The Annals of Applied Probability
2013, Vol. 23, No. 6, 2238–2258
DOI: 10.1214/12-AAP899

STICKY CENTRAL LIMIT THEOREMS ON OPEN BOOKS

By Thomas Hotz[1], Stephan Huckemann[2], Huiling Le, J. S. Marron[3], Jonathan C. Mattingly[4], Ezra Miller[5], James Nolen[6], Megan Owen[7], Vic Patrangenaru[8] and Sean Skwerer[3]

Ilmenau University of Technology, University of Göttingen, University of Nottingham, University of North Carolina at Chapel Hill, Duke University, Duke University, Duke University, University of Waterloo, Florida State University and University of North Carolina at Chapel Hill

Given a probability distribution on an open book (a metric space obtained by gluing a disjoint union of copies of a half-space along their boundary hyperplanes), we define a precise concept of when the Fréchet mean (barycenter) is *sticky*. This nonclassical phenomenon is quantified by a law of large numbers (LLN) stating that the empirical mean eventually almost surely lies on the (codimension 1 and hence measure 0) *spine* that is the glued hyperplane, and a central limit theorem (CLT) stating that the limiting distribution is Gaussian and supported on the spine. We also state versions of the LLN and CLT for the cases where the mean is nonsticky (i.e., not lying on the spine) and partly sticky (i.e., is, on the spine but not sticky).

Introduction. The mean of a finite set of points in Euclidean space moves slightly when one of the points is perturbed. This fluctuation is pervasive in classical probabilistic and statistical situations. In geometric contexts, the

Received February 2012; revised September 2012.
[1]Supported by DFG Grant CRC 803.
[2]Supported by DFG Grants CRC 755 and HU 1575/2.
[3]Supported by NSF Grant DMS-08-54908.
[4]Supported by NSF Grant DMS-08-54879.
[5]Supported by NSF Grants DMS-04-49102 = DMS-10-14112 and DMS-10-01437.
[6]Supported by NSF Grant DMS-10-07572.
[7]Supported by a desJardins Postdoctoral Fellowship in Mathematical Biology at the University of California, Berkeley.
[8]Supported by NSF Grants DMS-08-05977 and DMS-11-06935.
AMS 2000 subject classifications. 60B99, 60F05.
Key words and phrases. Fréchet mean, central limit theorem, law of large numbers, stratified space, nonpositive curvature.

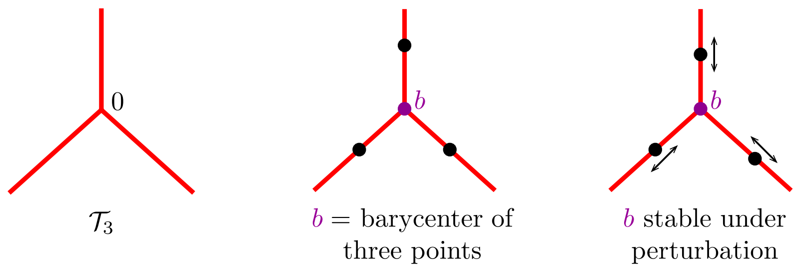

Fig. 1. *(left) The space of rooted phylogenetic trees with three leaves and fixed pendant edge lengths; (center) the probability distribution supported on three points in $\mathcal{T}_3$ equidistant from the vertex 0 has barycenter 0; (right) perturbing the distribution—and even macroscopically moving all three points a limited distance—leaves the barycenter fixed.*

barycenter (Fréchet mean [10], L^2-minimizer, least squares approximation), which minimizes the sum of the square distances to a given set of points, generalizes the notion of mean. Intuition from the Euclidean setting suggests that if the points are randomly sampled from a well-behaved probability distribution on a space M of dimension $d+1$, then the random variable that is the barycenter ought not be confined to a particular subspace of dimension d or less, if the distribution is generic. While this intuition has been made rigorous when M is a manifold [5, 12, 14, 15], it can fail when M has certain types of singularities, as we demonstrate here for an *open book* $\mathcal{O}$: a space obtained by gluing disjoint copies of a half-space along their boundary hyperplanes; see Section 1 for precise definitions.

Example 1. The simplest singular space is the 3-*spider*: a union $\mathcal{T}_3$ of three rays with their endpoints glued at a point 0 (Figure 1, left). This space $\mathcal{T}_3$ is the open book $\mathcal{O}$ of dimension 1 with three leaves. If three points are chosen equidistant from 0 on the different rays, then the barycenter lies at 0 by symmetry (Figure 1, center). The unexpected "sticky" phenomenon is that wiggling one or more of the points has no effect on the barycenter (Figure 1, right). For instance, if the points lie at radius r from 0, then the barycenter remains at 0 upon moving one of the points to radius at most $2r$.

Example 2. The name "open book" comes from the case of dimension 2, which looks like an ordinary open book, in the usual lay sense of the words; see Figure 2.

Our main goal is to define a precise concept of when a distribution on an open book has a *sticky mean* in Definition 2.10, and to quantify this highly nonclassical condition with a law of large numbers (LLN) in Theorem 4.3 and a central limit theorem (CLT) in Theorem 5.7. Roughly speaking, the

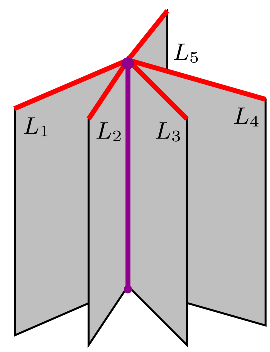

FIG. 2. *Open book of dimension 2 with five leaves. Ideally, the picture of this embedding would continue to infinity vertically, both up and down, as well as away from the spine on every leaf.*

sticky LLN says that in certain situations, empirical (sample) means almost surely eventually lie on the *spine*: the hyperplane shared by all of the glued half-spaces by virtue of the gluing. In Figure 1, the spine is the point 0. In Figure 2, the spine is the central line.

The phenomenon of the sticky mean contrasts with the classical LLN, where the empirical mean approaches the theoretical mean from all directions. The sticky CLT says that the limiting distribution is Gaussian and supported on the spine. Again, the nonclassical nature of this result contrasts with the classical CLT, in which the limiting distribution has full support rather than being supported on a thin (positive codimension and hence measure zero) subset of the sample space. Versions of the LLN and CLT are also stated in Theorems 4.3, 5.7 and 5.11 for the cases where the mean is:

- nonsticky—not lying on the spine—so the LLN and CLT behave classically; and
- partly sticky—on the spine but not sticky—so the LLN and CLT are hybrids of the sticky and nonsticky ones.

This paper is motivated by a desire to understand statistical sampling from topologically stratified spaces, including:

- shape spaces, representing equivalence classes of point configurations under operations such as rotation, translation, scaling, projective transformations, or other nonlinear transformations (e.g., see [9, 18, 19] for direct similarities, affine transformations, and projective transformations, resp.);
- spaces of covariance matrices, arising as data points in diffusion tensor imaging (see [1, 3, 6, 20, 21], e.g.); and
- tree spaces, representing metric phylogenetic trees on fixed sets of taxa (see [7, 16, 17], e.g.).

Open books are the simplest singular topologically stratified spaces. Roughly speaking, topologically stratified spaces decompose as finite disjoint unions of manifolds (*strata*) in such a way that the singularities of the total space are constant along each stratum (this is the structure described in [11], Section 1.4). Every topologically stratified space that is singular along a stratum of codimension 1 is, by definition of topological stratification, locally homeomorphic to an open book along that stratum. Therefore, to understand statistical sampling from arbitrary stratified spaces possessing singularities in maximal dimension, it is first necessary to understand sampling from open books.

