Foundations of Data Science

doi:10.3934/fods.2024035

ON THE LIMITS OF TOPOLOGICAL DATA ANALYSIS FOR STATISTICAL INFERENCE

Siddharth Vishwanath[*1], Kenji Fukumizu[2], Satoshi Kuriki[2] and Bharath K. Sriperumbudur[3]

[1]University of California San Diego, 9500 Gilman Dr, La Jolla, CA 92093, USA

[2]The Institute of Statistical Mathematics, 10-3 Midoricho, Tachikawa, Tokyo 190-8562, Japan

[3]The Pennsylvania State University, University Park, PA 16802, USA

(Communicated by Henry Adams)

Abstract. Topological data analysis has emerged as a powerful tool for extracting the metric, geometric and topological features underlying the data as a multi-resolution summary statistic, and has found applications in several areas where data arises from complex sources. In this paper, we examine the use of topological summary statistics through the lens of statistical inference. We investigate necessary and sufficient conditions under which *valid statistical inference* is possible using topological summary statistics. Additionally, we provide examples of models that demonstrate invariance with respect to topological summaries.

1. **Introduction.** Let $\mathbb{X}_n = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n\}$ be a collection of points observed i.i.d. at random from a probability distribution $\mathbb{P}$ on $\mathcal{X} \subseteq \mathbb{R}^d$. The core objective of the statistical inference is to employ a test statistic $\mathbf{T}(\mathbb{X}_n)$ to infer meaningful information underlying the data-generating mechanism $\mathbb{P}$. A good choice of the statistic $\mathbf{T}(\cdot)$ sheds light on some population quantity of interest, $\psi(\mathbb{P})$, which encodes the essential information underlying the sample $\mathbb{X}_n$. For instance, choosing $\mathbf{T}(\mathbb{X}_n) = (X_1 + X_2 + \cdots + X_n)/n$ to be the sample mean sheds light on the population mean $\psi(\mathbb{P}) = \int_{\mathcal{X}} \boldsymbol{x} \cdot d\mathbb{P}(\boldsymbol{x})$. In particular, for two classes of probability distributions $\mathcal{P}_0$ and $\mathcal{P}_1$, a good choice of $\mathbf{T}(\cdot)$ for $\psi(\mathbb{P})$ enables a practitioner to test hypotheses $H_0 : \mathbb{P} \in \mathcal{P}_0$ vs. $H_1 : \mathbb{P} \in \mathcal{P}_1$ related to the data generating mechanism and to perform valid statistical inference.

Recent years have witnessed the availability of high-dimensional data from unconventional sources such as text and images, and it has become increasingly important to employ statistical summaries that encode the subtle features that underlie the data. To this end, topological data analysis (TDA) has emerged as an important tool for extracting geometric and topological features underlying the data. Summary statistics obtained using TDA are particularly attractive since they are multi-resolution summaries that encode the metric, geometric *and* topological features underlying the data. Therefore, topological summary statistics, e.g., Betti numbers [12], Euler characteristic curves [33], persistence functions [43] and persistence diagrams [2],

2020 *Mathematics Subject Classification.* Primary: 62R40, 62F30; Secondary: 94A16.

Key words and phrases. Statistical invariance, topological data analysis, Betti numbers.

*Corresponding author: Siddharth Vishwanath.

have been employed in statistical tests for data arising from complex sources such as neuroscience [11], cosmology [1] and proteomics [20]. Despite the widespread adoption of TDA in data-analysis pipelines, a formal framework for statistical inference using these statistics is still limited.

More precisely, topological summary statistics, represented as $\mathbf{T} : \mathcal{X}^n \to \mathcal{S}$, are measurable mappings (w.r.t. $\mathbb{P}^{\otimes n}$) such that the data $\mathbb{X}_n$ is mapped to an element $\mathbf{T}(\mathbb{X}_n) \in \mathcal{S}$ in a suitable topological summary space. However, several topological summaries are challenging to analyze in a formal statistical context owing to the unusual mathematical structure of the summary space $\mathcal{S}$. For example, the summary space for persistence diagrams is an Alexandrov space with non-negative curvature bounded from below [34]. Working with topological representations on such summary spaces $\mathcal{S}$ are not amenable for employing classical tools of statistical inference owing to the bounded curvature of geodesics, non-uniqueness of Fréchet means, and absence of any Hilbertian structure [13, 47].

In contrast, well-behaved topological summaries[1]—such as Betti numbers and persistent Betti numbers—have been studied extensively in a probabilistic context. Based on the central ideas from Penrose [38], the large sample behavior of $\mathbf{T}(\mathbb{X}_n)$ branches into three qualitatively different regimes: the *sparse*, *thermodynamic* and *dense* regimes which depend on the resolution at which the observations are examined relative to the number of samples $n \to \infty$. In particular, recent advances in the discipline have established the existence of limiting quantities $\mu(\mathbb{P})$ and $\Sigma(\mathbb{P})$ such that

$$\frac{\mathbf{T}(\mathbb{X}_n)}{n} \xrightarrow{p} \mu(\mathbb{P}), \quad \text{and} \quad \frac{\mathbf{T}(\mathbb{X}_n) - \mu(\mathbb{P})}{\sqrt{n}} \overset{d}{\rightsquigarrow} \mathcal{N}\big(\mathbf{0}, \Sigma(\mathbb{P})\big),$$

as $n \to \infty$ under a suitable asymptotic regime. For example, the law of large numbers and central limit theorem for the *Betti numbers* of random geometric complexes is established in [27, 28, 49, 50, 22, 46, 7, 5]. Similar results for the *persistent Betti numbers* of random geometric complexes are established in [26, 31, 10]. In a similar vein, by regarding the topological summaries as stochastic processes, analogous convergence results also hold in the Skorohod metric [45, 37, 30]. These results establish that studying fundamental topological quantities in a random setting guarantee stability in a probabilistic sense, and pave the way to more detailed statistical investigation. In particular, methods for incorporating topological summaries in a formal statistical inference setting are described in [18, 3, 40, 14]. Notwithstanding, it is tempting to ask:

When are these topological summaries $\mathbf{T}(\mathbb{X}_n)$ *meaningful for statistical inference?*

The *central objective* of this work is to investigate a variant of this question: *when are the topological summaries insufficient for statistical inference?* To this end, we characterize conditions under which the limiting distribution of $\mathbf{T}(\mathbb{X}_n)$ fails to be injective.

1.1. Contributions.

In the deterministic setting, even without imposing the probabilistic structure, the injectivity of metric spaces via topological transforms is largely an open problem [36]. Therefore, we make a foray into this question using a slightly simplified approach. Given $\mathbb{X}_n$ sampled i.i.d. from $\mathbb{P}$ where $\mathrm{supp}(\mathbb{P}) = \mathcal{X} \subset \mathbb{R}^D$ and $\dim(\mathcal{X}) = d \leq D$, and for $0 \leq s, t \leq \infty$ the topological summary we investigate are

[1] e.g., when $\mathcal{S} = \mathbb{R}^k$, or, more generally, when $\mathcal{S}$ is a Hilbert space or a Banach space.

the *persistent Betti numbers*, $\beta_k^{s,t}\big(\mathcal{K}\big(n^{1/d}\mathbb{X}_n\big)\big)$ for each $0 \le k \le d$, i.e.,

$$\mathbf{T}(\mathbb{X}_n) = \Big(\beta_0^{s,t}\big(\mathcal{K}(n^{1/d}\mathbb{X}_n)\big), \beta_1^{s,t}\big(\mathcal{K}(n^{1/d}\mathbb{X}_n)\big), \ldots, \beta_d^{s,t}\big(\mathcal{K}(n^{1/d}\mathbb{X}_n)\big)\Big) \in \mathbb{Z}_+^d,$$

where, as outlined in Section 2, $\mathcal{K}(n^{1/d}\mathbb{X}_n)$ corresponds to the Čech complex constructed using $\mathbb{X}_n$ in the thermodynamic regime. Examining the behavior of the Betti numbers collectively serves as a stepping-stone to understanding the behavior of more complex topological invariants in the context of persistent homology.

Although the exact sampling distribution for $\mathbf{T}(\mathbb{X}_n)$ is difficult to characterize, we can investigate its usefulness in the following asymptotic sense. If $\mathbf{T}(\mathbb{P}) = (\mu(\mathbb{P}), \Sigma(\mathbb{P}))$ characterizes the limiting distribution of $\mathbf{T}(\mathbb{X}_n)$ i.e.,

$$n^{-1}\mathbf{T}(\mathbb{X}_n) \xrightarrow{p} \mu(\mathbb{P}), \quad \text{and} \quad n^{-1/2}\big(\mathbf{T}(\mathbb{X}_n) - \mu(\mathbb{P})\big) \overset{d}{\rightsquigarrow} \mathcal{N}\big(0, \Sigma(\mathbb{P})\big),$$

in this paper we investigate conditions under which the map $\mathbb{P} \mapsto \mathbf{T}(\mathbb{P})$ fails to be injective, i.e., for two different probability distributions $\mathbb{P} \neq \mathbb{Q}$, we have that $\mathbf{T}(\mathbb{P}) = \mathbf{T}(\mathbb{Q})$.

At this point, it becomes instructive to consider a class of distributions $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ indexed by a parameter set Θ. A statistic $\mathbf{T}(\mathbb{X}_n(\theta))$ is said to be sufficient for the model $\mathcal{P}$ when the probabilistic and statistical information underlying the observations $\mathbb{X}_n(\theta) \sim \mathbb{P}_\theta$ is faithfully encoded in the statistic $\mathbf{T}(\mathbb{X}_n(\theta))$, and plays a central role in statistical inference[2]. The injectivity of $\mathbf{T}$ is closely related to the notion of *sufficiency* of $\mathbf{T}$. In particular, when injectivity fails, the limiting distribution of $\mathbf{T}(\mathbb{X}_n(\theta))$ ultimately provides no information about the parameter θ underlying $\mathbb{X}_n(\theta)$. This qualitative behavior, which is complementary to the notion of sufficiency, is called *ancillarity*; our results can be viewed as characterizing conditions for $\mathcal{P}$ under which the topological summary statistic $\mathbf{T}(\mathbb{X}_n(\theta))$ is *asymptotically ancillary* for the model $\mathcal{P}$.

Our main contributions are the following:

(I) We introduce the notion of β-equivalence which characterizes distributions for which the mapping $\mathbf{T}$ fails to be injective. We examine conditions under which a parametric class of distributions $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ admits β-equivalence, by introducing an alternate equivalence relationship, called $\mathcal{F}$-equivalence, whereby distributions satisfying $\mathcal{F}$-equivalence are also guaranteed to satisfy β-equivalence (Lemma 3.5). As a consequence, when $\mathbb{P}$ and $\mathbb{Q}$ admit $\mathcal{F}$-equivalence, any statistical test for distinguishing between $\mathbb{X}_n \sim \mathbb{P}$ and $\mathbb{Y}_m \sim \mathbb{Q}$ using $\mathbf{T}(\mathbb{X}_n), \mathbf{T}(\mathbb{Y}_m)$ has vanishing power as $n, m \to \infty$.

(II) By imposing an algebraic structure on $\mathcal{P}$, and using the notion of *group maximal invariance* [48, 15], in Theorem 4.5, we provide the necessary and sufficient conditions for the family $\mathcal{P}$ to admit $\mathcal{F}$-equivalence. We illustrate this result through some supporting examples in Section 4.

(III) Next, in Theorem 5.4 we relax the algebraic assumptions on $\mathcal{P}$ and investigate sufficient conditions for $\mathcal{F}$-equivalence when the underlying space $\mathcal{X}$ admits a smooth fiber bundle structure. In contrast to (II), our method here is more constructive and is illustrated through several examples in Section 5.

(IV) Lastly, in Theorem 5.14, we present a necessary and sufficient condition for $\mathcal{F}$-equivalence to hold based on the geometry of the score function $\nabla_\theta \log f_\theta(\boldsymbol{x})$.

[2]Specifically, $\mathbf{T}(\mathbb{X}_n)$ is sufficient for the model $\mathcal{P}$ if the joint distribution of $\mathbb{X}_n(\theta)$ admits the factorization $f_{\mathbb{X}_n}(\boldsymbol{x}_{1:n}; \theta) = h(\boldsymbol{x}) \cdot g(\theta, \mathbf{T}(\boldsymbol{x}_{1:n}))$

In contrast to (II) and (III), this result provides a simple method to verify if a given class of distributions, $\mathcal{P}$, admits $\mathcal{F}$-equivalence.

The rest of the paper is organized as follows. In Section 2, we provide a background on the probabilistic, topological and statistical tools needed. Sections 3, 4 and 5 contain the main results, and missing proofs are collected to Section 6. In Section 7 we discuss possible extensions, and collect the supplementary results in the Appendix.

2. Background.

In this section, we provide background on probabilistic, topological, and statistical tools needed for the rest of the paper. We use $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$ to denote random variables from a probability space $(\Omega, \mathcal{F}, \pi)$ taking values in a measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, and we use $\boldsymbol{X} \perp \boldsymbol{Y}$ to indicate that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent. We assume that the space on which the random values are observed, $\mathcal{X}$ is sufficiently regular, i.e., either $\mathcal{X} \subset \mathbb{R}^d$ contains an open subset of $\mathbb{R}^d$, or $\mathcal{X}$ is a compact $\mathcal{C}^1$-manifold of dimension $d < D$. For all probability distributions $\mathbb{P}$, defined on $\mathcal{X}$, we assume the existence of a Radon-Nikodym derivative $f = d\mathbb{P}/d\mu$ w.r.t. a canonical choice of the base measure on μ on $\mathcal{X}$, e.g., the d-dimensional Lebesgue measure, denoted λ_d, when $\mathcal{X}$ contains an open subset of $\mathbb{R}^d$. We further assume that $f \in L^p(\mathcal{X}, \mu)$ for all $1 \leq p < \infty$. The Jacobian of a differentiable function f is denoted by $\mathbf{D}f$ and $\|\mathbf{D}f\| \stackrel{\bullet}{=} |\det(\mathbf{D}f)|$ denotes the absolute value of its determinant.

2.1. Betti numbers and persistent homology.

Given a collection of observations $\mathbb{X}_n = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ in a metric space $(\mathcal{X}, d)$, and for a given spatial resolution $r > 0$, the topology underlying the points at resolution r is encoded in a simplicial complex $\mathcal{K}(\mathbb{X}_n, r) \subseteq 2^{\mathbb{X}_n}$. The simplicial complex can be constructed in several ways (see, for example, [16]). In particular, the Čech complex is given by

$$\mathcal{K}(\mathbb{X}_n, r) \stackrel{\bullet}{=} \left\{ \sigma \subseteq \mathbb{X}_n : \bigcap_{\boldsymbol{x} \in \sigma} B(\boldsymbol{x}, r) \neq \varnothing \right\}, \tag{1}$$

where $B(\boldsymbol{x}, r) = \left\{ \boldsymbol{y} \in \mathbb{R}^D : \|\boldsymbol{x} - \boldsymbol{y}\| < r \right\}$. For $0 \leq k \leq d$, the k^{th}-*homology* of a simplicial complex $\mathcal{K}$, given by $H_k(\mathcal{K})$ is an algebraic object encoding its topology as a vector-space (over a fixed field k, typically taken to be $\mathbb{Z}_2$; [25]). Using the Nerve lemma [8], $H_k(\mathcal{K}(\mathbb{X}_n, r))$ is isomorphic to the homology of its union of r-balls, $H_k(\bigcup_{i=1}^n B(\boldsymbol{x}_i, r))$. The k^{th}-*Betti number* is defined as

$$\beta_k(\mathcal{K}(\mathbb{X}_n, r)) \stackrel{\bullet}{=} \dim(H_k(\mathcal{K}(\mathbb{X}_n, r))).$$

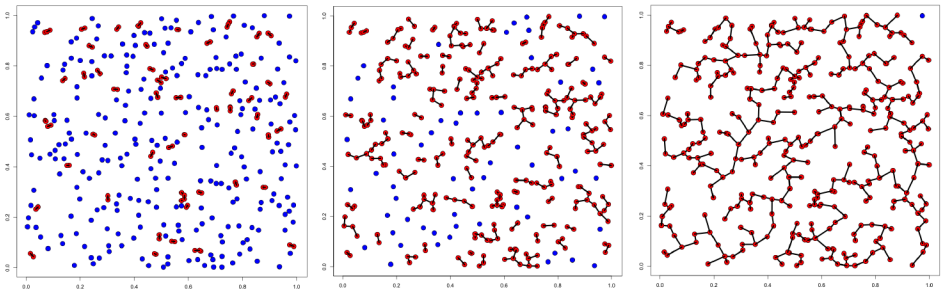
It counts the number of k-dimensional voids or *non-trivial k–cycles* in $\mathcal{K}(\mathbb{X}_n, r)$. The ordered sequence $\{\mathcal{K}(\mathbb{X}_n, r)\}_{r>0}$ forms a *filtration*, encoding the evolution of topological features over a spectrum of resolutions. For $0 < s < t$, $\bigcup_{i=1}^n B(\boldsymbol{x}_i, s) \subset \bigcup_{i=1}^n B(\boldsymbol{x}_i, t)$ and the simplicial complex $\mathcal{K}(\mathbb{X}_n, s)$ is a *sub-simplicial complex* of $\mathcal{K}(\mathbb{X}_n, t)$. Their homology groups are associated with the induced linear map

$$\iota_s^t : H_k(\mathcal{K}(\mathbb{X}_n, s)) \to H_k(\mathcal{K}(\mathbb{X}_n, t))$$

and the k^{th} order (r, s)-*persistent Betti number*, given by

$$\beta_k(\mathcal{K}(\mathbb{X}_n, s, t)) = \operatorname{rank}\left(\iota_s^t\right),$$

counts the number of non-trivial cycles which are born at or before s and have a death after t. The k^{th}-persistence diagram, denoted by $\mathsf{dgm}_k(\mathbb{X}_n)$, is defined as the collection $\{(b_i, d_i)\}_i$ of birth-death pairs associated with the non-trivial cycles from the filtration. We refer the reader to [25, 16] for a comprehensive introduction.



(A) Sparse regime (B) Thermodynamic regime (C) Dense regime

FIGURE 1. Illustration of different asymptotic regimes

2.2. **Asymptotic regimes.** A point process Φ is a locally finite, random counting measure defined on the measurable space $(\Omega, \mathcal{F})$ whose random elements take values in $\mathcal{X}$, and for a $B \in \mathcal{B}(\mathcal{X})$, the random variable $\Phi(B)$ measures the number of elements of Φ in B. $\mathbb{X}_n$ is called a Binomial point process if $\mathbb{X}_n = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n\}$ is i.i.d. with distribution $\mathbb{P}$. The Binomial point process can equivalently be represented by the random measure $\Phi_n = \sum_{i=1}^{n} \delta_{\boldsymbol{X}_i}$, from which it follows that $\Phi_n(B) = \text{Bin}(n, \mathbb{P}(B))$ for all $B \in \mathcal{B}(\mathcal{X})$.

In Section 2.1, the collection of points $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ were fixed points from a space $(\mathcal{X}, d)$. The probabilistic setting deals with a random collection of points $\mathbb{X}_n = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n\}$. The analysis of the asymptotic behavior of the topological quantities depends on how the radii of the balls for the Čech complex, r_n, decays relative to n. If the radii r_n decay too quickly, then the associated simplicial complexes fail to recognize the higher dimensional simplices, resulting in sparsely connected points. On the flip side, if the radii r_n decay too slowly, then all the points become densely connected. At a critical rate of decay for r_n one can observe a phase transition. This is illustrated in Figure 1.

Formally, the *thermodynamic* regime corresponds to the case when the expected number of points inside a ball of radius r_n has constant order, i.e., if $\mathbb{P} \ll \lambda_d$ then

$$\mathbb{E}\big(\Phi(B(\boldsymbol{x}, r_n))\big) = n\mathbb{P}(\boldsymbol{X} \in B(\boldsymbol{x}, r_n)) \propto n r_n^d \to c \quad \text{as} \quad n \to \infty,$$

whenever $r_n = cn^{-1/d}$. Faster rates of decay, $r_n = o(n^{-1/d})$, corresponds to the *sparse* regime, and slower rates of decay, $r_n = \omega(n^{-1/d})$, corresponds to the *dense* regime.

