can find a compact set $K$ with Lebesgue measure 0 in $[0,1]^2$ such that $\nu(K^{\mathrm{c}}) \le \epsilon$. By possibly increasing $N$, we may also assume that

$$\sum_{(j,k):Q_{j,k}\cap K\neq\emptyset} \mathrm{vol}(Q_{j,k}) \le \mathrm{vol}\Big(K + B\Big((0,0), \frac{1}{N}\Big)\Big) \le \epsilon \tag{9}$$

We write $\tilde{\nu}$ for the restriction of $\nu$ on $K$, i.e. $\tilde{\nu}(A) = \nu(A \cap K)$ for all Borel subset $A$ of $[0,1]^2$. Let $\tilde{P}^{(X,Y)}$ be defined such that

$$d\tilde{P}^{(X,Y)} = \frac{\tilde{h}\,d\mathrm{vol} + d\tilde{\nu}}{\int_{[0,1]^2} \tilde{h}\,d\mathrm{vol} + \tilde{\nu}([0,1]^2)}.$$

By construction, we have $d_{\mathrm{TV}}(P^{(X,Y)}, \tilde{P}^{(X,Y)}) \leq 2\epsilon$. In particular, the marginals $\tilde{P}^X$ and $\tilde{P}^Y$ of $\tilde{P}^{(X,Y)}$ also satisfies $\max\{d_{\mathrm{TV}}(P^X, \tilde{P}^X), d_{\mathrm{TV}}(P^Y, \tilde{P}^Y)\} \leq 2\epsilon$, so by the triangle inequality

$$d_{\mathrm{TV}}(P^X \otimes P^Y, \tilde{P}^X \otimes \tilde{P}^Y) \leq d_{\mathrm{TV}}(P^X \otimes P^Y, P^X \otimes \tilde{P}^Y) + d_{\mathrm{TV}}(P^X \otimes \tilde{P}^Y, \tilde{P}^X \otimes \tilde{P}^Y) \leq 4\epsilon.$$

By Lemma 16, we conclude that for $M = 1$ and $L = 1/(1 - e^{-1})$

$$\left|D_f(P^{(X,Y)} \,\|\, P^X \otimes P^Y) - D_f(\tilde{P}^{(X,Y)} \,\|\, \tilde{P}^X \otimes \tilde{P}^Y)\right| \leq (8M + 6L)\epsilon^{1/2}. \tag{10}$$

Let $\kappa_n^{\tilde{X},\tilde{Y}}$ be the empirical coverage correlation of samples $(\tilde{X}_i, \tilde{Y}_i)_{i\in[n]} \overset{\text{iid}}{\sim} \tilde{P}^{(X,Y)}$ with respect to reference points $(\tilde{U}_i, \tilde{V}_i) \overset{\text{iid}}{\sim} \mathrm{Unif}[0,1]^2$. By Proposition 9, there is a coupling between $(X_i, Y_i, U_i, V_i)_{i\in[n]}$ and $(\tilde{X}_i, \tilde{Y}_i, \tilde{U}_i, \tilde{V}_i)_{i\in[n]}$ such that

$$|\kappa_n^{X,Y} - \kappa_n^{\tilde{X},\tilde{Y}}| \leq \frac{2\epsilon}{1 - e^{-1}} + O_p(n^{-1/2}). \tag{11}$$

Now, since $\tilde{P}^{(X,Y)}$ has piecewise constant density in its absolutely continuous part, we may explicitly control $\kappa_n^{\tilde{X},\tilde{Y}}$. Specifically, writing $\tilde{R}_i$ for the bivariate ranks of $(\tilde{X}_i, \tilde{Y}_i)$ with respect to $(\tilde{U}_i)_{i\in[n]}$ and $(\tilde{V}_i)_{i\in[n]}$ defined as in (2), we have that

$$\begin{aligned}
(1 - e^{-1})\kappa_n^{\tilde{X},\tilde{Y}} &= 1 - e^{-1} - \mathrm{vol}\Big(\bigcup_{i=1}^{n} B\Big(\tilde{R}_i, \frac{1}{2\sqrt{n}}\Big)\Big) \\
&\leq 1 - e^{-1} - \sum_{(j,k):Q_{j,k}\cap K=\emptyset} \mathrm{vol}\Big(Q_{j,k} \cap \bigcup_{i:(\tilde{X}_i,\tilde{Y}_i)\in Q_{j,k}} B\Big(\tilde{R}_i, \frac{1}{2\sqrt{n}}\Big)\Big) \\
&\leq 1 - e^{-1} - (1 + o_p(1)) \sum_{j,k=1}^{N} \int_{(x,y)\in Q_{j,k}} (1 - e^{-\tilde{h}(x,y)}) + \epsilon \\
&= (1 + o_p(1)) \int_{(x,y)\in[0,1]^2} (e^{-\tilde{h}(x,y)} - e^{-1}) + \epsilon \\
&= (1 - e^{-1} + o_p(1)) D_f(\tilde{P}^{(X,Y)} \,\|P^X \otimes P^Y) + \epsilon
\end{aligned}$$

where the second inequality follows from Proposition 8 and (9). By a similar argument, we can bound $\kappa_n^{\tilde{X},\tilde{Y}}$ below as

$$\begin{aligned}
(1 - e^{-1})\kappa_n^{\tilde{X},\tilde{Y}} &= 1 - e^{-1} - \sum_{(j,k)} \mathrm{vol}\Big(Q_{j,k} \cap \bigcup_{i:(\tilde{X}_i,\tilde{Y}_i)\in Q_{j,k}} B\Big(\tilde{R}_i, \frac{1}{2\sqrt{n}}\Big)\Big) \\
&\geq (1 + o_p(1)) \int_{(x,y)\in[0,1]^2} (e^{-\tilde{h}(x,y)} - e^{-1}) - \mathrm{vol}\Big(\bigcup_{\substack{(j,k):Q_{j,k}\cap K\neq\emptyset \\ i:(\tilde{X}_i,\tilde{Y}_i)\in Q_{j,k}}} B\Big(\tilde{R}_i, \frac{1}{2\sqrt{n}}\Big)\Big) \\
&\geq (1 - e^{-1} + o_p(1)) D_f(\tilde{P}^{(X,Y)} \,\|P^X \otimes P^Y) - \mathrm{vol}\Big(K + B\Big(0, \frac{1}{N} + \epsilon + O_p(n^{-1/2})\Big)\Big),
\end{aligned} \tag{12}$$

where in the final step, we applied Lemma 17 to both $(\tilde{X}_i)_{i\in[n]}$, $(\tilde{Y}_i)_{i\in[n]}$ and $(\tilde{U}_i)_{i\in[n]}$, $(\tilde{V}_i)_{i\in[n]}$.

By another application of Lemma 16, $D_f(\tilde{P}^{(X,Y)} \,\|P^X \otimes P^Y)$ is at most $(8M + 2L)\sqrt{\epsilon}$ away from $D_f(\tilde{P}^{(X,Y)} \,\|\tilde{P}^X \otimes \tilde{P}^Y)$. Also, using the fact that $\mathrm{vol}(K) = 0$ and the upper continuity of

Lebesgue measure, by choosing $\epsilon$ sufficiently small (and consequently $N$ sufficiently large), the volume of the Minkowski dilation of $K$ of width $1/N + \epsilon + O_p(n^{-1/2})$ on the right-hand side of (12) can be made smaller than any positive number in probability. Therefore, we conclude that

$$\kappa_n^{\tilde{X},\tilde{Y}} \xrightarrow{\mathrm{p}} D_f(\tilde{P}^{(X,Y)} \| \tilde{P}^X \otimes \tilde{P}^Y). \tag{13}$$

The desired result then follows by combining (10), (11) and (13), since we can set $\epsilon$ arbitrarily small. □

### A.2 Proof of Proposition 2

*Proof of Proposition 2.* The first part of the proposition follows directly from Lemma 24, which implies that $\mathbb{E}(\mathcal{V}_n) \to e^{-1}$ and $\mathrm{Var}(\mathcal{V}_n) \to 0$.

Now we consider the case that $P^{(X,Y)}$ is singular with respect to $P^X \otimes P^Y$. By the same argument as in the proof of Lemma 15, there exist convex functions $\phi : \mathbb{R}^{d_X} \to \mathbb{R}$ and $\psi : \mathbb{R}^{d_Y} \to \mathbb{R}$ and random vectors $U \sim \mathrm{Unif}([0,1]^{d_X})$ and $V \sim \mathrm{Unif}([0,1]^{d_Y})$ such that $X = \nabla\phi(U)$ and $Y = \nabla\psi(V)$ defines respectively the optimal transport maps from $U$ to $X$ and from $V$ to $Y$. Let $T_X$ and $T_Y$ be the Markov transition kernel from $X$ to $U$ and from $Y$ to $V$, respectively, corresponding to the conditional distribution $U \mid X$ and $V \mid Y$. Then we have $T_X(X_1), \ldots, T_X(X_n) \overset{\mathrm{iid}}{\sim} \mathrm{Unif}([0,1]^{d_X})$ and $T_Y(Y_1), \ldots, T_Y(Y_n) \overset{\mathrm{iid}}{\sim} \mathrm{Unif}([0,1]^{d_Y})$. Since $(\nabla\phi(T_X(X_1)), \nabla\psi(T_Y(Y_1))) = (X_1, Y_1)$, we have that the joint distribution of $(T_X(X_1), T_Y(Y_1))$ is singular with respect to the product of the marginals (the preimage of the joint distribution of $(X_1, Y_1)$ under the mapping $(a, b) \mapsto (\nabla\phi(a), \nabla\phi(b))$ is has measure 1 under the joint distribution of $(T_X(X_1), T_Y(Y_1))$ and measure 0 under the product of the marginals). Fix $\epsilon > 0$. Since the joint distribution of $(T_X(X_1), T_Y(Y_1))$ is a Radon measure, and hence inner regular, we can find a compact subset $K$ of its support (so $K$ has Lebesgue measure 0) such that

$$\mathbb{P}((T_X(X_1), T_Y(Y_1)) \in K) \geq 1 - \epsilon.$$

Denote $\mathcal{I} := \{i : (T_X(X_1), T_Y(Y_1)) \notin K\}$. By the multiplicative Chernoff bound (Samworth and Shah, 2025+, Exercise 10.6.11), we have

$$\mathbb{P}(|\mathcal{I}| > 2\epsilon n) \leq e^{-3n\epsilon/8}.$$

By Fournier and Guillin (2015, Theorem 1), the empirical distribution of both $(R_i)_{i\in[n]}$ and $(T_X(X_i), T_Y(Y_i))_{i\in[n]}$ are at most $C_d\rho_n$ away from $\mathrm{Unif}([0,1]^d)$ in 1-Wasserstein distance where

$$\rho_n = \begin{cases} n^{-1/2}\log(en) & \text{if } d = 2 \\ n^{-1/d} & \text{if } d \geq 3 \end{cases}$$

and $C_d$ depends only on $d$. Hence by the triangle inequality and Markov's inequality, the 1-Wasserstein distance between $(R_i)_{i\in[n]}$ and $(T_X(X_i), T_Y(Y_i))_{i\in[n]}$ is bounded by $2C_d\rho_n^{1/2}$ with probability at least $1 - \rho_n^{1/2}$. Define

$$\mathcal{J} := \{i : \min_{j\in[n]} \|R_i - (T_X(X_j), T_Y(Y_j))\|_2 > 2\epsilon^{-1}C_d\rho_n^{1/2}\}.$$

Then $\mathbb{P}(|\mathcal{J}| > \epsilon n) \leq \rho_n^{1/2}$. Therefore, on an event with probability at least $1-\rho_n^{1/2}-e^{-3n\epsilon/8}$, we have for some $C'_d$ depending only on $d$ that

$$\begin{aligned}\mathrm{vol}\Big(\bigcup_{i\in[n]} B\Big(R_i, \frac{1}{2n^{1/d}}\Big)\Big) &\leq \mathrm{vol}\Big(\bigcup_{i\in\mathcal{I}\cup\mathcal{J}} B\Big(R_i, \frac{1}{2n^{1/d}}\Big)\Big) + \mathrm{vol}\Big(K + B(0, C'_d\epsilon^{-1}\rho_n^{1/2})\Big)\\ &\leq 3\epsilon + \mathrm{vol}\Big(K + B(0, C'_d\epsilon^{-1}\rho_n^{1/2})\Big)\end{aligned}$$

For each fixed $\epsilon$, the Minkowski dilation $K + B(0, C'_d\epsilon^{-1}\rho_n^{1/2})$ has Lebesgue measure converging to 0 (since $K$ is compact and Lebesgue null). Since $\epsilon$ is arbitrary, we must have the left-hand side of the above converging to 0 in probability, and consequently $\kappa_n^{X,Y} \xrightarrow{\mathrm{p}} 1$ as desired. $\square$

### A.3 Proof of Theorem 3

The proof strategy for Theorem 3 is similar to that of Theorem 1 and is as follows. We first reduce the problem to the case where the marginal distributions $P^X$ and $P^Y$ are both $\mathrm{Unif}[0,1]$, so that the joint distribution $P^{(X,Y)}$ is simply the copula. Next, we approximate the absolutely continuous part of $P^{(X,Y)}$ by a piecewise constant density on rectangular pieces and the singular part of $P^{(X,Y)}$ by a singular measure supported on a Lebesgue null compact set. Finally, we show that the contribution to the coverage from each rectangular piece with constant density $q$ is proportional to $e^{-q}$ and the contribution from the singular measure with Lebesgue null support is 0 to complete the argument. Following this strategy, we will lay down some preliminary results before presenting the overall proof.

The first preliminary result controls the contribution to the coverage from a rectangular region in the domain where the density is constant.

**Proposition 10.** *Suppose $P^{(X,Y)}$ is a probability measure on $[0,1]^2$ with $\mathrm{Unif}[0,1]$ marginals. Suppose that for some $0 \leq a_1 < a_2 \leq 1$ and $0 \leq b_1 < b_2 \leq 1$, $P^{(X,Y)}$ is equal to $q\cdot\mathrm{vol}$ when restricted to $[a_1,a_2]\times[b_1,b_2]$, where $\mathrm{vol}$ denotes the Lebesgue measure. Let $(X_1,Y_1),\ldots,(X_n,Y_n) \overset{\mathrm{iid}}{\sim} P^{(X,Y)}$ and let $R_i$ be defined as in (2) with respect to reference points $\boldsymbol{U} = \boldsymbol{V} = (1/n,\ldots,(n-1)/n, 1)$. Then*

$$\mathrm{vol}\Big(\bigcup_{i:(X_i,Y_i)\in(a_1,a_2]\times(b_1,b_2]} B\Big(R_i, \frac{1}{2\sqrt{n}}\Big)\Big) = e^{-q}(a_2-a_1)(b_2-b_1)(1+o_p(1)).$$

*Proof.* Define $\mathcal{I} := \{i \in [n] : a_1 < X_i \leq a_2\}$, $\mathcal{I}_- := \{i \in [n] : X_i \leq a_1\}$, $\mathcal{J} := \{j \in [n] : b_1 < Y_j \leq b_2\}$ and $\mathcal{J}_- := \{j \in [n] : Y_j \leq b_1\}$. We write $S_0 := |\mathcal{I}_-|$, $S_1 := |\mathcal{I}_- \cup \mathcal{I}|$, $T_0 := |\mathcal{J}_-|$ and $T_1 := |\mathcal{J}_- \cup \mathcal{J}|$. Also, let $M := |\mathcal{I}\cap\mathcal{J}|$. Let

$$\mathcal{R}_X = \{(S_0+1)/n,\ldots,S_1/n\} \quad\text{and}\quad \mathcal{R}_Y = \{(T_0+1)/n,\ldots,T_1/n\}.$$

For an arbitrary finite set $\mathcal{S}$ and integer $k > 0$, let $\mathcal{S}^{(k)} := \{A \subseteq \mathcal{S} \mid |A| = k\}$, e.g. the set of all subsets of size $k$ of $\mathcal{S}$. Define $\mathrm{Unif}(\mathcal{S}^{(k)})$ to be the uniform distribution over all subsets of size $k$ of $\mathcal{S}$. Let $\mathcal{F}$ be the $\sigma$-algebra generated by $S_0, S_1, T_0, T_1$, and $M$. Then using (8) we have

$$\{R_i^X : i \in \mathcal{I}\cap\mathcal{J}\} \mid \mathcal{F} \sim \mathrm{Unif}(\mathcal{R}_X^{(M)}) \quad\text{and}\quad \{R_j^Y : j \in \mathcal{I}\cap\mathcal{J}\} \mid \mathcal{F} \sim \mathrm{Unif}(\mathcal{R}_Y^{(M)}).$$

which are independent of each other, which gives us

$$(R_i : i \in \mathcal{I}\cap\mathcal{J}) \mid \mathcal{F} \sim \mathrm{Unif}(\mathcal{R}_X^{(M)}) \otimes \mathrm{Unif}(\mathcal{R}_Y^{(M)}).$$

This means that any matchings between any subset of size $M$ of $\mathcal{R}_X$ and any subset of size $M$ of $\mathcal{R}_Y$ are equally likely given $\mathcal{F}$.

By law of large numbers, there is an event $\Omega$ with probability 1 on which we have $M/n \to q(a_2 - a_1)(b_2 - b_1)$, $S_0/n \to a_1$, $S_1/n \to a_2$, $T_0/n \to b_1$, $T_1/n \to b_2$. We will work on this event henceforth.