The metrics on open books that appear as local pieces of arbitrary stratified spaces are arbitrary. However, sticky means on open books seem to stem from topological phenomena, rather than geometric ones, so we consider only the simplest metric on $\mathcal{O}$: each half-space has the Euclidean metric and the boundaries are glued isometrically. Although this restriction is substantial, these "Euclidean" open books occur in applications. For instance, the space $\mathcal{T}_3$ from the first example above parametrizes all rooted (metric) phylogenetic trees with three taxa and fixed pendant edge lengths. More generally, open books of arbitrary dimension and precisely three leaves reflect the local structure of phylogenetic tree space near any point on a stratum of codimension 1; such a point represents a tree possessing a node with nonbinary branching. Observations of "unresolved" (i.e., nonbinary) trees as barycenters of biologically meaningful samples (see [16], Examples 5.5 and 5.6, for descriptions of cases involving yeast phylogenies and brain arteries) constituted crucial motivation for the present study.

The relation between open books and tree spaces is that of local to global. After completing an early draft of this paper we found that Basrak [2] had independently and simultaneously proved a sticky CLT for certain global situations in dimension 1, namely arbitrary binary trees: connected graphs with no cycles where each node is incident to at most three edges. In contrast, our dimension 1 results are local, in that all edges meet, but there can be more than three incident to the intersection.

It bears mentioning that in contrast to their behavior in open books, barycenters do not stick to thin subspaces of shape spaces, or to thin subspaces of more general quotients of manifolds by isometric proper actions of Lie groups [13]. The differentiating property amounts to curvature: open books are, in a precise sense, negatively curved at the spine, whereas passing to the quotient in the construction of shape spaces adds positive curvature. Basrak's binary trees [2] are negatively curved in the same way that open books or spaces of trees are [7]: they are *globally nonpositively curved*. (We recommend Sturm's exposition of this condition [22], particularly for its clarity regarding connections between probability and geometry, which was both a theoretical starting point and a source of inspiration for our developments here.) It is a principal long-term goal of our investigations to tease out the connection between stickiness of means of probability distributions with values in metric spaces and notions of negative curvature.

1. Open books. Set $S = \mathbb{R}^d$, the real vector space of dimension d with the standard Euclidean metric. If $\mathbb{R}_{\geq 0} = [0, \infty)$ is the closed nonnegative ray in the real line, then the closed half-space

$$\overline{H}_+ = \mathbb{R}_{\geq 0} \times S$$

is a metric subspace of $\mathbb{R}^{d+1} = \mathbb{R} \times S$ with boundary S which we identify with $H = \{0\} \times S$, and interior $H_+ = \mathbb{R}_{>0} \times S$. The *open book* $\mathcal{O}$ is the quotient of the disjoint union $\overline{H}_+ \times \{1, \ldots, K\}$ of K closed half-spaces modulo the equivalence relation that identifies their boundaries. Therefore $p = (x, k) = (x^{(0)}, x^{(1)}, \ldots, x^{(d)}, k)$ is identified with $q = (y, j) = (y^{(0)}, y^{(1)}, \ldots, y^{(d)}, j)$ whenever $x^{(0)} = 0 = y^{(0)}$ and $x^{(i)} = y^{(i)}$ for all $i \in \{0, \ldots, d\}$, regardless of k and j. The following definition summarizes and introduces terminology.

DEFINITION 1.1 (Leaves and spine). The open book $\mathcal{O}$ consists of $K \geq 3$ *leaves* L_k, for $k = 1, \ldots, K$, each of dimension $d + 1$ and defined by

$$L_k = \overline{H}_+ \times \{k\}.$$

The leaves are joined together along the *spine* L_0 which comprises the equivalence classes in $\bigcup_{k=1}^{K} (H \times \{k\})$, that is, L_0 can be identified with the hyperplane $H = \{0\} \times S$ or with the space $S = \mathbb{R}^d$. Thus, the open book $\mathcal{O}$ is the disjoint union

$$\mathcal{O} = L_0 \cup L_1^+ \cup \cdots \cup L_K^+ \tag{1.1}$$

of the spine L_0 and the interiors $L_k^+ = L_k \smallsetminus L_0$ of the leaves, $k = 1, \ldots, K$. Figure 2 illustrates an open book with $d = 1$ and $K = 5$.

When we speak of the spine in the following, we make clear which of these three *instances of the spine* we have in mind. The following diagram gives an overview of these instances, spaces and mappings introduced further below in Definitions 2.4, 3.4, 5.2 and in the proof of Lemma 3.5.

$$\begin{array}{lcccccccc}
\mathcal{O} & \supseteq & L_k & \supseteq & L_0 & \subseteq & L_k & \subseteq & \mathcal{O} \\
\downarrow F_k & & \updownarrow & & \updownarrow F_k|_{L_0} & & & & \\
\mathbb{R}^{d+1} & \xrightarrow{\hat{P}} & \overline{H}_+ & \supseteq & H & & & P_S = \pi_S \circ F_k & \\
& & & & \updownarrow \pi_S|_H & & & & \\
& & & & S = \mathbb{R}^d & & & &
\end{array}$$

DEFINITION 1.2 (Reflection). For a given point $x \in \overline{H}_+$, let $Rx \in \overline{H}_- = \mathbb{R}_{\leq 0} \times \mathbb{R}^d = (-\infty, 0] \times \mathbb{R}^d$ denote its *reflection* across the hyperplane $\overline{H}_+ \cap \overline{H}_- = \{0\} \times S$.

The metric d on $\mathcal{O}$ is expressed in terms of reflection in a natural way: given two points $p, q \in \mathcal{O}$, with $p = (x, k)$ and $q = (y, j)$,

$$d(p,q) = \begin{cases} |x-y|, & \text{if } k = j, \\ |x - Ry|, & \text{if } k \neq j, \end{cases} \tag{1.2}$$

where $|x - y|$ denotes Euclidean distance on $\mathbb{R}^{d+1}$. Note that if $k \neq j$ in equation (1.2), then $d(p,q) = 0$ if and only if x and y lie on the spine and coincide. Our assumption $K \geq 3$ implies that $\mathcal{O}$ is not isometric to a subset of $\mathbb{R}^{d+1}$ (as it would be for $K \leq 2$).

The next lemma refers to *globally nonpositive curvature*. See [22] for a definition and background. The only times we apply this concept here are in noting the uniqueness of barycenters in our context (see Definition 3.1 and the line following it) and to obtain a quick proof of a strong law of large numbers (Lemma 4.2).

LEMMA 1.3. *The open book $(\mathcal{O}, d)$ is a Hausdorff metric space that is globally nonpositively curved, and its spine is isometric to $\mathbb{R}^d$.*

PROOF. [22], Example 3.3. □

REMARK 1.4. Although the open book $\mathcal{O}$ is not a vector space over $\mathbb{R}$, scaling by a positive constant $\lambda \in \mathbb{R}_{\geq 0}$ is defined in the natural way:

$$\lambda p = (\lambda x, k) \qquad \text{for all } p = (x,k) \in \mathcal{O}.$$

The open book also carries an action of the spine S, considered as an additive group, by translation, via the action of S on each leaf:

$$\mathcal{O} \ni p = (x^{(0)}, x^{(1)}, \dots, x^{(d)}, k) \xrightarrow{z} (x^{(0)}, x^{(1)} + z^{(1)}, \dots, x^{(d)} + z^{(d)}, k) \in \mathcal{O},$$

with $z = (z^{(1)}, \dots, z^{(d)}) \in S$. For the above right-hand side we write simply $z + p$.

2. Probability measures on the open book. Our goal is to understand the statistical behavior of points sampled randomly from $\mathcal{O}$. Suppose that μ is a Borel probability measure on $\mathcal{O}$. We assume throughout the paper that $d(0,q)$ has bounded expectation under the measure μ,

$$\int_{\mathcal{O}} d(0,q)\, d\mu(q) < \infty. \tag{2.1}$$

When explicitly stated, we also assume the stronger condition

$$\int_{\mathcal{O}} d(0,q)^2\, d\mu(q) < \infty, \tag{2.2}$$

of square integrability.