From the definition of the Čech complex in Eq. (1), it is easy to see that in the thermodynamic regime $r_n = tn^{-1/d}$, $\mathcal{K}(\mathbb{X}_n, r_n)$ is equivalently obtained by holding the radius fixed at t and rescaling the original points to get $\mathcal{K}(n^{1/d}\mathbb{X}_n, t)$. Specifically, taking $r_n = tn^{-1/d}$ if $\sigma \subset \mathbb{X}_n$ is a $k-$simplex of $\mathcal{K}(\mathbb{X}_n, r_n)$ then for $n^{1/d}\sigma \subset n^{1/d}\mathbb{X}_n$ it follows that

$$\bigcap_{\boldsymbol{y} \in n^{1/d}\sigma} B(\boldsymbol{y}, t) \neq \varnothing \quad \Leftrightarrow \quad \bigcap_{\boldsymbol{x} \in \sigma} B(\boldsymbol{x}, r_n) \neq \varnothing,$$

and, therefore, $\mathcal{K}(\mathbb{X}_n, r_n) \simeq \mathcal{K}(n^{1/d}\mathbb{X}_n, t)$ is a simplicial isomorphism. Similarly, when r_n, R_n are such that $r_n = sn^{-1/d}$ and $R_n = tn^{1/d}$ in the thermodynamic regime, the (r_n, R_n)-persistent Betti numbers are equivalently given by

$$\beta_k^{s,t}\Big(\mathcal{K}(n^{1/d}\mathbb{X}_n)\Big) \overset{\bullet}{=} \beta_k\Big(\mathcal{K}(n^{1/d}\mathbb{X}_n, s, t)\Big) \equiv \beta_k(\mathcal{K}(\mathbb{X}_n, r_n, R_n)).$$

3. **Injectivity of $\mathbf{T}(\mathbb{P})$, β-equivalence and $\mathcal{F}$-equivalence.** We begin by describing the setting in which we examine injectivity. Let $\mathbb{X}_n = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n\}$ be sampled i.i.d. from a probability distribution $\mathbb{P}$ with density f on $\mathcal{X} \subseteq \mathbb{R}^D$. The topological summary we are interested in is the collection of persistent Betti numbers for the random Čech complexes in the thermodynamic regime,

$$\mathbf{T}(\mathbb{X}_n) \stackrel{\bullet}{=} \Big(\beta_0^{s,t}(\mathcal{K}(n^{1/d}\mathbb{X}_n)), \beta_1^{s,t}(\mathcal{K}(n^{1/d}\mathbb{X}_n)), \ldots, \beta_d^{s,t}(\mathcal{K}(n^{1/d}\mathbb{X}_n))\Big),$$

for $0 \leq s < t$. The thermodynamic limit for each $\beta_k(\mathcal{K}(\mathbb{X}_n, r_n))$, $0 \leq k \leq d$ has been established in [28, 50, 46, 22, 31, 10]. In particular, we rely on the following characterization of the thermodynamic limit found in [31, Proposition 3.1 & Theorem 4.5] which follows from a slight restatement of the conditions in [22, Theorem 1.1].

Proposition 3.1 ([22, Theorem 1.1]; [31, Proposition 3.1 & Theorem 4.5])**.** *Let $\mathbb{X}_n \subset \mathcal{X}$ sampled i.i.d. from $\mathbb{P}$ with density f, such that $f \in L^p(\mu, \mathcal{X})$ for all $1 \leq p < \infty$. For fixed $0 \leq s < t \leq \infty$, there exist fixed functions γ_k, ς_k depending only on k such that*

$$\lim_{n\to\infty} \frac{\beta_k^{s,t}\big(\mathcal{K}(n^{1/d}\mathbb{X}_n)\big)}{n} \stackrel{\bullet}{=} \mu_k(\mathbb{P}; s, t) = \int_{\mathcal{X}} \gamma_k\Big(f(\boldsymbol{x})^{1/d}s, f(\boldsymbol{x})^{1/d}t\Big) f(\boldsymbol{x}) d\boldsymbol{x},$$

and

$$\frac{\beta_k^{s,t}\big(\mathcal{K}(n^{1/d}\mathbb{X}_n)\big) - \mu_k(\mathbb{P}; s, t)}{\sqrt{n}} \stackrel{d}{\rightsquigarrow} \mathcal{N}(0, \sigma_k^2(s, t)),$$

where $\sigma_k^2(s,t) \stackrel{\bullet}{=} \int_{\mathcal{X}} \varsigma_k(f(\boldsymbol{x})^{1/d}s, f(\boldsymbol{x})^{1/d}t) f(\boldsymbol{x}) d\boldsymbol{x}$.

Remark 3.2. While we focus on Betti numbers associated with random Čech complexes in the thermodynamic regime, as noted in [49, Section 1.3], the results in this paper will extend to the Vietoris-Rips, alpha and witness complexes as well. For example, the probabilistic results in [22, Theorem 1.1] and [46, Theorem 3.3] hinge on establishing moment bounds using the topological additivity property in [22, Lemma 2.2] and [46, Lemma 2.1] respectively—which extend to the aforementioned simplicial complexes. In particular, some general conditions an abstract simplicial complex needs to satisfy for the results to hold are described in [40, Section 4.2].

A key observation is to note that the limiting quantities in Proposition 3.1 can equivalently be written as the statistical functionals

$$\mu_k(\mathbb{P}; s, t) \stackrel{\bullet}{=} \int_{\mathcal{X}} \gamma_k\Big(f(\boldsymbol{x})^{1/d}s, f(\boldsymbol{x})^{1/d}t\Big) f(\boldsymbol{x}) d\boldsymbol{x},$$
$$\sigma_k^2(\mathbb{P}; s, t) \stackrel{\bullet}{=} \int_{\mathcal{X}} \varsigma_k(f(\boldsymbol{x})^{1/d}s, f(\boldsymbol{x})^{1/d}t) f(\boldsymbol{x}) d\boldsymbol{x},$$

where f is the density associated with $\mathbb{P}$. Therefore, the thermodynamic limit, $\beta_k(\mathbb{P}; s, t) \stackrel{\bullet}{=} \big(\mu_k(\mathbb{P}; s, t),\ \sigma_k^2(\mathbb{P}; s, t)\big)$, encodes the limiting behavior of $n^{-1/2}\beta_k^{s,t}(\mathcal{K}(n^{1/d}\mathbb{X}_n))$.

With this background, we define the notion of β_k-equivalence under which two distributions $\mathbb{P}, \mathbb{Q}$ admit the same thermodynamic limit $\beta_k(\mathbb{P}; s, t)$ for all $0 \leq s < t$. We are interested in families of distributions which admit β_k-equivalence for each k, i.e., $\beta_k(\mathbb{P}; s, t) = \beta_k(\mathbb{Q}; s, t)$ for each $k \geq 0$. We call such a family of distributions β-equivalent.

Definition 3.3 (β-equivalence)**.** Two distributions $\mathbb{P}, \mathbb{Q}$ are said to admit β_k-equivalence if $\beta_k(\mathbb{P}; s, t) = \beta_k(\mathbb{Q}; s, t)$ for all $0 \leq s < t$. Furthermore, $\mathbb{P}$ and $\mathbb{Q}$ are said to admit β-equivalence, denoted $\mathbb{P} \overset{\beta}{\sim} \mathbb{Q}$, if $\mathbb{P}$ and $\mathbb{Q}$ admit β_k-equivalence for all $k \geq 0$. Moreover, a family of distributions $\mathcal{P}$ admits β-equivalence if $\mathbb{P} \overset{\beta}{\sim} \mathbb{Q}$ for all $\mathbb{P}, \mathbb{Q} \in \mathcal{P}$.

Establishing β-equivalence for distributions directly is infeasible because the exact forms of the quantities γ_k and ς_k, as described in Proposition 3.1, are typically unknown. To circumvent this challenge, we introduce an alternative equivalence, termed $\mathcal{F}$-equivalence.

Definition 3.4 ($\mathcal{F}$-equivalence)**.** Consider two probability distributions $\mathbb{P}, \mathbb{Q}$ with probability density functions f and g, respectively. Let $\boldsymbol{X} \sim f$ and $\boldsymbol{Y} \sim g$ be two random variables with distribution $\mathbb{P}$ and $\mathbb{Q}$, respectively. Then, $\mathbb{P}$ and $\mathbb{Q}$ are said to be $\mathcal{F}$-equivalent, denoted $\mathbb{P} \overset{\mathcal{F}}{\sim} \mathbb{Q}$ (equivalently, $f \overset{\mathcal{F}}{\sim} g$) if $f(\boldsymbol{X}) \overset{d}{=} g(\boldsymbol{Y})$. A family of distributions $\mathcal{P}$ is said to admit $\mathcal{F}$-equivalence if $\mathbb{P} \overset{\mathcal{F}}{\sim} \mathbb{Q}$ for all $\mathbb{P}, \mathbb{Q} \in \mathcal{P}$.

The following result establishes the relationship between distributions that admit $\mathcal{F}$-equivalence and those that admit β-equivalence.

Lemma 3.5. *If $\mathcal{P}$ is a family of probability distributions that admit $\mathcal{F}$-equivalence, then $\mathcal{P}$ also admits β-equivalence.*

In other words, $\mathcal{F}$-equivalence is a sufficient condition for distributions to admit β-equivalence, i.e., distributions satisfying $\mathcal{F}$-equivalence are also guaranteed to satisfy β-equivalence.

Remark 3.6. We make the following observations regarding distributions which admit $\mathcal{F}$-equivalence.

(i) Let $\mathcal{P}(\Theta) = \{\mathbb{P}_{\boldsymbol{\theta}} : \theta \in \Theta\}$ be a parametric family of distributions. A statistic $\mathbf{T}(\mathbb{X}_n)$ is an *ancillary statistic* for the model $\mathcal{P}(\Theta)$ if the distribution of $\mathbf{T}(\mathbb{X}_n)$ (w.r.t. $\mathbb{P}_\theta^{\otimes n}$) does not depend on θ. In a similar vein, $\mathbf{T}(\mathbb{X}_n)$ is *approximately ancillary* for the model if the limiting distribution $\mathbf{T}(\mathbb{P}_\theta)$ does not depend on θ [42, Chapter 6.6]. This provides the following interpretation of $\mathcal{F}$-equivalence: if $\mathcal{P}(\Theta)$ admits $\mathcal{F}$-equivalence, then the topological summaries are approximately ancillary statistics for the model.

Given a fixed model $\mathcal{P}(\Theta)$, in general, there are no constructive techniques for determining ancillary statistics [32]. Our objective here is, however, somewhat complementary; given a fixed statistic $\mathbf{T}(\mathbb{X}_n)$, we investigate conditions under which the model $\mathcal{P}(\Theta)$ admits $T(\mathbb{X}_n)$ as an approximate ancillary statistic.

(ii) Notably, if we are given observations $\mathbb{X}_n(\theta) \sim \mathbb{P}_\theta$ from a β-equivalent family $\mathcal{P}(\Theta)$, any level-α hypothesis test for $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_1$ with $\Theta_0 \cap \Theta_1 = \varnothing$ using $\mathbf{T}(\mathbb{X}_n)$ as a test statistic will have negligible power. This is a simple consequence of the fact that, in the limit, for any rejection region $B \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned}\mathbb{P}_{H_1}\big(\mathbf{T}(\mathbb{X}_n) \in B\big) &\leq \mathbb{P}_{H_0}\big(\mathbf{T}(\mathbb{X}_n) \in B\big) + \Big|\mathbb{P}_{H_1}\big(\mathbf{T}(\mathbb{X}_n) \in B\big) - \mathbb{P}_{H_0}\big(\mathbf{T}(\mathbb{X}_n) \in B\big)\Big| \\ &\leq \alpha + o(1),\end{aligned}$$

where the last inequality follows from the fact that the α is the fixed type-I error rate and the second term follows from the fact that, due to $\mathcal{F}$-equivalence, as $n \to \infty$

the limiting distributions of $\mathbf{T}(\mathbb{X}_n)$ under the null and alternate hypothesis are the same, i.e., $\mathbf{T}(\mathbb{P}_{\theta_0}) \overset{d}{=} \mathbf{T}(\mathbb{P}_{\theta_1})$ for all $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta_1$.

We, however, emphasize that the approximate ancillarity of topological summaries $\mathbf{T}(\mathbb{X}_n)$ for such models does not invalidate their usefulness in statistical inference; it only limits their ability to be used as the primary test statistic when the model admits $\mathcal{F}$-equivalence. In fact, ancillary statistics play a pivotal role in deriving efficient and optimal testing procedures in the framework of conditional inference [17, 21].

(iii) Studying injectivity for Betti numbers collectively serves as a stepping-stone to understanding the behavior of several other topological invariants, e.g., the *Euler characteristic* is an important topological invariant and is given by the alternating sum of Betti numbers,

$$\chi(\mathcal{K}(\mathbb{X}_n, r)) \overset{\bullet}{=} \sum_{k=0}^{d} (-1)^k \beta_k(\mathcal{K}(\mathbb{X}_n, r)). \tag{2}$$

Unlike Betti numbers, the asymptotic behavior of the Euler characteristic of a random Čech complex exhibits interesting phenomena only in the thermodynamic regime [5, Corollary 4.2], i.e.,

$$\mathbf{T}_\chi(\mathbb{X}_n) = n^{-1/2} \chi\Big(\mathcal{K}(\mathbb{X}_n, tn^{-1/d})\Big).$$

By noting that $\beta_k(\mathcal{K}(\mathbb{X}_n, tn^{-1/d})) = \beta_k^{t,t}(\mathcal{K}(n^{1/d}\mathbb{X}_n))$ in Eq. (2) and invoking the continuous mapping theorem, it is easy to see that families of distributions which admit $\mathcal{F}$-equivalence in the thermodynamic regime will also admit identical asymptotic behavior of the Euler characteristic. In particular, this shows that the goodness-of-fit test proposed in [14] will fail to distinguish between β-equivalent point processes.

(iv) Furthermore, note that the persistence diagram, which (informally) is given by

$$\mathsf{dgm}_k(n^{1/d}\mathbb{X}_n) \overset{\bullet}{=} \Big\{(b_i, d_i) : \exists k\text{-dim feature with birth time } b_i \ \&\ \text{death time } d_i > b_i\Big\},$$

can be represented by a locally finite Radon measure ([26, 13])

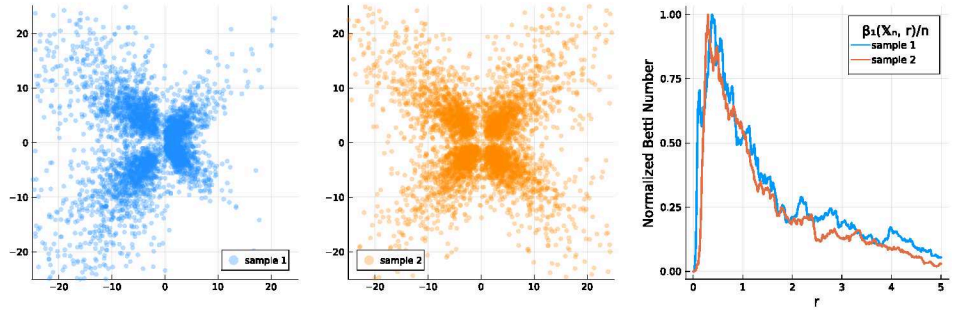
$$\Psi[\mathsf{dgm}(n^{1/d}\mathbb{X}_n)] = \sum_{(u,v)\in \mathsf{dgm}_k(n^{1/d}\mathbb{X}_n)} \delta_{(u,v)}.$$

Denoting by $B_{s,t} = [0, s] \times [t, \infty)$ the rectangular set in the upper-left half-plane, the persistent Betti numbers are equivalently given by

$$\begin{aligned}\beta_k^{s,t}(\mathcal{K}(n^{1/d}\mathbb{X}_n)) &= \Big\langle \mathbb{1}(B_{s,t}), \mathsf{dgm}_k(n^{1/d}\mathbb{X}_n)\Big\rangle \\ &= \int_{0\le u<v\le\infty} \mathbb{1}\big((u,v) \in B_{s,t}\big)\ d\Psi\big[\mathsf{dgm}(n^{1/d}\mathbb{X}_n)\big](u,v).\end{aligned}$$

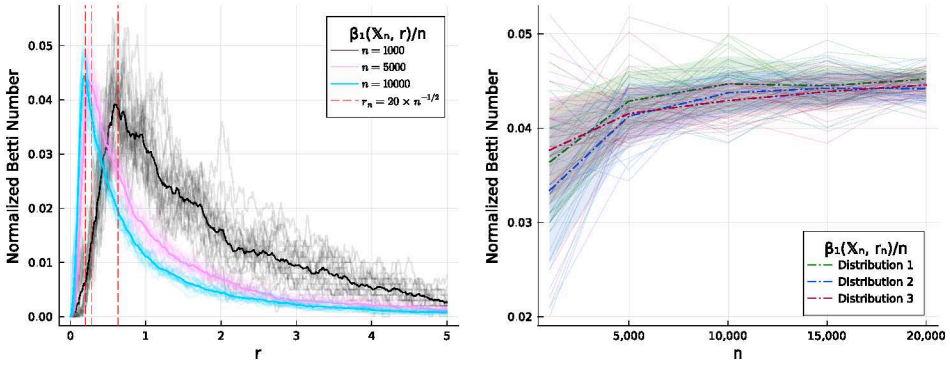
As a direct consequence, $\mathcal{F}$-equivalence also implies that the distribution of $\langle \phi, \mathsf{dgm}_k(n^{1/d}\mathbb{X}_n)\rangle$ is the same for all $\mathbb{P} \in \mathcal{P}$ and for all piecewise constant functions $\phi(\cdot)$ on the space of persistence diagrams.

We conclude this section with a numerical illustration to demonstrate the topological inference in a family of distributions admitting $\mathcal{F}$-equivalence (and, thereby, also β-equivalence).



(A) $\boldsymbol{\theta} = (0.46, 0.47, 0.03, 0.04)$ (B) $\boldsymbol{\theta} = (0.17, 0.29, 0.21, 0.24)$ (C) $r \mapsto \beta_1(\mathbb{X}_n, r)/n$

FIGURE 2. Scatterplot and Betti curve for point clouds $\mathbb{X}_n$ obtained from two different distributions in $\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$. (Left) $\mathbb{X}_n \sim f_{\boldsymbol{\theta}}$ for $\boldsymbol{\theta} = (0.46, 0.47, 0.03, 0.04)$ in blue, (Center) $\mathbb{X}_n \sim f_{\boldsymbol{\theta}}$ for $\boldsymbol{\theta} = (0.17, 0.29, 0.21, 0.24)$ in orange, and (Right) the (normalized) Betti curve $r \mapsto \beta_1(\mathbb{X}_n, r)/n$ for the two point clouds.



(A) $r \mapsto \beta_1(\mathbb{X}_n, r)/n$ for varying $\boldsymbol{\theta}$ but fixed n. (B) $n \mapsto \beta_1(\mathbb{X}_n, r_n)/n$ for fixed $\boldsymbol{\theta}$ and varying n.

FIGURE 3. Betti curves and the Betti numbers in the thermodynamic regime for $\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ from Example 3.7.

Example 3.7. Let $\Theta = \left\{(\theta_1, \theta_2, \theta_3, \theta_4) \in \mathbb{R}^4_+ : \theta_1 + \theta_2 + \theta_3 + \theta_4 = 4\right\}$, and let $f_{\boldsymbol{\theta}}$ be a probability density function on $\mathcal{X} = \mathbb{R}^2$, given by

$$f_\theta(x_1, x_2) = \begin{cases} g(\ \ x/\sqrt{\theta}_1, \ \ y/\sqrt{\theta}_1)/4, & \text{if } x \geq 0, y \geq 0 \\ g(-x/\sqrt{\theta}_2, \ \ y/\sqrt{\theta}_2)/4, & \text{if } x < 0, y \geq 0 \\ g(\ \ x/\sqrt{\theta}_3, -y/\sqrt{\theta}_3)/4, & \text{if } x \geq 0, y < 0 \\ g(-x/\sqrt{\theta}_4, -y/\sqrt{\theta}_4)/4, & \text{if } x < 0, y < 0 \end{cases}, \tag{3}$$

where $g \sim$ Biv-$\chi^2(10, 10, 0)$ is the probability density function of the bivariate χ^2 distribution with parameters $n = p = 10$ and $m = 0$ [24, Eq. 4.5]. In Example 5.13, we detail how $\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ admits $\mathcal{F}$-equivalence, and, therefore, from Lemma 3.5 also admits β-equivalence. Here, we examine the topological summaries of samples $\mathbb{X}_n$ obtained from this family of distributions. Figures 2 (A) and 2 (B) show the scatterplot of samples $\mathbb{X}_n$ obtained from two different distributions in this family.

The 1st order (normalized) Betti curves $r \mapsto \beta_1(\mathcal{K}(\mathbb{X}_n, r))/n$, in Figure 2 (C), show that that their topological summaries are almost indistinguishable.

In Figure 3 (A), for $k = 20$ random parametrizations $\{\boldsymbol{\theta}_i : 1 \leq i \leq k\}$ of $f_{\boldsymbol{\theta}}$, we plot the Betti curves $\mathsf{curve}_i : r \mapsto \beta_1(\mathbb{X}_n(\theta_i), r)/n$ for $\mathbb{X}_n(\boldsymbol{\theta}_i)$ when $n \in \{1000, 5000, 10000\}$. For fixed n, the (pointwise) mean Betti curve for the k individual curves is highlighted in bold. The plot shows that as the sample size n increases, the individual curves are harder to distinguish, and, shed light on the indistinguishability of the parameters in the asymptotic setting. In addition, the vertical dashed line in Figure 3 (A) highlights the region of maximum variability, which was empirically observed to correspond to $r_n = 20 \times n^{-1/2}$ in the thermodynamic regime.

To further investigate this region, we choose $k = 3$ random parametrizations $f_{\boldsymbol{\theta}_i}$ for $i \in \{1, 2, 3\}$. For n ranging from $1,000$ to $20,000$, and for each combination of n and $\boldsymbol{\theta}_i$ we generate 30 realizations of the samples from $f_{\boldsymbol{\theta}_i}$, i.e., $\mathbb{X}_n^{(j)}(\boldsymbol{\theta}_i)$ for $j \in \{1, 2, \ldots, 30\}$, and compute their normalized Betti numbers in the thermodynamic regime, $\mathsf{b}(n, i, j) := \beta_1(\mathcal{K}(\mathbb{X}_n^{(j)}(\boldsymbol{\theta}_i), r_n))/n$, with $r_n = 20 \times n^{-1/2}$. Figure 3 (B) shows the plot of n vs. $\mathsf{b}(n, i, j)$ for each combination of i and j. For each fixed $\boldsymbol{\theta}_i$, the mean curve is highlighted in a bold dotted line with the shaded region illustrating one standard deviation. The figure shows that the topological summaries, in this example, are insufficient to distinguish the three parameters in a formal statistical setting. The accompanying code for reproducing the experiments is available at https://github.com/sidv23/invariance. ♢

4. **$\mathcal{F}$-equivalence I: Algebraic perspective.** Having introduced the notion of $\mathcal{F}$-equivalence, and discussed its implications in the preceding section, a natural question arises: *when does a family of distributions admit $\mathcal{F}$-equivalence?* In this section, we impose an algebraic structure on the parametric model $\mathcal{P}(\Theta)$ in order to provide a general template for characterizing distributions that admit $\mathcal{F}$-equivalence. We begin by motivating the choice of imposing an algebraic structure by means of the following prototypical example.