As $n \to \infty$, the contribution of the covered area by points near the boundary of any rectangle is negligible (so we may ignore the periodic boundary condition), hence Lemma 26 and a linear rescaling, conditional on $\mathcal{F}$, we have

$$\mathbb{E}\Big\{\mathrm{vol}\Big(\bigcup_{i\in\mathcal{I}\cap\mathcal{J}} B\Big(R_i, \frac{1}{2\sqrt{n}}\Big)\Big) \Bigm| \mathcal{F}\Big\} \to e^{-q}(a_2 - a_1)(b_2 - b_1).$$

Then using Lemma 27

$$\begin{aligned}
&\mathrm{Var}\Big(\mathrm{vol}\Big(\bigcup_{i\in\mathcal{I}\cap\mathcal{J}} B\Big(R_i, \frac{1}{2\sqrt{n}}\Big)\Big) \Bigm| \mathcal{F}\Big) \\
&= \mathbb{E}\Big\{\Big(\mathrm{vol}\Big(\bigcup_{i\in\mathcal{I}\cap\mathcal{J}} B\Big(R_i, \frac{1}{2\sqrt{n}}\Big)\Big) - \mathbb{E}\Big\{\mathrm{vol}\Big(\bigcup_{i\in\mathcal{I}\cap\mathcal{J}} B\Big(R_i, \frac{1}{2\sqrt{n}}\Big)\Big) \mid \mathcal{F}\Big\}\Big)^2 \Bigm| \mathcal{F}\Big\} \\
&= \int_0^\infty \mathbb{P}\Big\{\Big(\mathrm{vol}\Big(\bigcup_{i\in\mathcal{I}\cap\mathcal{J}} B\Big(R_i, \frac{1}{2\sqrt{n}}\Big)\Big) - \mathbb{E}\Big\{\mathrm{vol}\Big(\bigcup_{i\in\mathcal{I}\cap\mathcal{J}} B\Big(R_i, \frac{1}{2\sqrt{n}}\Big)\Big) \mid \mathcal{F}\Big\}\Big)^2 \geq t \Bigm| \mathcal{F}\Big\} dt \to 0.
\end{aligned}$$

Using Lemma 27 and McDiarmid inequality we have

$$\mathbb{P}\Big\{\Big(\mathrm{vol}\Big(\bigcup_{i\in\mathcal{I}\cap\mathcal{J}} B\Big(R_i, \frac{1}{2\sqrt{n}}\Big)\Big) - \mathbb{E}\Big\{\mathrm{vol}\Big(\bigcup_{i\in\mathcal{I}\cap\mathcal{J}} B\Big(R_i, \frac{1}{2\sqrt{n}}\Big)\Big) \mid \mathcal{F}\Big\}\Big)^2 \geq t \Bigm| \mathcal{F}\Big\} \leq 2\exp(-Cnt),$$

which gives

$$\mathrm{Var}\Big(\mathrm{vol}\Big(\bigcup_{i\in\mathcal{I}\cap\mathcal{J}} B\Big(R_i, \frac{1}{2\sqrt{n}}\Big)\Big) \Bigm| \mathcal{F}\Big) \to 0.$$

By the Dominated Convergence Theorem, the same result holds unconditionally, which implies the desired result by an application of Chebyshev's inequality. $\square$

The next preliminary result shows that the coverage correlation coefficients of samples generated from two probability measures close in total variation distance are (stochastically) close to each other.

**Proposition 11.** *Let $\mu$ and $\nu$ be two probability measures on $\mathbb{R}^2$ with $d_{\mathrm{TV}}(\mu,\nu) \leq \epsilon$. Let $\boldsymbol{U} = \boldsymbol{V} = (1/n, \ldots, (n-1)/n, 1)$. Then there is a coupling between $\mu$ and $\nu$ such that $(X_i, Y_i)_{i\in[n]} \overset{\mathrm{iid}}{\sim} \mu$, $(\tilde{X}_i, \tilde{Y}_i) \overset{\mathrm{iid}}{\sim} \nu$*

$$\mathbb{P}\Big(\Big|\mathcal{V}\Big((X_i)_{i\in[n]}, (Y_i)_{i\in[n]}, \boldsymbol{U}, \boldsymbol{V}; \frac{1}{2\sqrt{n}}\Big) - \mathcal{V}\Big((\tilde{X}_i)_{i\in[n]}, (\tilde{Y}_i)_{i\in[n]}, \boldsymbol{U}, \boldsymbol{V}; \frac{1}{2\sqrt{n}}\Big)\Big| \geq 12\epsilon\sqrt{n}\Big) \leq e^{-n\epsilon/3}.$$

*Proof.* Since $d_{\mathrm{TV}}(\mu,\nu) \leq \epsilon$ we can construct a coupling between $\mu$ and $\nu$ such that for two i.i.d. samples $\{(X_i, Y_i)\}_{i=1}^n$ and $\{(\tilde{X}_i, \tilde{Y}_i)\}_{i=1}^n$ and for each $i \in [n]$ we have (e.g. Theorem 5.2 in Lindvall (2002))

$$\mathbb{P}\big((X_i, Y_i) \neq (\tilde{X}_i, \tilde{Y}_i)\big) = d_{\mathrm{TV}}(\mu,\nu)/2 \leq \epsilon/2.$$

Let $\mathcal{I}$ be the set of all indices $i \in [n]$ where $(X_i, Y_i)$ and $(\tilde{X}_i, \tilde{Y}_i)$ are different and let $K = |\mathcal{I}|$. The variable $K$ follows a binomial distribution $B(n, p)$ with $p \leq \epsilon/2$ which gives us $\mathbb{E}(K) \leq n\epsilon/2$. For any $i$ we have

$$R_i^X = n^{-1} \sum_{j=1}^{n} \mathbb{1}\{X_j \leq X_i\}, \qquad \tilde{R}_i^X = n^{-1} \sum_{j=1}^{n} \mathbb{1}\{\tilde{X}_j \leq \tilde{X}_i\}.$$

For $i \notin \mathcal{I}$, since $X_i = \tilde{X}_i$ and there are at most $K$ indices $j$ such that $\mathbb{1}\{\tilde{X}_j \leq \tilde{X}_i\} \neq \mathbb{1}\{X_j \leq \tilde{X}_i\}$ we have $|R_i^X - \tilde{R}_i^X| \leq K/n$. Using the same argument, we have $|R_i^Y - \tilde{R}_i^Y| \leq K/n$. Therefore $\|R_i - \tilde{R}_i\|_\infty \leq K/n$. Note that for $i \in \mathcal{I}$ we cannot provide any non-trivial bound. Let

$$S = \cup_{i=1}^{n} (R_i + B), \qquad \tilde{S} = \cup_{i=1}^{n} (\tilde{R}_i + B).$$

Then note that $|\text{vol}(S) - \text{vol}(\tilde{S})| \leq \text{vol}(S \Delta \tilde{S})$. Also we have

$$\begin{aligned} \text{vol}(S\Delta\tilde{S}) \leq \text{vol}\Big(\big(\cup_{i \notin \mathcal{I}} (R_i + B)\big)\Delta\big(\cup_{i \notin \mathcal{I}} (\tilde{R}_i + B)\big)\Big) \\ + \text{vol}\Big(\cup_{i \in \mathcal{I}} (R_i + B)\Big) + \text{vol}\Big(\cup_{i \in \mathcal{I}} (\tilde{R}_i + B)\Big). \end{aligned}$$

where

$$\text{vol}\Big(\cup_{i \in \mathcal{I}} (R_i + B)\Big) + \text{vol}\Big(\cup_{i \in \mathcal{I}} (\tilde{R}_i + B)\Big) \leq \frac{2K}{n}.$$

Then for the set of $i \notin \mathcal{I}$ we have

$$\text{vol}\Big(\big(\cup_{i \notin \mathcal{I}} (R_i + B)\big)\Delta\big(\cup_{i \notin \mathcal{I}} (\tilde{R}_i + B)\big)\Big) \leq \sum_{i \notin \mathcal{I}} \text{vol}\Big((R_i + B)\Delta(\tilde{R}_i + B)\Big) \leq \frac{4K(n-K)}{n\sqrt{n}}.$$

Putting these together, we get

$$|\text{vol}(S) - \text{vol}(\tilde{S})| \leq \frac{4K(n-K)}{n\sqrt{n}} + \frac{2K}{n} \leq \frac{6K}{\sqrt{n}}.$$

Therefore, using the Chernoff bound, we have

$$\mathbb{P}(|\text{vol}(S) - \text{vol}(\tilde{S})| \geq 6(1+\delta)\epsilon\sqrt{n}) \leq \mathbb{P}(K \geq (1+\delta)n\epsilon) \leq \exp\big(-\frac{n\epsilon\delta^2}{3}\big).$$

Therefore, by setting $\delta = 1$, we get the desired result. □

We are now in a position to prove Theorem 3.

*Proof of Theorem 3.* Consider the same set-up as in the proof of Theorem 1. Let $\kappa_n^{\tilde{X},\tilde{Y}}$ be the empirical coverage correlation of sample $(\tilde{X}_i, \tilde{Y}_i)_{i\in[n]} \overset{\text{iid}}{\sim} \tilde{P}^{(X,Y)}$ with $\tilde{P}^{(X,Y)}$ constructed as in proof of Theorem 1 and using the coupling argument in Proposition 11 for $n\epsilon \to \infty$ such that

$$|\kappa_n^{X,Y;grid} - \kappa_n^{\tilde{X},\tilde{Y};grid}| \leq O_p(\epsilon\sqrt{n}). \tag{14}$$

Now, since $\tilde{P}^{(X,Y)}$ has piecewise constant density in its absolutely continuous part, we may explicitly control $\kappa_n^{\tilde{X},\tilde{Y};grid}$. Specifically, writing $\tilde{R}_i$ for the bivariate ranks of $(\tilde{X}_i, \tilde{Y}_i)$ with respect to reference points $\boldsymbol{U} = \boldsymbol{V} = (1/n, \dots, (n-1)/n, 1)$ defined as in (2), we have

$$\begin{aligned}(1-e^{-1})\kappa_n^{\tilde{X},\tilde{Y}} &= 1 - e^{-1} - \mathrm{vol}\Big(\bigcup_{i=1}^{n} B\Big(\tilde{R}_i, \frac{1}{2\sqrt{n}}\Big)\Big) \\ &\leq 1 - e^{-1} - \sum_{(j,k):Q_{j,k}\cap K=\emptyset} \mathrm{vol}\Big(Q_{j,k} \cap \bigcup_{i:(\tilde{X}_i,\tilde{Y}_i)\in Q_{j,k}} B\Big(\tilde{R}_i, \frac{1}{2\sqrt{n}}\Big)\Big) \\ &\leq 1 - e^{-1} - (1+o_p(1)) \sum_{j,k=1}^{N} \int_{(x,y)\in Q_{j,k}} (1 - e^{-\tilde{h}(x,y)}) + \epsilon \\ &= (1+o_p(1)) \int_{(x,y)\in[0,1]^2} (e^{-\tilde{h}(x,y)} - e^{-1}) + \epsilon \\ &= (1 - e^{-1} + o_p(1)) D_f(\tilde{P}^{(X,Y)} \| P^X \otimes P^Y) + \epsilon\end{aligned}$$

where the second inequality follows from Proposition 10 and (9). By a similar argument, we can bound $\kappa_n^{\tilde{X},\tilde{Y};grid}$ below as

$$\begin{aligned}(1-e^{-1})\kappa_n^{\tilde{X},\tilde{Y};grid} &= 1 - e^{-1} - \sum_{(j,k)} \mathrm{vol}\Big(Q_{j,k} \cap \bigcup_{i:(\tilde{X}_i,\tilde{Y}_i)\in Q_{j,k}} B\Big(\tilde{R}_i, \frac{1}{2\sqrt{n}}\Big)\Big) \\ &\geq (1+o_p(1)) \int_{(x,y)\in[0,1]^2} (e^{-\tilde{h}(x,y)} - e^{-1}) - \mathrm{vol}\Big(\bigcup_{\substack{(j,k):Q_{j,k}\cap K\neq\emptyset \\ i:(\tilde{X}_i,\tilde{Y}_i)\in Q_{j,k}}} B\Big(\tilde{R}_i, \frac{1}{2\sqrt{n}}\Big)\Big) \\ &\geq (1 - e^{-1} + o_p(1)) D_f(\tilde{P}^{(X,Y)} \| P^X \otimes P^Y) - \mathrm{vol}\Big(K + B\Big(0, \frac{1}{N} + \epsilon + O_p(n^{-1/2})\Big)\Big),\end{aligned} \tag{15}$$

where in the final step, we applied Lemma 17 to $(\tilde{X}_i)_{i\in[n]}$, $(\tilde{Y}_i)_{i\in[n]}$.

By Lemma 16, $D_f(\tilde{P}^{(X,Y)} \| P^X \otimes P^Y)$ is at most $(8M + 2L)\sqrt{\epsilon}$ away from $D_f(\tilde{P}^{(X,Y)} \| \tilde{P}^X \otimes \tilde{P}^Y)$. Also, using the fact that $\mathrm{vol}(K) = 0$ and the upper continuity of Lebesgue measure, by choosing $\epsilon$ sufficiently small (and consequently $N$ sufficiently large), the volume of the Minkowski dilation of $K$ of width $1/N + \epsilon + O_p(n^{-1/2})$ on the right-hand side of (15) can be made smaller than any positive number in probability. Therefore, we conclude that

$$\kappa_n^{\tilde{X},\tilde{Y};grid} \xrightarrow{\mathrm{p}} D_f(\tilde{P}^{(X,Y)} \| \tilde{P}^X \otimes \tilde{P}^Y). \tag{16}$$

The desired result then follows by combining (10), (14) and (16), by choosing small $\epsilon$ such that $n\epsilon \to \infty$ and $\sqrt{n}\epsilon \to 0$. □

### A.4 Proof of Proposition 4

*Proof of Proposition 4.* Part (i) is true since $D_f(P^{(X,Y)} \| P^X \otimes P^Y) = 0$ if and only if $P^{(X,Y)} = P^X \otimes P^Y$ for strictly convex $f$. For part (ii), write $dP^{(X,Y)} = h\, d(P^X \otimes P^Y) + d\nu$ for $\nu$ singular with respect to $P^X \otimes P^Y$. Since $\lim_{t\to\infty} t^{-1} f(t) = 0$, we have

$$D_f(P^{(X,Y)} \| P^X \otimes P^Y) - 1 = \int_{[0,1]^2} \frac{e^{-h(x)} - 1}{1 - e^{-1}}\, d(P^X \otimes P^Y)(x) = 0$$

if and only if $h(x) = 0$ $P^X \otimes P^Y$ almost everywhere, i.e. $P^{(X,Y)} = \nu$ is singular with respect to $P^X \otimes P^Y$.

For part (iii), observe that conditional independence of $X$ and $Y$ given $Z$ means that we can generate both $P^X \otimes P^Y$ and $P^{(X,Y)}$ from $P^X \otimes P^Z$ and $P^{(X,Z)}$ respectively using the same Markov kernel (channel) $P^{Y|Z}$. Hence, we can apply the data processing inequality (Polyanskiy and Wu, 2025, Theorem 7.4) of the $f$-divergence to obtain that $\kappa(Z, Y) \geq \kappa(X, Y)$ as desired. Part (iv) follows from the lower semicontinuity of $f$-divergence (Polyanskiy and Wu, 2025, Theorem 4.9). Finally, part (v) is true since the definition of $\kappa(X, Y)$ is symmetric with respect to the two arguments by Fubini's theorem. □

### A.5 Proof of Theorem 5

*Proof of Theorem 5.* Our proof strategy is inspired by Hall (1985, Theorem 1). Write $\gamma := \frac{1}{2n^{1/d}}$ for notational simplicity and fix $\lambda \in \mathbb{N}$ for now. Define $L_1 := \lfloor \frac{1}{(\lambda+2)\gamma} \rfloor$. We can partition $[0,1)^d$ into $L := L_1^d$ small cubes $\prod_{j\in[d]} \bigl[\frac{k_j-1}{L_1}, \frac{k_j}{L_1}\bigr)$ for $k_1, \ldots, k_d \in [L_1]$. We call these small cubes $\mathcal{P}_1, \ldots, \mathcal{P}_L$. For each $\ell \in [L]$, we define $\mathcal{Q}_\ell$ to be the concentric cube in $\mathcal{P}_\ell$ with side length $\lambda\gamma$. Define $\mathcal{I}_\ell = \{i : R_i \in \mathcal{P}_\ell\}$ and $N_\ell := |\mathcal{I}_\ell|$. Writing

$$\mathcal{V}_{n,\ell}^{\mathrm{in}} := \mathrm{vol}\Bigl(\mathcal{Q}_\ell \setminus \bigcup_{i\in\mathcal{I}_\ell} B(R_i, \gamma)\Bigr) \; \forall\, \ell \in [n] \quad \text{and} \quad \mathcal{V}_n^{\mathrm{out}} := \mathrm{vol}\Bigl([0,1]^d \setminus \Bigl\{\bigcup_{\ell\in[L]} \mathcal{Q}_\ell \cup \bigcup_{i\in[n]} B(R_i, \gamma)\Bigr\}\Bigr), \tag{17}$$

we have

$$\mathcal{V}_n = \mathcal{V}_n^{\mathrm{in}} + \mathcal{V}_n^{\mathrm{out}}, \quad \text{where } \mathcal{V}_n^{\mathrm{in}} := \sum_{\ell=1}^{L} \mathcal{V}_{n,\ell}^{\mathrm{in}}.$$

The key observation here is that $(\mathcal{V}_{n,\ell}^{\mathrm{in}} : \ell \in [L])$ are conditionally independent given $(N_\ell : \ell \in [L])$. Our proof strategy here is to first establish a Berry–Esseen bound for $\mathcal{V}_n^{\mathrm{in}}$ conditional on $(N_\ell : \ell \in [L])$, then control the asymptotic behaviour of the mean and variance of $\mathcal{V}_n^{\mathrm{in}}$ to derive its unconditional central limit theorem, and finally show that $\mathcal{V}_n^{\mathrm{out}}$ has negligible contribution by choosing $\lambda$ sufficiently large. Define

$$M_n := \mathbb{E}(\mathcal{V}_n^{\mathrm{in}} \mid N_1, \ldots, N_L), \qquad S_n := \mathrm{Var}(\mathcal{V}_n^{\mathrm{in}} \mid N_1, \ldots, N_L). \tag{18}$$

By Proposition 13 we have

$$\sup_{t\in\mathbb{R}} \biggl| \mathbb{P}\biggl( \frac{\sqrt{n}(\mathcal{V}_n^{\mathrm{in}} - M_n)}{\sqrt{S_n}} \leq t \biggm| N_1, \ldots, N_L \biggr) - \Phi(t) \biggr| = O_p(n^{-1/2}), \tag{19}$$

where $\Phi$ is the standard Gaussian distribution function. Additionally Proposition 12 gives us

$$\sqrt{n}(M_n - \mathbb{E}(M_n)) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \alpha_\lambda^2), \qquad nS_n \xrightarrow{\mathrm{p}} \beta_\lambda^2. \tag{20}$$