LEMMA 2.1. *Any Borel probability measure μ on the open book $\mathcal{O}$ decomposes uniquely as a weighted sum of Borel probability measures μ_k on the* *open leaves L_k^+ and a Borel probability measure μ_0 on the spine L_0. More precisely, there are nonnegative real numbers $\{w_k\}_{k=0}^K$ summing to 1 such that, for any Borel set $A \subseteq \mathcal{O}$, the measure μ takes the value*

$$\mu(A) = w_0\mu_0(A \cap L_0) + \sum_{k=1}^{K} w_k\mu_k(A \cap L_k^+).$$

Proof. This follows from the decomposition (1.1) and the additivity of measures on disjoint sets. □

Remark 2.2. For $k \geq 1$, $w_k = \mu(L_k^+)$ is the probability that a random point lies in L_k^+, while $w_0 = \mu(L_0)$ is the probability that a point lies somewhere on the spine.

Assumption 2.3. Throughout this paper, assume the nondegeneracy condition

$$w_k = \mu(L_k^+) > 0 \qquad \text{for all } k \in \{1, \dots, K\}. \tag{2.3}$$

Otherwise, we would remove those leaves L_k for which $\mu(L_k^+) = 0$ from the open book. Nondegeneracy implies that $w_0 < 1$ and $0 < w_k < 1$ for all $k \geq 1$ in the decomposition from Lemma 2.1.

Definition 2.4 (Folding map). For $k \in \{1, \dots, K\}$ the kth *folding map* $F_k : \mathcal{O} \to \mathbb{R}^{d+1}$ sends $p \in \mathcal{O}$ to

$$F_k p = \begin{cases} x, & \text{if } p = (x, k) \in L_k, \\ Rx, & \text{if } p = (x, j) \in L_j \text{ and } j \neq k, \end{cases}$$

where the reflection operator R was defined in Definition 1.2.

Remark 2.5. In the definition of the folding map F_k, the leaf L_k is identified with the subset $\overline{H}_+ \subset \mathbb{R}^{d+1}$, by slight abuse of notation (again). The other leaves L_j are collapsed to the negative half-space $\overline{H}_- \subset \mathbb{R}^{d+1}$ via the reflection map. All of these identifications have the same effect on the spine S, which becomes the hyperplane $H = \{0\} \times \mathbb{R}^d \subset \mathbb{R}^{d+1}$. For example, F_4 takes the picture in Figure 2 to $\mathbb{R}^2$ as in Figure 3.

The notations H_+ and H_- (with no bars) are reserved for the *strictly positive* and *strictly negative* open half-spaces that are the interiors of $\overline{H}_+$ and $\overline{H}_-$, respectively.

Lemma 2.6. *Under the folding map F_k, the measure μ pushes forward to a measure $\tilde{\mu}_k = \mu \circ F_k^{-1}$ on $\mathbb{R}^{d+1}$ such that, given a Borel subset $A \subseteq \mathbb{R}^{d+1}$,*

$$\tilde{\mu}_k(A) = w_k\mu_k(A \cap \overline{H}_+) + w_0\mu_0(A \cap S) + \sum_{\substack{j \geq 1 \\ j \neq k}} w_j\mu_j(A \cap H_-).$$

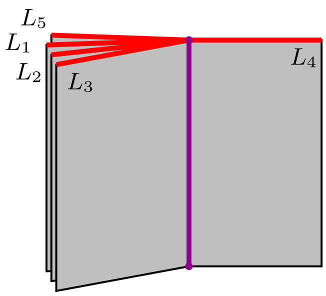

FIG. 3. *The 4th folding map identifies leaf L_4 with the half-space $\bar{H}_+$ and identifies all other leaves L_j for $j \neq k$ with the half-space $\bar{H}_-$.*

PROOF. Lemma 2.1. $\square$

DEFINITION 2.7 (First moment on a leaf). Let $x^{(0)}, \dots, x^{(d)}$ be the coordinate functions on $\mathbb{R}^{d+1}$. The *first moment* of the measure μ on the kth leaf L_k is the real number

$$m_k = \int_{\mathbb{R}^{d+1}} x^{(0)}\, d\tilde{\mu}_k(x) = \int_{\mathcal{O}} (\pi_0 F_k p)\, d\mu(p),$$

where $\pi_0 : \mathbb{R}^{d+1} \to \mathbb{R}$ is the orthogonal projection with kernel $H = \{0\} \times \mathbb{R}^d$.

REMARK 2.8. For any point $p \in \mathcal{O}$, the projection $\pi_0 F_k p$ is positive if $p \in L_k^+$ and negative if $p \in L_j^+$ for some $j \neq k$. Moreover, $|\pi_0 F_k p| = |x^{(0)}|$ is the distance of p from the spine. The integrability in equation (2.1) guarantees that the first moments of μ are all finite.

THEOREM 2.9. *Under integrability (2.1) and nondegeneracy (2.3), either:*

(1) *$m_j < 0$ for all indices $j \in \{1, \dots, K\}$, or there is exactly one index $k \in \{1, \dots, K\}$ such that $m_k \geq 0$, in which case either:*
(2) *$m_k > 0$, or*
(3) *$m_k = 0$.*

PROOF. For $k = 1, \dots, K$, let

$$v_k = \int_{H_+} x^{(0)}\, d\mu_k(x).$$

The nondegeneracy (2.3) implies that $v_k > 0$. Observe that

$$m_k = w_k v_k - \sum_{\substack{j \geq 1 \\ j \neq k}} w_j v_j.$$

For any $j \neq k \in \{1,\dots,K\}$,

$$m_j = w_j v_j - \sum_{\substack{\ell \geq 1 \\ \ell \neq j}} w_\ell v_\ell \leq w_j v_j - w_k v_k \leq \left(\sum_{\substack{\ell \geq 1 \\ \ell \neq k}} w_\ell v_\ell \right) - w_k v_k = -m_k,$$

since the weights w_ℓ are nonnegative. Therefore, if $m_k > 0$ for some k, then $m_j \leq -m_k < 0$ for all $j \neq k$. Also, if $m_k = 0$ for some index k, then $m_j \leq 0$ for all $j \neq k$.

Now suppose there are two indices $j, k \in \{1,\dots,K\}$ such that $j \neq k$ and $m_j = 0$ and $m_k = 0$. Then

$$0 = m_j = w_j v_j - w_k v_k - \sum_{\substack{\ell \geq 1 \\ \ell \neq j,k}} w_\ell v_\ell$$

and

$$0 = m_k = w_k v_k - w_j v_j - \sum_{\substack{\ell \geq 1 \\ \ell \neq j,k}} w_\ell v_\ell.$$

Adding these two equalities results in

$$0 = m_j + m_k = -2 \sum_{\substack{\ell \geq 1 \\ \ell \neq j,k}} w_\ell v_\ell.$$

Since $w_\ell v_\ell \geq 0$, it follows that $w_\ell v_\ell = 0$ for all $i \neq j, k$. Consequently, $\mu(L_\ell^+) = 0$ for all $\ell \neq j, k$. However, this contradicts nondegeneracy (2.3) and the fact that $K \geq 3$. Hence at most one of the numbers m_k can be nonnegative. $\square$

Motivated by Theorem 4.3 and Corollary 4.4, we use the following terms to describe the three mutually-exclusive conditions given in Theorem 2.9.

Definition 2.10. Under integrability (2.1) and nondegeneracy (2.3), we say that the mean of the measure μ is either:

(1) *sticky* if $m_j < 0$ for all indices $j \in \{1,\dots,K\}$, or
(2) *nonsticky* if $m_k > 0$ for some (unique) $k \in \{1,\dots,K\}$, or
(3) *partly sticky* if $m_k = 0$ for some (unique) $k \in \{1,\dots,K\}$.

