# A Survey of Riemannian Centres of Mass for Data

Wilfrid S Kendall
University of Warwick

Department of Statistics, University of Warwick, Coventry CV4 7AL, UK
w.s.kendall@warwick.ac.uk

Riemannian barycentres are centres of mass defined for data which are best represented as points on nonlinear metric spaces (typically Riemannian manifolds, such as spheres). Because of the non-linearity, uniqueness of the barycentre can no longer be taken for granted, though much can be said about when uniqueness will hold. The subject has a surprisingly long history, stretching back to Fréchet and Cartan and indeed its origin is nearly contemporaneous with that of mathematical probability; however there has recently been a major resurgence of interest, fuelled by the need to deal with the intricate and geometrical structures of problems arising in modern data analysis. This short paper mentions some applications (including statistical shape theory, fibre reconstruction from point patterns, and paths of hurricanes) and surveys some underlying theory (which is now quite well-developed, including laws of large numbers and central limit theory, and which has led to a surprising retrospective look at the classical central limit theorem of Lindeberg and Feller).

**Key Words:** Barycentre, shape, fibre reconstruction.

## 1. Introduction

Riemannian barycentres have a history almost as old as that of measure-theoretic probability: note Cartan (1929) and Fréchet (1948), and compare the tantalizing remarks about Kolmogorov's view of Fréchet in the Royal Society obituary of Kolmogorov (D.G.Kendall *et al*., 1991). A *barycentre* $\xi$ in a metric space $M$ minimizes "energy" $\int \text{dist}(x, \xi)^2\, \mu(\text{d}\, x)$ of a (probability) measure $\mu$ defined on $M$; this extends the classic characterization of the statistical expectation $\mathbf{E}[X]$ of a random variable $X$ as the minimizer $\xi$ of the mean-square-error $\mathbf{E}[(X-\xi)^2]$. We consider only the case of exponent 2. Motivated primarily by robustness considerations, Afsari (2010) and Arnaudon *et al.* (2010) have considered exponents $p \geq 1$, thus defining $p$-means.

Uniqueness of $\xi$ for general metric spaces $M$ cannot be guaranteed unless appropriate conditions are satisfied. In the case when $M$ is a Riemannian manifold (perhaps with boundary), conditions for uniqueness can be expressed in terms of the curvatures of the manifold. Karcher (1977) gives an early result motivated by pure geometry, of which the following is a special case; if $M$ is a closed subset of the unit sphere contained in a small open cap of angular radius $\pi/4$, then there is a unique barycentre $\xi$ lying in $M$ for any probability measure concentrated in $M$. Motivated by the completely different area of applied statistics, Ziezold (1977) established a strong law of large numbers for sets of *empirical barycentres*, barycentres (not necessarily unique) of empirical distributions determined by i.i.d. samples of random variables taking values in metric and quasi-metric spaces.

## 2. Applications

Ziezold was motivated by the data-analytic consideration of defining and computing mean values for random figures and shapes of random figures, arising from the study of biological data. This is now a major driver of contemporary interest in Riemannian and metric space barycentres. D.G.Kendall's theory of statistical shape (D.G. Kendall *et al.,* 1999; see also a recent survey by W.S.Kendall & Le, 2010) leads naturally to non-trivial problems in this area. Data arising from studies of growth or comparative anatomy can be summarized *via* sequences of coordinates of "landmarks" (clearly defined locations such as points of maximum curvature or anatomical significance on a skull). For each individual case, interest focusses on only those features of the sequence of coordinates which are invariant under shifts, rotations, and typically scale-changes. Thus each sequence is naturally parametrized by a single representative shape-point on a high-dimensional metric space (the shape-space), and data-analytic questions concern (for example) the empirical barycentre of the resulting sample of shape-points, and fluctuation theory. Even in the simplest cases, the shape-space has non-trivial geometry. For example, one of the first applications of shape theory (Kendall & Kendall, 1980) led to consideration of sequences of three points in the plane; each such sequence has a shape which is naturally parametrized by a shape-point lying on the sphere of radius ½. Biologically relevant cases involve points in 3-space, for which the resulting shape-space will typically have singularities.

A different context for mean values of interesting objects arises in various applications when considering tensors (typically, fields of symmetric positive-definite matrices). For example, Hill *et al.* (2012) study the recovery of curvilinear structure from point patterns. This leads naturally to consideration of symmetric positive-definite 2×2 matrices, one matrix for each point in the pattern. The principal eigenvector of the matrix (if unique) generates an orientation field which can be integrated to recover the curvilinear structure. However it is necessary to smooth and interpolate this field, and a natural way to do this is to take a weighted barycentre of the 2×2 matrices at neighbouring points. In practice a considerable computational simplification is available using an alternate definition of barycentre (Arsigny *et al.,* 2006, working on a similar application in diffusion tensor imaging): one transforms the matrices using a matrix-logarithm, then uses ordinary coordinate-wise averaging, then takes the matrix-exponential of the result.