Using Lemma 29, (19) and (20) gives us

$$\sqrt{n}(\mathcal{V}_n^{\mathrm{in}} - \mathbb{E}(\mathcal{V}_n^{\mathrm{in}})) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \alpha_\lambda^2 + \beta_\lambda^2).$$

Finally, using Proposition 14 and Chebyshev's inequality by $\lambda \to \infty$ we have

$$\sqrt{n}(\mathcal{V}_n^{\mathrm{out}} - \mathbb{E}(\mathcal{V}_n^{\mathrm{out}})) \xrightarrow{\mathrm{p}} 0.$$

By Lemma 30, as $\lambda \to \infty$, we have $\alpha_\lambda^2 \to 0$ and $\beta_\lambda^2 \to \beta^2$ for $\beta^2 > 0$. Consequently, we conclude that

$$\sqrt{n}(\mathcal{V}_n - \mathbb{E}(\mathcal{V}_n)) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \beta^2).$$

The proof is complete by combining the above distributional convergence with the variance calculation in Lemma 24. $\square$

**Proposition 12.** *Let $M_n$ and $S_n$ be defined as in (18), we have*

$$nS_n \xrightarrow{\mathrm{p}} \beta_\lambda^2,$$
$$\sqrt{n}(M_n - \mathbb{E}(M_n)) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \alpha_\lambda^2),$$

*such that*

$$\alpha_\lambda^2 := \frac{\lambda^{2d}}{e^2 2^d (\lambda+2)^d}\left(e^{(\lambda/2+1)^{-d}} - 1\right) - \frac{\lambda^{2d}}{e^2(\lambda+2)^{2d}}, \tag{21}$$
$$\beta_\lambda^2 := \left\{\frac{\lambda^d}{(\lambda+2)^d e^2} C_d + \frac{\lambda^{2d}}{2^d(\lambda+2)^d e^2}(1 - e^{(\lambda/2+1)^{-d}})\right\} + O(\lambda^{-1}), \tag{22}$$

*where $C_d = \sum_{k\geq 1} \frac{2^d}{k!(k+1)^d}$.*

*Proof.* To prove that $nS_n \xrightarrow{\mathrm{p}} \beta_\lambda^2$ we show that $\mathbb{E}(nS_n) \to \beta_\lambda^2$, and $\mathrm{Var}(nS_n) = o(1)$. Note that $(\mathcal{V}_{n,\ell}^{\mathrm{in}} : \ell \in [L])$ are conditionally independent given $(N_\ell : \ell \in [L])$. Therefore

$$S_n = \mathrm{Var}(\mathcal{V}_n^{\mathrm{in}} \mid N_1, \dots, N_L) = \mathrm{Var}\Big(\sum_{\ell=1}^{L} \mathcal{V}_{n,\ell}^{\mathrm{in}} \mid N_1, \dots, N_L\Big) = \sum_{\ell=1}^{L} \mathrm{Var}(\mathcal{V}_{n,\ell}^{\mathrm{in}} \mid N_\ell).$$

First, note that when $X$ and $Y$ are independent, the optimal permutations $\pi^X$ and $\pi^Y$ are independent, and this implies that for $i \in [n]$ we have $R_i \overset{\mathrm{iid}}{\sim} \mathrm{Unif}[0,1]^d$. Also note that given $i \in \mathcal{I}_\ell$, $R_i$ follows a uniform distribution on $\mathcal{P}_\ell$.

Without loss of generality, we fix $\mathcal{Q}_\ell$ and introduce the following functions, which will be used throughout the remainder of the proof. For any $x_1, x_2 \in [0,1]^d$ define

$$C(x_1, x_2) := \frac{\mathrm{vol}(B(x_1,\gamma) \cap B(x_1,\gamma))}{\mathrm{vol}(B(x_1,\gamma))} = n\mathrm{vol}(B(x_1,\gamma) \cap B(x_1,\gamma)).$$

For $x_1, x_2 \in \mathcal{Q}_\ell$ and random variable $W \sim \mathrm{Unif}(\mathcal{P}_\ell)$ define

$$v(x_1) := \mathbb{P}\big(W \notin B(x_1,\gamma)\big) = 1 - (\frac{\lambda}{2}+1)^{-d},$$
$$u(x_1,x_2) := \mathbb{P}\big(W \notin B(x_1,\gamma) \cup B(x_2,\gamma)\big) = 1 - 2(\frac{\lambda}{2}+1)^{-d} + C(x_1,x_2)(\frac{\lambda}{2}+1)^{-d}.$$

First, note that we have

$$\mathbb{E}[\mathcal{V}_{n,\ell}^{\mathrm{in}} \mid N_\ell] = \mathbb{E}\Big[\int_{Q_\ell} \mathbb{1}\{x \notin \bigcup_{i\in\mathcal{I}_\ell} B(R_i,\gamma)\}dx \;\Big|\; N_\ell\Big] = \int_{Q_\ell} v(x)^{N_\ell} dx,$$
$$\mathbb{E}[(\mathcal{V}_{n,\ell}^{\mathrm{in}})^2 \mid N_\ell] = \mathbb{E}\Big[\int_{Q_\ell^2} \mathbb{1}\{x_1, x_2 \notin \bigcup_{i\in\mathcal{I}_\ell} B(R_i,\gamma)\}dx_1 dx_2 \;\Big|\; N_\ell\Big] = \int_{Q_\ell^2} u(x_1,x_2)^{N_\ell} dx_1 dx_2.$$

Since $N_\ell \sim \text{Bin}(n, \text{vol}(\mathcal{P}_\ell))$, for $a > 0$ constant, we have $\mathbb{E}a^{N_\ell} = (1 + \text{vol}(\mathcal{P}_\ell)(a-1))^n$. Using this equality, we have

$$\begin{aligned}
&\mathbb{E}\big[\text{Var}\big(\mathcal{V}^{\text{in}}_{n,\ell} \mid N_\ell\big)\big] = \mathbb{E}\Big[\mathbb{E}\big((\mathcal{V}^{\text{in}}_{n,\ell})^2 \mid N_\ell\big) - \mathbb{E}\big(\mathcal{V}^{\text{in}}_{n,\ell} \mid N_\ell\big)^2\Big] \\
&\quad = \mathbb{E}\Big[\int_{\mathcal{Q}_\ell^2} u(x_1,x_2)^{N_\ell} - \int_{\mathcal{Q}_\ell^2} v(x_1)^{N_\ell} v(x_2)^{N_\ell} dx_1 dx_2\Big] \\
&\quad = \int_{\mathcal{Q}_\ell^2} \Big\{1 + \text{vol}(\mathcal{P}_\ell)(u(x_1,x_2) - 1)\Big\}^n - \Big\{1 + \text{vol}(\mathcal{P}_\ell)(v(x_1)v(x_2) - 1)\Big\}^n dx_1 dx_2 \\
&\quad = (1 + O(\frac{1}{n})) \int_{\mathcal{Q}_\ell^2} \Big\{ \exp\big((\frac{\lambda}{2}+1)^d (u(x_1,x_2)-1)\big) - \exp\big((\frac{\lambda}{2}+1)^d (v(x_1)v(x_2)-1)\big)\Big\} dx_1 dx_2 \\
&\quad = (1 + O(\frac{1}{n})) e^{-2} \int_{\mathcal{Q}_\ell^2} \Big\{ \exp\big(C(x_1,x_2)\big) - \exp\big((\frac{\lambda}{2}+1)^{-d}\big)\Big\} dx_1 dx_2 \\
&\quad = (1 + O(\frac{1}{n})) \Big\{\frac{\lambda^d}{2^d e^2 n^2} C_d + \frac{\lambda^{2d}}{2^{2d} e^2 n^2}(1 - e^{(\lambda/2+1)^{-d}})\Big\} \\
&\quad = \Big\{\frac{\lambda^d}{2^d e^2 n^2} C_d + \frac{\lambda^{2d}}{2^{2d} e^2 n^2}(1 - e^{(\lambda/2+1)^{-d}})\Big\} + O(\frac{\lambda^{2d}}{n^3} + \frac{\lambda^{d-1}}{2^d n^2})
\end{aligned}$$

where in the last line we have used the following equality

$$\int_{\mathcal{Q}_\ell^2} \exp(C(x_1,x_2)) dx_1 dx_2 = \text{vol}(\mathcal{Q}_\ell)^2 + \\ \text{vol}(\mathcal{Q}_\ell)(2\gamma)^d \int_{[-1,1]^d} (\exp(\prod_{i=1}^d \max\{(1 - |u_i|), 0\}) - 1) du (1 + O(\lambda^{-1})),$$

where

$$\int_{[-1,1]^d} (\exp(\prod_{i=1}^d \max\{(1-|u_i|),0\}) - 1) du = \sum_{k \ge 1} \frac{2^d}{k!(k+1)^d} = C_d.$$

Therefore

$$\begin{aligned}
\mathbb{E}[nS_n] &= n\mathbb{E}[\sum_{\ell=1}^L \text{Var}(\mathcal{V}^{\text{in}}_{n,\ell} \mid N_\ell)] \\
&= e^{-2}\Big\{C_d + \frac{\lambda^d}{2^d}(1 - e^{(\lambda/2+1)^{-d}})\Big\} + O(\lambda^{-1} + \frac{\lambda^d}{n}) \\
&= e^{-2}(C_d - 1) + O(\lambda^{-1} + \lambda^{-d} + \frac{\lambda^d}{n}). \qquad (23)
\end{aligned}$$

since $2^{-d}\lambda^d(1 - e^{(\lambda/2+1)^{-d}}) = -1 + O(\lambda^{-d})$.

We then work out $\text{Var}(S_n)$.

$$\text{Var}(S_n) = \text{Var}\Big(\sum_{\ell=1}^L \text{Var}(\mathcal{V}^{\text{in}}_{n,\ell} \mid N_\ell)\Big) = \sum_{\ell,k \in [L]} \text{Cov}\Big(\text{Var}(\mathcal{V}^{\text{in}}_{n,\ell} \mid N_\ell), \text{Var}(\mathcal{V}^{\text{in}}_{n,k} \mid N_k)\Big).$$

Since

$$\text{Var}(\mathcal{V}^{\text{in}}_{n,\ell} \mid N_\ell) = \int_{\mathcal{Q}_\ell^2} u(x_1,x_2)^{N_\ell} - v(x_1)^{N_\ell} v(x_2)^{N_\ell} dx_1 dx_2,$$

applying Fubini's theorem, we have

$$
\begin{aligned}
&\mathrm{Cov}\Big(\mathrm{Var}(\mathcal{V}_{n,\ell}^{\text{in}} \mid N_\ell), \mathrm{Var}(\mathcal{V}_{n,k}^{\text{in}} \mid N_k)\Big) \\
&\quad = \int_{\mathcal{Q}_1^4} \mathrm{Cov}(u(x_1,x_2)^{N_\ell}, u(x_3,x_4)^{N_k}) + \mathrm{Cov}(v(x_1)^{N_\ell} v(x_2)^{N_\ell}, v(x_3)^{N_k} v(x_4)^{N_k}) \\
&\qquad - \mathrm{Cov}(u(x_1,x_2)^{N_\ell}, v(x_3)^{N_k} v(x_4)^{N_k}) - \mathrm{Cov}(u(x_3,x_4)^{N_k}, v(x_1)^{N_\ell} v(x_2)^{N_\ell}) dx_1 dx_2 dx_3 dx_4 \\
&\quad = (\mathrm{vol}(\mathcal{Q}_1))^4 O(n^{-2}) = O(n^{-6}),
\end{aligned}
$$

where in the last line we have used Lemma 28 together with the following fact

$$
\max\{(u(x_1,x_2)-1), (u(x_3,x_4)-1), (v(x_1)v(x_2)-1), (v(x_3)v(x_4)-1)\} = O(1).
$$

As a result, we have

$$
\mathrm{Var}(nS_n) = O(n^{-2}). \tag{24}
$$

Combining (23) and (24), Markov's inequality implies that $nS_n \xrightarrow{\text{p}} \beta_\lambda^2$

We now turn to proving that $\sqrt{n}(M_n - \mathbb{E}(M_n)) \xrightarrow{\text{d}} \mathcal{N}(0, \alpha_\lambda^2)$. For $w \in \{0, 1, \ldots, n\}$, let $f(w) := \mathbb{E}[\mathcal{V}_{n,1}^{\text{in}} \mid N_1 = w]$. Let $W \sim \mathrm{Poi}(n\mathrm{vol}(\mathcal{P}_1))$. Define

$$
\tau^2 := L\mathrm{Var}(f(W)) - \frac{L^2}{n}\mathrm{Cov}^2(W, f(W)).
$$

Holst (1972, Theorem 1) implies that as $n \to \infty$

$$
\frac{1}{\tau}(M_n - \mathbb{E}(M_n)) \xrightarrow{\text{d}} \mathcal{N}(0,1).
$$

To finish the proof, it is therefore enough to show that $n\tau^2 \to \alpha^2$. For $X \sim \mathrm{Unif}[\mathcal{Q}_1]$ we have

$$
\begin{aligned}
nL\mathrm{Var}\big(f(W)\big) &= nL\mathrm{Var}\Big(\int_{\mathcal{Q}_1} v(x)^W dx\Big) \\
&= nL\mathrm{vol}^2(\mathcal{Q}_1)\mathrm{Var}\Big(\mathbb{E}\big(\{v(X)\}^W \mid W\big)\Big) \\
&= \frac{\lambda^{2d}}{(2\lambda+4)^d}\mathrm{Var}\Big(\big\{1-(\lambda/2+1)^{-d}\big\}^W\Big) = \frac{\lambda^{2d}}{(2\lambda+4)^d} e^{-2}\Big(e^{(\lambda/2+1)^{-d}} - 1\Big),
\end{aligned} \tag{25}
$$

where the final equality follows from the fact that $\mathbb{E}a^W = e^{n\mathrm{vol}(\mathcal{P}_1)(a-1)}$ for any constant $a > 0$. Similarly we derive

$$
\begin{aligned}
L\mathrm{Cov}\big(W, f(W)\big) &= L\mathrm{vol}(\mathcal{Q}_1)\Big(\mathbb{E}\Big[W\mathbb{E}\big(\{v(X)\}^W \mid W\big)\Big] - n\mathrm{vol}(\mathcal{P}_1)\mathbb{E}\Big[\mathbb{E}\big(\{v(X)\}^W \mid W\big)\Big]\Big) \\
&= \Big(\frac{\lambda}{\lambda+2}\Big)^d \Big(\mathbb{E}\big\{W(1-(\lambda/2+1)^{-d})^W\big\} - (\lambda/2+1)^d \mathbb{E}\big\{1-(\lambda/2+1)^{-d}\big\}^W\Big) \\
&= \Big(\frac{\lambda}{\lambda+2}\Big)^d \Big(n\mathrm{vol}(\mathcal{P}_1)\big(1-(\lambda/2+1)^{-d}\big)e^{-1} - (\lambda/2+1)^d e^{-1}\Big) = -\Big(\frac{\lambda}{\lambda+2}\Big)^d e^{-1},
\end{aligned} \tag{26}
$$

where we use the fact that $\mathbb{E}(Wa^W) = n\text{vol}(\mathcal{P}_1)\eta e^{np(a-1)}$ for constant $a > 0$ in the second to last equality. Combining (25) and (26) we have

$$n\tau^2 = \frac{\lambda^{2d}}{2^d e^2(\lambda+2)^d}\left(e^{(\lambda/2+1)^{-d}} - 1\right) - \frac{\lambda^{2d}}{e^2(\lambda+2)^{2d}} = \alpha_\lambda^2,$$

which finishes the proof. $\square$

**Proposition 13.** *Let $M_n$ and $S_n$ be defined as in (18), we have*

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\sqrt{n}(\mathcal{V}_n^{\text{in}} - M_n)}{\sqrt{S_n}} \leq t \;\middle|\; N_1,\ldots,N_L\right) - \Phi(t)\right| = O_p(n^{-1/2}),$$

*where $\Phi$ is the standard Gaussian distribution function.*

*Proof.* Since each point in $\mathcal{Q}_\ell$ lies at least $\gamma$ away from the boundary of $\mathcal{P}_\ell$, it follows that, conditional on $(N_\ell : \ell \in [L])$, the collections $(\mathcal{V}_{n,\ell}^{\text{in}} : \ell \in [L])$ are independent. Then using Berry-Esseen Theorem (Berry, 1941; Esseen, 1942) we have

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\sqrt{n}(\mathcal{V}_n^{\text{in}} - M_n)}{\sqrt{S_n}} \leq t \middle| N_1,\ldots,N_L\right) - \Phi(t)\right| \leq C_n,$$

where

$$C_n := C\frac{\sum_{\ell=1}^L \mathbb{E}\big(|\mathcal{V}_{n,\ell}^{\text{in}} - \mathbb{E}(\mathcal{V}_{n,\ell}^{\text{in}} \mid N_\ell)|^3 \mid N_\ell\big)}{S^{3/2}},$$

with $C$ a universal constant independent of $n$. Additionally note that for all $\ell \in [L]$

$$\mathcal{V}_{n,\ell}^{\text{in}} \leq \text{vol}(\mathcal{P}_\ell) = \frac{(\lambda/2+1)^d}{n}.$$

Therefore we have

$$C_n \leq \frac{(\lambda/2+1)^{3d}}{n^2 S_n^{3/2}}.$$

Using Proposition 12, we have $S_n = O_p(n^{-1})$; thus it follows that $C_n = O_p(n^{-1/2})$ which completes the proof. $\square$

**Proposition 14.** *Let $\mathcal{V}_n^{\text{out}}$ be defined as in (17). We have*

$$\mathcal{V}_n^{\text{out}} - \mathbb{E}(\mathcal{V}_n^{\text{out}}) = O_p(n^{-1/2}\lambda^{-1/2}).$$

*Proof.* First note that by Lemma 24 we have

$$\mathbb{E}(\mathcal{V}_n^{\text{out}}) = \left\{1 - \left(\frac{\lambda}{\lambda+2}\right)^d\right\}\left(1-\frac{1}{n}\right)^n. \tag{27}$$