Example 4.1 (Location and scale families)**.** For $\mathcal{X} = \mathbb{R}^d$ and a fixed density function $f_{\mathbf{0}}$ on $\mathbb{R}^d$, consider the location family of distributions, $\mathcal{P}_{loc} = \{f_{\boldsymbol{\theta}}(\boldsymbol{x}) = f_{\mathbf{0}}(\boldsymbol{x} - \boldsymbol{\theta}) : \boldsymbol{\theta} \in \mathbb{R}^d\}$, if $\boldsymbol{X}_\theta \sim f_{\boldsymbol{\theta}}$, by a standard transformation of random variables it follows that $(\boldsymbol{X}_\theta - \theta) \overset{d}{=} \boldsymbol{X}_{\mathbf{0}} \sim f_{\mathbf{0}}$. Therefore, $f_{\boldsymbol{\theta}}(\boldsymbol{X}_\theta) = f_{\mathbf{0}}(\boldsymbol{X}_\theta - \boldsymbol{\theta}) \overset{d}{=} f_{\mathbf{0}}(\boldsymbol{X}_{\mathbf{0}})$, which does not depend on θ. It follows that $\mathcal{P}_{loc}$ admits $\mathcal{F}$-equivalence.

However, the scale family of distributions, $\mathcal{P}_{\text{scale}} = \{g_\theta(\boldsymbol{x}) : \theta \in \mathbb{R}_+\}$, where $g_\theta(\boldsymbol{x}) = \theta^{-1} f_1(\boldsymbol{x}/\theta)$ for a fixed density function f_1, does not admit $\mathcal{F}$-equivalence. If $\boldsymbol{Y}_\theta \sim g_\theta$, then $\boldsymbol{Y}_\theta/\theta \overset{d}{=} \boldsymbol{Y}_1 \sim f_1$. But, the distribution of $g_\theta(\boldsymbol{Y}_\theta)$ is the same as the distribution of $\theta^{-1} f_1(\boldsymbol{Y}_1)$, which clearly depends on $\theta \in \mathbb{R}_+$. Therefore, $\mathcal{P}_{scale}$ does not admit $\mathcal{F}$-equivalence. ♢

It may be argued that it is natural to expect topological invariants to be insensitive to translations but sensitive to scaling and dilation. The next example illustrates a simple family of distributions which admit $\mathcal{F}$-equivalence.

Example 4.2 (Motivating example)**.** Let $\mathcal{X} = [0, 1]^2$ and $\Theta = [0, 2\pi]$, consider the family of distributions $\mathcal{P}(\Theta) = \{f_\theta : \theta \in \Theta\}$ given by

$$f_\theta(x, y) = \left(\cos\theta \cdot \Phi^{-1}(x) + \sin\theta \cdot \Phi^{-1}(y)\right)^2 \mathbb{1}((x, y) \in \mathcal{X}),$$

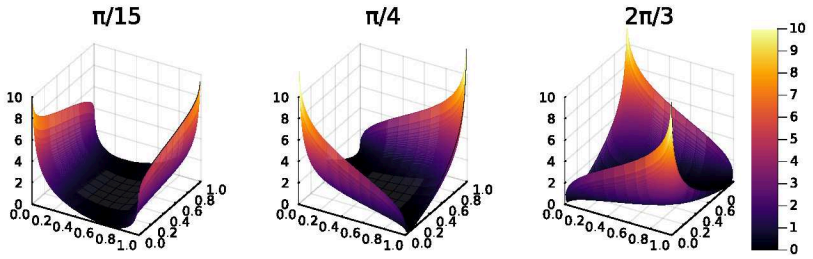


FIGURE 4. Illustration $f_\theta(x, y)$ from Example 4.2 for $\theta \in \{\pi/15, \pi/4, 2\pi/3\}$

where $\Phi^{-1}(\cdot)$ is the inverse CDF of a standard Normal distribution $\mathcal{N}(0, 1)$. See Figure 4 for an illustration. It is easy to verify that f_θ is a well-defined probability density function. Let $(X_\theta, Y_\theta) \sim f_\theta$ and consider the transformation $(x, y) \mapsto (\Phi^{-1}(x), \Phi^{-1}(y)) =: (u, v)$. A simple transformation shows that the density function for the transformed random vector $(U_\theta, V_\theta) = (\Phi^{-1}(X_\theta), \Phi^{-1}(Y_\theta))$ is given by

$$f_\theta(u, v) = (\cos\theta \cdot u + \sin\theta \cdot v)^2 (2\pi)^{-1} e^{-\frac{1}{2}(u^2+v^2)}.$$

By further transformation $(u, v) \mapsto \big(u\cos\theta + v\sin\theta, -u\sin\theta + v\cos\theta\big) =: (r, s)$, the density function for the random vector (R_θ, S_θ) becomes

$$f_\theta(r, s) = (2\pi)^{-1} r^2 e^{-r^2/2} \cdot e^{-s^2/2},$$

from which we can see that the distribution of (R_θ, S_θ) is free of θ and $R_\theta \perp S_\theta$. Therefore, by marginalizing (R_θ, S_θ) and then squaring, the desired distribution for $Z_\theta = f_\theta(X_\theta, Y_\theta) = R_\theta^2$ is

$$f_{Z_\theta}(z) = (z/2\pi)^{-1/2} e^{-z/2}.$$

Thus $Z_\theta \sim \Gamma(3/2, 1/2)$ for all $\theta \in \Theta$, and $\mathcal{P}(\Theta)$ admits $\mathcal{F}$-equivalence. ♢

A key step in establishing $\mathcal{F}$-equivalence in Example 4.2 is showing that the distribution of (R_θ, S_θ) is free of θ. In particular, the map $(u, v) \mapsto (r, s)$ can be represented by the action of an element of the rotation group, i.e., $(r, s) = g_\theta(u, v)$, where $g_\theta \in S\mathcal{O}(2)$; and the $(u^2 + v^2)$ term in $f_\theta(u, v)$ ensures that the action of g_θ remains constant on orbits of g_θ. This suggests that the algebraic structure underlying the family of distributions (if any) plays a key role in establishing their invariance. Lie groups provide a natural framework for studying such properties. We begin by reviewing some basic facts about group invariance in statistics. The presentation here closely follows the monographs by Wijsman [48] and Eaton [15].

Given a group $\mathcal{G} \overset{\bullet}{=} (\mathcal{G}, *)$ acting on a space $\mathcal{X}$, its *orbit* is given by $\mathcal{G}\boldsymbol{x} = \{g\boldsymbol{x} : g \in \mathcal{G}\}$; the *stabilizer* with respect to $\boldsymbol{x}$ is $\mathcal{G}_{\boldsymbol{x}} = \{g \in \mathcal{G} : g\boldsymbol{x} = \boldsymbol{x}\}$. When $\mathcal{G}$ acts bijectively on $\mathcal{X}$, a function $T : \mathcal{X} \to \mathcal{T}$ is $\mathcal{G}$-invariant if $T(\boldsymbol{x}) = T(g\boldsymbol{x})$ for all $g \in \mathcal{G}$ and a space $\mathcal{T}$.

Definition 4.3 (Maximal Invariant)**.** A function $T : \mathcal{X} \to \mathcal{T}$ is $\mathcal{G}$-maximal invariant if it is constant on orbits, i.e., $T(g\boldsymbol{x}) = T(\boldsymbol{x})$ for each $g \in \mathcal{G}$; and, it takes different values on different orbits, i.e., for each $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ such that $T(\boldsymbol{x}) = T(\boldsymbol{y})$, we have that $\boldsymbol{y} \in \mathcal{G}\boldsymbol{x}$.

It follows from this definition that if $J : \mathcal{T} \to \mathcal{J}$, from the space $\mathcal{T}$ to the space $\mathcal{J}$, is any injective map, then the composition $J \circ T$ will also be $\mathcal{G}$-maximal invariant. Thus,

maximal invariants for a group $\mathcal{G}$ are unique up to injective transformations. The relationship between a $\mathcal{G}$-invariant function and $\mathcal{G}$-maximal invariance is described in the following result.

Proposition 4.4 ([15]). *Suppose $T : \mathcal{X} \to \mathcal{T}$ is $\mathcal{G}$-maximal invariant. A function $\phi : \mathcal{X} \to \mathcal{Y}$ is $\mathcal{G}$-invariant if and only if there exists $\kappa : \mathcal{T} \to \mathcal{Y}$ such that $\phi = \kappa \circ T$.*

If $\Psi : \mathcal{X} \to \mathcal{Y}$ is a surjective function taking elements from $\mathcal{X}$ to the space $\mathcal{Y}$, then the action of $g \in \mathcal{G}$ on elements in $\mathcal{X}$ *induces* an action on elements in $\mathcal{Y}$ via Ψ. The *induced action* of $\mathcal{G}$ on $\mathcal{Y}$ is given by $g\boldsymbol{y} \stackrel{\bullet}{=} \Psi(g\boldsymbol{x})$ for every $\boldsymbol{x} \in \Psi^{-1}(\boldsymbol{y})$. This is well-defined whenever $\Psi(\boldsymbol{x}_1) = \Psi(\boldsymbol{x}_2)$ implies that $\Psi(g\boldsymbol{x}_1) = \Psi(g\boldsymbol{x}_2)$ for each $g \in \mathcal{G}$. We say that $\Psi : \mathcal{X} \to \mathcal{Y}$ is $\mathcal{G}$-*compatible* when the induced action of $\mathcal{G}$ on $\mathcal{Y}$ is well-defined.

With this background, let $\mathbb{P}$ be a fixed distribution on $\mathcal{X}$ and $\boldsymbol{X} \sim \mathbb{P}$. The action of an element $g \in \mathcal{G}$ on $\mathcal{X}$ induces a transformation on $\boldsymbol{X}$ to a new random variable $g\boldsymbol{X}$ taking values in $(g\mathcal{X}, \mathcal{B}(g\mathcal{X}))$. With a slight abuse of notation, let $g_{\#}\mathbb{P}$ denote the distribution of $g\boldsymbol{X}$. If the elements of $\mathcal{G}$ are indexed by a parameter $\theta \in \Theta$, i.e., $\mathcal{G}(\Theta) = \{g_\theta : \theta \in \Theta\}$, then the action of $\mathcal{G}$ on $\boldsymbol{X}$ induces a family of distributions $\mathcal{P}(\Theta) = \{\mathbb{P}_\theta = g_{\theta\#}\mathbb{P} : \theta \in \Theta\}$. The following result establishes necessary and sufficient conditions for $\mathcal{P}(\Theta)$ to admit $\mathcal{F}$-equivalence for a wide class of distributions.

Theorem 4.5 (Group Invariance). *Suppose $\Psi : \mathcal{X} \to \mathcal{Y}$ is differentiable and bijective, $\mathcal{G}(\Theta)$ is a $\mathcal{C}^1$-group of isometries acting on $\mathcal{Y}$, and $T : \mathcal{Y} \to \mathcal{T}$ is $\mathcal{G}(\Theta)$-maximal invariant. Define the family of distributions $\mathcal{P}(\Theta) = \{f_\theta : \theta \in \Theta\}$ on $\mathcal{X}$ by*

$$f_\theta(\boldsymbol{x}) = \phi(g_\theta \circ \Psi(\boldsymbol{x})),$$

where $g_\theta \in \mathcal{G}(\Theta)$ and $\phi : \mathcal{X} \to \mathbb{R}_{\geq 0}$ is some function which ensures that f_θ is a valid density. Then:

(i) $\mathcal{P}(\Theta)$ admits $\mathcal{F}$-equivalence if and only if there exists $\zeta : \mathcal{T} \to \mathbb{R}$ such that

$$\det\big(\mathbf{D}\Psi^{-1}(\boldsymbol{y})\big) = \zeta(T(\boldsymbol{y})).$$

(ii) If $\mathcal{G}(\Theta) = \times_{i=1}^m \mathcal{G}_i(\Theta_i)$ where $\Theta = \Theta_1 \times \Theta_2 \times \cdots \times \Theta_m$, then $\mathcal{P}(\Theta)$ admits $\mathcal{F}$-equivalence if and only if there exists a sequence of $\mathcal{G}_i(\Theta_i)$-compatible functions $T_i : \mathcal{Y}_{i-1} \to \mathcal{Y}_i$ with $\mathcal{Y}_0 = \mathcal{Y}$ and a function $\zeta : \mathcal{Y}_m \to \mathbb{R}$ such that each T_i is $\mathcal{G}_i(\Theta_i)$-maximal invariant and

$$\det\big(\mathbf{D}\Psi^{-1}(\boldsymbol{y})\big) = \zeta(T_m \circ T_{m-1} \circ \cdots \circ T_1(\boldsymbol{y})).$$

We defer the proof to Section 6.2 and illustrate some examples of Theorem 4.5 in the remainder of the section. We begin by illustrating a multivariate generalization of Example 4.2 in the context of Theorem 4.5.

Example 4.6. Let $\xi \sim F$ be random variable on $\mathbb{R}$ with density $f(x) = F'(x)$ such that $\mathbb{E}(\xi) = 0$ and $\mathbb{E}(\xi^2) = 1$. For $\mathcal{X} = [0,1]^d$ and $\Theta \in \mathbb{S}^{d-1}$, let $\mathcal{P}(\Theta) = \{f_{\boldsymbol{\theta}} : \theta \in \Theta\}$ be given by

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \big(\boldsymbol{\theta}^\top \boldsymbol{F}^{-1}(\boldsymbol{x})\big)^2 \mathbb{1}(\boldsymbol{x} \in \mathcal{X}), \tag{4}$$

where, for brevity, $\boldsymbol{F}^{-1}(\boldsymbol{x}) \stackrel{\bullet}{=} \big(F^{-1}(x_1), F^{-1}(x_2), \ldots, F^{-1}(x_d)\big)^\top \in \mathbb{R}^d$ is the "vectorized" inverse CDF of F. Then, the family of distributions $\mathcal{P}(\Theta)$ admits $\mathcal{F}$-equivalence if and only if $\xi \sim \mathcal{N}(0,1)$.

First note that f_θ is a well-defined density. To see this, taking $\boldsymbol{y} = \boldsymbol{F}^{-1}(\boldsymbol{x})$, we have

$$\begin{aligned}\int_{\mathcal{X}} f_\theta(\boldsymbol{x})d\boldsymbol{x} &= \int_{\mathbb{R}^d} (\theta^\top \boldsymbol{y})^2 \prod_{i=1}^d f(y_i)\ d\boldsymbol{y} \\ &= \sum_{j=1}^d \sum_{k=1}^d \int_{\mathbb{R}^d} \theta_j\theta_k y_j y_k \prod_{i=1}^d f(y_i)\ d\boldsymbol{y} \\ &= \sum_{j=1}^d \int_{\mathbb{R}^d} \theta_j^2 y_j^2 \prod_{i=1}^d f(y_i)\ d\boldsymbol{y} + \sum_{1\le j<k\le d} \int_{\mathbb{R}^d} 2\cdot\theta_j\theta_k y_j y_k \prod_{i=1}^d f(y_i)\ d\boldsymbol{y} \\ &\overset{\text{(i)}}{=} \sum_{j=1}^d \theta_j^2 + 0 = 1,\end{aligned}$$

where (i) uses the fact that $\mathbb{E}(\xi) = 0$ and $\mathbb{E}(\xi^2) = 1$.

Let $\mathcal{G} = S\mathcal{O}(d) = \left\{g_\theta \in GL(\mathbb{R}, d) : g_\theta^{-1} = g_\theta^\top, g_\theta^\top e_1 = \theta \in \mathbb{S}^{d-1}\right\}$ be the group of rotations on $\mathcal{Y} = \mathbb{R}^d$ where $e_1 = (1, 0, \ldots, 0)^\top \in \mathbb{R}^d$. The function $T(\boldsymbol{y}) = \|\boldsymbol{y}\|^2$ is $\mathcal{G}$-maximal invariant since $T(\boldsymbol{y}_1) = T(\boldsymbol{y}_2)$ if and only if $\boldsymbol{y}_2 = g_\theta \boldsymbol{y}_1$ for some $g_\theta \in \mathcal{G}$. As per Theorem 4.5, Eq. (4) is equivalently written as $f_\theta(\boldsymbol{x}) = \phi(g_\theta \circ \Psi(\boldsymbol{x}))$ where

$$\Psi(\boldsymbol{x}) = \boldsymbol{F}^{-1}(\boldsymbol{x}), \quad \textit{and} \quad \phi(\boldsymbol{y}) = (\boldsymbol{y}^\top e_1)^2.$$

The Jacobian for Ψ^{-1} is given by $\left\|\mathbf{D}\Psi^{-1}(\boldsymbol{y})\right\| = \prod_{i=1}^n f(y_i)$. From Theorem 4.5 it follows that $\mathcal{P}(\Theta)$ admits $\mathcal{F}$-equivalence if and only if there exists some function $\zeta_d : \mathbb{R}_+ \to \mathbb{R}_+$, which may implicitly depend on d, such that $\prod_{i=1}^n f(y_i) = \zeta_d(\|\boldsymbol{y}\|^2)$. We show, using Lemma A.1 in Appendix A, that this is satisfied only when $\xi \sim \mathcal{N}(0, 1)$. When $d = 2$, this recovers the family of distributions illustrated in Example 4.2. $\diamondsuit$

To illustrate a family of distributions generated by a subgroup of transformations, consider the following variant of Example 4.6.

Example 4.7. For $p + q = d$, consider $\mathcal{G} = S\mathcal{O}(p) \times S\mathcal{O}(q)$ acting as a subgroup of $S\mathcal{O}(d)$ on $\mathcal{Y} = \mathbb{R}^d$, i.e., for every $g_\theta \in \mathcal{G}$ it follows that $g_\theta^\top e_1 = \theta_p \oplus \theta_q$ where $\theta_p \in \mathbb{S}^{p-1}$ and $\theta_q \in \mathbb{S}^{q-1}$. Let $\boldsymbol{\Phi}_p^{-1} : [0,1]^p \to \mathbb{R}^p$ and $\boldsymbol{\Phi}_q^{-1} : [0,1]^q \to \mathbb{R}^q$ be the "vectorized" inverse CDFs of $\mathcal{N}(0, \sigma_p^2)$ and $\mathcal{N}(0, \sigma_q^2)$. For $\mathcal{X} = [0,1]^d$, consider $\mathcal{P}(\Theta) = \left\{f_\theta : \theta \in \mathbb{S}^{p-1} \times \mathbb{S}^{q-1}\right\}$ given by

$$f_\theta(\boldsymbol{x}) = \kappa(p,q)\left(\theta_p^\top \boldsymbol{\Phi}_p^{-1}(\boldsymbol{x}_p) + \theta_q^\top \boldsymbol{\Phi}_q^{-1}(\boldsymbol{x}_q)\right)^2 \mathbb{1}(\boldsymbol{x}_p \oplus \boldsymbol{x}_q \in \mathcal{X}), \tag{5}$$

where $\boldsymbol{x} = \boldsymbol{x}_p \oplus \boldsymbol{x}_q$ and $\kappa(p,q)$ is a fixed normalizing constant free of θ. As before, $f_\theta(\boldsymbol{x}) = \phi(g_\theta \circ \Psi(\boldsymbol{x}))$ where

$$\Psi(\boldsymbol{x}) = \left(\boldsymbol{\Phi}_p^{-1}(\boldsymbol{x}_p), \boldsymbol{\Phi}_q^{-1}(\boldsymbol{x}_q)\right),$$

and

$$\left\|\mathbf{D}\boldsymbol{\Psi}^{-1}(\boldsymbol{y})\right\| = \frac{1}{(2\pi)^{d/2}\sigma_p^p\sigma_q^q} \exp\left(-\frac{\|\boldsymbol{y}_p\|^2}{2\sigma_p^2} - \frac{\|\boldsymbol{y}_q\|^2}{2\sigma_q^2}\right).$$

In the context of Theorem 4.5, let $\mathcal{Y}_1 = \mathbb{R}^{p+2}$ and $\mathcal{Y}_2 = \mathbb{R}^2$, and consider $T_1 : \mathcal{Y} \mapsto \mathcal{Y}_1$ and $T_2 : \mathcal{Y}_1 \mapsto \mathcal{Y}_2$ given by

$$T_1 : \boldsymbol{y}_p \oplus \boldsymbol{y}_q \mapsto \boldsymbol{y}_p \oplus \|\boldsymbol{y}_q\|^2, \quad \textit{and} \quad T_2 : \boldsymbol{y}_p \oplus z \mapsto \|\boldsymbol{y}_p\|^2 \oplus z.$$

Clearly, T_1 is $S\mathcal{O}(q)$-maximal invariant and T_2 is a maximal invariant w.r.t. the induced action of $S\mathcal{O}(p)$ on $\mathcal{Y}_1$. Therefore, for $\zeta : \mathbb{R}^2 \to \mathbb{R}_+$ given by

$$\zeta(x,y) = \frac{1}{(2\pi)^{d/2}\sigma_p^p\sigma_q^q} \exp\left(-\frac{x}{2\sigma_p^2} - \frac{y}{2\sigma_q^2}\right),$$

it follows that $\left\|\mathbf{D}\boldsymbol{\Psi}^{-1}(\boldsymbol{y})\right\| = \zeta(T_2 \circ T_1(\boldsymbol{y}))$, implying that $\mathcal{P}(\Theta)$ admits $\mathcal{F}$-equivalence. ♢

The preceding two examples provide an alternate characterization of the Normal distribution through the lens of $\mathcal{F}$-equivalence w.r.t. rotational transformations. The next example illustrates $\mathcal{F}$-equivalence for a family of distributions generated by an unconventional group of transformations.