Remark 2.11. If square integrability (2.2) also holds, the first moment m_k may be identified with the partial derivative

$$m_k = -\frac{\partial \Gamma_k}{\partial x^{(0)}}(x)\Big|_{x^{(0)}=0}$$

where $\Gamma_k : \mathbb{R}^{d+1} \to \mathbb{R}$ is defined by

$$\Gamma_k(x) = \frac{1}{2} \int_{\mathbb{R}^{d+1}} |x - y|^2 \, d\tilde{\mu}_k(y).$$

Observe that $-\frac{\partial \Gamma_k}{\partial x^{(0)}}(x)$ depends on $x^{(0)}$, but not on $(x^{(1)}, \dots, x^{(d)})$.

3. Sample means. For any finite collection of points $\{p_n\}_{n=1}^N \subset \mathcal{O}$, the Fréchet mean is a natural generalization of the arithmetic mean in Euclidean space:

DEFINITION 3.1. The *Fréchet mean*, or *barycenter*, of a set $\{p_n\}_{n=1}^N \subset \mathcal{O}$ of points is

$$b(p_1, \dots, p_N) = \operatorname*{arg\,min}_{p \in \mathcal{O}} \left(\sum_{n=1}^{N} d(p, p_n)^2 \right).$$

By Lemma 1.3 and [22], Proposition 4.3, the barycenter $b(p_1, \dots, p_N) \in \mathcal{O}$ exists and is unique.

DEFINITION 3.2. For fixed $k \in \{1, \dots, K\}$, the point $\eta_{k,N} \in \mathbb{R}^{d+1}$ defined by

$$\eta_{k,N} = \frac{1}{N} \sum_{n=1}^{N} F_k p_n \tag{3.1}$$

is the kth *folded average*: the barycenter of the pushforward under the kth folding map.

For a set of points $\{p_n\}_{n=1}^N \subset \mathcal{O}$, the condition $b(p_1, \dots, p_N) \in L_0$ does not necessarily imply $\eta_{k,N} \in H$. Nevertheless, the following lemma establishes an important relationship between $b(p_1, \dots, p_N)$ and $\eta_{k,N}$. Specifically, taking barycenters commutes with the kth folding in two cases: if the barycenter lies off the spine in L_k^+; or if the kth folded average lies in the closure of the positive half-space.

LEMMA 3.3. *Let $\{p_n\}_{n=1}^N \subset \mathcal{O}$ and $b_N = b(p_1, \dots, p_N)$. If $b_N \in L_k^+$, then $\eta_{k,N} \in H_+$ and $\eta_{k,N} = F_k b_N$. If $\eta_{k,N} \in \overline{H}_+$, then $b_N \in L_k$ and $F_k b_N = \eta_{k,N}$ (i.e. $b_N = (\eta_{k,N}, k)$).*

PROOF. Let $k, \ell \in \{1, \dots, K\}$. If $p \in L_k$, then $d(p, p_n) = |F_k p - F_k p_n|$. Therefore, if $b_N \in L_k^+$, then

$$b_N = \operatorname*{arg\,min}_{p \in \mathcal{O}} \sum_{n=1}^{N} d(p, p_n)^2 = \operatorname*{arg\,min}_{p \in L_k^+} \sum_{n=1}^{N} |F_k p - F_k p_n|^2.$$

Since F_k is continuously bijective from L_k to $\overline{H}_+$, this implies that the function

$$z \mapsto \sum_{n=1}^{N} |z - F_k p_n|^2$$

attains a local minimum in the open set H_+. However, this functional has only one local minimizer, which must be the unique global minimizer $\eta_{k,N}$,

$$\eta_{k,N} = \operatorname*{arg\,min}_{z \in \mathbb{R}^{d+1}} \sum_{n=1}^{N} |z - F_k p_n|^2.$$

Consequently, $\eta_{k,N} \in H_+$ and hence $F_k b_N = \eta_{k,N}$.

If $b_N \notin L_k$, then $b_N \in L_\ell^+$ for some $\ell \neq k$. Hence $\eta_{\ell,N} = F_\ell b_N$, as we have shown. In particular, $\eta_{\ell,N} \in H_+$ and $\pi_0 \eta_{\ell,N} > 0$. Hence

$$\sum_{p_n \in L_\ell^+} \pi_0 F_\ell p_n > - \sum_{p_n \notin L_\ell^+} \pi_0 F_\ell p_n \geq - \sum_{p_n \in L_k} \pi_0 F_\ell p_n = \sum_{p_n \in L_k} \pi_0 F_k p_n. \tag{3.2}$$

Observe that

$$\begin{aligned}
\pi_0 \eta_{k,N} = \frac{1}{N} \sum_{n=1}^{N} \pi_0 F_k p_n &\leq \frac{1}{N} \sum_{p_n \in L_k} \pi_0 F_k p_n + \frac{1}{N} \sum_{p_n \in L_\ell^+} \pi_0 F_k p_n \\
&= \frac{1}{N} \sum_{p_n \in L_k} \pi_0 F_k p_n - \frac{1}{N} \sum_{p_n \in L_\ell^+} \pi_0 F_\ell p_n.
\end{aligned}$$

Because of equation (3.2), this last expression is negative. Hence, we have shown that $b_N \notin L_k$ implies $\eta_{k,N} \in H_-$. Therefore, if $\eta_{k,N} \in \overline{H}_+$ it must be that $b_N \in L_k$. Consequently, as above,

$$\begin{aligned}
b_N &= \operatorname*{arg\,min}_{p \in \mathcal{O}} \sum_{n=1}^{N} d(p, p_n)^2 \\
&= \operatorname*{arg\,min}_{p \in L_k} \sum_{n=1}^{N} |F_k p - F_k p_n|^2 \\
&= F_k^{-1} \left(\operatorname*{arg\,min}_{z \in \overline{H}_+} \sum_{n=1}^{N} |z - F_k p_n|^2 \right) \\
&= F_k^{-1} \eta_{k,N}.
\end{aligned}$$

Note that $F_k^{-1} \eta_{k,N}$ is well defined, since $\eta_{k,N} \in \overline{H}_+$. $\square$

DEFINITION 3.4. Given a point $p=(x,j)=(x^{(0)},x^{(1)},\dots,x^{(d)},j)\in\mathcal{O}$,

$$P_S p=(x^{(1)},\dots,x^{(d)})\in S$$

is the orthogonal projection of p onto the spine S.

The following lemma shows that taking barycenters commutes with projection to the spine.

LEMMA 3.5. *If* $\{p_n\}_{n=1}^N\subset\mathcal{O}$ *and*

$$\bar{y}_N=\frac{1}{N}\sum_{n=1}^N P_S p_n,$$

then $\bar{y}_N=P_S b(p_1,\dots,p_N)$.

PROOF. Let $\pi_S:\mathbb{R}^{d+1}\to\mathbb{R}^d$ be the orthogonal projection onto the last d coordinates. Let $b_N=b(p_1,\dots,p_N)$. If $b_N\in L_k^+$ for some k, then $\eta_{k,N}=F_k b_N$ by Lemma 3.3. Therefore, since $P_S p=\pi_S F_k p$ for all $p\in\mathcal{O}$,

$$P_S b_N=\pi_S F_k b_N=\pi_S\eta_{k,N}=\frac{1}{N}\sum_{n=1}^N\pi_S F_k p_n=\frac{1}{N}\sum_{n=1}^N P_S p_n=\bar{y}_N.$$

On the other hand, if $b_N\in L_0$ then by definition of b_N,

$$b_N=\operatorname*{arg\,min}_{p\in L_0}\sum_{n=1}^N d(p,p_n)^2=\operatorname*{arg\,min}_{p\in S}\sum_{n=1}^N(|\pi_0 p_n|^2+|p-P_S p_n|^2).$$

Therefore $P_S b_N=\arg\min_{y\in\mathbb{R}^d}\sum_{n=1}^N|y-P_S p_n|^2=\frac{1}{N}\sum_{n=1}^N P_S p_n=\bar{y}_N$, as desired. $\square$

4. Random sampling and the law of large numbers. We now consider points $\{p_n\}_{n=1}^N$ sampled independently at random from a Borel probability measure μ on $\mathcal{O}$; we wish to understand the statistical behavior of their barycenter for large N. More precisely, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, and for each integer $n\ge1$ let $p_n(\omega):\Omega\to\mathcal{O}$ for fixed $\omega\in\Omega$ be a random point in $\mathcal{O}$.