Take $Z_1, Z_1 \overset{\text{iid}}{\sim} \text{Unif}\big[[0,1]^d \setminus \bigcup_{\ell\in[L]} \mathcal{Q}_\ell\big]$. Using Lemma 24, we have

$$\begin{aligned}\mathbb{E}[(\mathcal{V}_n^{\text{out}})^2] &= \left\{1 - \left(\frac{\lambda}{\lambda+2}\right)^d\right\}^2 \mathbb{E}\left\{1 - \frac{2}{n} + \text{vol}(C(Z_1,Z_2))\right\}^n \\ &= \left\{1 - \left(\frac{\lambda}{\lambda+2}\right)^d\right\}^2 \sum_{k=0}^n \binom{n}{k}\left(1-\frac{2}{n}\right)^{n-k} \mathbb{E}(\text{vol}(C(Z_1,Z_2)))^k. \end{aligned} \tag{28}$$

Note that for $1 \le k \le n$ we have

$$\begin{aligned}
\mathbb{E}(\mathrm{vol}(C(Z_1, Z_2)))^k &= \left\{1 - \left(\frac{\lambda}{\lambda+2}\right)^d\right\}^{-2} \int_{\left([0,1]^d \setminus \bigcup_{\ell\in[L]} \mathcal{Q}_\ell\right)^2} \mathrm{vol}\big(C(z_1, z_2)\big)^k \, dz_1 dz_2 \\
&\le \left\{1 - \left(\frac{\lambda}{\lambda+2}\right)^d\right\}^{-2} \int_{[0,1]^d \setminus \bigcup_{\ell\in[L]} \mathcal{Q}_\ell} \int_{[0,1]^d} n^{-k} \mathbb{1}\{z_1 \in B_\infty^d(z_2, 2\gamma)\} \, dz_2 dz_1 \\
&= \left\{1 - \left(\frac{\lambda}{\lambda+2}\right)^d\right\}^{-1} 2^d n^{-(k+1)}. \qquad (29)
\end{aligned}$$

Therefore putting together (27),(28) and (29) we have

$$\begin{aligned}
\mathrm{Var}(\mathcal{V}_n^{\mathrm{out}}) &\le \frac{2^d}{n} \left\{1 - \left(\frac{\lambda}{\lambda+2}\right)^d\right\} \sum_{k=1}^n \binom{n}{k} \left(1 - \frac{2}{n}\right)^{n-k} \frac{1}{n^k} \\
&\qquad + \left\{1 - \left(\frac{\lambda}{\lambda+2}\right)^d\right\}^2 \left((1 - \frac{2}{n})^n - (1 - \frac{1}{n})^{2n}\right) \\
&= \frac{2^d}{n} \left\{1 - \left(\frac{\lambda}{\lambda+2}\right)^d\right\} \left((1 - \frac{1}{n})^n - (1 - \frac{2}{n})^n\right) \\
&\qquad + \left\{1 - \left(\frac{\lambda}{\lambda+2}\right)^d\right\}^2 \left((1 - \frac{2}{n})^n - (1 - \frac{1}{n})^{2n}\right) \\
&= \frac{2^d}{n} \left\{1 - \left(\frac{\lambda}{\lambda+2}\right)^d\right\} (e^{-1} - e^{-2}) - \left\{1 - \left(\frac{\lambda}{\lambda+2}\right)^d\right\}^2 \frac{e^{-2}}{n} + o(\frac{1}{n}).
\end{aligned}$$

Note that for large $\lambda$ we have

$$1 - \left(\frac{\lambda}{\lambda+2}\right)^d = O(\frac{1}{\lambda}),$$

therefore allowing $\lambda$ to grow in infinity as $n \to \infty$ we have

$$\mathrm{Var}(\mathcal{V}_n^{\mathrm{out}}) = O(\frac{1}{n\lambda}).$$

This completes the proof. □

## A.6 Proof of Theorem 6

*Proof of Theorem 6.* As in the proof of Lemma 21, we generate $P_1, \ldots, P_{n+1}, Q_1, \ldots, Q_{n+1} \overset{\mathrm{iid}}{\sim}$ Beta$(1/2, 1/2)$ independent of all other randomness in the problem and define $G_i := U_{(i)} - (U_{(i)} - U_{(i-1)})P_i$ and $H_i := V_{(i)} - (V_{(i)} - V_{(i-1)})Q_i$ for $i \in [n+1]$ (with the convention that $U_{(0)} = V_{(0)} = 0$ and $U_{(n+1)} = V_{(n+1)} = 1$).

Given any deterministic $\boldsymbol{x} = (x_i)_{i\in[n]}$ and $\boldsymbol{y} = (y_i)_{i\in[n]}$ and $i_0 \in [n]$, define

$$\boldsymbol{x}^{-i_0} = (x_1, \ldots, x_{i_0-1}, x_{i_0+1}, \ldots, x_n), \qquad \boldsymbol{y}^{-i_0} = (y_1, \ldots, y_{i_0-1}, y_{i_0+1}, \ldots, y_n).$$

By the proof of Lemma 21, there exists $\tilde{\boldsymbol{U}}^{[r_x]} = (\tilde{U}_i^{[r_x, r_y]})_{i\in[n-1]}$ defined in terms of $\boldsymbol{U} = (U_i)_{i\in[n]}$, $P_1, \ldots, P_{n+1}$ and $r_x := \sum_{i\in[n]} \mathbb{1}_{\{x_{i_0} \ge x_i\}}$, and $\tilde{\boldsymbol{V}}^{[r_y]} = (\tilde{U}_i^{[r_y]})_{i\in[n-1]}$ defined in terms of $\boldsymbol{V} =$

$(V_i)_{i\in[n]}$, $Q_1,\ldots,Q_{n+1}$ and $r_y := \sum_{i\in[n]}\mathbb{1}_{\{y_{i_0}\geq y_i\}}$, such that $\tilde{\boldsymbol{U}}^{[r_x]}$ and $\tilde{\boldsymbol{V}}^{[r_y]}$ each have iid Unif$[0,1]$ entries and

$$\Big|\mathcal{V}\Big(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{U},\boldsymbol{V};\frac{1}{2\sqrt{n}}\Big)-\mathcal{V}\Big(\boldsymbol{X}^{-i_0},\boldsymbol{Y}^{-i_0},\tilde{\boldsymbol{U}}^{[r_x]},\tilde{\boldsymbol{V}}^{[r_y]};\frac{1}{2\sqrt{n-1}}\Big)\Big|$$
$$\leq 5\max\{G_{(r_x+1)}-G_{(r_x)},H_{(r_y+1)}-H_{(r_y)},n^{-1}\}.$$

Thus, if $\boldsymbol{X}'$ and $\boldsymbol{Y}'$ differ from $\boldsymbol{X}$ and $\boldsymbol{Y}$ only in the $i_0$th entry respectively, we must have

$$\Big|\mathcal{V}\Big(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{U},\boldsymbol{V};\frac{1}{2\sqrt{n}}\Big)-\mathcal{V}\Big(\boldsymbol{X}',\boldsymbol{Y}',\boldsymbol{U},\boldsymbol{V};\frac{1}{2\sqrt{n}}\Big)\Big|$$
$$\leq 10\max\{G_{(r_x+1)}-G_{(r_x)},H_{(r_y+1)}-H_{(r_y)},n^{-1}\}.$$

Since $(G_{(r+1)}-G_{(r)})_{r\in[n]}$ and $(H_{(r+1)}-H_{(r)})_{r\in[n]}$ are have Beta$(1,n)$ entries, we have by Lemma 25 that on an event $\Omega$ with probability at least $1-2ne^{-t^2/(2+4t/3)}$ that

$$\max_{r\in[n]}\max\{G_{(r+1)}-G_{(r_x)},H_{(r+1)}-H_{(r)},n^{-1}\}\leq\frac{1+t}{n}.$$

Thus, we can apply McDiarmid's inequality conditional on the event $\Omega$ to obtain that

$$\begin{aligned}\mathbb{P}(|\mathcal{V}_n-\mathbb{E}(\mathcal{V}_n)|\geq s)&\leq\mathbb{P}\big(|\mathcal{V}_n-\mathbb{E}(\mathcal{V}_n)|\geq s\;\big|\;\Omega\big)\mathbb{P}(\Omega)+\mathbb{P}(\Omega^{\mathrm{c}})\\&\leq 2\exp\Big\{-\frac{ns^2}{50(1+t)^2}\Big\}+2n\exp\Big\{-\frac{t^2}{2+4t/3}\Big\}\end{aligned}$$

Setting $t=\frac{1}{5}\min\{(ns^2)^{1/3},(ns^2)^{1/2}\}$, we then have

$$\mathbb{P}(|\mathcal{V}_n-\mathbb{E}(\mathcal{V}_n)|\geq s)\leq 2(n+1)\exp\Big\{-\frac{1}{72}\min\{ns^2,(ns^2)^{1/3}\}\Big\},$$

as desired. □

### A.7 Proof of Proposition 7

*Proof of Proposition 7.* The proof of this proposition is an immediate application of Lemma 27 and McDiarmid's inequality (McDiarmid et al., 1989). □

## B Ancillary results

The following lemma shows that $f$-divergence is between the joint distribution and product of marginal distributions is preserved under the Monge–Kantorovich rank transform.

**Lemma 15.** *Let $(X,Y)$ be a pair of jointly distributed random variables on $\mathbb{R}^{d_X}\times\mathbb{R}^{d_Y}$. Let $U$ and $V$ be continuous random variables with distribution $P^U$ on $\mathbb{R}^{d_X}$ and $P^V$ on $\mathbb{R}^{d_Y}$ respectively, chosen such that $U\perp\!\!\!\perp V\mid(X,Y)$ and that*

$$U\in\operatorname*{arg\,min}_{\tilde{U}\sim P^U}\mathbb{E}\|X-\tilde{U}\|_2^2\quad\text{and}\quad V\in\operatorname*{arg\,min}_{\tilde{V}\sim P^V}\mathbb{E}\|Y-\tilde{V}\|_2^2.\tag{30}$$

*Let $P^{(X,Y)}$ be the joint distribution of $(X,Y)$ with marginals $P^X$ and $P^Y$, and similarly $P^{(U,V)}$, $P^U$, $P^V$ the joint and marginal distributions of $(U,V)$. Then for any convex function $f:\mathbb{R}\to\mathbb{R}\cup\{\infty\}$ such that $f(1)=0$ (cf. Definition 2, we have*

$$D_f(P^{(X,Y)}\,\|\,P^X\otimes P^Y)=D_f(P^{(U,V)}\,\|\,P^U\otimes P^V).$$

*Proof.* Let $\pi$ be the joint coupling (joint distribution) of $(X, U)$ and $\gamma$ the coupling of $(Y, V)$ defined through the solution of the optimal transport problem in (30). Let $P^{U|X=x}$ and $P^{V|Y=y}$ be the corresponding conditional distributions of $U$ given $X = x$ and $V$ given $Y = y$ respectively. Note that these conditional distributions are well-defined up to a $P^X$-measure 0 set of $x$-values and $P^Y$-measure 0 set of $y$-values. The fact that $U \perp\!\!\!\perp V \mid (X, Y)$ means that $P^{U|X=x} \otimes P^{V|Y=y}$ is the conditional distribution of $(U, V)$ given $(X, Y) = (x, y)$. Therefore, we have

$$
\begin{aligned}
P^{(U,V)} &= \int_{\mathbb{R}^{d_X} \times \mathbb{R}^{d_Y}} P^{U|X=x} \otimes P^{V|Y=y} \, dP^{(X,Y)}(x, y) \\
P^U \otimes P^Y &= \int_{\mathbb{R}^{d_X} \times \mathbb{R}^{d_Y}} P^{U|X=x} \otimes P^{V|Y=y} \, d(P^X \otimes P^Y)(x, y).
\end{aligned}
$$

By the data processing inequality (Polyanskiy and Wu, 2025, Theorem 7.4), we thus have

$$
D_f(P^{(X,Y)} \,\|\, P^X \otimes P^Y) \geq D_f(P^{(U,V)} \,\|\, P^U \otimes P^V).
$$

On the other hand, since $U$ is absolutely continuous with respect to the Lebesgue measure, by Brenier's Theorem (see, e.g. Villani, 2021, Theorem 2.12), there exists a convex function $\phi : \mathbb{R}^{d_X} \to \mathbb{R}$ such that $d\pi(x, u) = dP^U(u)\delta_{\{y = \nabla\phi(u)\}}$. In other words, the the optimal transport from $U$ to $X$ is the function $\nabla\phi$ (which is $P^U$-almost everywhere uniquely defined), and so $X = \nabla\phi(U)$. Similarly, we have $Y = \nabla\psi(V)$ for some convex function $\psi : \mathbb{R}^{d_Y} \to \mathbb{R}$. Consequently, we have that conditional on $(U, V)$, $X$ and $Y$ are deterministic, so in particular, conditionally independent. This allows us to run a symmetric argument with the conditional distribution of $(X, Y)$ given $(U, V)$ to obtain a data processing inequality in the reverse direction, thus establishing the desired equality. $\square$

The next lemma shows the stability of the $f$-divergence $D_f(P \,\|\, Q)$ with respect to total-variation perturbation of $P$ and $Q$ when the generator function $f$ is bounded Lipschitz in $[0, \infty)$.

**Lemma 16.** *Suppose $f : [0, \infty) \to [0, M]$ is convex and $L$-Lipschitz. Then for any probability measures $P, Q, P', Q'$ such that $Q'$ is absolutely continuous with respect to $Q$, we have*

$$
|D_f(P \,\|\, Q) - D_f(P' \,\|\, Q')| \leq 2L \, d_{\mathrm{TV}}(P, P') + (4M + L)\sqrt{d_{\mathrm{TV}}(Q, Q')}.
$$

*Proof.* For notational convenience, we write $\epsilon_P := d_{\mathrm{TV}}(P, P')$ and $\epsilon_Q := d_{\mathrm{TV}}(Q, Q')$. Since $f$ is convex and bounded, it must be decreasing on $[0, \infty)$ with $\lim_{t\to\infty} t^{-1} f(t) \to 0$, so singular components of $P$ and $P'$ has no contribution in the $f$ divergence. Let $P_{\mathrm{ac}}$ and $P'_{\mathrm{ac}}$ be the absolutely continuous part of $P$ and $P'$ with respect to $Q$, and let $p, p', q, q'$ be densities of $P_{\mathrm{ac}}, P'_{\mathrm{ac}}, Q, Q'$ with respect to $Q$ (note $q \equiv 1$). Then we have

$$
D_f(P \,\|\, Q) = \int f\left(\frac{p(x)}{q(x)}\right) q(x) \, dQ(x) \quad \text{and} \quad D_f(P' \,\|\, Q') = \int f\left(\frac{p'(x)}{q'(x)}\right) q'(x) \, dQ(x),
$$

where $f(\infty)$ is interpreted as $\lim_{t\to\infty} f(t)$, which exists since $f$ is decreasing and bounded from below. Thus, we have

$$
\begin{aligned}
D_f(P \,\|\, Q) - D_f(P' \,\|\, Q') &= \int f\left(\frac{p(x)}{q(x)}\right) (q(x) - q'(x)) \, dQ(x) \\
&\quad + \int \left\{ f\left(\frac{p(x)}{q(x)}\right) - f\left(\frac{p(x)}{q'(x)}\right) \right\} q'(x) \, dQ(x) \\
&\quad + \int \left\{ f\left(\frac{p(x)}{q'(x)}\right) - f\left(\frac{p'(x)}{q'(x)}\right) \right\} q'(x) \, dQ(x) =: I_1 + I_2 + I_3.
\end{aligned}
$$

For the first terms on the right-hand side, we have

$$|I_1| \le M \int |q(x) - q'(x)|\, dQ(x) = 2M\epsilon_Q.$$

For the last term, using the fact that $f$ is Lipschitz, we have

$$|I_3| \le L \int \left| \frac{p(x)}{q'(x)} - \frac{p'(x)}{q'(x)} \right| q'(x)\, dQ(x) \le 2L\epsilon_P.$$

For the second term, writing $\mathcal{A} := \{|q'(x) - q(x)| \le \epsilon_Q^{1/2}\}$ and using the fact that $q(x) \equiv 1$, we have

$$\begin{aligned}
|I_2| &\le L \int \left| \frac{p(x)}{q(x)} - \frac{p(x)}{q'(x)} \right| q'(x) \mathbb{1}_{\mathcal{A}}\, dQ(x) + M \int q'(x) \mathbb{1}_{\mathcal{A}^c}\, dQ(x) \\
&\le L \int |q(x) - q'(x)| p(x) \mathbb{1}_{\mathcal{A}}\, dQ(x) + M \int |q'(x) - q(x)|\, dQ(x) + M \int q(x) \mathbb{1}_{\mathcal{A}^c}\, dQ(x) \\
&\le L\epsilon_Q^{1/2} + M\epsilon_Q + M\epsilon_Q^{1/2},
\end{aligned}$$

where we used Markov's inequality in the final step. The desired result is obtained by combining the bounds for $I_1$, $I_2$ and $I_3$. □

The following lemma studies locations of order statistics of a sample drawn from a distribution close in total variation to $\mathrm{Unif}[0,1]$.

**Lemma 17.** *Suppose $P$ is a distribution on $[0,1]$ such that $d_{\mathrm{TV}}(P, \mathrm{Unif}[0,1]) \le \epsilon$. For observations $X_1, \ldots, X_n \overset{\mathrm{iid}}{\sim} P$, let $X_{(1)} \le \cdots \le X_{(n)}$ denote their order statistics. Then for any $t > 0$, we have*

$$\mathbb{P}\biggl( \max_{i \in [n]} \biggl| X_{(i)} - \frac{i}{n} \biggr| \ge \epsilon + t \biggr) \le 2e^{-2nt^2}.$$

*Proof.* Let $F$ be the distribution function of $P$ and $F_n$ the empirical c.d.f. of $X_1, \ldots, X_n$. We then have

$$\begin{aligned}
\max_{i \in [n]} \biggl| X_{(i)} - \frac{i}{n} \biggr| &\le \max_{i \in [n]} \Bigl| X_{(i)} - F(X_{(i)}) \Bigr| + \max_{i \in [n]} \Bigl| F(X_{(i)}) - F_n(X_{(i)}) \Bigr| \\
&\le \sup_{x \in [0,1]} |x - F(x)| + \sup_{x \in [0,1]} |F(x) - F_n(x)| \\
&\le d_{\mathrm{TV}}(P, \mathrm{Unif}[0,1]) + \sup_{x \in [0,1]} |F(x) - F_n(x)|.
\end{aligned}$$

Therefore, we have

$$\mathbb{P}\biggl( \max_{i \in [n]} \biggl| X_{(i)} - \frac{i}{n} \biggr| \ge \epsilon + t \biggr) \le \mathbb{P}\biggl( \sup_{x \in [0,1]} \Bigl| F(x) - F_n(x) \Bigr| \ge t \biggr) \le 2e^{-2nt^2},$$

where the final bound uses the Dvoretzky–Kiefer–Wolfowitz–Massart–Reeve inequality (Dvoretzky et al., 1956; Massart, 1990; Reeve, 2024). □

The following lemma shows that we can construct disjoint (randomised) intervals around each order statistic of a uniform sample $U_1, \ldots, U_n$ on $[0,1]$, such that after deleting a subset of these intervals, the remaining $U_i$'s are uniformly distributed in the carved out set.