Example 4.8. For a fixed shape parameter $\kappa = 0.5$, consider the family of bivariate Weibull distributions on $\mathbb{R}^2_{++}$ given by

$$f_\theta(x,y) = \frac{1}{4\sqrt{xy}} \exp\left(-\theta\sqrt{x} - \frac{\sqrt{y}}{\theta}\right).$$

For $\Theta = \mathbb{R}_+$, it follows that $\mathcal{P}(\Theta) = \{f_\theta(x,y) : \theta \in \Theta\}$ admits $\mathcal{F}$-equivalence. We verify this as per Theorem 4.5. The functions Ψ and ϕ are $\Psi(x,y) = (\sqrt{x}, \sqrt{y})$, and $\phi(x,y) = \exp(-(x+y))/4xy$. The group action may be identified as follows: let $\mathcal{G}$ be a subgroup of $GL(\mathbb{R}, 2)$ consisting of elements

$$g_\theta = \begin{pmatrix} \theta & 0 \\ 0 & 1/\theta \end{pmatrix},$$

for $\theta \in \mathbb{R}_+$. It is easy to verify that $T(x,y) = xy$ is $\mathcal{G}$-maximal invariant. The density can now be expressed in the form $f_\theta(x,y) = \phi(g_\theta \circ \Psi(x,y))$. It follows that $\Psi^{-1}(x,y) = (x^2, y^2)$ and

$$\det\big(\mathbf{D}\Psi^{-1}(x,y)\big) = 4xy = 4T(x,y).$$

Hence, by Theorem 4.5, $\mathcal{P}(\Theta)$ admits $\mathcal{F}$-equivalence. ♢

5. **$\mathcal{F}$-equivalence II: General cases.** While the results in Section 4 examined necessary and sufficient conditions for $\mathcal{F}$-equivalence by enforcing some algebraic structure on the family of distributions, the objective of this section is to relax these requirements and, instead, exploit the structure underlying the support of the distributions, $\mathcal{X}$. Before we present the main results, we introduce the main tools we employ:

(i) The *modular character of a measure* μ, denoted by Ψ_μ, and
(ii) A fiber bundle representation $\mathscr{E} = (\mathcal{X}, \mathcal{Z}, \mathcal{Y}, \pi, G)$ of the underlying space $\mathcal{X}$.

Modular Character. Consider the set of diffeomorphisms $\Delta(\mathcal{X})$, given by

$$\Delta(\mathcal{X}) \overset{\bullet}{=} \{\phi \in \mathrm{Diff}(\mathcal{X}) : \|\mathbf{D}\phi(\boldsymbol{x})\| = \|\mathbf{D}\phi(\boldsymbol{x}')\|,\ \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}\}.$$

In other words, $\Delta(\mathcal{X})$ comprises of smooth maps from $\mathcal{X}$ to itself such that the Jacobian of the map does not depend on the specific location where the transformation is made, e.g., when $\mathcal{X} = \mathbb{R}^d$ the set $\Delta(\mathcal{X})$ is the group of rigid transformations on $\mathbb{R}^d$, $E(d)$. The elements of $\Delta(\mathcal{X})$ form a subgroup of transformations with respect to $\mathrm{Diff}(\mathcal{X})$. The change in measure induced by diffeomorphic transformations of a space $\mathcal{X}$ w.r.t. the measure μ is given by its modular character.

Definition 5.1 (Modular Character)**.** Given a measure μ on the space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, a function Ψ is defined to be the modular character of μ if for each $\phi \in \Delta(\mathcal{X})$ with $\boldsymbol{y} = \phi(\boldsymbol{x})$ we have that

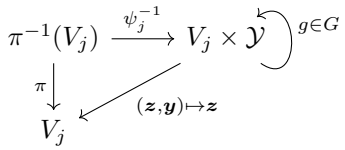
$$\mu(d\boldsymbol{y}) = \Psi(\|\mathbf{D}\phi\|)\mu(d\boldsymbol{x}).$$

For example, when $\mathcal{X} = \mathbb{R}^d$ and $\mu = \lambda_d$, the d-dimensional Lebesgue measure, the modular character of μ is given by $\Psi(x) = x$. Observe that for any full-rank linear map $\phi \in GL(\mathbb{R}, d)$ such that $\boldsymbol{y} = \phi(\boldsymbol{x})$, we have $\mu(d\boldsymbol{y}) = d\boldsymbol{y} = \|\mathbf{D}\phi\|d\boldsymbol{x} = \|\mathbf{D}\phi\|\mu(d\boldsymbol{x})$. A nontrivial example of the modular character is illustrated in Example 5.7.

Remark 5.2. The modular character is closely related to the notion of tensor-density (see, for instance, [41]). When $G \leq \Delta(\mathcal{X})$ is a locally compact group continuously acting on $\mathcal{X}$ from the left, i.e., $(g, \boldsymbol{x}) \mapsto g \cdot x$ for every $g \in G$, then Ψ_μ is also called the relatively invariant multiplier [15].

Fiber Bundle Representation. Following the convention in [44], when $\mathcal{X}$ is a compact d-dimensional $\mathcal{C}^1$-manifold, suppose $\mathcal{X}$ admits a fiber bundle representation $\mathscr{E} \overset{\bullet}{=} (\mathcal{X}, \mathcal{Z}, \mathcal{Y}, \pi, G)$, i.e., $\mathcal{X}$ is the total space, $\mathcal{Z}$ is the base space with the bundle projection π defined by the continuous surjective map $\pi : \mathcal{X} \to \mathcal{Z}$, and G is a topological group[3] which acts on the canonical fiber space $\mathcal{Y}$. A collection $\{(V_j, \psi_j) : j \in J\}$ is called a *local trivialization* of $\mathscr{E}$ and serves as the coordinate charts for the fiber bundle, i.e., $\{V_j : j \in J\}$ is an open cover of $\mathcal{Z}$ and for each $j \in J$ the map $\psi_j : V_j \times \mathcal{Y} \to \pi^{-1}(V_j)$ is a diffeomorphism which guarantees that locally, in the neighborhood $V_j \subset \mathcal{Z}$, the fiber $\pi^{-1}(V_j)$ looks like the product $V_j \times \mathcal{Y}$. In particular, the map $\psi_{j,\boldsymbol{z}} = \psi_j(\boldsymbol{z}, \cdot) : \mathcal{Y} \to \mathcal{Y}_{\boldsymbol{z}} := \pi^{-1}(\boldsymbol{z})$ is a diffeomorphism for each $\boldsymbol{z} \in V_j \subseteq \mathcal{Z}$.

Furthermore, for every $i, j \in J$ and all $\boldsymbol{z} \in V_i \cap V_j$, the map $\psi_{j,\boldsymbol{z}}^{-1} \circ \psi_{i,\boldsymbol{z}} : \mathcal{Y} \to \mathcal{Y}$ should coincide with an element $g_{ji,\boldsymbol{z}}$ of the structure group G, and $\mathscr{E}$ is represented by the following commutative diagram:



From [44, Section 2.4], if $\{(U_i, \eta_i) : i \in I\}$ is another local trivialization for $\mathscr{E}$, then the two local trivializations are *equivalent* in the sense that for every $\boldsymbol{z} \in U_i \cap V_j$ there exists $g_{ji,\boldsymbol{z}} \in G$ such that

$$g_{ji,\boldsymbol{z}} = \psi_{j,\boldsymbol{z}}^{-1} \circ \eta_{i,\boldsymbol{z}}. \tag{6}$$

In other words, the structure group G determines the change of coordinates from $\{(V_j, \psi_j) : j \in J\}$ to $\{(U_i, \eta_i) : i \in I\}$.

If the base space $\mathcal{Z}$ and the fiber $\mathcal{Y}$ are endowed with measures μ and ν respectively, they induce a *local product measure* $\lambda = \nu \otimes_{loc} \mu$ on the space $\mathcal{X}$ [23]. Specifically, for a measurable set $A \subset \mathcal{X}$ such that $\pi(A) \subset V_j \subseteq \mathcal{Z}$, the induced measure λ is given by

$$\lambda(A) = (\nu \otimes_{\text{loc}} \mu)(\psi_j^{-1}(A)) = \int_{\pi(A)} \nu_{\boldsymbol{z}}(\pi^{-1}(\boldsymbol{z}) \cap A)\mu(d\boldsymbol{z}), \tag{7}$$

[3]We use G to distinguish the structure group, in this section, from the groups, $\mathcal{G}$, in Section 4.

where for $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{z} = \pi(\boldsymbol{x}) \in V_j \subseteq \mathcal{Z}$, the measure $\nu_{\boldsymbol{z}} \stackrel{\bullet}{=} (\psi_{j,\boldsymbol{z}})_{\#}\nu$ is the pushforward of ν on the space $\pi^{-1}(\boldsymbol{z})$ via the diffeomorphism $\psi_{j,\boldsymbol{z}}$, i.e., $\nu_{\boldsymbol{z}}(B) = \nu(\psi_{j,\boldsymbol{z}}^{-1}(B))$ for all measurable $B \subseteq \pi^{-1}(\boldsymbol{z})$. Furthermore, for $f \in L^1(\mathcal{X}, \lambda)$, the following version of Fubini's theorem holds [23]:

$$\int_{\mathcal{X}} f(\boldsymbol{x})\lambda(d\boldsymbol{x}) = \int_{\mathcal{Z}} \int_{\pi^{-1}(\boldsymbol{z})} f(\boldsymbol{w})\ \nu_{\boldsymbol{z}}(d\boldsymbol{w})\mu(d\boldsymbol{z}). \tag{8}$$

From [23, Theorem 1], the measure λ in Eq. (7) exists only when ν is G-invariant. To see this, consider two local trivializations $\{(V_j, \psi_j) : j \in J\}$ and $\{(U_i, \eta_i) : i \in I\}$; for $\boldsymbol{z} \in U_i \cap V_j$ and a measurable set $B \subseteq \pi^{-1}(\boldsymbol{z})$ from the fiber over $\boldsymbol{z}$, the induced measure $\nu_{\boldsymbol{z}}(B)$ should be invariant to the specific choice of the local trivialization, i.e., $\nu\big(\eta_{i,\boldsymbol{z}}^{-1}(B)\big) = \nu\big(\psi_{j,\boldsymbol{z}}^{-1}(B)\big)$. From Eq. (6), this implies that

$$\nu\big(\eta_{i,\boldsymbol{z}}^{-1}(B)\big) = \nu\big(g_{ji,\boldsymbol{z}}^{-1} \circ \psi_{j,\boldsymbol{z}}^{-1}(B)\big) = \nu\big(\psi_{j,\boldsymbol{z}}^{-1}(B)\big),$$

from which it follows that ν is G-invariant.

It is worth pointing out that when $\mathscr{E}$ is a flat bundle, i.e., $\mathcal{X} = \mathcal{Z} \times \mathcal{Y}$, then $G = \{\mathrm{id}_{\mathcal{Y}}\}$ and ν is always well-defined.

Excess mass function. The final ingredient we require is an alternate characterization of $\mathcal{F}$-equivalence using *excess mass functions*. They are defined as follows: For a probability distribution $\mathbb{P}$ with density f and $\boldsymbol{X} \sim \mathbb{P}$, the *excess mass function*, $\widehat{f}$, is given by

$$\widehat{f}(t) = \mathbb{P}\big(f(\boldsymbol{X}) \geq t\big) = \int_{\mathcal{X}} \mathbb{1}(f(\boldsymbol{x}) \geq t) f(\boldsymbol{x}) d\boldsymbol{x}, \qquad \text{for all } t \geq 0.$$

Excess mass functions have been employed for geometric inference in nonparametric statistics [35, 39], and to characterize the Betti-0 function [9]. The next lemma supplements Lemma 3.5 using excess mass functions.

Lemma 5.3. *For two probability density functions f and g, $f \stackrel{\mathcal{F}}{\sim} g$ if and only if $\widehat{f} = \widehat{g}$. Moreover, if $\mathcal{P}$ is a family of distributions such that $\widehat{f} = \widehat{g}$ for all $f, g \in \mathcal{P}$, then $\mathcal{P}$ admits β-equivalence.*

In other words, for $\boldsymbol{X} \sim f$ and $\boldsymbol{Y} \sim g$ we may replace the condition $f(\boldsymbol{X}) \stackrel{d}{=} g(\boldsymbol{Y})$ in Definition 3.4 with the condition $\widehat{f} = \widehat{g}$. Lemma 3.5 formalizes the intuition that the topological summaries in the thermodynamic limit are capturing, at most, local information encoded in the superlevel sets of the probability density function.

With this background, the following result provides a sufficient characterization for distributions to admit $\mathcal{F}$-equivalence and is generated using structure underlying $\mathcal{X}$.

Theorem 5.4. *Let $\mathscr{E} = (\mathcal{X}, \mathcal{Z}, \mathcal{Y}, \pi, G)$ be a fiber bundle representation of $\mathcal{X}$ with a local trivialization $\{(V_j, \psi_j) : j \in J\}$ and compact $\mathcal{Y}$. For each $\boldsymbol{x} \in \mathcal{X}$ with $\boldsymbol{z} = \pi(\boldsymbol{x}) \in V_j$, let f_ϕ be given by*

$$f_\phi(\boldsymbol{x}) = C \cdot g\Big(\phi\big(\psi_{j,\boldsymbol{z}}^{-1}(\boldsymbol{x}), \boldsymbol{z}\big)\Big), \tag{9}$$

where

(i) $C = 1/\nu(\mathcal{Y})$ is the density of a uniform distribution w.r.t. a measure ν on $\mathcal{Y}$,

(ii) g is the density of a probability distribution on $\mathcal{Z}$ w.r.t. a base measure μ with modular character Ψ_μ,

(iii) $\phi : \mathcal{Y} \times \mathcal{Z} \to \mathcal{Z}$ is such that $\phi_{\boldsymbol{y}} \stackrel{\bullet}{=} \phi(\boldsymbol{y}, \cdot) \in \Delta(\mathcal{Z})$ for each $\boldsymbol{y} \in \mathcal{Y}$.

Then $\{f_\phi : \phi \in \Phi\}$ *admits* $\mathcal{F}$*-equivalence when*

$$\Phi \subseteq \left\{ \phi : \int_{\mathcal{Y}} C \cdot \Psi_\mu\big(\big\|\mathbf{D}\phi^{-1}(\boldsymbol{y}, \cdot)\big\|\big)\ d\nu(\boldsymbol{y}) = 1 \right\}. \tag{10}$$

The proof is deferred to Section 6.4, and we provide the intuition here. The density g on the base space $\mathcal{Z}$ together with the uniform distribution on the fiber $\mathcal{Y}$, induce a probability density $C \cdot g(\boldsymbol{z})$ locally in $\mathcal{X}$. The map $\phi(\boldsymbol{y}, \cdot)$ is free to move mass on the base space $\mathcal{Z}$, which, in turn, induces a different mass in the fiber over $\boldsymbol{z}$. By Lemma 5.3, $\mathcal{F}$-equivalence holds when the excess mass function is preserved; this corresponds to the Jacobian constraint in Eq. (10). Theorem 5.4 demonstrates that, in the thermodynamic limit, the topological summaries are effectively capturing, at most, only local information underlying the probability distribution.

While the density function f_ϕ in Theorem 5.4 seems to depend on the choice of the local trivialization $\{(V_j, \psi_j) : j \in J\}$, the following result shows that for a fixed $\phi \in \Phi$, the density function f_ϕ doesn't depend on the choice of the local trivialization used.

Proposition 5.5. *Under the conditions of Theorem 5.4, let* $\{(V_j, \psi_j) : j \in J\}$ *and* $\{(U_i, \eta_i) : i \in I\}$ *be two local trivializations of* $\mathscr{E} = (\mathcal{X}, \mathcal{Z}, \mathcal{Y}, \pi, G)$*. For a fixed* ϕ *and for* $\boldsymbol{x} \in \mathcal{X}$ *with* $\boldsymbol{z} = \pi(\boldsymbol{x}) \in U_i \cap V_j$*, let* $f_\phi(\boldsymbol{x})$ *be the density given by* Eq. (9)*, and let* $\widetilde{f}_\phi(\boldsymbol{x})$ *be given by*

$$\widetilde{f}_\phi(\boldsymbol{x}) = C \cdot g\Big(\phi\big(\eta_{i,\boldsymbol{z}}^{-1}(\boldsymbol{x}), \boldsymbol{z}\big)\Big). \tag{11}$$

Then, for every measurable $A \subset \mathcal{X}$*,*

$$\int_A f_\phi d\lambda = \int_A \widetilde{f}_\phi d\lambda.$$

In other words, Proposition 5.5 ensures that two equivalent coordinate representations of the $\mathscr{E}$ induce the same probability density function f_ϕ on $\mathcal{X}$ for a fixed ϕ, and the collection Φ in Theorem 5.4 cannot be simply obtained by a coordinate transformation of $\mathscr{E}$.

Remark 5.6. We highlight some salient observations regarding Theorem 5.4 below.

(i) When $\mathcal{Y} = \mathcal{G}$ is a Lie group, $\mathscr{X}$ simplifies to become a principal $\mathcal{G}$-bundle and the bundle projection $\pi : \mathcal{X} \to \mathcal{X}/\mathcal{G}$ projects each element in $\mathcal{X}$ to its orbit. Principal $\mathcal{G}$-bundles admit local cross-sections. Furthermore, when it admits a global-cross section, $\mathcal{X}$ admits the factorization $\mathcal{X} = \mathcal{X}/\mathcal{G} \times \mathcal{G}$, and the factorization of measure on $\mathcal{X}$ simplifies to the product of an invariant measure μ and an equivariant measure ν (see, for example, [29]). Lastly, when $\mathcal{X} = \mathcal{Y} \times \mathcal{Z}$ is globally trivial, then the bundle charts simply become $\psi_\alpha = \mathrm{id}_X$, such that the induced-measure $\nu_{\boldsymbol{z}} = \nu$ for each $\boldsymbol{z} \in \mathcal{Z}$, and Eq. (8) reduces to the familiar setting of Fubini's theorem.

(ii) Suppose $\mathbb{P}$ is a distribution on $\mathcal{X}$ with density f w.r.t. a dominating measure λ, and $A \subset \mathcal{X}$ is a set of $\lambda-$measure zero. Let $\mathbb{P}\big|_{\mathcal{X}\setminus A}$ be the restriction of $\mathbb{P}$ on $\mathcal{X} \setminus A$ and let $f_{\mathcal{X}\setminus A}$ be its resulting density. Then, for all $t \geq 0$,

$$\widehat{f}(t) = \int_{\mathcal{X}} \mathbb{1}(f(\boldsymbol{x}) \geq t) f(\boldsymbol{x}) d\lambda(\boldsymbol{x}) \overset{(i)}{=} \int_{\mathcal{X}\setminus A} \mathbb{1}(f(\boldsymbol{x}) \geq t) f(\boldsymbol{x}) d\lambda(\boldsymbol{x}) = \widehat{f}_{\mathcal{X}\setminus A}(t),$$

where (i) follows by noting that $\lambda(A) = 0$. Therefore, $\widehat{f} = \widehat{f}_{\mathcal{X}\setminus A}$, indicating that $\mathcal{F}$-equivalence for families of distributions can be analyzed by excluding a set of measure zero from the space $\mathcal{X}$. In particular, if $\mathscr{E} = (\mathcal{X}, \mathcal{Z}, \mathcal{Y}, \pi, G)$ is a fiber bundle, then we may omit sets of measure zero from both $\mathcal{Z}$ and $\mathcal{Y}$ since the dominating measure λ on $\mathcal{X}$ is admits the disintegration $\lambda = \nu \otimes_{\text{loc}} \mu$, for dominating measures ν and μ on $\mathcal{Y}$ and $\mathcal{Z}$, respectively.

It follows from Theorem 5.4 that a fairly large family of distributions admit $\mathcal{F}$-equivalence. However, the exact representation of these families using the index ϕ is not entirely obvious. Nevertheless, the elements in Φ may be indexed by a more well-behaved set Θ such that the family $\{f_\theta : \theta \in \Theta\}$ admits $\mathcal{F}$-equivalence. This is made precise in the following examples. The following example illustrates $\mathcal{F}$-equivalence when ν has a non-trivial modular character but $\mathcal{X}$ admits global cross sections.