Assume for all $n\ge1$ that $p_1,\dots,p_n$ are independent random variables and that for any Borel set $A\subseteq\mathcal{O}$,

$$\mathbb{P}(p_n\in A)=\mathbb{P}(\{\omega\in\Omega\mid p_n(\omega)\in A\})=\mu(A).$$

The sample space Ω may be constructed as the set of infinite sequences $(p_1,p_2,p_3,\dots)$ of points in $\mathcal{O}$ endowed with the product measure $\mathbb{P}=\prod_{n=1}^\infty\mu(p_n)$ on the σ-algebra $\mathcal{F}$ generated by cylinder sets. Observe that the folded points $\{F_k p_n(\omega)\}_{n=1}^\infty\subset\mathbb{R}^{d+1}$ are independent, each distributed according to $\tilde{\mu}_k$.

DEFINITION 4.1. For any positive integer N, let $b_N(\omega) = b(p_1, \ldots, p_N)$ denote the barycenter of the random sample $\{p_1(\omega), \ldots, p_N(\omega)\}$. This random point in $\mathcal{O}$ is the *empirical mean* of the distribution μ. Similarly, for $k \in \{1, \ldots, K\}$, the random point $\eta_{k,N}(\omega) \in \mathbb{R}^{d+1}$ denotes the kth *folded average* of the random sample $\{p_1(\omega), \ldots, p_N(\omega)\}$, as defined by (3.1).

The goal is to understand the statistical behavior of empirical means b_N as $N \to \infty$.

LEMMA 4.2 (Strong law of large numbers). *There is a unique point $\bar{b} \in \mathcal{O}$ such that*

$$\lim_{N\to\infty} b_N(\omega) = \bar{b}$$

holds $\mathbb{P}$-almost surely. If the square integrability condition (2.2) also holds, the limit $\bar{b}$ is the Fréchet mean *(or* barycenter*) of μ,*

$$\bar{b} = \operatorname*{arg\,min}_{p\in\mathcal{O}} \int_{\mathcal{O}} d(p,q)^2 \, d\mu(q).$$

PROOF. This is a special case of [22], Proposition 6.6, whose generality occurs in the context of distributions on globally nonpositively curved spaces. (An elementary proof from scratch is also possible, using arguments similar to the proof of Theorem 4.3. In general on metric spaces, there can be more than one Fréchet mean, and there are corresponding set-valued strong laws [4, 23].) $\square$

THEOREM 4.3 (Sticky LLN). *Assume nondegeneracy (2.3).*

(1) *If the moment m_j satisfies $m_j < 0$, then there is a random integer $N^*(\omega)$ such that $b_N(\omega) \notin L_j^+$ for all $N \geq N^*(\omega)$ holds $\mathbb{P}$-almost surely. Furthermore, $\bar{b} \notin L_j^+$.*

(2) *If the moment m_k satisfies $m_k > 0$, then there is a random integer $N^*(\omega)$ such that $b_N(\omega) \in L_k^+$ for all $N \geq N^*(\omega)$ holds $\mathbb{P}$-almost surely. Furthermore, $\bar{b} \in L_k^+$.*

(3) *If the moment m_k satisfies $m_k = 0$, then there is a random integer $N^*(\omega)$ such that $b_N(\omega) \in L_k$ for all $N \geq N^*(\omega)$ holds $\mathbb{P}$-almost surely. Furthermore, $\bar{b} \in L_0$.*

PROOF. By the usual strong law of large numbers,

$$\lim_{N\to\infty} \eta_{k,N} = \bar{\eta}_k = \int_{\mathbb{R}^{d+1}} x \, d\tilde{\mu}_k(x)$$

holds $\mathbb{P}$-almost surely. Observe that $m_k = \pi_0 \bar{\eta}_k$. Therefore, if $m_k > 0$, $\bar{\eta}_k \in H_+$ and $\eta_{k,N} \in H_+$ for all sufficiently large N. In that case, $b_N \in L_k^+$ for all sufficiently large N by Lemma 3.3. In fact, $\pi_0 b_N = \pi_0 \eta_{k,N} > m_k/2 > 0$ for N sufficiently large, so by virtue of Lemma 4.2, $\bar{b} \in L_k^+$. The same argument starting with $m_k \geq 0$ proves the case $m_k = 0$. On the other hand, if $m_j < 0$, then $\eta_{j,N} \in H_-$ for all sufficiently large N; Lemma 3.3 implies that $b_N \notin L_j^+$ for all sufficiently large N, and $\bar{b} \notin L_j^+$. $\square$

As a consequence, if the mean of μ is sticky then the empirical mean b_N sticks to the spine $L_0 \subset \mathcal{O}$ for all sufficiently large N, in the following sense.

COROLLARY 4.4. *If the mean of μ is sticky, then there is a random integer $N^*(\omega)$ such that $b_N(\omega) \in L_0$ for all $N \geq N^*(\omega)$ holds $\mathbb{P}$-almost surely. Moreover, $\bar{b} \in L_0$. If the mean of μ is partly sticky, with $m_k = 0$, then then there is a random integer $N^*(\omega)$ such that $b_N(\omega) \in L_k$ for all $N \geq N^*(\omega)$ holds $\mathbb{P}$-almost surely. Moreover, $\bar{b} \in L_0$.*

Recall that P_S is the orthogonal projection onto the spine S. The measure μ pushes forward along the projection to a measure $\mu_S = \mu \circ P_S^{-1}$ on S,

$$\mu_S(A) = \mu(P_S^{-1} A)$$

for any Borel set $A \subseteq \mathbb{R}^d$. Note that $\mu_0(A) \leq \mu_S(A)$ for all Borel sets $A \subseteq S$, but $\mu_S \neq \mu_0$ by Assumption 2.3.

COROLLARY 4.5. *In all cases (sticky, nonsticky, partly sticky), the limit $\bar{b} \in \mathcal{O}$ satisfies*

$$P_S \bar{b} = \int_S y \, d\mu_S(y). \tag{4.1}$$

PROOF. By Lemma 3.5 and Theorem 4.3,

$$P_S \bar{b} = P_S \lim_{N \to \infty} b_N = \lim_{N \to \infty} \bar{y}_N$$

holds almost surely. By the strong law of large numbers for $\bar{y}_N \in S = \mathbb{R}^d$, the last limit is (4.1). $\square$

5. Central limit theorems. In this section we consider fluctuations of the empirical mean $b_N(\omega)$ about the asymptotic limit $\bar{b}$, within the tangent cone at $\bar{b}$. We have shown that if the mean is either sticky or partly sticky, then $\bar{b} \in S$, and the tangent cone at $\bar{b}$ is an open book $\mathcal{O}$. On the other hand, if the mean is nonsticky, with $m_k > 0$, then $\bar{b}$ is in the interior of the leaf L_k^+ and the tangent cone at $\bar{b}$ is the vector space $\mathbb{R}^{d+1}$. We treat these two scenarios separately.

These facts essentially follow from Theorem 4.3 which shows that in the sticky cases with probability one the fluctuations away from the mean in certain directions stop as more random variables are added to the empirical mean. In particular, this implies that the correctly normalized limit of the fluctuation from the mean cannot, in the sticky case, converge to a Gaussian random variable as one would have in the standard central limit theorem.

Since the fluctuations in some directions are exactly zero at some point along each sequence of random variables, it is not all together surprising that limiting measure has mass concentrated on a lower dimensional set. This is the content of Theorem 5.7 which is the principal result of this section.

5.1. *The sticky central limit theorem.* Throughout this section, assume $m_j \leq 0$ for all $j \in \{1, \dots, K\}$. Hence $\bar{b} \in L_0$, and the mean is either sticky or partially sticky. In the partially sticky case, denote by k the unique index satisfying $m_k = 0$. The central limit theorem involves a centered and rescaled empirical mean.