**Lemma 18.** *Suppose $U_1, \dots, U_n \overset{\text{iid}}{\sim} \text{Unif}([0,1])$ and $P_1, \dots, P_{n+1} \overset{\text{iid}}{\sim} \text{Beta}(1/2, 1/2)$ are independent. Let $U_{(1)}, \dots, U_{(n)}$ be order statistics of $(U_i)_{i\in[n]}$ with the convention that $U_{(0)} := 0$ and $U_{(n+1)} := 1$ and set $G_i := U_{(i)} - (U_{(i)} - U_{(i-1)})P_i$. Given $\mathcal{I} \subseteq [n]$, define a mapping*

$$g : [0,1] \setminus \bigcup_{i\in\mathcal{I}} [G_i, G_{i+1}) \to [0,1]$$

*by*

$$g(x) = \frac{x - \sum_{i\in\mathcal{I}}(G_{i+1} - G_i)\mathbb{1}\{x \ge G_{i+1}\}}{1 - \sum_{i\in\mathcal{I}}(G_{i+1} - G_i)}.$$

*Then, $\{g(U_{(i)}) : i \notin \mathcal{I}\}$ are order statistics of an independent and identically distributed sample from* Unif$[0,1]$.

*Proof.* Let $M \sim \text{Gamma}(n+1, 1)$ be independent from other randomness in the problem. Write $S_i := U_{(i)}M$ and and $T_i := G_iM$, then

$$T_1 - S_0, S_1 - T_1, T_2 - S_1, S_2 - T_2, \dots, T_{n+1} - S_n, S_{n+1} - T_{n+1}$$

are $2n + 2$ independent Gamma$(1/2, 1)$ random variables. Moreover,

$$D := [S_{(0)}, S_{(n+1)}] \setminus \bigcup_{i\in\mathcal{I}} [T_i, T_{i+1}),$$

is exactly the domain of $g$ scaled by $M$. For every $i \in [n+1]$, define its predecessor $\text{pred}(i) := \max\{i' \le i : i' \notin I\}$. Let $L_i$ denote the Lebesgue measure of $[T_{(\text{pred}(i)+1)}, T_{(i+1)}] \cap D$. We note that each $L_i = (S_i - T_i) + (T_{\text{pred}(i)+1} - S_{\text{pred}(i)})$ is the sum of two independent Gamma$(1/2, 1)$ increments and distinct increments are used to compute $L_i$ for different $i$. Hence, $L_i : i \notin \mathcal{I}$ are independent Exp$(1)$ random variables, which is equivalent to the desired result after rescaling. □

The next two lemmas show how the covered area changes under simple operations. First, we show that the covered area decreases by a small amount when we delete narrow horizontal and vertical strips in $[0,1]^2$.

**Lemma 19.** *For $(x_1, y_1), \dots, (x_n, y_n) \in [0,1]^2$, let $[a, b), [c, d) \subseteq [0,1]$ be intervals such that $[a, b) \cap \{x_1, \dots, x_n\} = x_n$ and $[c, d) \cap \{y_1, \dots, y_n\} = y_n$. Define $g : [0,1] \setminus [a, b) \to [0, 1 - (b - a)]$ by $g(x) := x - (b-a)\mathbb{1}\{x \ge b\}$ and $h : [0,1] \setminus [c, d) \to [0, 1 - (d - c)]$ by $h(y) := y - (d-c)\mathbb{1}\{y \ge d\}$. Then*

$$0 \le \text{vol}\left(\bigcup_{i=1}^{n} B_\infty((x_i, y_i), r)\right) - \text{vol}\left(\bigcup_{i=1}^{n-1} B_\infty((g(x_i), h(y_i)), r)\right) \le (b - a) + (d - c) + 4r^2.$$

*Proof.* We define four sets of unions $\mathcal{A}_1 := \cup_{i=1}^{n} B_\infty((x_i, y_i), r)$, $\mathcal{A}_2 := \cup_{i=1}^{n-1} B_\infty((x_i, y_i), r)$, $\mathcal{A}_3 := \cup_{i=1}^{n-1} B_\infty((g(x_i), y_i), r)$ and $\mathcal{A}_4 := \cup_{i=1}^{n-1} B_\infty((g(x_i), h(y_i)), r)$. It is easy to see that $\text{vol}(\mathcal{A}_1) - \text{vol}(\mathcal{A}_2) \in [0, 4r^2]$, hence it suffices to show that $\text{vol}(\mathcal{A}_2) - \text{vol}(\mathcal{A}_3) \in [0, b - a]$ and $\text{vol}(\mathcal{A}_3) - \text{vol}(\mathcal{A}_4) \in [0, d - c]$. We will prove the former, and the latter follows by an essentially identical argument.

Let $S(y) := \{x : (x, y) \in \mathcal{A}_2\}$ and $\tilde{S}(y) := \{x : (x, y) \in \mathcal{A}_3\}$ be horizontal 'slices' of $\mathcal{A}_2$ and $\mathcal{A}_3$ respectively. By Fubini's theorem, it suffices to check that $\lambda(S(y)) - \lambda(\tilde{S}(y)) \in [0, b - a]$ for all $y$, where $\lambda$ denotes the Lebesgue measure. Observe that

$$S(y) = \bigcup_{i\in[n-1]:|y-y_i|\le r, x_i<a} [x_i - r, x_i + r] \cup \bigcup_{i\in[n-1]:|y-y_i|\le r, x_i\ge b} [x_i - r, x_i + r] =: S_1(y) \cup S_2(y).$$

and

$$\tilde{S}(y) = \bigcup_{i\in[n-1]:|y-y_i|\le r, x_i<a} [x_i - r, x_i + r] \cup \bigcup_{i\in[n-1]:|y-y_i|\le r, x_i\ge b} [x_i - (b-a) - r, x_i - (b-a) + r]$$
$$= S_1(y) \cup \{S_2(y) + (a-b)\},$$

where $S_2(y) + (a-b)$ denotes translation of the set $S_2(y)$ by $a-b$. If $S_1(y) \cap S_2(y) = \emptyset$, then $\lambda(S_1(y) \cap \{S_2(y) + (a-b)\}) \le b-a$, so $\lambda(S(y)) - \lambda(\tilde{S}(y)) \in [0, b-a]$. On the other hand, if $S_1(y) \cap S_2(y) \ne \emptyset$, let $i_1 = \arg\max_{i\in[n-1],|y-y_i|\le r, x_i<a} x_i$ and $i_2 = \arg\min_{i\in[n-1],|y-y_i|\le r, x_i\ge b} x_i$. We must have $x_{i_1} + r \ge x_{i_2} - r$. Observe that $x_{i_2} - (b-a) - r \ge x_{i_1} - r$ and $x_{i_1} + r \le x_{i_2} - (b-a) + r$, together with $[x_{i_1} - r, x_{i_1} + r] \subseteq S_1(y)$ and $[x_{i_2} - r, x_{i_2} + r] \subseteq S_2(y)$, we deduce that $S_2(y) + (a-b) \cap [0,a] \subseteq S_1(y)$ and $S_1(y) \cap [a, 1-(b-a)] \subseteq S_2(y) + (a-b)$. In particular, we have $\lambda(\tilde{S}(y)) = \lambda(S(y)) - (b-a)$. This establishes the desired result. $\square$

The following lemma controls the extent of change in the coverage area when we scale both the domain and the radius of each $\ell_\infty$ ball.

**Lemma 20.** *Fix $a, b \in (0,1)$ and let $f : [0,a] \times [0,b] \to [0,1]^2$ be defined such that $f(x,y) = (x/a, y/b)$. For $(x_1, y_1), \ldots, (x_n, y_n) \in [0,a] \times [0,b]$ and $r, r' \in (0, 1/2)$, we have*

$$-\left(\frac{1}{ab} - 1\right) nr^2 - 4n\left\{(r')^2 - \min\left(\frac{r}{a}, r'\right)\min\left(\frac{r}{b}, r'\right)\right\}$$
$$\le \mathrm{vol}\left(\bigcup_{i\in[n]} B_\infty((x_i,y_i), r)\right) - \mathrm{vol}\left(\bigcup_{i\in[n]} B_\infty(f(x_i,y_i), r')\right)$$
$$\le 4n\left\{\max\left(\frac{r}{a}, r'\right)\max\left(\frac{r}{b}, r'\right) - (r')^2\right\}.$$

*Proof.* Let $\mathcal{A}_1 := \cup_{i\in[n]} B_\infty((x_i,y_i), r)$, $\mathcal{A}_2 := \cup_{i\in[n]} [x_i/a - r/a, x_i/a + r/a] \times [y_i/b - r/b, y_i/b + r/b]$ and $\mathcal{A}_3 := \cup_{i\in[n]} B_\infty(f(x_i,y_i), r')$. We will control $\mathrm{vol}(\mathcal{A}_1) - \mathrm{vol}(\mathcal{A}_3)$ by controlling separately $\mathrm{vol}(\mathcal{A}_1) - \mathrm{vol}(\mathcal{A}_2)$ and $\mathrm{vol}(\mathcal{A}_2) - \mathrm{vol}(\mathcal{A}_3)$. For the former, we observe that $\mathcal{A}_2$ is simply $f(\mathcal{A}_1)$, so

$$0 \le \mathrm{vol}(\mathcal{A}_2) - \mathrm{vol}(\mathcal{A}_1) \le \left(\frac{1}{ab} - 1\right)\mathrm{vol}(\mathcal{A}_1) \le \left(\frac{1}{ab} - 1\right) nr^2.$$

For the latter, we have

$$\mathrm{vol}(\mathcal{A}_2) - \mathrm{vol}(\mathcal{A}_3) \le \mathrm{vol}(\mathcal{A}_2 \setminus \mathcal{A}_3) \le 4n\big\{\max(r/a, r')\max(r/b, r') - (r')^2\big\}$$

and

$$\mathrm{vol}(\mathcal{A}_3) - \mathrm{vol}(\mathcal{A}_2) \le \mathrm{vol}(\mathcal{A}_3 \setminus \mathcal{A}_2) \le 4n\big\{(r')^2 - \min(r/a, r')\min(r/b, r')\big\}$$

The desired result follows by combining the two bounds. $\square$

The following lemma provides an upper bound on the change in vacancy area when we delete a few points from a sample of size $n$. We recall the definition of the vacancy volume in (3).

**Lemma 21.** *Given $n, m \in \mathbb{N}$ with $m \le n/2$, let $\boldsymbol{X} := (X_i)_{i\in[n]}$ and $\boldsymbol{Y} := (Y_i)_{i\in[n]}$ be fixed and suppose $(U_i, V_i)_{i\in[n]} \overset{\mathrm{iid}}{\sim} \mathrm{Unif}[0,1]^2$ and $(\tilde{U}_i, \tilde{V}_i)_{i\in[n-m]} \overset{\mathrm{iid}}{\sim} \mathrm{Unif}[0,1]^2$. Write $\boldsymbol{U} := (U_i)_{i\in[n]}$, $\boldsymbol{V} := (V_i)_{i\in[n]}$, $\tilde{\boldsymbol{U}} := (\tilde{U}_i)_{i\in[n-m]}$, $\tilde{\boldsymbol{V}} := (\tilde{V}_i)_{i\in[n-m]}$, $\tilde{\boldsymbol{X}} := (X_i)_{i\in[n-m]}$ and $\tilde{\boldsymbol{Y}} := (Y_i)_{i\in[n-m]}$.*

*There exists a coupling between $(\boldsymbol{U}, \boldsymbol{V})$ and $(\tilde{\boldsymbol{U}}, \tilde{\boldsymbol{V}})$ such that for every $t \in [0, \frac{n}{2m}-1]$ the following holds with probability at least $1 - 2e^{-t^2 m/(2+4t/3)}$:*

$$\frac{(5+3t)m}{n} \leq \mathcal{V}\Big(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{V}; \frac{1}{2\sqrt{n}}\Big) - \mathcal{V}\Big(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{Y}}, \tilde{\boldsymbol{U}}, \tilde{\boldsymbol{V}}; \frac{1}{2\sqrt{n-m}}\Big) \leq \frac{(5+5t)m}{n},$$

*where, by convention, $U_{(0)} = 0$ and $U_{(n+1)} = 1$.*

*Proof.* Let $P_1, \dots, P_{n+1}, Q_1, \dots, Q_{n+1} \overset{\text{iid}}{\sim} \text{Beta}(1/2, 1/2)$ be independent from other randomness in the problem and define $G_i := U_{(i)} - (U_{(i)} - U_{(i-1)})P_i$ and $H_i := V_{(i)} - (V_{(i)} - V_{(i-1)})Q_i$. Let $r_i^X := \sum_{i'=1}^n \mathbb{1}\{X_{i'} \leq X_i\}$ and $r_i^Y := \sum_{i'=1}^n \mathbb{1}\{Y_{i'} \leq Y_i\}$ for $i \in [n]$. For $r \in [m]$, define

$$g_r : [0,1] \setminus \bigcup_{i=n-r+1}^{n} [G_{r_i^X}, G_{r_i^X+1}) \to \Big[0, 1 - \sum_{i=n-r+1}^{n} (G_{r_i^X+1} - G_{r_i^X})\Big]$$
$$h_r : [0,1] \setminus \bigcup_{i=n-r+1}^{n} [H_{r_i^Y}, H_{r_i^Y+1}) \to \Big[0, 1 - \sum_{i=n-r+1}^{n} (H_{r_i^Y+1} - H_{r_i^Y})\Big]$$

by

$$g_r(x) := x - \sum_{i=n-r+1}^{n} (G_{r_i^X+1} - G_{r_i^X}) \mathbb{1}\{x \geq G_{r_i^X+1}\},$$
$$h_r(y) := y - \sum_{i=n-r+1}^{n} (H_{r_i^Y+1} - H_{r_i^Y}) \mathbb{1}\{y \geq H_{r_i^Y+1}\}.$$

Now, define

$$g : [0,1] \setminus \bigcup_{i=n-m+1}^{n} [G_{r_i^X}, G_{r_i^X+1}) \to [0,1] \quad \text{and} \quad h : [0,1] \setminus \bigcup_{i=n-m+1}^{n} [H_{r_i^Y}, H_{r_i^Y+1}) \to [0,1]$$

via

$$g(x) := g_m(x)/g_m(1) \quad \text{and} \quad h(y) := h_m(y)/h_m(1).$$

Intuitively, $g_r$ can be seen as the bijection that compresses the carved out interval $[0,1] \setminus \cup_{i=n-r+1}^n [G_{r_i^X}, G_{r_i^X+1})$ to a contiguous interval and $g$ further rescales the compressed interval to $[0,1]$ after deleting intervals associated with $X_{n-m+1}, \dots, X_n$. Similarly, $h_r$ and $h$ represent compression of the carved out $y$-interval and its rescaled version.

By Lemma 18, $\tilde{U}_i := g(U_i)$ and $\tilde{V}_i := h(V_i)$ for $i \in [n-m]$ satisfies $(\tilde{U}_i, \tilde{V}_i)_{i \in [n-m]} \overset{\text{iid}}{\sim} \text{Unif}[0,1]^2$. We will establish the desired bound under this coupling.

Let $R_i \equiv (R_i^X, R_i^Y) := (U_{r_i^X}, V_{r_i^Y})$ for $i \in [n]$, then

$$1 - \mathcal{V}\Big(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{V}; \frac{1}{2\sqrt{n}}\Big) = \text{vol}\Big(\bigcup_{i \in [n]} B_\infty\Big(R_i, \frac{1}{2\sqrt{n}}\Big)\Big).$$

Under the present coupling, we also have that

$$1 - \mathcal{V}\Big(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{Y}}, \tilde{\boldsymbol{U}}, \tilde{\boldsymbol{V}}; \frac{1}{2\sqrt{n-m}}\Big) = \text{vol}\Big(\bigcup_{i \in [n-m]} B_\infty\Big((g(R_i^x), h(R_i^y)), \frac{1}{2\sqrt{n-m}}\Big)\Big).$$

To establish the desired result, we will construct a few sets whose volume interpolates the two volumes on the right-hand side of the two previous displays. Specifically, define

$$\mathcal{A}_0 := \bigcup_{i\in[n]} B_\infty\Big(R_i, \frac{1}{2\sqrt{n}}\Big),$$
$$\mathcal{A}_r := \bigcup_{i\in[n-r]} B_\infty\Big((g_r(R_i^X), h_r(R_i^Y)), \frac{1}{2\sqrt{n}}\Big)$$
$$\mathcal{B} := \bigcup_{i\in[n-m]} B_\infty\Big((g(R_i^X), h(R_i^Y)), \frac{1}{2\sqrt{n-m}}\Big).$$

By Lemma 19,

$$0 \le \mathrm{vol}(\mathcal{A}_0) - \mathrm{vol}(\mathcal{A}_m) = \sum_{r=1}^{m}\big\{\mathrm{vol}(\mathcal{A}_{r-1}) - \mathrm{vol}(\mathcal{A}_r)\big\} \le (1-g_m(1)) + (1-h_m(1)) + \frac{m}{n}.$$

By Lemma 20,

$$\min\Big\{\frac{1}{g_m(1)}, \sqrt{\frac{n}{n-m}}\Big\}\min\Big\{\frac{1}{h_m(1)}, \sqrt{\frac{n}{n-m}}\Big\} - \frac{n}{n-m} - \Big(\frac{1}{g_m(1)h_m(1)} - 1\Big)$$
$$\le \mathrm{vol}(\mathcal{A}_m) - \mathrm{vol}(\mathcal{B}) \le \max\Big\{\frac{1}{g_m(1)}, \sqrt{\frac{n}{n-m}}\Big\}\max\Big\{\frac{1}{h_m(1)}, \sqrt{\frac{n}{n-m}}\Big\} - \frac{n}{n-m}.$$

Since $1-g_m(1), 1-h_m(1) \overset{\mathrm{iid}}{\sim} \mathrm{Beta}(m, n+1-m)$, by Lemma 25, there is an event $\Omega$ with probability at least $1-2e^{-t^2m/(2+4t/3)}$ on which

$$\max\{1-g_m(1), 1-h_m(1)\} \le \frac{(1+t)m}{n} =: \delta.$$

Writing $\delta' := m/n$, so that $\delta' \le \delta \le 1/2$ under the assumption. We have the event $\Omega$ that

$$\mathrm{vol}(\mathcal{A}_0) - \mathrm{vol}(\mathcal{B}) = \mathrm{vol}(\mathcal{A}_0) - \mathrm{vol}(\mathcal{A}_m) + \mathrm{vol}(\mathcal{A}_m) - \mathrm{vol}(\mathcal{B}) \le 2\delta + \delta' + \frac{1}{(1-\delta)^2} - \frac{1}{1-\delta'} \le 5\delta.$$

Similarly, on $\Omega$, we also have

$$\mathrm{vol}(\mathcal{B}) - \mathrm{vol}(\mathcal{A}_0) = \mathrm{vol}(\mathcal{A}_m) - \mathrm{vol}(\mathcal{A}_0) + \mathrm{vol}(\mathcal{B}) - \mathrm{vol}(\mathcal{A}_m) \le \frac{1}{(1-\delta)^2} - 1 + \frac{1}{1-\delta'} - 1 \le 3\delta + 2\delta'.$$

The desired result follows from combining the two bounds above. $\square$

We will often employ the 'Poissonisation trick' to analyse the vacancy volume, i.e. studying the vacancy associated with $N \sim \mathrm{Poi}(n)$ instead of $n$ data points. The following lemma studies the difference in vacancy due to Poissonisation.