Example 5.7. Consider the spherical decomposition $\boldsymbol{x} = (r, \boldsymbol{\theta})$ of $\mathbb{R}^d \setminus \{\mathbf{0}\}$ where $r \in \mathbb{R}_+$ and $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$. Let μ be a measure on $\mathcal{Z} = \mathbb{R}_+$ with density g w.r.t. η such that

$$\eta(dr) = d(r^d) = r^{d-1}dr. \tag{12}$$

For a nonnegative valued function $\xi : \mathbb{S}^{d-1} \to \mathbb{R}_+$, and for $\boldsymbol{x} \in \mathcal{X}$ with spherical coordinates $(r, \boldsymbol{\theta})$, let f_ξ given by

$$f_\xi(\boldsymbol{x}) = C_d \cdot g\big(r \cdot \xi(\boldsymbol{\theta})\big), \tag{13}$$

where $1/C_d = 2\pi^{d/2}/\Gamma(d/2)$ is the surface area of $\mathbb{S}^{d-1}$. In the context of Theorem 5.4, consider the fiber bundle representation of $\mathcal{X} = \mathbb{R}^d \setminus \{\mathbf{0}\}$ given by $\mathscr{E} = (\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \pi, G)$ where $\mathcal{Y} = \mathbb{S}^{d-1}$ denotes the typical fiber, the base space $\mathcal{Z} = \mathbb{R}_+$, and the projection map $\pi(\boldsymbol{x}) = \|\boldsymbol{x}\|$. Taking $V = \mathcal{Z} = \mathbb{R}_+$, and $\psi^{-1} : \mathcal{X} \to V \times \mathcal{Y}$ to be the map given by

$$\psi^{-1}(\boldsymbol{x}) = (\|\boldsymbol{x}\|, \boldsymbol{\theta}) \quad \text{where} \quad \boldsymbol{\theta} \stackrel{\bullet}{=} \boldsymbol{x}/\|\boldsymbol{x}\| \in \mathcal{Y},$$

provides a local trivialization $\{(V, \psi)\}$ for $\mathcal{X}$. Therefore, the density in Eq. (14) is equivalently given by

$$f_\xi(\boldsymbol{x}) = C_d \cdot g\big(\phi(\psi_r^{-1}(\boldsymbol{x}), r)\big),$$

where $r = \pi(\boldsymbol{x})$ and $\phi(\boldsymbol{\theta}, r) = r \cdot \xi(\boldsymbol{\theta})$. It is easy to verify that the map $\phi(\boldsymbol{\theta}, \cdot) \in \Delta(\mathcal{Z})$ for all $\boldsymbol{\theta} \in \mathbb{S}^{d-1}$, and $\phi^{-1}(\boldsymbol{\theta}, r) = r/\xi(\boldsymbol{\theta})$. Importantly, from Eq. (12), the modular character of η is $\Psi(t) = t^d$. If we consider the set Ξ, satisfying the constraint in Eq. (10),

$$\Xi = \left\{\xi : \int_{\mathbb{S}^{d-1}} C_d \cdot \xi(\boldsymbol{\theta})^{-d} \nu(d\boldsymbol{\theta}) = 1\right\},$$

then from Theorem 5.4 it follows that $\mathcal{P}(\Xi) = \{f_\xi : \xi \in \Xi\}$ admits $\mathcal{F}$-equivalence. Moreover, from Remark 5.6 (ii), $f_\xi(\boldsymbol{x})$ also satisfies $\mathcal{F}$-equivalence on $\mathbb{R}^d$. ♢

Remark 5.8. While seemingly contrived, the preceding example arises naturally in several situations. When ν is the surface measure on $\mathbb{S}^{d-1}$ (which is invariant under $G = S\mathcal{O}(d)$) and η is the measure given by Eq. (12), then μ on $\mathcal{Z} = \mathbb{R}_+$ and ν on $\mathcal{Y} = \mathbb{S}^{d-1}$ induce the standard Lebesgue measure λ_d locally in $\mathbb{R}^d$, $\lambda_d = \eta \otimes_{\text{loc}} \nu$, i.e., for $\boldsymbol{x} = (r, \boldsymbol{\theta})$ in local coordinates $\lambda_d(d\boldsymbol{x}) = \eta(dr) \cdot \nu(d\boldsymbol{z})$. In other words the non-trivial modular character in η arises naturally when the standard Lebesgue measure on $\mathbb{R}^d$ is decomposed in spherical coordinates.

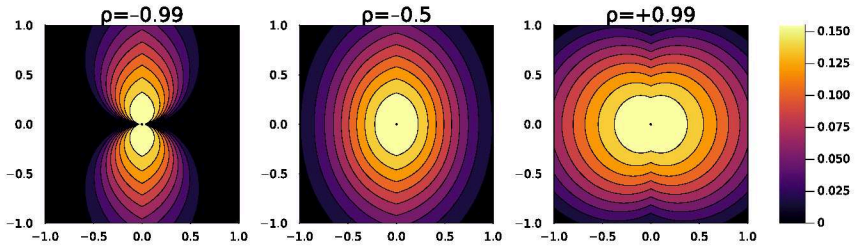


FIGURE 5. The sets $\{\boldsymbol{x} \in \mathbb{R}^2 : f_\rho(\boldsymbol{x}) \geq t\}$ for $\rho \in \{-0.99, -0.5, 0.99\}$ respectively. For a fixed level t, all three of them have the same mass. In general, $\mathbb{P}_\rho(\{\boldsymbol{x} \in \mathbb{R}^2 : f_\rho(\boldsymbol{x}) \geq t\})$ is the same for each $|\rho| < 1$.

For a more concrete version of Example 5.7, let $d = 2$ such that $\mathcal{Z} = \mathbb{S}^1 \simeq [0, 2\pi]/\sim$ where $\{0\} \sim \{2\pi\}$, and for $\boldsymbol{x} = (r\cos\theta, r\sin\theta)$ we have $\lambda_2(d\boldsymbol{x}) = r\ dr\ d\theta$. Consider

$$f_\rho(\boldsymbol{x}) = \frac{g\big(r \cdot \xi_\rho(\theta)\big)}{2\pi},$$

where $g(r)$ and $\xi_\rho(\theta)$ from Example 5.7 are given by

$$g(r) = \pi^{-1} \cdot \exp\big(-r^2/4\pi\big), \quad \text{and} \quad \xi_\rho(\theta) = (1 + \rho\cos\theta)^{-1/2}. \tag{14}$$

Note that $\xi_\rho(\theta)$ is well-defined whenever $|\rho| < 1$, and

$$\int_0^{2\pi} \xi_\rho^{-2}(\theta) \frac{d\theta}{2\pi} = 1.$$

Therefore $\{f_\rho(\boldsymbol{x}) : |\rho| < 1\}$ admits $\mathcal{F}$-equivalence. Figure 5 illustrates the superlevel sets $\big\{\boldsymbol{x} \in \mathbb{R}^2 : f_\rho(\boldsymbol{x}) \geq t\big\}$ of f_ρ.

The next example illustrates $\mathcal{F}$-equivalence for a family of distributions supported on the surface of a Möbius band with a full twist.

Example 5.9. Let $\mathcal{X}$ be the surface of a Möbius band with a full twist given by

$$\mathcal{X} = \Big\{\boldsymbol{x}(\theta, t) \in \mathbb{R}^3 : \theta \in [0, 2\pi),\ t \in [-0.5, 0.5]\Big\},$$

where

$$\boldsymbol{x}(\theta, t) = \Big(\cos\theta(1 + t\cos\theta),\ \sin\theta(1 + t\cos\theta),\ t\sin\theta\Big),$$

as illustrated in Figure 8 (A). For $\alpha \in \mathbb{R}$, let f_α be a density function on $\mathcal{X}$ given by

$$f_\alpha(\boldsymbol{x}) = f_\alpha(\boldsymbol{x}(\theta, t)) = \frac{\exp\left(\kappa \cdot \cos\left(\alpha t + \theta\right)\right)}{2\pi I_0(\kappa)}, \tag{15}$$

where $\kappa \in \mathbb{R}_+$ is a fixed parameter and $I_0(\kappa)$ is the modified Bessel function of the first kind. Then $\{f_\alpha : \alpha \in \mathbb{R}\}$ admits $\mathcal{F}$-equivalence. This follows by noting that $\mathcal{X}$ can be written as a fiber bundle with base space $\mathcal{Z} = \mathbb{S}^1$ and typical fiber $\mathcal{Y} = [-0.5, 0.5]$. By identifying $\mathcal{Z} \simeq [0, 2\pi]/\sim$ where $\{0\} \sim \{2\pi\}$, the projection map is given by $\pi(\boldsymbol{x}) = \arctan(x_2/x_1)$. The collection $\{(V_j, \psi_j) : j = 1, 2\}$ provide the local trivialization for $\mathcal{X}$, where

$$V_1 = (0, 2\pi) \setminus \{\pi\}, \quad \text{and} \quad V_2 = [0, \epsilon) \cup (\pi - \epsilon, \pi + \epsilon)$$

for sufficiently small $0 < \epsilon < \pi/2$, and for $\theta = \pi(\boldsymbol{x}) \in [0, 2\pi)$

$$\psi_{1,\theta}^{-1}(\boldsymbol{x}) = \frac{x_3}{\sin\theta}, \quad \text{and} \quad \psi_{2,\theta}^{-1}(\boldsymbol{x}) = \frac{1}{\cos\theta}\Big(\frac{x_1}{\cos\theta} - 1\Big).$$

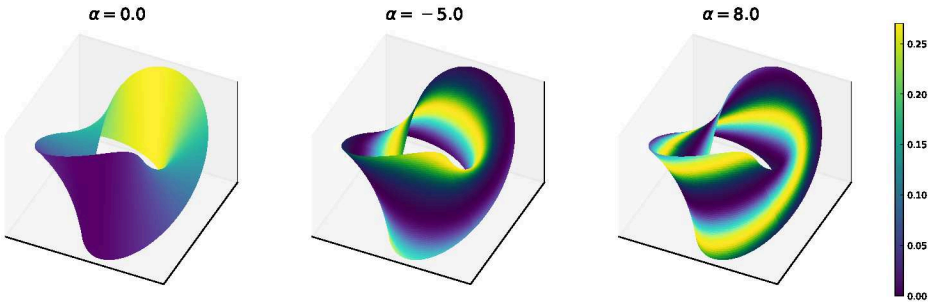


FIGURE 6. Superlevel sets of the probability density function f_α on $\mathcal{X}$ from Example 5.9 when $\alpha \in \{0, -5, 8\}$. For a fixed level $t \geq 0$, the mass of the superlevel set $\{\boldsymbol{x} \in \mathcal{X} : f_\alpha(\boldsymbol{x}) \geq t\}$ is the same for each α.

A more detailed description is provided in Appendix A.2. The density in Eq. (15) may be rewritten as

$$f_\alpha(\boldsymbol{x}) = g\Big(\phi_\alpha(\psi_{j,\theta}^{-1}(\boldsymbol{x}), \theta)\Big),$$

where $g(\theta) = \exp(\kappa \cos\theta)/2\pi I_0(\kappa)$ is the density of the von Mises distribution on $[0, 2\pi]/\sim$ with parameter κ, and modular character $\Psi(u) = u$. The map $\phi_\alpha(\theta, t) = \alpha t + \theta \mod 2\pi$ is such that $\phi_\alpha(\cdot, t) \in \Delta(\mathcal{Z})$ for all $t \in [-0.5, 0.5]$. The Jacobian constraint in Eq. (10),

$$\int_{-0.5}^{0.5} \big\|\mathbf{D}\phi_\alpha^{-1}(\cdot, t)\big\| d\nu(t) = \int_{-0.5}^{0.5} \left|\frac{d}{d\theta}(\theta - \alpha t \mod 2\pi)\right| d\nu(t) = \int_{-0.5}^{0.5} 1 \cdot dt = 1,$$

is satisfied for all $\alpha \in \mathbb{R}$. Therefore, by Theorem 5.4, $\{f_\alpha : \alpha \in \mathbb{R}\}$ admits $\mathcal{F}$-equivalence. Figure 6 shows the superlevel sets $\{\boldsymbol{x} \in \mathcal{X} : f_\alpha(\boldsymbol{x}) \geq t\}$ when $\alpha \in \{0, -5, 8\}$. ♢

For the fiber bundle representation $\mathscr{E} = (\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \pi, G)$, when the typical fiber space $\mathcal{Y}$ is discrete, Theorem 5.4 simplifies to the following result.

Corollary 5.10. *Let $\mathcal{X} \subset \mathbb{R}^d$, μ be a measure on $\mathcal{X}$, and $\pi : \mathcal{X} \to A$ be a given continuous surjection from $\mathcal{X}$ to $A \subset \mathcal{X}$. Suppose $\mathcal{X}$ admits the representation $\mathcal{X} = \big(\sqcup_{i=1}^N A_i\big) \bigsqcup \big(\cup_{j=1}^m B_i\big)$ where each A_i is diffeomorphic to A and $\mu(B_j) = 0$ for each j. For a collection of maps $\underline{\phi} \overset{\bullet}{=} \{\phi_1, \phi_2, \ldots, \phi_N\}$ such that each $\phi_i \in \Delta(A)$, let $f_{\underline{\phi}}$ be a probability density function on $\mathcal{X}$ given by*

$$f_{\underline{\phi}}(\boldsymbol{x}) = \frac{1}{N} \sum_{i=1}^N g\big(\phi_i(\pi(\boldsymbol{x}))\big) \cdot \mathbb{1}(\boldsymbol{x} \in A_i), \tag{16}$$

where g is a probability density function with $\operatorname{supp}(g) = A$. Then, $\Big\{f_{\underline{\phi}} : \underline{\phi} \in \mathbf{\Phi}_N\Big\}$ admits $\mathcal{F}$-equivalence for

$$\mathbf{\Phi}_N \overset{\bullet}{=} \left\{\underline{\phi} = \{\phi_1, \phi_2, \ldots, \phi_N\} : \frac{1}{N} \sum_{i=1}^N \Psi_\mu\big(\big\|\mathbf{D}\phi_i^{-1}\big\|\big) = 1\right\}.$$

Remark 5.11. For $t \in \mathbb{R}$, the excess mass function for $f_{\underline{\phi}}$ is given by $\widehat{f_{\underline{\phi}}}(t) = \widehat{g}(Nt)$. Therefore, $N \in \mathbb{Z}_+$ is a fixed value in the family of distributions $\big\{f_{\underline{\phi}} : \underline{\phi} \in \mathbf{\Phi}_N\big\}$. In

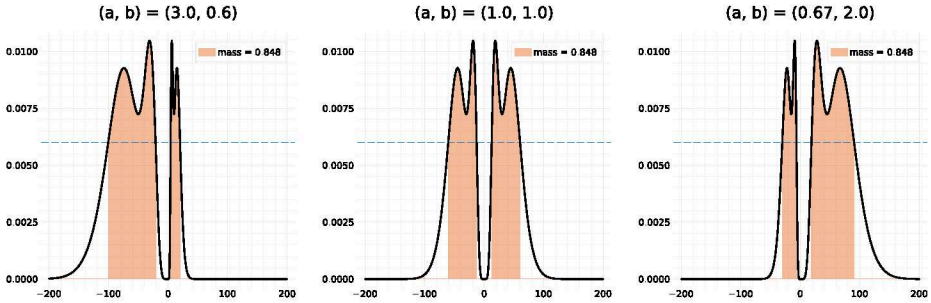


FIGURE 7. Illustration of $\mathcal{F}$-equivalence when $g \sim 0.5\Gamma(10,5) + 0.5\Gamma(10,2)$ is a mixture of Gamma distributions. For different values of $\theta \in \Theta$, the excess mass functions $\widehat{f}_\theta(t)$ are identical, as shaded in orange for $t = 0.006$.

other words, if $f_{\underline{\phi}_1}, f_{\underline{\phi}_2}$ admit the representation in Eq. (16) and $f_{\underline{\phi}_1} \overset{\mathcal{F}}{\sim} f_{\underline{\phi}_2}$, then $|\underline{\phi}_1| = |\underline{\phi}_2|$.

For completeness, the proof is provided in Section 6.6. The following two examples illustrate $\mathcal{F}$-equivalence for distributions which admit the representation in Corollary 5.10.

Example 5.12. For $\boldsymbol{\theta} = (\theta_1, \theta_2)$, let $\Theta \overset{\bullet}{=} \{(\theta_1, \theta_2) \in \mathbb{R}^2_{++} : 1/\theta_1 + 1/\theta_2 = 2\}$, and consider the family of distributions $\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ on $\mathcal{X} = \mathbb{R}$ given by

$$f_{\boldsymbol{\theta}}(x) = \begin{cases} g(\theta_1 x)/2, & \text{if } x > 0 \\ g(-\theta_2 x)/2, & \text{if } x \leq 0 \end{cases}, \tag{17}$$

where g is a probability density function on $\mathbb{R}_+$. Then, $\{f_\theta : \theta \in \Theta\}$ admits $\mathcal{F}$-equivalence. This can be seen by representing $\mathcal{X} = \mathbb{R}_+ \sqcup \mathbb{R}_- \sqcup \{0\}$, where $A_1 = \mathbb{R}_+$ and $A_2 = \mathbb{R}_-$ are both diffeomorphic to $A = \mathbb{R}_+$, and $\{0\}$ has measure zero. By taking $\pi(\boldsymbol{x}) = |x|$ to be the continuous surjection from $\mathcal{X}$ to A, and the maps $\phi_1, \phi_2 \in \Delta(\mathbb{R}_+)$ to be

$$\phi_1(z) = \theta_1 z \quad \text{and} \quad \phi_2(z) = \theta_2 z,$$

the density function, $f_{\boldsymbol{\theta}}$, is equivalently expressed as

$$f_{\boldsymbol{\theta}}(x) = \frac{1}{2} \sum_{i=1}^{2} g\Big(\phi_i(|x|)\Big) \cdot \mathbb{1}(x \in A_i).$$

Since $\underline{\boldsymbol{\phi}} = \{\phi_1, \phi_2\}$ is completely specified by $\boldsymbol{\theta} = (\theta_1, \theta_2)$, from Corollary 5.10, the Jacobian constraint is satisfied when

$$\sum_{i=1}^{2} \frac{1}{2} \cdot \Psi_\mu\big(\|\mathbf{D}\phi_i^{-1}\|\big) = \frac{1}{2\theta_1} + \frac{1}{2\theta_2} = 1.$$

Therefore, $\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ admits $\mathcal{F}$-equivalence from Corollary 5.10. See Figure 7 for an illustration when g is the density function associated with the mixture of Gamma distributions: $0.5 \cdot \Gamma(10,5) + 0.5 \cdot \Gamma(10,2)$. ♢

Example 5.13. For $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4) \in \mathbb{R}^4_{++}$ consider the probability density $f_{\boldsymbol{\theta}}$ on $\mathcal{X} = \mathbb{R}^2$ from Eq. (3) in Example 3.7. To see that $\{f_{\boldsymbol{\theta}} : \theta_1 + \theta_2 + \theta_3 + \theta_4 = 4\}$

admits $\mathcal{F}$-equivalence, consider the representation of $\mathcal{X} = \mathbb{R}^2$ as

$$\mathcal{X} = \left\{ \bigsqcup_{i=1}^{4} A_i \right\} \bigsqcup \underbrace{\Big\{ (x_1, x_2) : x_1 = 0 \text{ or } x_2 = 0 \Big\}}_{=:B}, \tag{18}$$

where B is a set of measure zero, and A_1, A_2, A_3, A_4 are the four (open) quadrants of $\mathbb{R}^2$ which are each diffeomorphic to $A = \mathbb{R}^2_{++} = \big\{(z_1, z_2) \in \mathbb{R}^2 : z_1, z_2 > 0\big\}$. By taking $\pi(x_1, x_2) = (|x_1|, |x_2|)$ to be the continuous surjection from $\mathcal{X}$ to A, consider the maps $\phi_i \in \Delta(\mathbb{R}^2_{++})$ to be given by $\phi_i(\boldsymbol{z}) = R(\theta_i)^\top \boldsymbol{z}$ where

$$R(\theta_i) = \begin{pmatrix} 1/\sqrt{\theta_i} & 0 \\ 0 & 1/\sqrt{\theta_i} \end{pmatrix}.$$

Then, $f_{\boldsymbol{\theta}}$ is equivalently expressed as

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \frac{1}{4} \sum_{i=1}^{4} g\Big(\phi_i(\pi(\boldsymbol{x}))\Big) \cdot \mathbb{1}(\boldsymbol{x} \in A_i).$$

Similar to Example 5.12, $\underline{\boldsymbol{\phi}} = \{\phi_1, \phi_2, \phi_3, \phi_4\}$ is completely specified by $\boldsymbol{\theta} \in \mathbb{R}^4$, and the Jacobian constraint in Eq. (10) is satisfied when

$$\sum_{i=1}^{4} \frac{1}{4} \big\| \mathbf{D}\phi_i^{-1} \big\| = \sum_{i=1}^{4} \frac{1}{4} \cdot \det\big(R(\theta_i)^{-1}\big) = \frac{1}{4} \cdot \sum_{i=1}^{4} \theta_i = 1,$$

and, therefore, the family $\{f_{\boldsymbol{\theta}} : \theta \in \Theta\}$ admits $\mathcal{F}$-equivalence. ♢

While Theorem 4.5 and Theorem 5.4 provide conditions for $\mathcal{P}(\Theta)$ to admit $\mathcal{F}$-equivalence when f_θ has a specific form, we might ask: *What happens to distributions that do not conform to the templates mentioned above*? The next result provides a necessary and sufficient geometric constraint (in the space of probability distributions) which $\mathcal{P}(\Theta)$ needs to satisfy for $\mathcal{F}$-equivalence.

Theorem 5.14. *For an open set $\Theta \subseteq \mathbb{R}^p$, let $\mathcal{P}(\Theta) = \{f_\theta : \boldsymbol{\theta} \in \Theta\}$ be a family of distributions on $\mathcal{X}$. Then $\mathcal{P}(\Theta)$ admits $\mathcal{F}$-equivalence if and only if for all $1 \le i \le p$,*

$$\frac{\partial}{\partial \theta_i} \left(\int_{\mathcal{X}} f_{\boldsymbol{\theta}}^{k+1}(\boldsymbol{x}) d\boldsymbol{x} \right) = 0, \text{ for all } k \in \mathbb{N}_0. \tag{19}$$

Moreover, if the gradient $\nabla_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}$ exists a.e.-λ_d and there exists a function $M \in L^1(\mathcal{X})$ such that for each $1 \le i \le p$ and for all $\theta \in \Theta$, $\left| \frac{\partial}{\partial \theta_i} f_{\boldsymbol{\theta}}(\boldsymbol{x}) \right| \le M(\boldsymbol{x})$ a.e.-λ_d, then $\mathcal{P}(\Theta)$ admits $\mathcal{F}$-equivalence if and only if

$$\Big\langle f_{\boldsymbol{\theta}}^k, \nabla_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}} \Big\rangle_{L^2(\mathcal{X})} = \mathbf{0}, \quad \text{for all } k \in \mathbb{N}_0. \tag{20}$$

We collect the proof in Section 6.7, and illustrate Theorem 5.14 by verifying $\mathcal{F}$-equivalence for the family of distributions already studied in Example 5.12.