Definition 5.1 (Rescaled empirical mean). Assume that $P_S \bar{b} = 0$ (after the action of $-P_S \bar{b} \in S$ on $\mathcal{O}$ as explained in Remark 1.4 if necessary). The *rescaled empirical mean* is the random variable $\sqrt{N} b_N \in \mathcal{O}$. Write ν_N for its induced probability law on $\mathcal{O}$,

$$\mathbb{P}(\{\omega \mid \sqrt{N} b_N(\omega) \in A\}) = \int_{\mathcal{O} \cap A} d\nu_N(p)$$

for all Borel sets $A \subseteq \mathcal{O}$.

Since in sticky settings, we need to collapse fluctuations in some directions back to the spine, it is convenient to define the following projection.

Definition 5.2. The convex projection $\hat{P}$ of $\mathbb{R}^{d+1}$ onto $\overline{H}_+$ is

$$\hat{P}x = \begin{cases} (0, x^{(1)}, \dots, x^{(d)}), & \text{if } x^{(0)} < 0, \\ (x^{(0)}, x^{(1)}, \dots, x^{(d)}), & \text{if } x^{(0)} \geq 0. \end{cases}$$

We now define measures which we will see shortly describe the limiting behaviors of ν_N as $N \to \infty$. In short, they are the limiting measures in the central limit theorem given in Theorem 5.7 below.

Definition 5.3. Assume square integrability (2.2) and assume that $P_S \bar{b} = 0$.

(1) The *spinal limit measure* g_S is the law of a multivariate normal random variable on the spine $S \cong \mathbb{R}^d$ with mean zero and covariance matrix

$$C_S = \int_{\mathbb{R}^d} y y^T \, d\mu_S(y) = \int_{\mathcal{O}} (P_S p)(P_S p)^T \, d\mu(p).$$

(2) The kth *costal* [9] *limit measure* g_k is the law of a multivariate normal random variable on $\mathbb{R}^{d+1}$ with mean zero and covariance matrix

$$C_k = \int_{\mathbb{R}^{d+1}} x x^T \, d\tilde{\mu}_k(x) = \int_{\mathcal{O}} (F_k p)(F_k p)^T \, d\mu(p).$$

[9] Adjective: of or pertaining to the ribs, in anatomy.

(3) The kth *spinocostal*[10] *limit measure* h_k on the closed leaf $L_k \cong \overline{H}_+$ is defined by

$$h_k(A) = h_k^0(F_k(A) \cap H) + g_k(F_k(A) \cap \overline{H}_+)$$

for Borel sets $A \subseteq L_k$, where the *semispinal limit measure* h_k^0 on L_0 is defined by

$$h_k^0((P_S|_{L_0})^{-1}B) = g_S(B) - g_k((0,\infty) \times B)$$

for Borel sets $B \subseteq S$. (A possibly more natural definition of h_k is given in Proposition 5.6 below.)

REMARK 5.4. Square integrability (2.2) implies that the covariance matrices are finite.

REMARK 5.5. The semispinal limit measure is generally not Gaussian. Although the orthogonal projection to $\mathbb{R}^d$ of any Gaussian measure on $\mathbb{R}^{d+1}$ is Gaussian, h_k^0 is the projection of only half of a Gaussian; this is implied by Proposition 5.6, an alternate direct description of h_k interpolating between the first two parts of Definition 5.3.

PROPOSITION 5.6. *The spinocostal limit measure is the pushforward of the costal limit measure g_k under convex projection: $h_k = g_k \circ \hat{P}^{-1} \circ F_k$.*

PROOF. Since the measures agree on L_k outside of L_0 by definition, it is enough to show that

$$h_k^0((P_S|_{L_0})^{-1}B) = g_k(\hat{P}^{-1} \circ (\pi_S|_H)^{-1}B) \tag{5.1}$$

for any Borel set $B \subseteq S$. For any vectors $w, w' \in \mathbb{R}^{d+1}$ that lie on the spine $H \subseteq \mathbb{R}^{d+1}$, considering them as vectors in $z = \pi_S(w), z' = \pi_S(w') \in S = \mathbb{R}^d$ results in quantities $z^T C_S z'$, and $w^T C_k w'$. The integrals in Definition 5.3 directly imply that $z^T C_S z' = w^T C_k w'$. Consequently, the matrix C_S is a submatrix of C_k; the action of C_k on the subspace H is given by C_S. Thus $g_S(B) = g_k((-\infty,\infty) \times B)$, and hence by definition

$$\begin{aligned} h_k^0(B) &= g_k((-\infty,\infty) \times B) - g_k((0,\infty) \times B) = g_k((-\infty,0] \times B) \\ &= g_k(\hat{P}^{-1} \circ (\pi_S|_H)^{-1}B) \end{aligned}$$

for any Borel set $B \subseteq S$. $\square$

Now we come to the primary result in the paper: as the sample size N becomes large, the law ν_N of the rescaled empirical mean converges weakly

[10] Adjective: spanning the ribs and spine, in anatomy.

to the appropriate measure from Definition 5.3, according to how sticky the mean is. (We have included a forward reference to the nonsticky case in Theorem 5.7 to preserve the numbering of items 1, 2 and 3, which corresponds precisely to the numbering elsewhere, namely Theorem 2.9, Definition 2.10, Theorem 4.3, and Definition 5.3.) When the mean is:

(1) sticky, ν_N converges weakly to the spinal limit measure g_S;
(2) nonsticky, ν_N converges weakly to the costal limit measure g_j supported on the tangent space $\mathbb{R}^{d+1}$ to the leaf L_j containing the mean;
(3) partly sticky, ν_N converges weakly to the spinocostal limit measure g_j supported on the (unique) leaf L_k with moment $m_k = 0$.

As discussed at the start of the section, the fact that the limiting distribution is supported on the spine S when the mean is sticky follows from Theorem 4.3, since then the first moments m_j are strictly negative for all j.

THEOREM 5.7 (Sticky CLT). *Let μ be a nondegenerate (2.3) probability distribution on the open book $\mathcal{O}$ with finite second moment (2.2).*

(1) *If the mean of μ is sticky, then for any continuous, bounded function $\phi : \mathcal{O} \to \mathbb{R}$,*

$$\lim_{N\to\infty} \int_{\mathcal{O}} \phi(p)\, d\nu_N(p) = \int_S \phi \circ (P_S|_{L_0})^{-1}(q)\, dg_S(q).$$

(2) *If the mean of μ is nonsticky, then see Theorem 5.11.*
(3) *If the mean of μ is partly sticky, with first moment $m_k = 0$, then for any continuous bounded function $\phi : \mathcal{O} \to \mathbb{R}$,*

$$\lim_{N\to\infty} \int_{\mathcal{O}} \phi(p)\, d\nu_N(p) = \int_{\overline{H}_+} \phi \circ F_k^{-1}(q)\, dh_k(q).$$

PROOF. The proof works by decomposing the relevant measures—the empirical mean on the open book and its pushforward to $\mathbb{R}^{d+1}$ under folding—into pieces corresponding to the leaves and the spine.