**Lemma 22.** *Let $P^{(X,Y)}$ be a probability measure on $\mathbb{R}^2$. Let $(X_i, Y_i)_{i\in\mathbb{N}} \overset{\mathrm{iid}}{\sim} P^{(X,Y)}$ and $N \sim \mathrm{Poi}(n)$. Suppose we have $(U_i, V_i)_{i\in\mathbb{N}} \overset{\mathrm{iid}}{\sim} \mathrm{Unif}[0,1]^2$ and $(\tilde{U}_i, \tilde{V}_i)_{i\in\mathbb{N}} \overset{\mathrm{iid}}{\sim} \mathrm{Unif}[0,1]^2$. There is a coupling between $(U_i, V_i)_{i\in[n]}$ and $(N, (\tilde{U}_i, \tilde{V}_i)_{i\in[N]})$ such that*

$$\mathcal{V}\Big((X_i)_{i\in[n]}, (Y_i)_{i\in[n]}, (U_i)_{i\in[n]}, (V_i)_{i\in[n]}; \frac{1}{2\sqrt{n}}\Big)$$
$$- \mathcal{V}\Big((X_i)_{i\in[N]}, (Y_i)_{i\in[N]}, (\tilde{U}_i)_{i\in[N]}, (\tilde{V}_i)_{i\in[N]}; \frac{1}{2\sqrt{N}}\Big) = O_p(n^{-1/2})$$

*Proof.* We define the coupling conditionally on $N$ as follows. If $N \geq n$, we can view $(X_i, Y_i)_{i\in[n]}$ as obtained from $(X_i, Y_i)_{i\in N}$ by deleting the last $N - n$ points, and $(U_i, V_i)_{i\in[n]}$ can be defined conditionally on $(\tilde{U}_i, \tilde{V}_i)_{i\in[N]}$ by Lemma 21. If $N < n$, we view $(X_i, Y_i)_{i\in N}$ as obtained from $(X_i, Y_i)_{i\in[n]}$ by deleting the last $n - N$ points and again obtain a conditional joint distribution between $(\tilde{U}_i, \tilde{V}_i)_{i\in[N]}$ and $(U_i, V_i)_{i\in[n]}$ by Lemma 21.

Now, given any $\epsilon > 0$, we can find $C > 0$ such that $\mathbb{P}(|N - n| \geq Cn^{1/2}) \leq \epsilon/2$. Conditional on $N = n'$ for $n' \in [n - Cn^{1/2}, n + Cn^{1/2}]$, by Lemma 21 under the current coupling, we have with (conditional) probability at least $1 - 2e^{-3Cn^{1/2}/10}$ that

$$\Big|\mathcal{V}\Big((X_i)_{i\in[n]}, (Y_i)_{i\in[n]}, (U_i)_{i\in[n]}, (V_i)_{i\in[n]}; \frac{1}{2\sqrt{n}}\Big) - \mathcal{V}\Big((X_i)_{i\in[N]}, (Y_i)_{i\in[N]}, (\tilde{U}_i)_{i\in[N]}, (\tilde{V}_i)_{i\in[N]}; \frac{1}{2\sqrt{N}}\Big)\Big| \leq \frac{10}{n^{1/2}}.$$

Choosing $n$ sufficiently large so that $2e^{-3Cn^{1/2}/10} < \epsilon/2$, and integrating over $\mathbb{N}$, we have with probability at least $1 - \epsilon$ that the desired vacancy volumes have difference bounded by $10n^{-1/2}$. $\square$

For any finite measure $\lambda$ on $\mathbb{R}^2$, we use $\mathrm{PP}(\lambda)$ to denote the Poisson point process with intensity $\lambda$. Recall that for finite measure, the Poisson point process can be identified as $N \sim \mathrm{Poi}(\lambda(\mathbb{R}^2))$ points drawn independently (conditionally on $N$) from the probability measure $\lambda/\lambda(\mathbb{R}^2)$. The following lemma studies the effect on vacancy due to thinning the Poisson point process.

**Lemma 23.** *Given two finite measures $\mu$ and $\nu$ on $\mathbb{R}^2$ such that $\mu \leq \nu$, $\nu(\mathbb{R}^2) = n$ and $\mu(\mathbb{R}^2) = (1 - \epsilon)n$. There exists a coupling between $((X_i, Y_i)_{i\in[M]}, (U_i, V_i)_{i\in\mathbb{N}}) \sim \mathrm{PP}(\mu) \otimes \mathrm{Unif}[0, 1]^{\otimes\mathbb{N}}$ and $((\tilde{X}_i, \tilde{Y}_i)_{i\in[N]}, (\tilde{U}_i, \tilde{V}_i)_{i\in\mathbb{N}}) \sim \mathrm{PP}(\nu) \otimes \mathrm{Unif}[0, 1]^{\otimes\mathbb{N}}$ such that*

$$\Big|\mathcal{V}\Big((X_i)_{i\in[M]}, (Y_i)_{i\in[M]}, (U_i)_{i\in[M]}, (V_i)_{i\in[M]}; \frac{1}{2\sqrt{M}}\Big) - \mathcal{V}\Big((\tilde{X}_i)_{i\in[N]}, (\tilde{Y}_i)_{i\in[N]}, (\tilde{U}_i)_{i\in[N]}, (\tilde{V}_i)_{i\in[N]}; \frac{1}{2\sqrt{N}}\Big)\Big| \leq 10\epsilon + O_p(n^{-1/2}).$$

*Proof.* Since $\mu \leq \nu$, we have that $\mu$ is absolutely continuous with respect to $\nu$ and the Radon–Nikodym derivative satisfies $0 \leq d\mu/d\nu \leq 1$ $\nu$-almost everywhere. We can define a coupling between $\mathrm{PP}(\mu)$ and $\mathrm{PP}(\nu)$ via Poisson thinning, i.e. we can define $(X_1, Y_1), \ldots, (X_M, Y_M)$ to be a subset of $(\tilde{X}_1, \tilde{Y}_1), \ldots, (\tilde{X}_N, \tilde{Y}_N)$ such that conditional on $(\tilde{X}_i, \tilde{Y}_i)_{i\in[N]}$, each point $(\tilde{X}_i, \tilde{Y}_i)$ is selected into the subset with probability $d\mu/d\nu(\tilde{X}_i, \tilde{Y}_i)$. Note in particular that under such a coupling, $M \leq N$ and $N - M \sim \mathrm{Poi}(\epsilon n)$. We can then define a conditional joint distribution (conditional on $(X_i, Y_i)_{i\in[M]}$ and $(\tilde{X}_i, \tilde{Y}_i)_{i\in[M]}$) between $(U_i, V_i)_{i\in\mathbb{N}}$ and $(\tilde{U}_i, \tilde{V}_i)_{i\in\mathbb{N}}$ by first constructing a conditional coupling between $(U_i, V_i)_{i\in[M]}$ and $(\tilde{U}_i, \tilde{V}_i)_{i\in[N]}$ as in Lemma 21 and then set $(U_i, V_i)_{i>M}$ and $(\tilde{U}_i, \tilde{V}_i)_{i>N}$ to be independent.

Under this coupling, we have $N \sim \mathrm{Poi}(n)$ and $N - M \sim \mathrm{Poi}(n\epsilon)$. So given any $\delta > 0$, there exists $C > 0$ (depending on $\delta$ and $\epsilon$) and an event with probability at least $1 - \delta/2$ on which

$$N \in [n - C\sqrt{n}, n + C\sqrt{n}] \quad \text{and} \quad N - M \in [\epsilon n - C\sqrt{n}, \epsilon n + C\sqrt{n}].$$

Conditional on any realisation of $(X_i, Y_i)_{i\in[M]}$ and $(\tilde{X}_i, \tilde{Y}_i)_{i\in[N]}$ on this event, we can apply Lemma 21 to obtain that under the current coupling, we have with (conditional) probability at least $1 - 2e^{-3Cn^{1/2}/10}$ that

$$\begin{aligned}\Big|\mathcal{V}\Big((X_i)_{i\in[M]}, (Y_i)_{i\in[M]}, &(U_i)_{i\in[M]}, (V_i)_{i\in[M]}; \frac{1}{2\sqrt{M}}\Big) \\ &- \mathcal{V}\Big((\tilde{X}_i)_{i\in[N]}, (\tilde{Y}_i)_{i\in[N]}, (\tilde{U}_i)_{i\in[N]}, (\tilde{V}_i)_{i\in[N]}; \frac{1}{2\sqrt{N}}\Big)\Big| \le \frac{10(\epsilon n + C\sqrt{n})}{n - C\sqrt{n}},\end{aligned}$$

which can be further upper bounded by $10\epsilon + 2Cn^{-1/2}$ for sufficiently large $n$. Also, for large $n$, we have $1 - 2e^{-3Cn^{1/2}/10} \le \delta/2$, the desired result follows after integrating over realisation of $(X_i, Y_i)_{i\in[M]}$ and $(\tilde{X}_i, \tilde{Y}_i)_{i\in[N]}$. □

**Lemma 24.** *Suppose* $R_1, \ldots, R_n \overset{\text{iid}}{\sim} \mathrm{Unif}([0,1]^d)$. *For* $\delta \in [0, 1/2)$, *let* $B$ *be a compact set symmetric around* 0 *with* $\mathrm{vol}(B) = \delta$ *and define*

$$\mathcal{V} := \mathrm{vol}\Big([0,1]^d \setminus \bigcup_{i=1}^n (R_i + B)\Big),$$

*where* $R_i + B$ *denotes the Minkowski sum of the sets. If* $n\delta \to q$, *then as* $n \to \infty$, *we have*

$$\mathbb{E}(\mathcal{V}) = (1-\delta)^n \to e^{-q} \quad \textit{and} \quad \mathrm{Var}(\mathcal{V}) \to 0.$$

*If in addition,* $B = B(0, \gamma)$ *(i.e.* $\delta = (2\gamma)^d$*) for some* $\gamma < 1/4$, *then we have*

$$\mathrm{Var}(\mathcal{V}) = \sum_{r=2}^n \binom{n}{r}(1-2\delta)^{n-r}\Big\{\frac{2^d\delta^{r+1}}{(r+1)^d} - \delta^{2r}\Big\} = (1+o(1))\frac{qe^{-2q}}{n}\sum_{r=2}^\infty \frac{q^r}{r!}\Big(\frac{2}{r+1}\Big)^d.$$

*Proof.* Draw $W_1, W_2 \overset{\text{iid}}{\sim} \mathrm{Unif}[0,1]^d$ independent of all other randomness. We have

$$\mathbb{E}(\mathcal{V}) = \mathbb{P}\Big(W_1 \notin \bigcup_{i=1}^n (R_i + B)\Big) = \mathbb{E}\Big\{\mathbb{P}\Big(\bigcap_{i=1}^n \{R_i \notin (W_1 + B)\} \;\Big|\; W_1\Big)\Big\} = (1-\delta)^n \to e^{-q}.$$

To control the variance, we start with

$$\begin{aligned}\mathbb{E}(\mathcal{V}^2) =& \mathbb{P}\Big(W_j \notin \bigcup_{i=1}^n (R_i + B)\ \forall j \in \{1,2\}\Big) = \mathbb{E}\Big[\mathbb{P}\Big(\bigcap_{i=1}^n \{R_i \notin (W_1 + B) \cup (W_2 + B)\} \;\Big|\; W_1, W_2\Big)\Big] \\ =& \mathbb{E}\Big[\big\{1 - 2\delta + \mathrm{vol}\big((W_1 + B) \cap (W_2 + B)\big)\big\}^n\Big]. \end{aligned} \tag{31}$$

Since $\mathrm{vol}\big((W_1 + B) \cap (W_2 + B)\big) \xrightarrow{\text{a.s.}} 0$, by the Dominated Convergence Theorem, we have

$$\lim_{n\to\infty} \mathbb{E}(\mathcal{V}^2) = \lim_{n\to\infty} \{1 - 2\delta\}^n = e^{-2q}.$$

Consequently $\mathrm{Var}(\mathcal{V}) = \mathbb{E}(\mathcal{V}^2) - (\mathbb{E}\mathcal{V})^2 \to 0$.

When $B = B(0, \gamma)$ is a cube, we can obtain an explicit expression of the variance by examining moments of $H := \mathrm{vol}\big((W_1 + B) \cap (W_2 + B)\big)$ in detail. Specifically, noting that $H$

measures the volume of a hypercube with independent side lengths each having density distribution $4\gamma\mathrm{Unif}[0, 2\gamma] + (1-4\gamma)\delta_0$, we get for any $r \in \mathbb{N}$ that

$$\mathbb{E}(H^r) = \left\{2\int_0^{2\gamma} t^r \, dt\right\}^d = \left\{\frac{2(2\gamma)^{r+1}}{r+1}\right\}^d = \frac{2^d\delta^{r+1}}{(r+1)^d}.$$

Consequently, from (31), we have

$$\mathbb{E}(\mathcal{V}^2) = \sum_{r=0}^{n}\binom{n}{r}(1-2\delta)^{n-r}\mathbb{E}(H^r) = (1-2\delta)^n + \sum_{r=1}^{n}\binom{n}{r}\frac{(1-2\delta)^{n-r}2^d\delta^{r+1}}{(r+1)^d}. \tag{32}$$

On the other hand, we have

$$\{\mathbb{E}(\mathcal{V})\}^2 = (1-\delta)^{2n} = \sum_{r=0}^{n}\binom{n}{r}(1-2\delta)^{n-r}\delta^{2r}. \tag{33}$$

Combining (32) and (33), we have

$$\mathrm{Var}(\mathcal{V}) = \sum_{r=2}^{n}\binom{n}{r}(1-2\delta)^{n-r}\left\{\frac{2^d\delta^{r+1}}{(r+1)^d} - \delta^{2r}\right\}.$$

We next compute the asymptotic variance. Since $\sum_{r=2}^{n}\binom{n}{r}(1-2\delta)^{n-r}\delta^{2r} \le \sum_{r=2}^{\infty}(n\delta^2)^r = O(n^{-2})$, we have

$$\begin{aligned}
\delta^{-1}\mathrm{Var}(\mathcal{V}) &= \sum_{r=2}^{n}\binom{n}{r}(1-2\delta)^{n-r}\left(\frac{2}{r+1}\right)^d\delta^r + O(n^{-1}) \\
&= (1-2\delta)^n\sum_{r=2}^{n}\binom{n}{r}\frac{\delta^r}{(1-2\delta)^r}\left(\frac{2}{r+1}\right)^d + O(n^{-1}) \\
&= (1+o(1))e^{-2q}\sum_{r=2}^{n}\frac{q^r}{r!}\left(\frac{2}{r+1}\right)^d = (1+o(1))e^{-2q}\sum_{r=2}^{\infty}\frac{q^r}{r!}\left(\frac{2}{r+1}\right)^d
\end{aligned}$$

which implies the desired limit since $n\delta \to q$. $\square$

The following lemma provides a multiplicative Chernoff bound for a Beta distribution.