Example 5.15. Consider the family of distributions on $\mathbb{R}^2$ from Example 5.7 in Eq. (14) given by,

$$f_\rho(\boldsymbol{x}) = g(r \cdot \xi_\rho(\theta)) = \frac{1}{2\pi} \exp\left(\frac{-r^2}{2(1 + \rho\cos(\theta))} \right),$$

where $r = \|\boldsymbol{x}\|$ and $\tan(\theta) = x_2/x_1$. First, observe that $\rho \in [-1, 1]$ contains an open set in $\mathbb{R}$. Moreover, f_ρ clearly satisfies the stochastic regularity assumptions in Theorem 5.14 by taking

$$M(\boldsymbol{x}) = \frac{r^2|\cos(\theta)|}{4\pi(1-|\cos(\theta)|)} \exp\big(-r^2/4\big).$$

The derivative of f_ρ is given by

$$\frac{\partial}{\partial\rho} f_\rho(\boldsymbol{x}) = \frac{r^2\cos(\theta)}{2(1+\rho\cos(\theta))^2} \cdot \frac{1}{2\pi} \exp\left(\frac{-r^2}{2(1+\rho\cos(\theta))}\right).$$

Then, for any $k \geq 0$, Eq. (20) now becomes

$$\begin{aligned}\left\langle f_\rho^k, \frac{\partial}{\partial\rho} f_\rho \right\rangle_{L^2(\mathbb{R}^2)} &= \int_0^\infty \int_0^{2\pi} \frac{r^2\cos(\theta)}{2(1+\rho\cos(\theta))^2(2\pi)^{k+1}} \exp\left(-\frac{(k+1)r^2}{2(1+\rho\cos(\theta))}\right) \cdot r dr d\theta \\ &\overset{\text{(i)}}{=} \int_0^{2\pi} \frac{\cos(\theta)d\theta}{(k+1)^2 \cdot (2\pi)^{k+1}} = 0,\end{aligned}$$

where (i) follows from making the substitution $t = r^2$. It follows from Theorem 5.14 that $\{f_\rho : |\rho| < 1\}$ admits $\mathcal{F}$-equivalence. ♢

Example 5.16. For $\Theta = (1, \infty)$, consider the family of distributions on $\mathbb{R}$ given by

$$f_\theta(x) = \frac{1}{2}\, g(\theta x)\mathbb{1}(x \geq 0) + \frac{1}{2}\, g\left(\frac{\theta x}{1-2\theta}\right)\mathbb{1}(x < 0), \tag{21}$$

where g is any density on $\mathbb{R}_+$ satisfying the assumptions of Theorem 5.14. Then, $\{f_\theta : \theta \in \Theta\}$ admits $\mathcal{F}$-equivalence. Vis-à-vis Example 5.12, the density in Eq. (21) is a reparametrization of the density in Eq. (17) to ensure that $\Theta \subset \mathbb{R}$ is an open set. Observe that

$$\int_{\mathbb{R}} f_\theta(x)dx = \frac{1}{2\theta} + \frac{2\theta - 1}{2\theta} = 1,$$

implying that f_θ is a well defined density function for all $\theta > 1$. In order to verify the condition in Eq. (19) note that

$$\begin{aligned}&\frac{\partial}{\partial\theta}\int_{\mathbb{R}} f_\theta^{k+1}(x)dx \\ &\quad= \frac{1}{2^{k+1}} \cdot \frac{\partial}{\partial\theta}\left(\int_{\mathbb{R}} g^{k+1}(\theta x)\mathbb{1}(x \geq 0)dx + \int_{\mathbb{R}} g^{k+1}\left(\frac{\theta x}{1-2\theta}\right)\mathbb{1}(x < 0)dx\right) \\ &\quad\overset{\text{(i)}}{=} \frac{1}{2^{k+1}} \cdot \frac{\partial}{\partial\theta}\left(\left(\frac{1}{\theta} - \frac{1-2\theta}{\theta}\right) \cdot \int_0^\infty g^{k+1}(t)dt\right) = \frac{\partial}{\partial\theta}\int_0^\infty \frac{1}{2^k} \cdot g^{k+1}(t)dt = 0,\end{aligned}$$

where (i) follows from taking $t = \theta x$ in the first integral and $t = \theta x/1 - 2\theta$ in the second integral. By Theorem 5.14, this implies that $\{f_\theta : \theta \in \Theta\}$ admits $\mathcal{F}$-equivalence. ♢

6. **Proofs.** In this section, we present the proofs for the main results of this paper.

6.1. **Proof of Lemma 3.5.** Consider $\mathbb{P}, \mathbb{Q} \in \mathcal{P}$ with their respective probability density functions f and g,and consider $\boldsymbol{X} \sim f$ and $\boldsymbol{Y} \sim g$. For every fixed function γ_k, the equivalence $f \overset{\mathcal{F}}{\sim} g$ implies that

$$\gamma_k(sf(\boldsymbol{X})^{1/d}, tf(\boldsymbol{X})^{1/d}) \overset{d}{=} \gamma_k(sg(\boldsymbol{Y})^{1/d}, tg(\boldsymbol{Y})^{1/d}), \quad \text{for all} \quad 0 \leq s < t.$$

It follows that $\mu_k(\mathbb{P}; s, t) = \mu_k(\mathbb{Q}; s, t)$ for all $0 \leq s < t$. A similar argument also holds for ς_k. This implies that $\beta_k(\mathbb{P}; s, t) \overset{d}{=} \beta_k(\mathbb{Q}; s, t)$. Since this holds for each $k \geq 0$ we have that $\mathbb{P} \overset{\beta}{\sim} \mathbb{Q}$, and the result follows. □

6.2. **Proof of Theorem 4.5.** *Part (i).* Suppose $\boldsymbol{X}_\theta$ is a random variable with density f_θ. Consider the random variable $\boldsymbol{Y}_\theta \overset{\bullet}{=} g_\theta(\Psi(\boldsymbol{X}_\theta))$ as a transformation of $\boldsymbol{X}_\theta$, such that $f_\theta(\boldsymbol{X}_\theta) = \phi(\boldsymbol{Y}_\theta)$. By Lemma 3.5, if we can show that the distribution of $\boldsymbol{Y}_\theta$ does not depend on the parameter θ if and only if $\det\big(\mathbf{D}\Psi^{-1}(\boldsymbol{y})\big) = \zeta(T(\boldsymbol{y}))$ for some function $\zeta : \mathcal{T} \to \mathbb{R}$, then $\mathcal{F}$-equivalence for the family of distributions $\mathcal{P}(\Theta)$ follows.

The inverse transformation for $\boldsymbol{Y}_\theta$ is given by $\boldsymbol{x} = \Psi^{-1} \circ g_\theta^{-1}(\boldsymbol{y})$. The existence of g_θ^{-1} is guaranteed by the group $\mathcal{G}$. The Jacobian for the inverse transformation can be simplified using the multivariable chain-rule,

$$\mathbf{D}\big(\Psi^{-1} \circ g_\theta^{-1}\big)(\boldsymbol{y}) = \mathbf{D}\Psi^{-1}\big(g_\theta^{-1}(\boldsymbol{y})\big) \cdot \mathbf{D}g_\theta^{-1}(\boldsymbol{y}).$$

Since $\mathcal{G}$ is a group of isometries, we have that, $\big|\det\big(\mathbf{D}g_\theta^{-1}(\boldsymbol{y})\big)\big| = 1$. The density of $\boldsymbol{Y}_\theta$ is expressed as

$$h_\theta(\boldsymbol{y}) = \phi(\boldsymbol{y}) \cdot \big\|\mathbf{D}\Psi^{-1}\big(g_\theta^{-1}(\boldsymbol{y})\big)\big\|.$$

It follows that density h_θ does not depend on θ if and only if $\det\big(\mathbf{D}\Psi^{-1}\big(g_\theta^{-1}(\boldsymbol{y})\big)\big)$ does not depend on θ, i.e., $\det\big(\mathbf{D}\Psi^{-1}(\boldsymbol{y})\big)$ is $\mathcal{G}$-invariant. By Proposition 4.4, this holds if and only if there exists some function $\zeta : \mathcal{T} \to \mathbb{R}$ such that

$$\det\big(\mathbf{D}\Psi^{-1}(\boldsymbol{y})\big) = \zeta(T(\boldsymbol{y})),$$

where T is $\mathcal{G}$-maximal invariant. Therefore, the distribution of $\boldsymbol{Y}_\theta$ doesn't depend on $\boldsymbol{\theta}$ if and only if the condition in Eq. (22) holds. Since ϕ is a fixed function and $f_\theta(\boldsymbol{X}_\theta) = \phi(\boldsymbol{Y}_\theta)$, this implies that Eq. (22) is also a necessary and sufficient condition for the distribution of $f_\theta(\boldsymbol{X}_\theta)$ to not depend on θ. This concludes the proof for the first part of the Theorem 4.5.

Part (ii). For ease of notation, let $\mathcal{G} = \times_{i=1}^m \mathcal{G}_i$ act on the space $\mathcal{Y}$. If $T_i : \mathcal{Y}_{i-1} \to \mathcal{Y}_i$ is a sequence of $\mathcal{G}_i$-compatible maximal invariants, then, for the second claim, from part (i), it suffices to show that

$$T(\boldsymbol{y}) = T_m \circ T_{m-1} \circ \dots T_1(\boldsymbol{y})$$

is $\mathcal{G}$-maximal invariant. The proof follows from induction. For the case $m = 1$, $T(\boldsymbol{y}) = T_1(\boldsymbol{y})$ by definition, so the property holds trivially. Assume that the property holds for $m > 1$. Then, $T' = T_m \circ T_{m-1} \circ \cdots \circ T_1$ is $\mathcal{G}'$-maximal invariant, where $\mathcal{G}' = \times_{i=1}^m \mathcal{G}_i$.

Let $\mathcal{G}_{m+1}$ be a group acting on $\mathcal{Y}_m$ such that $T_{m+1} : \mathcal{Y}_m \to \mathcal{Y}_{m+1}$ is $\mathcal{G}_{m+1}$-maximal invariant. From the assumption that T_{m+1} is $\mathcal{G}_{m+1}$-compatible, we also have that T' is $\mathcal{G}_{m+1}$ compatible; therefore, we only need to show that $T = T_{m+1} \circ T'$ is $\mathcal{G}$-maximal invariant, where $\mathcal{G} = \mathcal{G}' \times \mathcal{G}_{m+1}$. Each element $g \in \mathcal{G}$ is given by $g = (g', g_{m+1})$ where $g' \in \mathcal{G}'$ and $g_{m+1} \in \mathcal{G}_{m+1}$. We can write g as

$$g = (g', g_{m+1}) = (g', e_{m+1}) * (e', g_{m+1}) = \tilde{g}' * \tilde{g}_{m+1},$$

where e' and e_{m+1} are the identity elements of $\mathcal{G}'$ and $\mathcal{G}_{m+1}$ respectively, and $\tilde{g}'$ and $\tilde{g}_{m+1}$ are the representations for the group action of $\mathcal{G}'$ and $\mathcal{G}_{m+1}$ on $\mathcal{Y}$ as subgroups of $\mathcal{G}$. First, we examine that T is $\mathcal{G}$-invariant. For each $\boldsymbol{y} \in \mathcal{Y}$ we have that

$$T(g\boldsymbol{y}) = T_{m+1} \circ T'(g\boldsymbol{y}) = T_{m+1} \circ T'(\tilde{g}' * \tilde{g}_{m+1}\boldsymbol{y}). \tag{22}$$

By the definition of the group action, we can write $\tilde{g}' * \tilde{g}_{m+1}\boldsymbol{y} = \tilde{g}'\boldsymbol{z}$, where $\boldsymbol{z} = \tilde{g}_{m+1}\boldsymbol{y}$. Since T' is $\mathcal{G}'$-maximal invariant, we have that $T'(\tilde{g}'\boldsymbol{z}) = T'(\boldsymbol{z}) = T'(\tilde{g}_{m+1}\boldsymbol{y})$. Additionally, since T' is $\mathcal{G}_{m+1}$-compatible, it follows that $T'(\tilde{g}_{m+1}\boldsymbol{y}) = g^*_{m+1}T'(\boldsymbol{y})$, where g^*_{m+1} is the induced action of $\tilde{g}_{m+1}$ on $\mathcal{Y}_m$ via T'. Lastly, using the fact that T_{m+1} is $\mathcal{G}_{m+1}$-maximal invariant, Eq. (22) becomes

$$T(g\boldsymbol{y}) = T_{m+1}\big(g^*_{m+1}T'(\boldsymbol{y})\big) = T_{m+1} \circ T'(\boldsymbol{y}) = T(\boldsymbol{y}).$$

Next, let $\boldsymbol{x}$ and $\boldsymbol{y}$ be such that $T(\boldsymbol{x}) = T_{m+1} \circ T'(\boldsymbol{x}) = T_{m+1} \circ T'(\boldsymbol{y}) = T(\boldsymbol{y})$. Since T_{m+1} is maximally invariant, there exists $g_{m+1} \in \mathcal{G}_{m+1}$ such that $g^*_{m+1}T'(\boldsymbol{x}) = T'(\boldsymbol{y})$. From the $\mathcal{G}_{m+1}$-compatibility of T' we have that $g^*_{m+1}T'(\boldsymbol{x}) = T'(\tilde{g}_{m+1}\boldsymbol{x})$, giving us $T'(\boldsymbol{y}) = T'(\tilde{g}_{m+1}\boldsymbol{x})$. Lastly, since T' is $\mathcal{G}'$-maximal invariant, there exists $\tilde{g}'$ such that $\tilde{g}'(\tilde{g}_{m+1}\boldsymbol{x}) = \boldsymbol{y}$. This implies that there exists $g \in \mathcal{G}$ such that

$$g\boldsymbol{x} = \tilde{g}' * \tilde{g}_{m+1}\boldsymbol{x} = \tilde{g}'(\tilde{g}_{m+1}\boldsymbol{x}) = \boldsymbol{y}.$$

Therefore, $T(\boldsymbol{x}) = T(\boldsymbol{y})$ if and only if $\boldsymbol{x} \in \mathcal{G}\boldsymbol{y}$, from which it follows that $T = T_{m+1} \circ T'$ is $\mathcal{G}$-maximal invariant. □

6.3. **Proof of Lemma 5.3.** Consider $\boldsymbol{X} \sim f$ and $\boldsymbol{Y} \sim g$ for $f, g \in \mathcal{P}$, and let $Z_{\boldsymbol{X}} = f(\boldsymbol{X})$ and $Z_{\boldsymbol{Y}} = g(\boldsymbol{Y})$ be the transformation of $\boldsymbol{X}$ and $\boldsymbol{Y}$ under their own density. Then, for $t \geq 0$

$$\widehat{f}(t) = \mathbb{P}(f(\boldsymbol{X}) \geq t) = \mathbb{E}(\mathbb{1}(f(\boldsymbol{X}) > t)) = \mathbb{E}(\mathbb{1}(Z_{\boldsymbol{X}} > t)) = 1 - F_{Z_{\boldsymbol{X}}}(t),$$

where $F_{Z_{\boldsymbol{X}}}(t)$ is the cumulative distribution function of $Z_{\boldsymbol{X}}$. Similarly, $\widehat{g}(t) = 1 - F_{Z_{\boldsymbol{Y}}}(t)$.

For the first claim, note that if $Z_{\boldsymbol{X}} \overset{d}{=} Z_{\boldsymbol{Y}}$, then $F_{Z_{\boldsymbol{X}}}(t) = F_{Z_{\boldsymbol{Y}}}(t)$ for all $t \geq 0$, which implies that $\widehat{f} = \widehat{g}$. Conversely, if $\widehat{f} = \widehat{g}$, then $F_{Z_{\boldsymbol{X}}} = F_{Z_{\boldsymbol{Y}}}$, which implies that $Z_{\boldsymbol{X}} \overset{d}{=} Z_{\boldsymbol{Y}}$. Therefore, $f \overset{\mathcal{F}}{\sim} g$ if and only if $\widehat{f} = \widehat{g}$. The second claim now follows from Lemma 3.5. □

6.4. **Proof of Theorem 5.4.** Let $\lambda = \nu \otimes_{\text{loc}} \mu$ be the local product measure induced in $\mathcal{X}$, and let $C = 1/\nu(\mathcal{Y})$. First, we verify that f_ϕ as defined in Eq. (9) is a well-defined probability density function for each $\phi \in \Phi$. From [23, Eq. 6]

$$\begin{aligned}
\int_{\mathcal{X}} f_\phi d\lambda &= \int_{\mathcal{Z}} \int_{\pi^{-1}(\boldsymbol{z})} C \cdot g\Big(\phi\big(\psi^{-1}_{j,\boldsymbol{z}}(\boldsymbol{w}), \boldsymbol{z}\big)\Big) d\nu_{\boldsymbol{z}}(\boldsymbol{w}) d\mu(\boldsymbol{z}) \\
&\overset{\text{(i)}}{=} \int_{\mathcal{Z}} \int_{\psi^{-1}_{j,\boldsymbol{z}} \circ \pi^{-1}(\boldsymbol{z})} C \cdot g\big(\phi(\boldsymbol{y}, \boldsymbol{z})\big) d\nu_{\boldsymbol{z}}\big(\psi_{j,\boldsymbol{z}}(\boldsymbol{y})\big) d\mu(\boldsymbol{z}) \\
&\overset{\text{(ii)}}{=} \int_{\mathcal{Z}} \int_{\mathcal{Y}} C \cdot g(\phi_{\boldsymbol{y}}(\boldsymbol{z})) d\nu(\boldsymbol{y}) d\mu(\boldsymbol{z}) \\
&\overset{\text{(iii)}}{=} \int_{\mathcal{Z}} \int_{\mathcal{Y}} C \cdot g(\boldsymbol{u}) d\nu(\boldsymbol{y}) d\mu\big(\phi^{-1}_{\boldsymbol{y}}(\boldsymbol{u})\big) \\
&\overset{\text{(iv)}}{=} \int_{\mathcal{Z}} \int_{\mathcal{Y}} C \cdot g(\boldsymbol{u}) d\nu(\boldsymbol{y}) \Psi\big(\big\|\mathbf{D}\phi^{-1}_{\boldsymbol{y}}\big\|\big) d\mu(\boldsymbol{u})
\end{aligned}$$

$$\overset{(v)}{=} \int_{\mathcal{Y}} C \cdot \Psi\big(\big\|\mathbf{D}\phi_{\boldsymbol{y}}^{-1}\big\|\big)\Big\{\int_{\mathcal{Z}} g(\boldsymbol{u})d\mu(\boldsymbol{u})\Big\}d\nu(\boldsymbol{y})$$
$$= \int_{\mathcal{Y}} C \cdot \Psi\big(\big\|\mathbf{D}\phi_{\boldsymbol{y}}^{-1}\big\|\big)d\nu(\boldsymbol{y}) = 1,$$

where (i) follows from making the substitution $\boldsymbol{y} = \psi_{j,\boldsymbol{z}}^{-1}(\boldsymbol{w})$, (ii) follows from noting that $\psi_{j,\boldsymbol{z}}^{-1}\big(\pi^{-1}(\boldsymbol{z})\big) = \mathcal{Y}$ by definition of the local trivialization and the pushforward measure $\nu_{\boldsymbol{z}}$ is defined to be $\nu_{\boldsymbol{z}}(B) = \nu(\psi_{j,\boldsymbol{z}}^{-1}(B))$ for all measurable sets $B \subseteq \pi^{-1}(\boldsymbol{z})$. Similarly, (iii) follows by making the substitution $\boldsymbol{u} = \phi_{\boldsymbol{y}}(\boldsymbol{z})$, (iv) follows from the fact that $\phi_{\boldsymbol{y}} \in \Delta(\mathcal{Z})$ for every $\boldsymbol{y} \in \mathcal{Y}$ and the modular character of μ is Ψ, and (v) follows from Tonelli's theorem [19, Theorem 2.7]. Next, it remains to verify that $\widehat{f}_\phi(t)$ does not depend on ϕ, and we use the same machinery as before. Consider,

$$\widehat{f}_\phi(t) = \int_{\mathcal{X}} \mathbb{1}(f_\phi(\boldsymbol{x}) \geq t)\ f_\phi(\boldsymbol{x})d\lambda(\boldsymbol{x})$$
$$= \int_{\mathcal{Z}}\int_{\pi^{-1}(\boldsymbol{z})} \mathbb{1}\Big(C{\cdot}g\Big(\phi(\psi_{j,\boldsymbol{z}}^{-1}(\boldsymbol{w}),\boldsymbol{z})\Big) \geq t\Big)\ \ C{\cdot}\ g\Big(\phi(\psi_{j,\boldsymbol{z}}^{-1}(\boldsymbol{w}),\boldsymbol{z})\Big)d\nu_{\boldsymbol{z}}(\boldsymbol{w})d\mu(\boldsymbol{z}).$$