Suppose that the mean is partly sticky with first moment $m_k = 0$. Let $\eta_N = \eta_{k,N}$ as in (3.1), and let $\nu_{\eta,N}(x)$ denote the law of $\sqrt{N}\eta_N$ on $\mathbb{R}^{d+1}$. By Lemma 3.3, $\nu_N(A) = \nu_{\eta,N}(F_k A)$ for any Borel set $A \subseteq L_k$, and if ϕ is a continuous and bounded function, then

$$\begin{aligned}\int_{\mathcal{O}} \phi(p)\, d\nu_N(p) &= \int_{L_k^+} \phi(p)\, d\nu_N(p) + \int_{\mathcal{O}\smallsetminus L_k^+} \phi(p)\, d\nu_N(p)\\ &= \int_{H_+} \phi((F_k^{-1}|H_+)^{-1}(y))\, d\nu_{\eta,N}(y) + \int_{\mathcal{O}\smallsetminus L_k^+} \phi(p)\, d\nu_N(p).\end{aligned}$$

The standard CLT in $\mathbb{R}^{d+1}$ (e.g., [8], Theorem 11.10) implies that the random variable $\sqrt{N}\eta_N$ converges in distribution to a centered Gaussian with covariance C_k. Therefore,

$$\lim_{N\to\infty}\int_{H_+}\phi((F_k^{-1}|H_+)^{-1}(y))\,d\nu_{\eta,N}(y)=\int_{H_+}\phi((F_k^{-1}|H_+)^{-1}(y))\,dg_k(y).$$

LEMMA 5.8. *If the jth first moment satisfies $m_j<0$, then $\nu_N(L_j^+)\to 0$ and*

$$\lim_{N\to\infty}\int_{L_j^+}\phi(p)\,d\nu_N(p)=0.$$

PROOF. Theorem 4.3(1). □

Resuming the proof of the theorem, consider the term

$$\int_{\mathcal{O}\setminus L_k^+}\phi(p)\,d\nu_N(p)=\int_{L_0}\phi(p)\,d\nu_N(p)+\int_{L_k^-}\phi(p)\,d\nu_N(p),$$

where $L_k^-=\mathcal{O}\setminus L_k=\bigcup_{j\neq k}L_j^+$, which excludes the spine L_0. With the projection $P_0:\mathcal{O}\to L_0,(x^{(0)},x,j)\mapsto(0,x,j)$ the function $p\mapsto\phi(P_0p)$ is again continuous and bounded, Lemma 5.8 implies that

$$\lim_{N\to\infty}\int_{L_k^-}\phi(P_0p)\,d\nu_N(p)=0. \tag{5.2}$$

Therefore,

$$\begin{aligned}\lim_{N\to\infty}\int_{L_0}\phi(p)\,d\nu_N(p)&=\lim_{N\to\infty}\int_{L_0}\phi(P_0p)\,d\nu_N(p)\\&=\lim_{N\to\infty}\biggl(\int_{L_k^-}\phi(P_0p)\,d\nu_N(p)+\int_{L_0}\phi(P_0p)\,d\nu_N(p)\biggr)\\&=\lim_{N\to\infty}\biggl(\int_{\mathcal{O}}\phi(P_0p)\,d\nu_N(p)-\int_{L_k^+}\phi(P_0p)\,d\nu_N(p)\biggr).\end{aligned}$$

Observe that

$$\int_{\mathcal{O}}\phi(P_0p)\,d\nu_N(p)=\int_S\phi\circ(P_S|_{L_0})^{-1}(y)\,d\gamma_N(y),$$

where $\gamma_N=\nu_N\circ P_S^{-1}$ which is the law of $\sqrt{N}\bar{y}_N$ on S, where $\bar{y}_N$ is the projected barycenter from Lemma 3.5. Therefore, setting $\hat{\phi}=\phi\circ(P_S|_{L_0})^{-1}$ and applying the usual CLT to $\sqrt{N}\bar{y}_N\in\mathbb{R}^d$,

$$\lim_{N\to\infty}\int_{\mathcal{O}}\phi(P_0p)\,d\nu_N(p)=\lim_{N\to\infty}\int_S\hat{\phi}(y)\,d\gamma_N(y)=\int_S\hat{\phi}(y)\,dg_S(y).$$

We cannot apply the same argument to

$$\lim_{N\to\infty}\int_{L_k^+}\phi(P_0p)\,d\nu_N(p)=\lim_{N\to\infty}\int_{L_k^+}\hat{\phi}(y)\,d\tau_N(y)$$

with $\tau_N=\nu\circ(P_S|_{L_k^+})^{-1}$ because there is no CLT for τ_N. We have, however, above derived a CLT for $\nu_N\circ F_k^{-1}=\nu_{\eta,N}$ on $H_+=F_k(L_k^+)$:

$$\lim_{N\to\infty}\int_{L_k^+}\phi(P_0p)\,d\nu_N(p)=\lim_{N\to\infty}\int_{H_+}\widetilde{\phi}(q)\,d\nu_{\eta,N}(q)=\lim_{N\to\infty}\int_{H_+}\widetilde{\phi}(q)\,dg_k(q),$$

where $\widetilde{\phi}=\phi\circ P_0\circ F_k\circ\hat{P}^{-1}$. In summary, we have shown that

$$\begin{aligned}
&\lim_{N\to\infty}\int_{\mathcal{O}}\phi(p)\,d\nu_N(p)\\
&\quad=\int_{H_+}\phi\circ F_k^{-1}(q)\,dg_k(q)+\int_S\hat{\phi}(y)\,dg_S(y)-\int_{H_+}\widetilde{\phi}(q)\,dg_k(q)\\
&\quad=\int_{H_+}\phi\circ F_k^{-1}(q)\,dg_k(q)+\int_H\phi\circ F_k^{-1}(q)\,dh_k^0(q)\\
&\quad=\int_{\overline{H}_+}\phi\circ F^{-1}(q)\,dh_k(q),
\end{aligned}$$

where the second equality uses the fact that $\widetilde{\phi}=\phi\circ F_k^{-1}$ on H and the final equality the fact that g_k has no mass supported on the spine H, so the integral of $\phi\circ F^{-1}\,dg_k$ over H_+ can just as well be taken over $\overline{H}_+$.

The sticky case proceeds in much the same way as the partly sticky case does, except that instead of equation (5.2), the simpler statement

$$\lim_{N\to\infty}\int_{\mathcal{O}\smallsetminus S}\phi(P_0p)\,d\nu_N(p)=0$$

holds. From that, the next step results in

$$\lim_{N\to\infty}\int_{L_0}\phi(p)\,d\nu_N(p)=\lim_{N\to\infty}\int_{\mathcal{O}}\phi(P_0p)\,d\nu_N(p),$$

and then the usual CLT applied to $\sqrt{N}\bar{y}_N\in\mathbb{R}^d$ proves the desired result. □

5.2. *The nonsticky central limit theorem.* If the mean is nonsticky with first moment $m_k>0$, then the limit $\bar{b}$ is in the interior of L_k^+. In this case, the tangent cone at $\bar{b}$ is the vector space $\mathbb{R}^{d+1}$, and the fluctuations of b_N about the limit $\bar{b}$ are qualitatively similar to what is described in the classical central limit theorem.

Definition 5.9. In this section we let $\widetilde{\nu}_N$ be the law on $\mathbb{R}^{d+1}$ of the random variable $\sqrt{N}(F_k b_N - F_k \bar{b})$,

$$\mathbb{P}(\{\omega | \sqrt{N}(F_k b_N - F_k \bar{b}) \in A\}) = \widetilde{\nu}_N(A)$$

for all Borel sets $A \subseteq \mathbb{R}^{d+1}$.