**Lemma 25.** *For $0 < m \le n/2$ and $X \sim \mathrm{Beta}(m, n-m)$, we have for any $t \ge 0$ that*

$$\mathbb{P}\left(X \ge \frac{(1+t)m}{n}\right) \le \exp\left\{-\frac{t^2 m}{2+4t/3}\right\}.$$

*Proof.* Set

$$v = \frac{m(n-m)}{n^2(n+1)} \qquad c = \frac{2(n-2m)}{n(n+2)}.$$

By Skorski (2023, Theorem 1), we have

$$\mathbb{P}\left(X \ge \frac{(1+t)m}{n}\right) \le \exp\left\{-\frac{t^2m^2/n^2}{2v + 2ctm/(3n)}\right\} \le \exp\left\{-\frac{t^2 m}{2+4t/3}\right\}$$

as desired. $\square$

**Lemma 26.** *Suppose $\sigma \in \mathcal{S}_n$ is a random permutation on $[n]$. Write $R_i = (i/n, \sigma(i)/n)$ for $i \in [n]$. For $\delta \in [0, 1/2)$, let $B$ be a compact set symmetric around 0 with* $\mathrm{vol}(B) = \delta$ *and define*

$$\mathcal{V} := \mathrm{vol}\Big([0,1]^d \setminus \bigcup_{i=1}^{n} (R_i + B)\Big),$$

*where $R_i + B$ denotes the Minkowski sum of the sets. If $n\delta \to q$, then as $n \to \infty$, we have*

$$\mathbb{E}(\mathcal{V}) \to e^{-q}.$$

*Proof.* For any $x \in [0,1]^2$, let $N_x := |\{i : (i/n, \sigma(i)/n) \in x + B\}|$. Note that $N_x$ follows a Hypergeometric distribution $\mathrm{Hyper}(n^2, K, n)$ where $K = |\{(i,j) : (i/n, j/n) \in x + B\}|$. Note that $K = n^2\delta + o(n)$. Therefore $\lambda_n := \mathbb{E}[N_x] = n\delta + o(1)$. Then note that we have

$$d_{TV}(\mathrm{Hyper}(n^2, K, n), \mathrm{Pois}(\lambda_n)) = O(\frac{1}{n})$$

Then note that

$$\mathcal{V} = \int_{[0,1]^2} \mathbb{1}\{N_x = 0\} dx,$$

therefore

$$\mathbb{E}[\mathcal{V}] = \int_{[0,1]^2} \mathbb{P}(N_x = 0) dx = e^{-n\delta} + O(\frac{1}{n}) \to e^{-q}.$$

as desired. □

**Lemma 27.** *Let $\{(x_i, y_i)\}_{i=1}^n$ be $n$ points in $\mathbb{R}^2$ and let $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$. Assume that $x_i \neq x_j$ and $y_i \neq y_j$ for all $i \neq j$. For any $z \in \mathbb{R}^n$ let $r_i^z = n^{-1} \sum_{j=1}^n \mathbb{1}\{z_j \leq z_i\}$. Define $r_i = (r_i^x, r_i^y)$ for $i \in [n]$. Let $B = [-1/2\sqrt{n}, 1/2\sqrt{n}]^2$ and define*

$$\mathcal{V} := \mathrm{vol}\Big([0,1]^2 \setminus \bigcup_{i=1}^{n} (r_i + B)\Big),$$

*where $r_i + B$ denotes the Minkowski sum of the sets. Let $\{(x_i', y_i')\}_{i=1}^n$ be another set of points in $\mathbb{R}^2$ and define $x^k, y^k \in \mathbb{R}^n$ where $x_i^k = x_i$ and $y_i^k = y_i$ for all $i \in [n] \setminus \{k\}$ and $x_k^k = x_k'$ and $y_k^k = y_k'$.*

$$\mathcal{V}^k := \mathrm{vol}\Big([0,1]^d \setminus \bigcup_{i=1}^{n} (r_i^k + B)\Big),$$

*where $r_i^k = (r_i^{x^k}, r_i^{y^k})$. Then*

$$|\mathcal{V} - \mathcal{V}^k| \leq 10/n.$$

*Proof.* First note that $|\mathcal{V} - \mathcal{V}^k| = 0$ if $x_k = x_k'$ and $y_k = y_k'$, thus the result holds automatically. Assume $x_k \neq x_k'$ or $y_k \neq y_k'$. Define

$$\begin{aligned}
\mathcal{I}_0 &:= \{i \in [n] \setminus \{k\} : (x_i - x_k)(x_i - x_k') \geq 0 \text{ and } (y_i - y_k)(y_i - y_k') \geq 0\}, \\
\mathcal{I}_1 &:= \{i \in [n] \setminus \{k\} : (x_i - x_k)(x_i - x_k') < 0 \text{ and } (y_i - y_k)(y_i - y_k') < 0\}, \\
\mathcal{I}_2 &:= \{i \in [n] \setminus \{k\} : (x_i - x_k)(x_i - x_k') < 0 \text{ and } (y_i - y_k)(y_i - y_k') \geq 0\}, \\
\mathcal{I}_3 &:= \{i \in [n] \setminus \{k\} : (x_i - x_k)(x_i - x_k') \geq 0 \text{ and } (y_i - y_k)(y_i - y_k') < 0\}.
\end{aligned}$$

Note that when replacing $x_k$ by $x'_k$ and $y_k$ by $y'_k$, both $\{x_i : i \in \mathcal{I}_0\}$ and $\{y_i : i \in \mathcal{I}_0\}$ maintain their original ranks, and both $\{x_i : i \in \mathcal{I}_1\}$ and $\{y_i : i \in \mathcal{I}_1\}$ have a shift of $1/n$ in either direction. The ranks of $\{x_i : i \in \mathcal{I}_2\}$ have a similar shift of $1/n$ while the ranks of $\{y_i : i \in \mathcal{I}_2\}$ remain the same. Conversely, the ranks of $\{x_i : i \in \mathcal{I}_3\}$ remains the same while ranks of $\{y_i : i \in \mathcal{I}_3\}$ have a shift of $1/n$ in either direction. Specifically, we have

$$r_i^k = \begin{cases} r_i, & \text{if } i \in \mathcal{I}_0 \\ r_i + (\pm\frac{1}{n}, \pm\frac{1}{n}), & \text{if } i \in \mathcal{I}_1 \\ r_i + (\pm\frac{1}{n}, 0), & \text{if } i \in \mathcal{I}_2 \\ r_i + (0, \pm\frac{1}{n}), & \text{if } i \in \mathcal{I}_3. \end{cases} \tag{34}$$

Let $\mathcal{U}_j = \bigcup_{i \in \mathcal{I}_j}(r_i + B)$ and $\mathcal{U}_j^k = \bigcup_{i \in \mathcal{I}_j}(r_j^k + B)$ for $j = 0, 1, 2, 3$, we have the following decompositions

$$\mathcal{C} := \bigcup_{i \in [n]} (r_i + B) = \Big(\bigcup_{j=0}^{3} \mathcal{U}_j\Big) \cup (r_k + B),$$

$$\mathcal{C}^k := \bigcup_{i \in [n]} (r_i^k + B) = \Big(\bigcup_{j=0}^{3} \mathcal{U}_j^k\Big) \cup (r_k^k + B).$$

By (34) we have $\mathrm{vol}\big(\mathcal{U}_0 \Delta \mathcal{U}_0^k\big) = 0$. For $i \in \mathcal{I}_2$, $\mathcal{U}_2^k$ is a shift of $\mathcal{U}_2$ by $1/n$ to the right(or left). Therefore, for "almost" every $x \in \mathcal{U}_2 \setminus \mathcal{U}_2^k$ there is $y \in \mathcal{U}_2^k \setminus \mathcal{U}_2$ except for those $x \in [0,1]^2$ with $x \in \big([0, 1/n] \cup [1 - 1/n, 1]\big) \times [0, 1]$. A similar argument holds for $i \in \mathcal{I}_3$, with the difference that the shift is up/down. Hence, for $j = 2, 3$

$$|\mathrm{vol}\big(\mathcal{U}_j\big) - \mathrm{vol}\big(\mathcal{U}_j^k\big)| \leq 2/n.$$

For $i \in \mathcal{I}_1$, boxes shift both right/left and up/down and hence we have

$$|\mathrm{vol}\big(\mathcal{U}_1\big) - \mathrm{vol}\big(\mathcal{U}_1^k\big)| \leq 4/n.$$

Therefore, we have

$$|\mathcal{V} - \mathcal{V}^k| \leq \sum_{j=0}^{3} |\mathrm{vol}\big(\mathcal{U}_j\big) - \mathrm{vol}\big(\mathcal{U}_j^k\big)| + \mathrm{vol}\Big((r_k + B)\Delta(r_k^k + B)\Big) \leq 10/n.$$

as claimed. $\square$

**Lemma 28.** *For $n, L \in \mathbb{N}$, let $p := 1/L$ and suppose $(N_1, \ldots, N_L) \sim \mathrm{Multin}(n; (p, \ldots, p))$. Consider the asymptotic regime where $n \to \infty$ and $L$ is fixed. Suppose $a, b \geq 1$ satisfies $p(a-1) = O(1/n)$ and $p(b-1) = O(1/n)$, then for any and $\ell, k \in [L]$, we have*

$$\mathrm{Cov}(a^{N_\ell}, b^{N_k}) = O(n^{-2}).$$

*Proof.* We first assume that $\ell \neq k$. We write $\alpha = p(a-1)$ and $\beta = p(b-1)$ for simplicity. Using the moment generating function of the Multinomial distribution, we observe that,

$$\mathbb{E}(a^{N_\ell}) = (1 + \alpha)^n$$
$$\mathbb{E}(a^{N_\ell} b^{N_k}) = (1 + \alpha + \beta)^n.$$

Using the above identities and the Taylor expansion

$$\begin{aligned}|\mathrm{Cov}(a^{N_\ell}, b^{N_k})| &= |(1+\alpha+\beta)^n - (1+\alpha+\beta+\alpha\beta)^n| \\ &= \alpha\beta \sum_{i=0}^{n-1} (1+\alpha+\beta)^{n-1-i}(\alpha\beta)^i \\ &\leq \alpha\beta(1+\alpha+\beta)^n \sum_{i=0}^{n-1} (\alpha\beta)^i = O(\alpha\beta) = O(n^{-2}).\end{aligned}$$

It remains to check the case where $\ell = k$. For this, define $\eta = p(ab-1)$,

$$\begin{aligned}|\mathrm{Cov}(a^{N_\ell}, b^{N_k})| &= \{1+p(ab-1)\}^n - \{1+p(a-1)\}^n\{1+p(b-1)\}^n \\ &= (1+\alpha\beta/p+\alpha+\beta)^n - (1+\alpha+\beta+\alpha\beta)^n \\ &= \alpha\beta(1/p-1)\sum_{i=0}^{n-1}(1+\alpha+\beta+\alpha\beta)^{n-1-i}(\alpha\beta/p-\alpha\beta)^i \\ &\leq \alpha\beta(1/p-1)(1+\alpha)^n(1+\beta)^n \sum_{i=0}^{n-1}(\alpha\beta/p-\alpha\beta)^i = O(\alpha\beta) = O(n^{-2}),\end{aligned}$$

as desired. □

**Lemma 29.** *Let $(M_n)_n$ and $(L)_n$ be sequences of random variables such that $M_n \xrightarrow{\mathrm{d}} \mathcal{N}(\mu, \alpha^2)$ and $L_n \xrightarrow{\mathrm{p}} \beta^2$. Let $\mathcal{F}_n$ be the sigma-algebra generated by $(M_i)_{i\leq n}$ and $(L_i)_{i\leq n}$. If $(X_n)_n$ is a sequence of random variables such that*

$$\mathbb{E} \sup_{-\infty<x<\infty} \left| \mathbb{P}\left( \frac{X_n - M_n}{\sqrt{L_n}} \leq x \;\middle|\; \mathcal{F}_n \right) - \Phi(x) \right| \to 0, \quad n \to +\infty. \tag{35}$$

*Then we have $X_n \xrightarrow{\mathrm{d}} \mathcal{N}(\mu, \alpha^2+\beta^2)$.*

*Proof.* For any $x \in \mathbb{R}$, we have

$$\begin{aligned}&\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left( X_n \leq \mu + x\sqrt{\alpha^2+\beta^2} \right) - \Phi(x) \right| \\ &\quad = \sup_{x\in\mathbb{R}} \left| \mathbb{E}\left\{ \mathbb{P}\left( \frac{X_n - M_n}{\sqrt{L_n}} \leq \frac{\mu - M_n + x\sqrt{\alpha^2+\beta^2}}{\sqrt{L_n}} \;\middle|\; \mathcal{F}_n \right) \right\} - \Phi(x) \right| \\ &\quad = \sup_{x\in\mathbb{R}} \left| \mathbb{E}\left\{ \Phi\left( \frac{\mu - M_n + x\sqrt{\alpha^2+\beta^2}}{\sqrt{L_n}} \right) \right\} - \Phi(x) \right| + o(1),\end{aligned} \tag{36}$$

where we use condition (35) in the final step. By Slutsky's theorem, we have for each $x \in \mathbb{R}$

$$\frac{\mu - M_n + x\sqrt{\alpha^2+\beta^2}}{\sqrt{L_n}} \xrightarrow{\mathrm{d}} \mathcal{N}\left( \frac{x\sqrt{\alpha^2+\beta^2}}{\beta}, \frac{\alpha^2}{\beta^2} \right).$$

Consequently, we have for $Z \sim \mathcal{N}(0,1)$ independent of all other randomness in the lemma that

$$\mathbb{E}\left\{ \Phi\left( \frac{\mu - M_n + x\sqrt{\alpha^2+\beta^2}}{\sqrt{L_n}} \right) \right\} = \mathbb{P}\left( Z - \frac{\mu - M_n + x\sqrt{\alpha^2+\beta^2}}{\sqrt{L_n}} \leq 0 \right) = \Phi(x) + o(1).$$

The conclusion holds by combining the above with (36), and using Chow and Teicher (1988, Lemma 3, pp.265). □

**Lemma 30.** *For $\alpha_\lambda^2$ and $\beta_\lambda^2$ defined in (21) and (22), as $\lambda \to \infty$, we have*

$$\alpha_\lambda^2 \to 0, \qquad \beta_\lambda^2 \to \beta^2.$$

*where $\beta^2$ is a positive constant.*

*Proof of Lemma 30.* Note that $\alpha_\lambda^2 = \alpha_{1,\lambda}^2 - \alpha_{2,\lambda}^2$ where

$$\alpha_{1,\lambda}^2 = \frac{\lambda^{2d}}{e^2 2^d (\lambda+2)^d}\left(e^{(\lambda/2+1)^{-d}} - 1\right), \qquad \alpha_{2,\lambda}^2 = \frac{\lambda^{2d}}{e^2(\lambda+2)^{2d}}.$$

We show that as $\lambda \to \infty$, $\alpha_\lambda^2$ converges to a positive constant. First note that $\alpha_{1,\lambda}^2 = e^{-2} + O(\lambda^{-d})$ and $\alpha_{2,\lambda}^2 = e^{-2} + O(\lambda^{-1})$. Therefore $\alpha_\lambda^2 \to 0$ as $\lambda \to \infty$.

For $\beta_\lambda^2$, note that as $\lambda \to \infty$ we have

$$\beta_\lambda^2 \to e^{-2}(C_d - 1) = \beta^2 > 0$$

as desired. □

# C Algorithm and simulation settings

## C.1 Algorithmic implementation details

The computation of the coverage correlation coefficient involves evaluating the volume of the union of $n$ axis-aligned hypercubes, each of volume $1/n$, in the unit cube $[0,1]^d$, with edge wrapping. This is a special case of Klee's measure problem (Klee, 1977), which concerns computing the volume of the union of arbitrary axis-aligned hyperrectangles. When $d = 2$, Bentley's algorithm solves this in $O(n \log n)$ time by sweeping along one axis and maintaining the union of intervals along the other using a segment tree (Ben-Or, 1983). In higher dimensions, the time complexity of Bentley's algorithm becomes $O(n^{d-1} \log n)$, while the best known theoretical bound is Chan's $O(n^{d/2})$ algorithm (Chan, 2013). However, for moderate dimensions (e.g., $d \leq 10$), Bentley's approach remains more practical due to its smaller constant factors, better memory behaviour, and simpler implementation. Moreover, in our setting, we are able to exploit the uniform size of the input small hypercubes to make computational gains using Bentley's algorithm alone. Specifically, we partition $[0,1]^d$ into $m^d$ grid blocks with $m \approx n^{1/d}$, and compute the union volume by summing contributions from individual blocks. In cases where the small hypercubes are spread out uniformly at random, we expect $O_p(\log n)$ small hypercubes intersecting each block, thus making the entire algorithm run in an average-case complexity of $O(n \log^{d-1} n)$. We outline the recursive union volume computation using Bentley's algorithm in Algorithm 1 and the full coverage correlation computation algorithm in Algorithm 2.

## C.2 Simulation supplement

We show in Figure 5 scatter plots of the six simulation settings used in Section 3.1 at different noise levels. Also, Table 2 shows the running time of the six algorithms under comparison for $n \in \{10, 100, 1000, 10000\}$ and $d_X = d_Y \in \{1, 2\}$. Algorithm timing was performed on an 8-core 3.2 GHz laptop CPU, averaged over 10 repetitions. We see that when $d_X = d_Y = 1$, both the coverage correlation and Chatterjee's correlation scale approximately linearly and the other

**Algorithm 1:** `UnionVolume`$(\mathcal{R}, d)$: Bentley's algorithm for union volume for $d \geq 2$

**Input:** List $\mathcal{R}$ of axis-aligned rectangles in $\mathbb{R}^d$, $d \geq 2$
**Output:** Volume of the union of rectangles

```
if d = 2 then
    Initialize an empty event list E;
    foreach rectangle (x_min, x_max, y_min, y_max) ∈ R do
        Add events (x_min, +1, [y_min, y_max]) and (x_max, −1, [y_min, y_max]) to E;
    end
    Sort E by x-coordinate;
    Initialize active multiset A ← ∅, total_area ← 0, x_prev ← undefined;
    foreach event (x, type, [y_min, y_max]) ∈ E do
        if x_prev is defined then
            Let height ← total length of union of intervals in A;
            total_area ← total_area + (x − x_prev) · height;
        end
        if type = +1 then
            Insert interval [y_min, y_max] into A;
        else
            Remove interval [y_min, y_max] from A;
        end
        x_prev ← x;
    end
    return total_area;
else
    Extract all unique coordinates along the first axis and sort them as x_1 < ··· < x_k;
    Initialize total_volume ← 0;
    foreach interval [x_i, x_{i+1}] do
        Let S ← {rectangles in R that span [x_i, x_{i+1}] in axis 1};
        Project each rectangle in S to the remaining d − 1 dimensions to obtain S′;
        total_volume ← total_volume + (x_{i+1} − x_i) · UnionVolume(S′, d − 1);
    end
    return total_volume;
```

**Algorithm 2:** Pseudocode for computing the coverage correlation coefficient

**Input:** Two samples $X_1, \ldots, X_n \in \mathbb{R}^{d_X}$ and $Y_1, \ldots, Y_n \in \mathbb{R}^d_Y$
**Output:** coverage correlation $\kappa_n^{X,Y}$ and the corresponding p-value $p_\kappa$

Draw $U_1, \ldots, U_n \overset{\text{iid}}{\sim} \text{Unif}([0,1]^{d_X})$ and $V_1, \ldots, V_n \overset{\text{iid}}{\sim} \text{Unif}[0,1]^{d_Y}$;
Compute Monge–Kantorovich ranks $R_1, \ldots, R_n \in [0,1]^d$ for $d = d_X + d_Y$ as in (2);
Initialize empty list $\mathcal{R}$;
**for** $i$ *in* $1, \ldots, n$ **do**
  Split $B(R_i, \frac{1}{2n^{1/d}})$ along wrapped axes to get up to $2^d$ axis-aligned hyperrectangles within $[0,1]^d$;
  Add all resulting non-wrapping rectangles to $\mathcal{R}$;
**end**
Partition $[0,1]^d$ into $m^d$ grid blocks $G_1, \ldots, G_{m^d}$ for $m := \lfloor n^{1/d} \rfloor$. ;
**if** $d = 2$ **then**
  $\mathcal{V}_n \leftarrow 1 - \text{CoveredVolume}(\mathcal{R})$
**else**
  Initialise $\mathcal{V}_n \leftarrow 1$;
  **foreach** *grid block* $G_k$, $k \in [m^d]$ **do**
    Define $\mathcal{R}'_k := \{A \cap G_k : A \in \mathcal{R}\}$ $\mathcal{V}_n \leftarrow \mathcal{V}_n - \text{CoveredVolume}(\mathcal{R}'_k)$
  **end**
**end**
Compute $\mu \leftarrow (1 - 1/n)^n$ and

$$\sigma^2 := \sum_{k=2}^{n} \binom{n}{k} \Big(1 - \frac{2}{n}\Big)^{n-k} \Big\{ \Big(\frac{2}{k+1}\Big)^d n^{-k-1} - n^{-2k} \Big\}.$$

**return** $\kappa_n^{X,Y} := (\mathcal{V}_n - \mu)/\mu$ and $p_\kappa := 1 - \Phi(\sqrt{n}(\mathcal{V}_n - \mu)/\sigma)$.

algorithms scale quadratically in $n$. When $d_X = d_Y = 2$, the coverage correlation has a quadratic scaling in $n$, mostly driven by the Monge–Kantorovich rank computation.