Again, substituting $\boldsymbol{y} = \psi_{j,\boldsymbol{z}}^{-1}(\boldsymbol{w})$, and then taking $\boldsymbol{u} = \phi_{\boldsymbol{y}}(\boldsymbol{z})$ we get

$$\widehat{f}_\phi(t) = \int_{\mathcal{Z}}\int_{\mathcal{Y}} \mathbb{1}(C{\cdot}g(\boldsymbol{u}) \geq t)\ C{\cdot}g(\boldsymbol{u})\ \Psi(\big\|\mathbf{D}\phi_{\boldsymbol{y}}^{-1}\big\|) \cdot d\nu(\boldsymbol{y})d\mu(\boldsymbol{u})$$
$$= \int_{\mathcal{Z}} \mathbb{1}(C{\cdot}g(\boldsymbol{u}) \geq t)\ g(\boldsymbol{u})\ \Big\{\int_{\mathcal{Y}}\ C{\cdot}\ \Psi(\big\|\mathbf{D}\phi_{\boldsymbol{y}}^{-1}\big\|) \cdot d\nu(\boldsymbol{y})\Big\}d\mu(\boldsymbol{u})$$
$$= \int_{\mathcal{Z}} \mathbb{1}(C{\cdot}g(\boldsymbol{u}) \geq t)\ g(\boldsymbol{u})d\mu(\boldsymbol{u}) = \widehat{g}(t/C),$$

which does not depend on the choice of $\phi \in \Phi$, and $\mathcal{F}$-equivalence follows. □

6.5. **Proof of Proposition 5.5.** For a fixed $\phi : \mathcal{Y} \times \mathcal{Z} \to \mathcal{Z}$, let f_ϕ and $\widetilde{f}_\phi$ be the density functions given by Eq. (9) and Eq. (11):

$$f_\phi(\boldsymbol{x}) = \ C{\cdot}\ g\Big(\phi\big(\psi_{j,\boldsymbol{z}}^{-1}(\boldsymbol{x}),\boldsymbol{z}\big)\Big), \quad \text{and} \quad \widetilde{f}_\phi(\boldsymbol{x}) = \ C{\cdot}\ g\Big(\phi\big(\eta_{i,\boldsymbol{z}}^{-1}(\boldsymbol{x}),\boldsymbol{z}\big)\Big),$$

corresponding to the local trivializations $\{(V_j, \psi_j) : j \in J\}$ and $\{(U_i, \eta_i) : i \in I\}$, respectively. Without loss of generality, we may take $C = 1$ for ease of notation. For a measurable set $A \subset \mathcal{X}$, and for $i \in I$ and $j \in J$ such that $\boldsymbol{z} \in \pi(A) \cap U_i \cap V_j$, using Eq. (8) we get

$$\int_A f_\phi d\lambda = \int_{\pi(A)}\int_{A_{\boldsymbol{z}}} f_\phi d\nu_{\boldsymbol{z}}d\mu, \tag{23}$$

where $A_{\boldsymbol{z}} = A \cap \pi^{-1}(\boldsymbol{z})$. Therefore, in order to establish the claim, it suffices to show that

$$\int_{A_{\boldsymbol{z}}} \widetilde{f}_\phi d\widetilde{\nu}_{\boldsymbol{z}} = \int_{A_{\boldsymbol{z}}} f_\phi d\nu_{\boldsymbol{z}}.$$

To this end, we have

$$\int_{A_{\boldsymbol{z}}} \widetilde{f}_\phi d\widetilde{\nu}_{\boldsymbol{z}} \overset{(i)}{=} \int_{A_{\boldsymbol{z}}} g\big(\eta_{i,\boldsymbol{z}}^{-1}(\boldsymbol{w}),\boldsymbol{z}\big)d\nu_{\boldsymbol{z}}(\boldsymbol{w})$$
$$= \int_{\eta_{i,\boldsymbol{z}}^{-1}(A_{\boldsymbol{z}})} g\big(\phi(\boldsymbol{y},\boldsymbol{z})\big)d\nu(\boldsymbol{y})$$

$$= \int_{g_{ji,\boldsymbol{z}}^{-1} \circ \eta_{i,\boldsymbol{z}}^{-1}(A_{\boldsymbol{z}})} g\big(\phi(\boldsymbol{u}, \boldsymbol{z})\big) d\nu(\boldsymbol{u}),$$

where (i) follows from the substitution $\boldsymbol{y} = \eta_{i,\boldsymbol{z}}^{-1}(\boldsymbol{w})$, and the last equality follows by taking $\boldsymbol{u} = g_{ji,\boldsymbol{z}}^{-1}(\boldsymbol{y})$ for $g_{ji,\boldsymbol{z}} \in G$ and by noting that ν is G-invariant. Using the "change of coordinates" in Eq. (6), we have $\psi_{j,\boldsymbol{z}}^{-1} = g_{ji,\boldsymbol{z}}^{-1} \circ \eta_{i,\boldsymbol{z}}^{-1}$, and therefore,

$$\begin{aligned}
\int_{A_{\boldsymbol{z}}} \widetilde{f}_{\phi} d\widetilde{\nu}_{\boldsymbol{z}} &= \int_{\psi_{j,\boldsymbol{z}}^{-1}(A_{\boldsymbol{z}})} g\big(\phi(\boldsymbol{u}, \boldsymbol{z})\big) d\nu(\boldsymbol{u}) \\
&= \int_{A_{\boldsymbol{z}}} g\big(\phi(\psi_{j,\boldsymbol{z}}^{-1}(\boldsymbol{w}), \boldsymbol{z})\big) d\big((\psi_{j,\boldsymbol{z}})_{\#}\nu\big)(\boldsymbol{w}) \\
&= \int_{A_{\boldsymbol{z}}} g\big(\phi(\psi_{j,\boldsymbol{z}}^{-1}(\boldsymbol{w}), \boldsymbol{z})\big) d\nu_{\boldsymbol{z}}(\boldsymbol{w}) = \int_{A_{\boldsymbol{z}}} f_{\phi} d\nu_{\boldsymbol{z}},
\end{aligned}$$

which proves the claim. □

6.6. **Proof of Corollary 5.10.** Let $\mathcal{X}^{\circ} = \sqcup_{i=1}^{n} A_i = \mathcal{X} \setminus \cup_{j=1}^{m} B_j$. First, note that since $\mu(B_j) = 0$ for each $j \in \{1, 2, \ldots, m\}$, it follows that $\mu\big(\cup_{j=1}^{m} B_j\big) = 0$. Therefore, from Remark 5.6 (ii), omitting $\cup_{j=1}^{m} B_j$ doesn't affect $\mathcal{F}$-equivalence, and it suffices to show that the claim holds for $\mathcal{X}^{\circ}$.

To this end, observe that $\mathcal{X}^{\circ}$ can be represented as the fiber bundle $\mathscr{E} = (\mathcal{X}^{\circ}, \mathcal{Y}, \mathcal{Z}, \pi, G)$ where $\mathcal{Z} = A$, $\mathcal{Y} = \{1, 2, \ldots, N\}$ and G is isomorphic to $\mathbb{Z}_N$ acting on $\mathcal{Y}$ by addition modulo N. Indeed, since π is assumed to be a continuously surjective map from $\mathcal{X}$ to A, it suffices to show that $\mathcal{X}$ is locally trivializable. Let $V = A$, and for $\pi^{-1}(A) = \sqcup_{i=1} A_i$ let $\psi : A \times \mathcal{Y} \to \pi^{-1}(A)$ be given by $\psi(A \times \{i\}) = \pi^{-1}(A) \cap A_i = A_i$ for each $i = 1, 2, \ldots, N$. Since each A_i is diffeomorphic to A, it follows that ψ is a diffeomorphism. Additionally, for $\boldsymbol{a} \in A$, the map $\psi_{\boldsymbol{a}}^{-1}(\boldsymbol{x}) = \{i : \boldsymbol{x} \in A_i\}$. Therefore, $\{(V, \psi)\}$ is a local trivialization, and $\mathscr{E}$ is a fiber bundle.

Let ν be the counting measure on $\mathcal{Y}$, and $C = 1/N = 1/\nu(\mathcal{Y})$. For $\boldsymbol{z} = \pi(\boldsymbol{x})$, we may write the density function in Eq. (16)

$$f_{\underline{\boldsymbol{\phi}}}(\boldsymbol{x}) = C \cdot g\Big(\phi\big(\psi_{\boldsymbol{z}}^{-1}(\boldsymbol{x}), \boldsymbol{z}\big)\Big),$$

where the map $\phi(i, \boldsymbol{a}) = \phi_i(\boldsymbol{a})$. Furthermore, the Jacobian constraint Eq. (10) requires that

$$\int_{\mathcal{Y}} C \cdot \Psi_{\mu}\big(\big\|\mathbf{D}\phi(i, \cdot)^{-1}\big\|\big) d\nu(i) = \sum_{i=1}^{N} \frac{1}{N} \cdot \Psi_{\mu}\big(\big\|\mathbf{D}\phi_i^{-1}\big\|\big) = 1.$$

Therefore, from Theorem 5.4, it follows that $\Big\{f_{\underline{\boldsymbol{\phi}}} : \underline{\boldsymbol{\phi}} \in \boldsymbol{\Phi}\Big\}$ admits $\mathcal{F}$-equivalence. □

6.7. **Proof of Theorem 5.14.** For each $\boldsymbol{\theta} \in \Theta$, let $\boldsymbol{X}_{\boldsymbol{\theta}}$ be a random variable with density $f_{\boldsymbol{\theta}}$. From Lemma 3.5, we know that if the distribution of $Z_{\boldsymbol{\theta}} \overset{\bullet}{=} f_{\boldsymbol{\theta}}(\boldsymbol{X}_{\boldsymbol{\theta}})$ does not depend on $\boldsymbol{\theta}$, $\mathcal{F}$-equivalence follows. The characteristic function for $Z_{\boldsymbol{\theta}}$ is given by

$$\varphi_{\boldsymbol{\theta}}(t) = \mathbb{E}_{Z_{\boldsymbol{\theta}}}\big(e^{itZ_{\boldsymbol{\theta}}}\big) = \mathbb{E}_{\boldsymbol{X}_{\theta}}\Big(e^{itf_{\boldsymbol{\theta}}(\boldsymbol{X}_{\theta})}\Big) = \int_{\mathcal{X}} e^{itf_{\boldsymbol{\theta}}(\boldsymbol{x})} f_{\boldsymbol{\theta}}(\boldsymbol{x}) d\boldsymbol{x},$$

for $t \in \mathbb{R}$. Using Euler's formula, we can write

$$\varphi_{\boldsymbol{\theta}}(t) = \int_{\mathcal{X}} \Big(\cos(t f_{\boldsymbol{\theta}}(\boldsymbol{x})) + i \cdot \sin(t f_{\boldsymbol{\theta}}(\boldsymbol{x}))\Big) \cdot f_{\boldsymbol{\theta}}(\boldsymbol{x}) d\boldsymbol{x}.$$

Using the Taylor series representation for $\cos(t f_{\boldsymbol{\theta}}(\boldsymbol{x}))$ and $\sin(t f_{\boldsymbol{\theta}}(\boldsymbol{x}))$ we get,

$$\begin{aligned}
\varphi_{\boldsymbol{\theta}}(t) &= \int_{\mathcal{X}} \left(\sum_{k=0}^{\infty} \frac{(-1)^k t^{2k}}{(2k)!} f_{\boldsymbol{\theta}}^{2k}(\boldsymbol{x}) + i \cdot \sum_{k=0}^{\infty} \frac{(-1)^k t^{2k+1}}{(2k+1)!} f_{\boldsymbol{\theta}}^{2k+1}(\boldsymbol{x})\right) \cdot f_{\boldsymbol{\theta}}(\boldsymbol{x}) d\boldsymbol{x} \\
&\overset{\text{(i)}}{=} \sum_{k=0}^{\infty} \frac{(-1)^k t^{2k}}{(2k)!} \int_{\mathcal{X}} f_{\boldsymbol{\theta}}^{2k+1}(\boldsymbol{x}) d\boldsymbol{x} + i \cdot \sum_{k=0}^{\infty} \frac{(-1)^k t^{2k+1}}{(2k+1)!} \int_{\mathcal{X}} \sum_{k=0}^{\infty} f_{\boldsymbol{\theta}}^{2k+2}(\boldsymbol{x}) d\boldsymbol{x} \\
&= \sum_{k=0}^{\infty} \frac{(-1)^k t^{2k}}{(2k)!} h_{2k}(\boldsymbol{\theta}) + i \cdot \sum_{k=0}^{\infty} \frac{(-1)^k t^{2k+1}}{(2k+1)!} h_{2k+1}(\boldsymbol{\theta}),
\end{aligned}$$

where $h_k(\boldsymbol{\theta}) \overset{\bullet}{=} \int_{\mathcal{X}} f_{\boldsymbol{\theta}}^{k+1}(\boldsymbol{x}) d\boldsymbol{x}$ and (i) follows from Fubini's theorem. It follows from this that the characteristic function $\varphi_{\boldsymbol{\theta}}$ does not depend on $\boldsymbol{\theta}$ if and only if $\mathrm{Re}(\varphi_{\boldsymbol{\theta}}(t))$ and $\mathrm{Im}(\varphi_{\boldsymbol{\theta}}(t))$ do not depend on $\boldsymbol{\theta}$ for all $t \in \mathbb{R}$. Therefore, $\varphi_{\boldsymbol{\theta}}$ does not depend on $\boldsymbol{\theta}$ if and only if the function $h_k(\boldsymbol{\theta})$ does not depend on $\boldsymbol{\theta}$ for each $k \in \mathbb{N}_0$. Equivalently, for each $k \in \mathbb{N}_0$ we must have that $\frac{\partial}{\partial \theta_i} h_k(\boldsymbol{\theta}) = 0$ for all $1 \le i \le p$, i.e.,

$$\frac{\partial}{\partial \theta_i} h_k(\boldsymbol{\theta}) = \frac{\partial}{\partial \theta_i} \left(\int_{\mathcal{X}} f_{\boldsymbol{\theta}}^{k+1}(\boldsymbol{x}) d\boldsymbol{x}\right) = 0.$$

Under the additional stochastic regularity conditions, using the Lebesgue-dominated convergence theorem we have

$$\frac{\partial}{\partial \theta_i} \left(\int_{\mathcal{X}} f_{\boldsymbol{\theta}}^{k+1}(\boldsymbol{x}) d\boldsymbol{x}\right) = \int_{\mathcal{X}} \frac{\partial}{\partial \theta_i} f_{\boldsymbol{\theta}}^{k+1}(\boldsymbol{x}) d\boldsymbol{x} = \left\langle f_{\boldsymbol{\theta}}^{k}, \frac{\partial}{\partial \theta_i} f_{\boldsymbol{\theta}} \right\rangle_{L^2(\mathcal{X})} = 0,$$

from which the second claim follows. □

7. **Discussion.** In this work, we have studied the framework of topological inference through the lens of classical statistical theory. In the parametric setup, we have investigated cases when the parameters of the statistical model are not sufficient for statistical inference based on their asymptotic limit in the thermodynamic regime. In our case, this is analogous to the property of β-equivalence. We have characterized several conditions under which a parametric family of distributions admits $\mathcal{F}$-equivalence, which also guarantees β-equivalence. When the distributions share an algebraic structure, we are able to describe necessary and sufficient conditions under which this asymptotic identifiability fails. In the absence of the underlying algebraic structure, we have shown that when the distributions satisfy a certain Jacobian constraint, they admit $\mathcal{F}$-equivalence. Lastly, in the absence of any of the above, when the distributions are stochastically regular (as is most often the case), we have shown that if the density function shares a certain geometry with its gradient, then $\mathcal{F}$-equivalence follows.

As noted in Remark 3.6, studying injectivity for Betti numbers collectively serves as a stepping-stone to understanding the behavior of more complex topological invariants, and we have focused on the phenomenon of $\mathcal{F}$-equivalence in the thermodynamic regime. Analogous asymptotic behavior for Betti numbers in the sparse regime has been the focus in [27, 7, 49]. For fixed $1 \le k \le d-1$, in the regime

that $nr_n^d \to 0$ and $n^{(k+2)} r_n^{d(k+1)} \to \infty$, the following limit theorem is established for random Čech complexes in [27, Theorem 3.2] and [7, Theorem 4.1]:

$$\frac{\beta_k(\mathcal{K}(\mathbb{X}_n, r_n)) - \overline{\mu}_k}{\sqrt{n^{(k+2)} r_n^{d(k+1)}}} \overset{d}{\rightsquigarrow} \mathcal{N}(0, \overline{\sigma}_k^2), \tag{24}$$

where, for fixed quantities ζ_k and ξ_k, the functionals μ_k and σ_k^2 are given by

$$\overline{\mu}_k = \frac{1}{(k+2)!} \int_{\mathcal{X}} \zeta_k f(\boldsymbol{x})^{k+2} d\boldsymbol{x}, \quad \textit{and} \quad \overline{\sigma}_k^2 = \frac{1}{(k+1)!} \int_{\mathcal{X}} \xi_k f(\boldsymbol{x})^{k+1} d\boldsymbol{x}.$$

From Theorem 5.14, we can see that if a family of distributions admit β-equivalence, then for each $1 \leq k \leq d-1$, in the regime that $nr_n^d \to 0$ and $n^{(k+2)} r_n^{d(k+1)} \to \infty$, the limit on the r.h.s. of Eq. (24) is identical. A similar conclusion for the Betti numbers of random Rips complexes also follows from [27, Theorem 3.1] in the regime that $n^{(2k+2)} r_n^{d(2k+1)} \to \infty$. As noted in [6, page 344] this difference in the regimes stems from the fact that the smallest nontrivial H_k cycle in the Rips complex is supported on $2k+2$ vertices as opposed to the $k+2$ vertices required for the Čech complex. The results in [4] suggest a useful direction for pursuing this line of investigation and is left for future work.

Lastly, it is important to note that the conditions characterized in this work are purely statistical in nature, and hold in the asympotic setting. However, there is still hope that topological summaries for statistical inference from $\mathcal{F}$-equivalent families of distributions may be useful in finite samples and is a promising direction for future work. We hope that this work will serve as a stepping-stone for further investigations in this direction.

Appendix A. Supplementary results.

Lemma A.1. *Suppose f is a probability density function with mean 0 and variance 1. For each $d \in \mathbb{Z}_+$, let $\boldsymbol{x} \in \mathbb{R}^d$ and $\zeta_d : \mathbb{R}_+ \to \mathbb{R}_+$. Then $\prod_{i=1}^d f(x_i) = \zeta_d(\|\boldsymbol{x}\|^2)$ holds for each $d \in \mathbb{N}_0$ if and only if $f(x) = \exp\left(-x^2/2\right)/\sqrt{2\pi}$, for each $x \in \mathbb{R}$.*

Proof. The sufficient condition follows unambiguously by plugging in the value for $f(x_i)$, i.e., when $f(x) = \exp\left(-x^2/2\right)/\sqrt{2\pi}$ it follows that

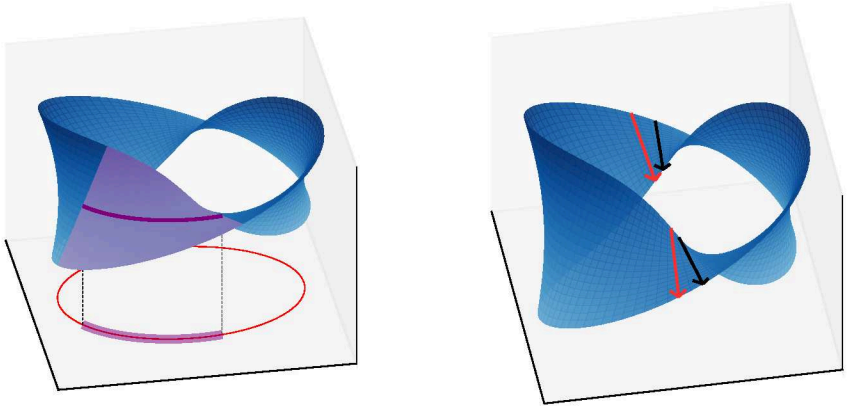
$$\prod_{i=1}^d f(x_i) = (2\pi)^{-d/2} \exp\left(-\|\boldsymbol{x}\|^2/2\right) \overset{\bullet}{=} \zeta_d(\|\boldsymbol{x}\|^2).$$

It remains to verify the necessary condition. To begin, consider the case when $d = 1$. For $x \in \mathbb{R}$, we have $f(x) = \zeta_1(x^2)$. Define $k \overset{\bullet}{=} f(0) = \zeta_1(0)$. We now proceed to consider the case when $d = 2$. For $(x, 0) \in \mathbb{R}^2$, we have $f(x) \cdot f(0) = \zeta_2(x^2)$. Since $f(x) = \zeta_1(x^2)$ and $f(0) = k$, it follows that $\zeta_2(x^2) = k\zeta_1(x^2)$ for all $x \in \mathbb{R}$. By induction, for each $d \in \mathbb{Z}_+$, we have

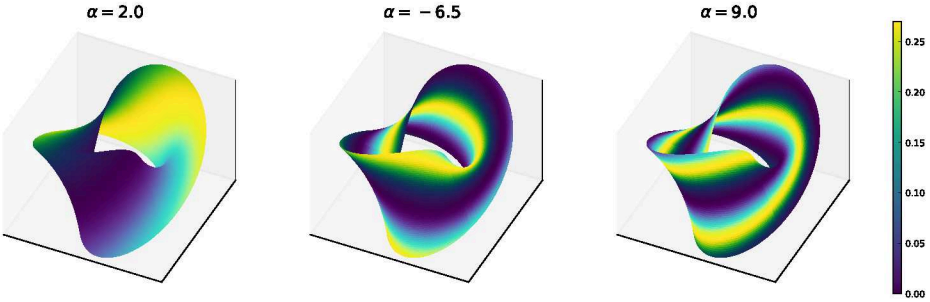
$$\zeta_d(x^2) = k^{d-1} \zeta_1(x^2). \tag{A.1}$$

Thus, for any $\boldsymbol{x} \in \mathbb{R}^d$, Eq. (A.1) implies

$$\prod_{i=1}^d f(x_i) = \zeta_d(x_1^2 + x_2^2 + \cdots + x_d^2) = k^{d-1} \zeta_1(x_1^2 + x_2^2 + \cdots + x_d^2).$$



(A) An illustration of the space X—the Möbius band with the full twist. (Left) The projection map π onto a neighborhood $U \subset [0, 2\pi]/\sim = \mathcal{Z}$. (Right) The map $\psi_\theta(t)$ for $t \in [-0.5, 0.5]$ preserves length and orientation, unlike the Möbius band with a half-twist. Therefore, the measure induced on the fiber over θ, $\pi^{-1}(\theta)$, is also a unit uniform distribution.