Definition 5.10. Assume $m_k > 0$. Let $\tilde{g}_k$ be the law of a multivariate normal random variable on $\mathbb{R}^{d+1}$ with mean zero and covariance matrix

$$\tilde{C}_k = \int_{\mathbb{R}^{d+1}} (x - F_k \bar{b})(x - F_k \bar{b})^T \, d\tilde{\mu}_k(x).$$

In contrast to the case of a sticky or partly sticky mean, the weak limit of ν_N is that of a nondegenerate Gaussian on $\mathbb{R}^{d+1}$:

Theorem 5.11 (Nonsticky CLT). *Assume $m_k > 0$. Then for any continuous bounded function $\phi : \mathbb{R}^{d+1} \to \mathbb{R}$,*

$$\lim_{N\to\infty} \int_{\mathbb{R}^{d+1}} \phi(x) \, d\widetilde{\nu}_N(x) = \int_{\mathbb{R}^{d+1}} \phi(x) \, d\tilde{g}_k(x).$$

Proof. Since $m_k > 0$, $\bar{b} \in L_k^+$ and Lemma 3.3 implies $F_k \bar{b} = \bar{\eta} = \int_{\mathbb{R}^{d+1}} x \, d\tilde{\mu}_k(x)$. Also,

$$\sqrt{N}(F_k b_N(\omega) - F_k \bar{b}) = \sqrt{N}(\eta_{k,N}(\omega) - \bar{\eta}) \qquad \forall N \geq N^*(\omega)$$

holds with probability one. Therefore, for any Borel set

$$|\widetilde{\nu}_N(A) - \mathbb{P}(\{\omega | \sqrt{N}(\eta_{k,N}(\omega) - \bar{\eta}) \in A\})| \leq R_N,$$

where $R_N = \mathbb{P}(\{\omega | N < N^*(\omega)\})$. By the classical central limit theorem, the random variable $\sqrt{N}(\eta_{k,N}(\omega) - \bar{\eta})$ converges in law to a centered, multivariate Gaussian on $\mathbb{R}^{d+1}$ with covariance C_k as $N \to \infty$. Consequently,

$$\limsup_{N\to\infty} \left| \int_{\mathbb{R}^{d+1}} \phi(x) \, d\widetilde{\nu}_N(x) - \int_{\mathbb{R}^{d+1}} \phi(x) \, d\tilde{g}_k(x) \right| \leq 2 \limsup_{N\to\infty} R_N \|\phi\|_\infty = 0. \qquad \square$$

Acknowledgments. Our thanks go to Seth Sullivant, for bringing our attention to stickiness for means of phylogenetic trees. This paper is output from a Working Group that ran under the Statistical and Applied Mathematical Sciences Institute (SAMSI) 2010–2011 program on Analysis of Object Data. We are grateful to the other members of the SAMSI Working Group on sampling from stratified spaces, not listed among the authors, who facilitated the development of this research program. We are indebted to SAMSI itself, for sponsoring many of the authors' visits to the Research Triangle, for hosting the Working Group meetings, and for stimulating this cross-disciplinary research in its unique way.

REFERENCES

[1] Arsigny, V., Fillard, P., Pennec, X. and Ayache, N. (2007). Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **29** 328–347 (electronic). MR2288028

[2] Basrak, B. (2010). Limit theorems for the inductive mean on metric trees. *J. Appl. Probab.* **47** 1136–1149. MR2752884

[3] Basser, P. J. and Pierpaoli, C. (1996). Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *J. Magn. Reson. B* **111** 209–219.

[4] Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. I. *Ann. Statist.* **31** 1–29. MR1962498

[5] Bhattacharya, R. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds. II. *Ann. Statist.* **33** 1225–1259. MR2195634

[6] Bhattacharya, R. N., Buibas, M., Dryden, I. L., Ellingson, L. A., Groisser, D., Hendriks, H., Huckemann, S., Le, H., Liu, X., Marron, J. S., Osborne, D. E., Patrângenaru, V., Schwartzman, A., Thompson, H. W. and Wood, A. T. A. (2011). Extrinsic data analysis on sample spaces with a manifold stratification. In *Advances in Mathematics, Invited Contributions at the Seventh Congress of Romanian Mathematicians, Brasov, 2011* (L. Beznea, V. Brîzanescu, M. Iosifescu, G. Marinoschi, R. Purice and D. Timotin. eds.) 148–156. Publishing House of the Romanian Academy.

[7] Billera, L. J., Holmes, S. P. and Vogtmann, K. (2001). Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.* **27** 733–767. MR1867931

[8] Breiman, L. (1992). *Probability. Classics in Applied Mathematics* **7**. SIAM, Philadelphia, PA. MR1163370

[9] Dryden, I. L. and Mardia, K. V. (1998). *Statistical Shape Analysis*. Wiley, Chichester. MR1646114

[10] Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré* **10** 215–310. MR0027464

[11] Goresky, M. and MacPherson, R. (1988). *Stratified Morse Theory. Ergebnisse der Mathematik und Ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]* **14**. Springer, Berlin. MR0932724

[12] Hendriks, H. and Landsman, Z. (1996). Asymptotic behavior of sample mean location for manifolds. *Statist. Probab. Lett.* **26** 169–178. MR1381468

[13] Huckemann, S. (2012). On the meaning of mean shape: Manifold stability, locus and the two sample test. *Ann. Inst. Statist. Math.* **64** 1227–1259.

[14] Huckemann, S. (2011). Inference on 3D Procrustes means: Tree bole growth, rank deficient diffusion tensors and perturbation models. *Scand. J. Stat.* **38** 424–446. MR2833839

[15] Jupp, P. (1988). Residuals for directional data. *J. Appl. Stat.* **15** 137–147.

[16] Miller, E., Owen, M. and Provan, J. S. (2012). Averaging metric phylogenetic trees. Unpublished manuscript, available at `http://arxiv.org/abs/1211.7046`.

[17] Owen, M. and Provan, J. S. (2011). A fast algorithm for computing geodesic distance in tree space. *ACM/IEEE Transactions on Computational Biology and Bioinformatics* **8** 2–13.

[18] Patrangenaru, V., Liu, X. and Sugathadasa, S. (2010). A nonparametric approach to 3D shape analysis from digital camera images. I. *J. Multivariate Anal.* **101** 11–31. MR2557615

[19] Patrangenaru, V. and Mardia, K. V. (2003). Affine shape analysis and image analysis. In *Proc. of the 2003 Leeds Annual Statistics Research Workshop* 57–62, available at http://www1.maths.leeds.ac.uk/Statistics/workshop/lasr2003/proceedings/patrangenaru.pdf.

[20] Schwartzman, A., Dougherty, R. F. and Taylor, J. E. (2008). False discovery rate analysis of brain diffusion direction maps. *Ann. Appl. Stat.* **2** 153–175. MR2415598

[21] Schwartzman, A., Mascarenhas, W. F. and Taylor, J. E. (2008). Inference for eigenvalues and eigenvectors of Gaussian symmetric matrices. *Ann. Statist.* **36** 2886–2919. MR2485016

[22] Sturm, K.-T. (2003). Probability measures on metric spaces of nonpositive curvature. In *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces (Paris, 2002). Contemp. Math.* **338** 357–390. Amer. Math. Soc., Providence, RI. MR2039961

[23] Ziezold, H. (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the Eighth European Meeting of Statisticians (Tech. Univ. Prague, Prague, 1974), Vol. A* 591–602. Reidel, Dordrecht. MR0501230

T. Hotz
Institute of Mathematics
Ilmenau University of Technology
Weimarer Strasse 25, 98693 Ilmenau
Germany
E-mail: thomas.hotz@tu-ilmenau.de

S. Huckemann
Institute for Mathematical Stochastics
University of Göttingen
Goldschmidtstrasse 7, 37077 Göttingen
Germany
E-mail: huckeman@math.uni-goettingen.de

H. Le
School of Mathematical Sciences
University of Nottingham
University Park
Nottingham, NG7 2RD
United Kingdom
E-mail: huiling.le@nottingham.ac.uk

J. S. Marron
S. Skwerer
Department of Statistics
and Operations Research
University of North Carolina
Chapel Hill, North Carolina 27599
USA
E-mail: marron@email.unc.edu
sskwerer@unc.edu

J. C. Mattingly
J. Nolen
E. Miller
Mathematics Department
Duke University
Durham, North Carolina 27708
USA
E-mail: jonm@math.duke.edu
nolen@math.duke.edu
URL: www.math.duke.edu/~ezra

M. Owen
Cheriton School of Computer Science
University of Waterloo
Waterloo
ON N2L 3G1 Canada
URL: www.cs.uwaterloo.ca/~m2owen

V. Patrangenaru
Department of Statistics
Florida State University
Tallahassee, Florida 32306
USA
URL: www.stat.fsu.edu/~vic