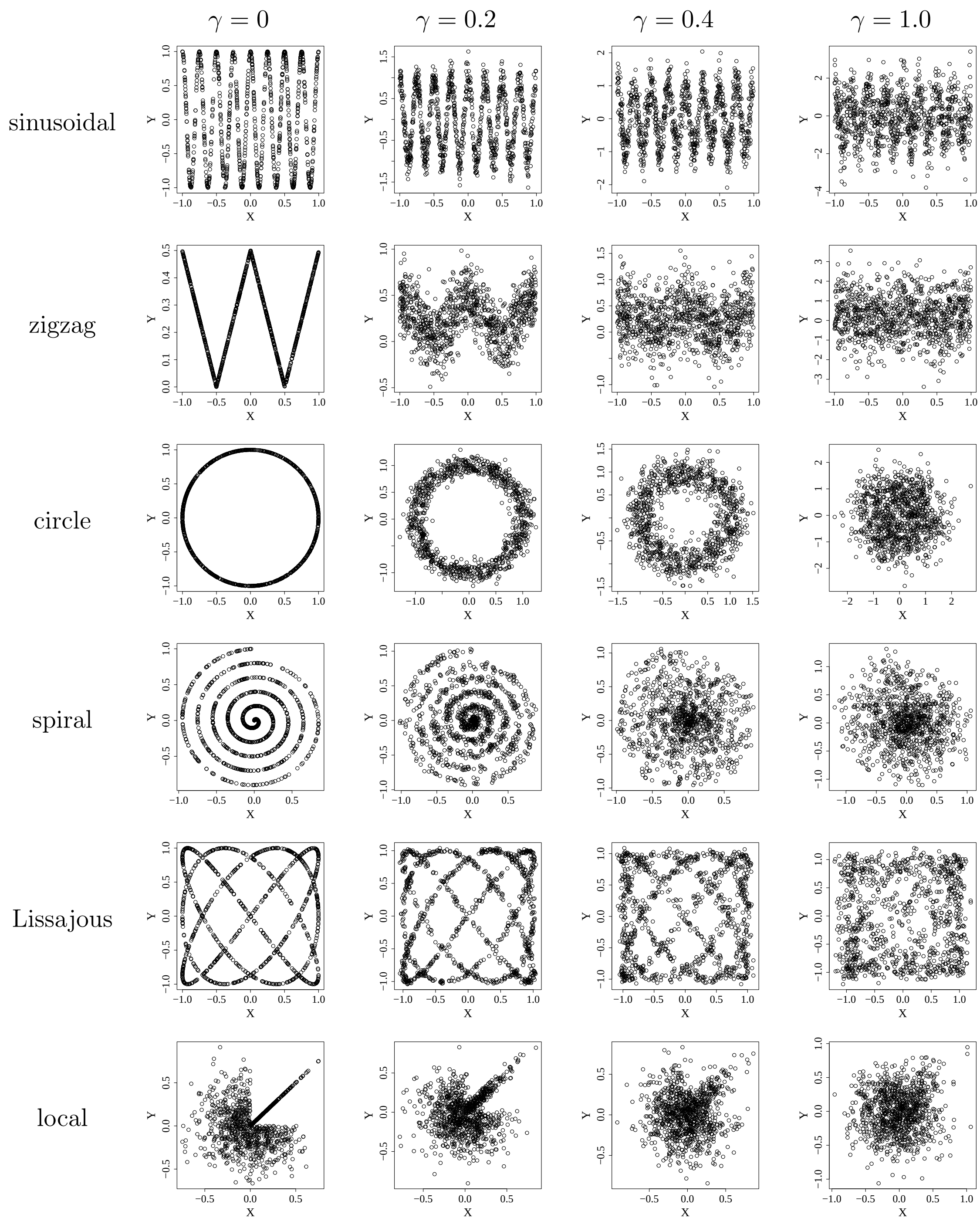


Figure 5: Scatter plot of data from different simulation settings at different noise levels.

| $n$ | $d_X$ | $\kappa_n^{X,Y}$ | $\xi_n^{X,Y}$ | dCor | HSIC | KMAc | USP |
|---|---|---|---|---|---|---|---|
| 125 | 1 | 0.001 | 0.001 | 0.008 | 0.014 | 0.553 | 0.010 |
| 250 | 1 | 0.001 | 0.001 | 0.010 | 0.047 | 1.06 | 0.043 |
| 500 | 1 | 0.002 | 0.001 | 0.037 | 0.192 | 2.50 | 0.182 |
| 1000 | 1 | 0.003 | 0.001 | 0.130 | 1.01 | 7.35 | 0.781 |
| 2000 | 1 | 0.005 | 0.001 | 0.498 | 4.23 | 26.3 | 2.98 |
| 4000 | 1 | 0.010 | 0.002 | 2.01 | 21.6 | - | 10.8 |
| 8000 | 1 | 0.019 | 0.003 | 7.95 | - | - | - |
| 125 | 2 | 0.034 | - | 0.004 | 0.011 | 0.514 | 0.042 |
| 250 | 2 | 0.076 | - | 0.014 | 0.042 | 1.05 | 0.164 |
| 500 | 2 | 0.177 | - | 0.052 | 0.186 | 2.52 | 0.720 |
| 1000 | 2 | 0.567 | - | 0.176 | 0.975 | 7.50 | 3.17 |
| 2000 | 2 | 1.93 | - | 0.694 | 4.45 | 27.5 | 10.8 |
| 4000 | 2 | 6.16 | - | 2.77 | 21.5 | - | 43.9 |
| 8000 | 2 | 24.4 | - | 11.4 | - | - | - |

Table 2: Average running time of coverage correlation, Chatterjee's correlation, distance correlation, HSIC, KMAc and USP for various $n$ and $d_X = d_Y$ values under the null. Chatterjee's correlation is only computed for $d_X = d_Y = 1$. Time is shown in seconds, and time larger than 60 seconds is not displayed.

# References


Azadkia, M. and Chatterjee, S. (2021) A simple measure of conditional dependence. *Ann. Statist.*, **49**, 3070–3102.

Azadkia, M. and Roudaki, P. (2025) A new measure of dependence: Integrated $R^2$. *arXiv preprint*, arxiv:2505.18146.

Ben-Or, M. (1983) Lower bounds for algebraic computation trees. In *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, 80–86.

Bergsma, W. and Dassios, A. (2014) A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli*, **20**, 1006–1028.

Berrett, T. B., Kontoyiannis, I. and Samworth, R. J. (2021) Optimal rates for independence testing via U-statistic permutation tests. *Ann. Statist.*, **49**, 2457–2490.

Berrett, T. B. and Samworth, R. J. (2019) Nonparametric independence testing via mutual information. *Biometrika*, **106**, 547–566.

Berry, A. C. (1941) The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.*, **49**, 122–136.

Bickel, P. J. (1965) On some asymptotically nonparametric competitors of Hotelling's $T^2$. *Ann. Math. Statist.*, **36**, 160–173.

Blum, J. R., Kiefer, J. and Rosenblatt, M. (1961) Distribution free tests of independence based on the sample distribution function. *Ann. Math. Statist.*, **32**, 485–498.

Borgonovo, E., Figalli, A., Ghosal, P., Plischke, E. and Savaré, G. (2025) Convexity and measures of statistical association. *J. Roy. Statist. Soc., Ser. B*, qkaf040.

Breiman, L. and Friedman, J. H. (1985) Estimating optimal transformations for multiple regression and correlation. *J. Amer. Statist. Assoc.*, **80**, 580–619.

Chan, T. M. (2013) Klee's measure problem made easy. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, 410–419. IEEE.

Chatterjee, S. (2021) A new coefficient of correlation. *J. Amer. Statist. Assoc.*, **116**, 2009–2022.

Chaudhuri, P. (1996) On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.*, **91**, 862–872.

Chernozhukov, V., Galichon, A., Hallin, M. and Henry, M. (2017) Monge–Kantorovich depth, quantiles, ranks and signs. *Ann. Statist.*, **45**, 223–256.

Chow, Y. and Teicher, H. (1988) *Probability Theory: Independence, Interchangeability, Martingales.* Springer, New York.

David, H. A. and Nagaraja, H. N. (2004) *Order Statistics.* John Wiley & Sons, Hoboken.

Deb, N., Ghosal, P. and Sen, B. (2020) Measuring association on topological spaces using kernels and geometric graphs. *arXiv preprint*, arxiv:2010.01768.

Deb, N. and Sen, B. (2023) Multivariate rank-based distribution-free nonparametric testing using measure transportation. *J. Amer. Statist. Assoc.*, **118**, 192–207.

Dette, H., Siburg, K. F. and Stoimenov, P. A. (2013) A copula-based non-parametric measure of regression dependence. *Scand. J. Stat.*, **40**, 21–41.

Dong, M., Wang, B., Wei, J., de O. Fonseca, A. H., Perry, C. J., Frey, A., Ouerghi, F., Foxman, E. F., Ishizuka, J. J., Dhodapkar, R. M. and Dijk, D. (2023) Causal identification of single-cell experimental perturbation effects with CINEMA-OT. *Nature Methods*, **20**, 1769–1779.

Drton, M., Han, F. and Shi, H. (2020) High-dimensional consistent independence testing with maxima of rank correlations. *Ann. Statist.*, **48**, 3206–3227.

Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956) Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.*, **27**, 642–669.

Esseen, C.-G. (1942) On the Liapunov limit error in the theory of probability. *Arkiv för matematik, astronomi och fysik*, **28**, 1–19.

Fournier, N. and Guillin, A. (2015) On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields*, **162**, 707–738.

Friedman, J. H. and Rafsky, L. C. (1983) Graph-theoretic measures of multivariate association and prediction. *Ann. Statist.*, **11**, 377–391.

Gebelein, H. (1941) Das statistische problem der korrelation als variationsund eigenwertproblem und sein zusammenhang mit der ausgleichsrechnung. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik*, **21**, 364–379.

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

Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schölkopf, B. and Smola, A. J. (2008) A kernel statistical test of independence. In *Advances in Neural Information Processing Systems*, 585–592.

Griessenberger, F., Junker, R. R. and Trutschnig, W. (2022) On a multivariate copula-based dependence measure and its estimation. *Electron. J. Statist.*, **16**, 2206–2251.

Hall, P. (1985) Three limit theorems for vacancy in multivariate coverage problems. *J. Multivariate Anal.*, **16**, 211–236.

Hall, P. (1988) *Introduction to the Theory of Coverage Processes.* Wiley, New York.

Hallin, M., Del Barrio, E., Cuesta-Albertos, J. and Matrán, C. (2021) Distribution and quantile functions, ranks and signs in dimension $d$: A measure transportation approach. *Ann. Statist.*, **49**, 1139–1165.

Hallin, M. and Paindaveine, D. (2002a) Multivariate signed ranks: Randles' interdirections or Tyler's angles? In Y. Dodge (ed.), *Statistical Data Analysis Based on the $L_1$-Norm and Related Methods*, 271–282. Birkhäuser Basel, Basel.

Hallin, M. and Paindaveine, D. (2002b) Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.*, **30**, 1103–1133.

Heller, R., Heller, Y. and Gorfine, M. (2013) A consistent multivariate test of association based on ranks of distances. *Biometrika*, **100**, 503–510.

Hirschfeld, H. O. (1935) A connection between correlation and contingency. *Math. Proc. Cambridge Philos. Soc.*, **31**, 520–524.

Hodges, J. L. (1955) A bivariate sign test. *Ann. Math. Statist.*, **26**, 523–527.

Hoeffding, W. (1948) A non-parametric test of independence. *Ann. Math. Statist.*, **19**, 546–557.

Holst, L. (1972) Asymptotic normality and efficiency for certain goodness-of-fit tests. *Biometrika*, **59**, 137–145.

Kendall, M. G. (1938) A new measure of rank correlation. *Biometrika*, **30**, 81–93.

Klee, V. (1977) Can the measure of $\cup_1^n [a_i, b_i]$ be computed in less than $O(n \log n)$ steps? *Amer. Math. Monthly*, **84**, 284–285.

Koltchinskii, V. I. (1997) M-estimation, convexity and quantiles. *Ann. Statist.*, **25**, 435–477.

Kraskov, A., Stögbauer, H. and Grassberger, P. (2004) Estimating mutual information. *Phys. Rev. E*, **69**, 066138.

Lindvall, T. (2002) *Lectures on the Coupling Method.* Dover Publications, Mineola.

Linfoot, E. H. (1957) An informational measure of correlation. *Inform. Control*, **1**, 85–89.

Liu, R. Y. and Singh, K. (1993) A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.*, **88**, 252–260.

Massart, P. (1990) The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *Ann. Probab.*, **18**, 1269–1283.

McDiarmid, C. et al. (1989) On the method of bounded differences. *Surveys in Combinatorics*, **141**, 148–188.

Móri, T. F. and Székely, G. J. (2019) Four simple axioms of dependence measures. *Metrika*, **82**, 1–16.

Möttönen, J. and Oja, H. (1995) Multivariate spatial sign and rank methods. *J. Nonparametr. Stat.*, **5**, 201–213.

Pearson, K. (1920) Notes on the history of correlation. *Biometrika*, **13**, 25–45.

Pfister, N., Bühlmann, P., Schölkopf, B. and Peters, J. (2018) Kernel-based tests for joint independence. *J. Roy. Statist. Soc., Ser. B*, **80**, 5–31.

Polyanskiy, Y. and Wu, Y. (2025) *Information Theory: From Coding to Learning.* Cambridge University Press, Cambridge.

Reeve, H. W. (2024) A short proof of the Dvoretzky–Kiefer–Wolfowitz–Massart inequality. *arXiv preprint*, arxiv:2403.16651.

Rényi, A. (1959) On measures of dependence. *Acta Math. Acad. Sci. Hungar.*, **10**, 441–451.

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M. and Sabeti, P. C. (2011) Detecting novel associations in large data sets. *Science*, **334**, 1518–1524.

Rubenstein, P., Bousquet, O., Djolonga, J., Riquelme, C. and Tolstikhin, I. O. (2019) Practical and consistent estimation of $f$-divergences. In *Advances in Neural Information Processing Systems.*

Samworth, R. J. and Shah, R. D. (2025+) *Modern Statistical Methods and Theory.* Cambridge University Press, Cambridge.

Sansalone, G., Wroe, S., Coates, G., Attard, M. R. and Fruciano, C. (2024) Unexpectedly uneven distribution of functional trade-offs explains cranial morphological diversity in carnivores. *Nature Communications*, **15**, 3275.

Schweizer, B. and Wolff, E. F. (1981) On nonparametric measures of dependence for random variables. *Ann. Statist.*, **9**, 879–885.

Sen, A. and Sen, B. (2014) Testing independence and goodness-of-fit in linear models. *Biometrika*, **101**, 927–942.

Sklar, M. (1959) Fonctions de répartition à n dimensions et leurs marges. *Annales de l'ISUP*, **8**, 229–231.

Skorski, M. (2023) Bernstein-type bounds for beta distribution. *Modern Stochastics: Theory and Applications*, **10**, 211–228.

Spearman, C. (1904) The proof and measurement of association between two things. *Amer. J. Psychol.*, **15**, 72–101.

Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D. and Futcher, B. (1998) Comprehensive identification of cell cycle–regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol. Biol. Cell*, **9**, 3273–3297.

Suo, C., Dann, E., Goh, I., Jardine, L., Kleshchevnikov, V., Park, J.-E., Botting, R. A., Stephenson, E., Engelbert, J., Tuong, Z. K. et al. (2022) Mapping the developing human immune system across organs. *Science*, **376**, eabo0510.

Suo, C., Polanski, K., Dann, E., Lindeboom, R. G., Vilarrasa-Blasi, R., Vento-Tormo, R., Haniffa, M., Meyer, K. B., Dratva, L. M., Tuong, Z. K. et al. (2024) Dandelion uses the single-cell adaptive immune receptor repertoire to explore lymphocyte developmental origins. *Nature Biotechnology*, **42**, 40–51.

Székely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007) Measuring and testing dependence by correlation of distances. *Ann. Statist.*, **35**, 2769–2794.

Tukey, J. W. (1975) Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians*, volume 2, 523–531.

Villani, C. (2021) *Topics in Optimal Transportation.* American Mathematical Society, Providence.

Wiesel, J. C. (2022) Measuring association with Wasserstein distances. *Bernoulli*, **28**, 2816–2832.

Zhang, K. (2019) BET on independence. *J. Amer. Statist. Assoc.*, **114**, 1620–1637.

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

Zuo, Y. and Serfling, R. (2000) General notions of statistical depth function. *Ann. Statist.*, **28**, 461–482.