(B) Superlevel sets of the probability density function f_α on $\mathcal{X}$ for $\alpha \in 2.0, -6.5, 9.0$.

FIGURE 8. Illustration of the family of distributions from Example 5.9.

We also have that $\prod_{i=1}^d f(x_i) = \prod_{i=1}^d \zeta_1(x_i^2)$. Define $g(x) = \frac{\zeta_1(x)}{k}$, which implies g satisfies

$$\prod_{i=1}^{d} g(x_i^2) = g(x_1^2 + x_2^2 + \cdots + x_d^2). \tag{A.2}$$

Eq. (A.2) holds if and only if $g(x^2) = e^{\beta x^2}$, for some fixed $\beta \in \mathbb{R}$, which implies

$$f(x) = \zeta_1(x^2) = k \cdot g(x^2) = k \cdot e^{\beta x^2}.$$

However, f is a probability density function on $\mathbb{R}$ with mean 0 and variance 1, i.e, $\int_{\mathbb{R}} f(x)dx = 1$, $\int_{\mathbb{R}} xf(x)dx = 0$ and $\int_{\mathbb{R}} x^2 f(x)dx = 1$. This yields $\beta = -\frac{1}{2}$, $k = \frac{1}{\sqrt{2\pi}}$ and the result follows. □

Example A.2 (Example 5.9 Continued)**.** The space $\mathcal{X}$ is the surface of a Möbius band with a full twist, $\mathcal{X}$, given by

$$\mathcal{X} = \Big\{ \boldsymbol{x}(\theta, t) \in \mathbb{R}^3 : \theta \in [0, 2\pi),\ t \in [-0.5, 0.5] \Big\},$$

where

$$\boldsymbol{x}(\theta,t)=\Big(\cos\theta(1+t\cos\theta),\ \sin\theta(1+t\cos\theta),\ t\sin\theta\Big).$$

This can be represented as a fiber bundle $\mathscr{E}=(\mathcal{X},\mathcal{Z},\mathcal{Y},\pi,G)$, with the base space $\mathcal{Z}=\mathcal{S}^1$, the typical fiber $\mathcal{Y}=[-0.5,0.5]$, and the projection map given by $\pi(\boldsymbol{x})=\arctan(x_2/x_1)$ by identifying $\mathcal{Z}\simeq[0,2\pi]/\sim$ where $\{0\}\sim\{2\pi\}$.

Consider the sets $V_1=(0,2\pi)\setminus\{\pi\}$, and $V_2=[0,\epsilon)\cup(\pi-\epsilon,\pi+\epsilon)$ for $0<\epsilon<\pi/2$, and for $\theta=\pi(\boldsymbol{x})\in[0,2\pi)$

$$\psi_{1,\theta}^{-1}(\boldsymbol{x})=\frac{x_3}{\sin\theta},\quad\text{and}\quad\psi_{2,\theta}^{-1}(\boldsymbol{x})=\frac{1}{\cos\theta}\Big(\frac{x_1}{\cos\theta}-1\Big),$$

The collection $\{(V_j,\psi_j):j=1,2\}$ provides a local trivialization for $\mathcal{X}$. Additionally, the structure group $G=\{\mathrm{id}\}$ is simply the identity element since $\psi_{i,\theta}^{-1}\circ\psi_{j,\theta}=\mathrm{id}$ for all $\theta\in V_i\cap V_j$, and $i,j\in\{1,2\}$. It is straightforward to verify that for $\theta\in V_j\subset[0,2\pi)$, $t\in[-0.5,0.5]$, and $\boldsymbol{x}=\boldsymbol{x}(\theta,t)$, it follows that $\pi(\boldsymbol{x})=\theta$ and $\psi_{j,\theta}^{-1}(\boldsymbol{x})=t$.

Let $\nu\sim\mathrm{Unif}([-0.5,0.5])$ be the uniform distribution on $\mathcal{Y}=[-0.5,0.5]$ with $d\nu(t)=1\cdot dt$ for all $t\in[-0.5,0.5]$. For a fixed $\theta\in V_j$, the map $\psi_{j,\theta}$ pushes forward the measure ν_θ to the fiber $\pi^{-1}(\theta)$. The image $\psi_{j,\theta}([-0.5,0.5])$ is a line segment in $\mathbb{R}^3$ of length

$$\|\psi_{j,\theta}(0.5)-\psi_{j,\theta}(-0.5)\|=\big\|(\cos^2\theta,\sin\theta\cos\theta,\sin\theta)\big\|=1,$$

as shown in Figure 8 (A). Therefore, the pushforward measure, $\nu_\theta=(\psi_{j,\theta})_{\#}\nu$, is also a uniform distribution with $d\nu_\theta(\boldsymbol{x})=1$ for all $\boldsymbol{x}\in\pi^{-1}(\theta)$. Hence, the density f_α in Eq. (15) can be faithfully represented as

$$f_\alpha(\boldsymbol{x})=g\Big(\phi_\alpha(\psi_{j,\theta}^{-1}(\boldsymbol{x}),\theta)\Big)d\nu_\theta(\boldsymbol{x}).$$

The $\mathcal{F}$-equivalence of $\{f_\alpha:\alpha\in\mathbb{R}\}$ now follows from Example 5.9. ♢

Example A.3 (Family of distributions on $\mathbb{S}^2$)**.** Let $\mathcal{X}=\mathbb{S}^2\setminus\{+\boldsymbol{p},-\boldsymbol{p}\}$ be the surface of the unit sphere embedded in $\mathbb{R}^3$, excluding the two poles. This can be represented in polar coordinates as

$$\mathcal{X}=\Big\{\boldsymbol{x}(\theta,\varphi)=(\sin\theta\cos\varphi,\sin\theta\sin\varphi,\cos\theta)\in\mathbb{R}^3:\theta\in(0,\pi),\quad\textit{and}\quad\varphi\in[0,2\pi)\Big\}.$$

Let f_α be a probability density function on $\mathcal{X}$ given by

$$f_\alpha(\boldsymbol{x})=f_\alpha\big(\boldsymbol{x}(\varphi,\theta)\big)=\frac{1}{2\pi}\cdot g\bigg(\frac{\theta+\alpha\varphi/2\mod\pi}{\pi}\bigg),\tag{A.3}$$

where g is a fixed probability density function with $\mathrm{supp}(g)=(0,1)$ w.r.t. the Lebesgue measure. The family $\{f_\alpha:\alpha\in\mathbb{Z}\}$ admits $\mathcal{F}$-equivalence.

This is similar to Example 5.9, and can be understood by considering the fiber bundle representation of $\mathcal{X}$ with $\mathcal{Z}=(0,\pi)$, $\mathcal{Y}=\mathbb{S}^1\simeq[0,2\pi]/\sim$ and the projection map $\pi(\boldsymbol{x})=\arctan(x_2/x_1)$. The local trivialization is given by $\{(V,\psi)\}$, where

$$V=(0,\pi),\ \psi^{-1}(\boldsymbol{x})=\Big(\arctan(x_2/x_1),\arccos(x_3)\Big),\ \text{and}\ \psi_\theta^{-1}(\boldsymbol{x})=\arccos(x_3).$$

The uniform measure on $\mathcal{Y}$ is $d\nu(\varphi)=1/2\pi$, and for $\boldsymbol{x}\in\mathcal{X}$ with $\theta=\pi(\boldsymbol{x})$ the density in Eq. (A.3) can be rewritten as

$$f_\alpha(\boldsymbol{x})=C\cdot g\big(\phi\big(\psi_\theta^{-1}(\boldsymbol{x}),\theta\big)\big),$$

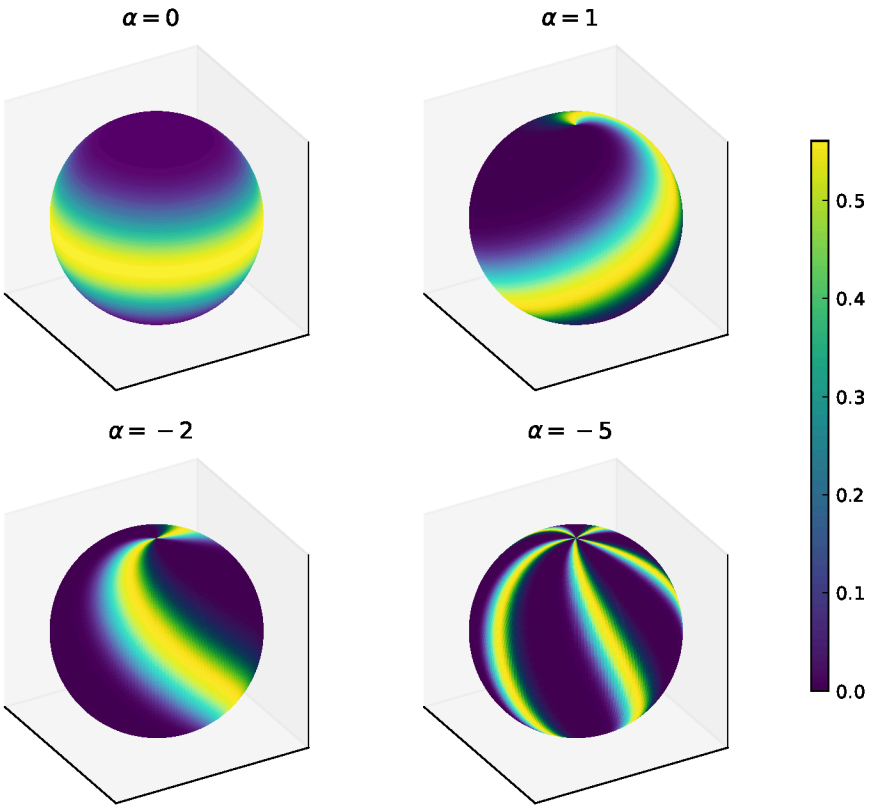


FIGURE 9. Superlevel sets of the density function f_α on $\mathcal{X} = \mathbb{S}^2 \setminus \{\boldsymbol{p}, -\boldsymbol{p}\}$ from Example A.3 when $g \sim \text{Beta}(10, 10)$ is the density function of a Beta distribution on $(0, 1)$, and for $\alpha \in \{0, 1, -2, -5\}$

where $C = 1/2\pi$, and the map $\phi(\varphi, \theta) = \theta + \alpha\varphi/2$ satisfies $\phi(\varphi, \cdot) \in \Delta((0, \pi))$ for all $\varphi \in [0, 2\pi)$. Since $\mathbb{S}^1 \simeq [0, 2\pi]/\sim$ requires $0 \sim 2\pi$, we have

$$\phi(0, \theta) = \phi(2\pi, \theta) \quad \iff \quad \pi\alpha \mod \pi = 0,$$

which holds whenever $\alpha \in \mathbb{Z}$. As g is a density w.r.t. the Lebesgue measure, the modular character is the identity map, and the Jacobian constraint in Eq. (10),

$$\begin{aligned}\int_0^{2\pi} C \cdot \left\|\mathbf{D}\phi^{-1}(\varphi, \cdot)\right\| d\nu(\varphi) &= \int_0^{2\pi} \left|\frac{d}{d\theta}(\theta - \alpha\varphi/2 \mod \pi)\right| \frac{1}{2\pi} d\varphi \\ &= \int_0^{2\pi} 1 \cdot \frac{1}{2\pi} d\varphi = 1,\end{aligned}$$

is satisfied for all $\alpha \in \mathbb{Z}$. From Theorem 5.4, we conclude that the family of distributions $\{f_\alpha : \alpha \in \mathbb{Z}\}$ admits $\mathcal{F}$-equivalence.

Furthermore, according to Remark 5.6 (ii), this family of distributions can be continuously extended to a family of distributions supported on $\mathbb{S}^2$. Figure 9 illustrates the superlevel sets $\{\boldsymbol{x} \in \mathcal{X} : f_\alpha(\boldsymbol{x}) \geq t\}$ for four different values of α when $g \sim \text{Beta}(10, 10)$ is the density function of a Beta distribution on $(0, 1)$. ♢

Acknowledgments. Kenji Fukumizu is supported in part by JST CREST JPMJCR15D3 and JSPS Grant-in-Aid for Transformative Research Areas (A) 22H05106. Satoshi Kuriki is partially supported by JSPS KAKENHI Grant Number JP16H02792. Bharath Sriperumbudur is partially supported by the National Science Foundation (NSF) grant DMS–1713011 and CAREER award DMS–1945396.

REFERENCES

[1] R. J. Adler, S. Agami and P. Pranav, Modeling and replicating statistical topology and evidence for CMB nonhomogeneity, *Proceedings of the National Academy of Sciences*, **114** (2017), 11878-11883.
[2] P. Bendich, J. S. Marron, E. Miller, A. Pieloch and S. Skwerer, Persistent homology analysis of brain artery trees, *The Annals of Applied Statistics*, **10** (2016), 198-218.
[3] C. A. N. Biscio, N. Chenavier, C. Hirsch and A. M. Svane, Testing goodness of fit for point processes via topological data analysis, *Electronic Journal of Statistics*, **14** (2020), 1024-1074.
[4] O. Bobrowski, Homological connectivity in random Čech complexes, *Probability Theory and Related Fields*, **183** (2022), 715-788.
[5] O. Bobrowski and R. J. Adler, Distance functions, critical points, and the topology of random Čech complexes, *Homology, Homotopy and Applications*, **16** (2014), 311-344.
[6] O. Bobrowski and M. Kahle, Topology of random geometric complexes: A survey, *Journal of Applied and Computational Topology*, **1** (2018), 331-364.
[7] O. Bobrowski and S. Mukherjee, The topology of probability distributions on manifolds, *Probability Theory and Related Fields*, **161** (2015), 651-686.
[8] K. Borsuk, On the imbedding of systems of compacta in simplicial complexes, *Fundamenta Mathematicae*, **35** (1948), 217-234.
[9] P. Bubenik and P. T. Kim, A statistical approach to persistent homology, *Homology, Homotopy and Applications*, **9** (2007), 337-362.
[10] V. H. Can and K. D. Trinh, Random connection models in the thermodynamic regime: Central limit theorems for add-one cost stabilizing functionals, *Electronic Journal of Probability*, **27** (2022), 1-40.
[11] E. G. Z. Centeno, G. Moreni, C. Vriend, L. Douw and F. A. N. Santos, A hands-on tutorial on network and topological neuroscience, *Brain Structure and Function*, **227** (2022), 741-762.
[12] M. K. Chung, S.-G. Huang, A. Gritsenko, L. Shen and H. Lee, Statistical inference on the number of cycles in brain networks, in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, IEEE, 2019, 113-116.
[13] V. Divol and T. Lacombe, Understanding the topology and the geometry of the space of persistence diagrams via optimal partial transport, *Journal of Applied and Computational Topology*, **5** (2021), 1-53.
[14] P. Dłotko, N. Hellmer, Ł. Stettner and R. Topolnicki, Topology-driven goodness-of-fit tests in arbitrary dimensions, *Statistics and Computing*, **34** (2024), 34, 23 pp.
[15] M. L. Eaton, *Group Invariance Applications in Statistics*, Institute of Mathematical Statistics and the American Statistical Association, Hayward, CA and Alexandria, VA, 1989.
[16] H. Edelsbrunner and J. L. Harer, *Computational Topology: An Introduction*, American Mathematical Society, Providence, RI, 2010.
[17] B. Efron and D. V. Hinkley, Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information, *Biometrika*, **65** (1978), 457-483.
[18] B. T. Fasy, F. Lecci, A. Rinaldo, L. Wasserman, S. Balakrishnan and A. Singh, Confidence sets for persistence diagrams, *The Annals of Statistics*, **42** (2014), 2301-2339.
[19] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, John Wiley & Sons, New York, 1999.
[20] M. Gameiro, Y. Hiraoka, S. Izumi, M. Kramar, K. Mischaikow and V. Nanda, A topological measurement of protein compressibility, *Japan Journal of Industrial and Applied Mathematics*, **32** (2015), 1-17.
[21] M. Ghosh, N. Reid and D. A. S. Fraser, Ancillary statistics: A review, *Statistica Sinica*, **20** (2010), 1309-1332.
[22] A. Goel, K. D. Trinh and K. Tsunoda, Strong law of large numbers for Betti numbers in the thermodynamic regime, *Journal of Statistical Physics*, **174** (2019), 865-892.
[23] A. Goetz, On measures in fibre bundles, in *Colloquium Mathematicum*, **7** (1959), 11-18.

[24] R. F. Gunst and J. T. Webster, Density functions of the bivariate chi-square distribution, *Journal of Statistical Computation and Simulation*, **2** (1973), 275-288.
[25] A. Hatcher, *Algebraic Topology*, Cambridge University Press, Cambridge, UK, 2002.
[26] Y. Hiraoka, T. Shirai and K. D. Trinh, Limit theorems for persistence diagrams, *The Annals of Applied Probability*, **28** (2018), 2740-2780.
[27] M. Kahle, Random geometric complexes, *Discrete & Computational Geometry*, **45** (2011), 553-573.
[28] M. Kahle and E. Meckes, Limit theorems for Betti numbers of random simplicial complexes, *Homology Homotopy Appl.*, **15** (2013), 343-374.
[29] H. Kamiya, A. Takemura and S. Kuriki, Star-shaped distributions and their generalizations, *Journal of Statistical Planning and Inference*, **138** (2008), 3429-3447.
[30] J. Krebs, B. Roycraft and W. Polonik, On approximation theorems for the Euler characteristic with applications to the bootstrap, *Electronic Journal of Statistics*, **15** (2021), 4462-4509.
[31] J. Krebs and W. Polonik, On the asymptotic normality of persistent Betti numbers, arXiv preprint, arXiv:1903.03280.
[32] C. J. Lloyd, E. J. Williams and P. S. Yip, Ancillary Statistics-II, *Encyclopedia of Statistical Sciences*, **1**.
[33] K. Meng, J. Wang, L. Crawfordc and A. Eloyan, Randomness of shapes and statistical inference on shapes via the smooth Euler characteristic transform, *Journal of the American Statistical Association*, (2024), 1-25.
[34] Y. Mileyko, S. Mukherjee and J. Harer, Probability measures on the space of persistence diagrams, *Inverse Problems*, **27** (2011), 124007.
[35] D. W. Müller and G. Sawitzki, Excess mass estimates and tests for multimodality, *Journal of the American Statistical Association*, **86** (1991), 738-746.
[36] S. Oudot and E. Solomon, Inverse problems in topological persistence, in *Topological Data Analysis: The Abel Symposium 2018*, Springer International Publishing, 2020, 405-433.
[37] T. Owada and A. M. Thomas, Limit theorems for process-level Betti numbers for sparse and critical regimes, *Advances in Applied Probability*, **52** (2020), 1-31.
[38] M. Penrose, *Random Geometric Graphs*, vol. 5, Oxford University Press, Oxford, UK, 2003.
[39] W. Polonik, Measuring mass concentrations and estimating density contour clusters-an excess mass approach, *The Annals of Statistics*, **23** (1995), 855-881.
[40] B. Roycraft, J. Krebs and W. Polonik, Bootstrapping persistent Betti numbers and other stabilizing statistics, *The Annals of Statistics*, **51** (2023), 1484-1509.
[41] J. A. Schouten, *Ricci–Calculus: An Introduction to Tensor Analysis and Its Geometrical Applications*, Springer–Verlag, Berlin, Germany, 1954.
[42] T. A. Severini, *Likelihood Methods in Statistics*, Oxford University Press, 2000.
[43] S. S. Sørensen, C. A. N. Biscio, M. Bauchy, L. Fajstrup and M. M. Smedskjaer, Revealing hidden medium-range order in amorphous materials using topological data analysis, *Science Advances*, **6** (2020), eabc2320.
[44] N. Steenrod, *The Topology of Fibre Bundles*, vol. 44, Princeton University Press, 1999.
[45] A. M. Thomas and T. Owada, Functional limit theorems for the Euler characteristic process in the critical regime, *Advances in Applied Probability*, **53** (2021), 57-80.
[46] K. D. Trinh, A remark on the convergence of Betti numbers in the thermodynamic regime, *Pacific Journal of Mathematics for Industry*, **9** (2017), 4, 7 pp.
[47] K. Turner, Y. Mileyko, S. Mukherjee and J. Harer, Fréchet means for distributions of persistence diagrams, *Discrete & Computational Geometry*, **52** (2014), 44-70.
[48] R. A. Wijsman, *Invariant Measures on Groups and Their Use in Statistics*, Institute of Mathematical Statistics, Hayward, CA, 1990.
[49] D. Yogeshwaran and R. J. Adler, On the topology of random complexes built over stationary point processes, *The Annals of Applied Probability*, **25** (2015), 3338-3380.
[50] D. Yogeshwaran, E. Subag and R. J. Adler, Random geometric complexes in the thermodynamic regime, *Probability Theory and Related Fields*, **167** (2017), 107-142.

Received June 2023; 1st revision February 2024; 2nd revision May 2024; early access August 2024.