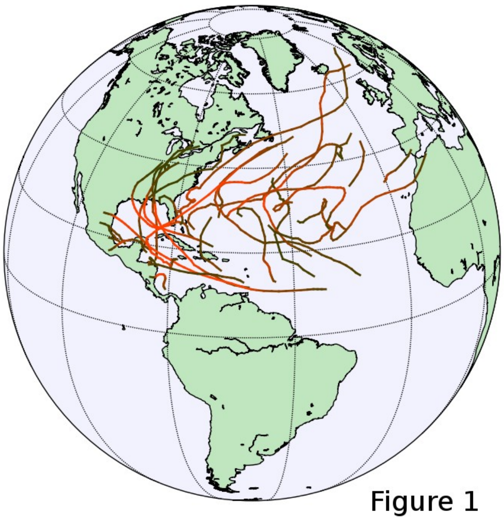

Figure 1

A third example concerns a remarkable publicly available dataset, composed of tracks of hurricanes dating back as far as 1851 (see Figure 1 for hurricanes in 2005). MacManus (2011) supplies preliminary data analysis. Interest centres on whether paths exhibit any "track memory". The scale of the tracks is such that they should be viewed as paths on a sphere; moreover, even simple studies of empirical means must take into account the fact that one cannot identify an objective start-time for each path. Distances between paths should be computed in a way which allows for this ambiguity (remniscent of the need in shape theory to discount location, scale and orientation). It is hoped to report on application of barycentre techniques to this dataset at a later stage; the current account uses this example as an illustrative theme.

These three contexts for the use of barycentres are representative of a much wider range: note for example applications to functional MRI (Fournel *et al.* 2012) and computer graphics (Buss & Fillmore, 2001). In a more theoretical context, barycentres have found application in a probabilistic and constructive approach to the theory of harmonic maps (Kendall, 1990).

## *3. Some theory*

The continuity of the energy functional $\int \text{dist}(x, \xi)^2 \mu(\text{d}\, x)$ guarantees existence of barycentres within a compact set. We have noted that uniqueness cannot be guaranteed in general: Émery & Mokobodzki (1991) therefore establish a theory of *set-valued* Riemannian barycentres, using the notion of convexity. However a careful convexity-based analysis can be used to establish uniqueness if the measure $\mu$ in question is supported within a so-called "regular geodesic ball" (defined in Kendall, 1990, 1991). For our purposes it suffices to note that the essence both of definition and uniqueness proof is contained in the special case when $\mu$ is supported on a "small hemisphere" of a Euclidean sphere, where "small" means, a closed spherical cap contained within an open strict hemisphere. The key point is that such a small hemisphere $M$ can be shown to possess *convex* geometry; namely, a convex function $\Phi{:}M{\times}M{\to}[0,1]$ vanishing only on the diagonal of $M{\times}M$. Afsari (2011) improves usefully on this result by noting that uniqueness then holds for potential barycentre locations varying over the entire sphere. D.G.Kendall *et al.* (1999) show that uniqueness also holds when $\mu$ has a density depending in a monotonically decreasing way on distance from a fixed point.

One might hope that convex geometry, and therefore uniqueness of barycentres, would hold for all Riemannian manifolds such that point pairs are connected by unique geodesics. Sadly, this is not the case (Kendall, 1992). It is possible to give necessary and sufficient conditions in case $M$ is a circle: see Charlier (2011) and the useful review of Kumar (2012).

It is instructive to consider the application of these results to the hurricane track example. Certainly the tracks can be regarded as lying in a small hemisphere $M$. They are more properly represented as maps from time intervals $[0,T]$ to $M$; However the above theory has an immediate consequence: it is straightforward to show that if $M$ has convex geometry then so does $M{\times}M{\times}\ldots{\times}M$ when endowed with the product Riemannian metric. The same is also true for the space of measurable maps from $[0,T]$ to $M$, and so one can define unique Riemannian barycentres for measures on such a space of hurricane tracks. On the other hand, it is necessary to consider the quotient metric under arbitrary time shifts, and hence also to determine the distance between tracks which are not defined over the same time interval – this could be viewed as a censored data problem. The most direct way to address this would be penalty-based; replace the distance from $x$ in $M$ to an unspecified point by the largest possible distance, namely $\pi/2+\text{dist}(x,\boldsymbol{n})$, where $\boldsymbol{n}$ is the "north pole" of the small hemisphere $M$. This contrasts with the Su *et al.*, (2012) approach to tracking in a variety of contexts, which takes a much more detailed statistical approach, with attention given to regularity of curves, but assumes that temporal registration is available.

If $M$ is a small hemisphere viewed as part of a unit Euclidean sphere, then the energy functional $\int \text{dist}(x, \xi)^2 \mu(\text{d}\, x)$ can be replaced by $\int (1\text{-}\cos \text{dist}(x, \xi))\, \mu(\text{d}\, x)$ (this could be justified by an extreme version of the robustness considerations motivating $p$-means!). In this case the barycentre can be computed using a simple extrinsic construction, as opposed to the iterative Newton methods required for energy-based barycentres: take the vector-valued expectation of $\mu$ viewed as a measure on the Euclidean space within which the unit sphere is defined, and project this expectation onto the unit sphere. In this case the barycentre is unique except when the vector-valued expectation is zero.

Finally, note that both kinds of barycentres can be viewed as maximum likelihood estimators for samples from various probability distributions on spheres as discussed *eg* in Mardia & Jupp (2009), and there is also considerable work on implementation of standard statistical theory in the context of Riemannian barycentres (for example, Bhattacharya & Patrangenaru, 2003, 2005; Ginestet, 2013). However here we choose to take a non-parametric view.

## 4. Limit theorems

We have already mentioned the Ziezold (1977) strong law of large numbers. For purposes of data analysis, for example of the hurricane tracks problem, we need results on the convergence of specific sequences of empirical barycentres $\mathcal{E}(X_1, ..., X_n)$, where $X_1, ..., X_n$ are independent random variables taking values in a specified metric space. This section provides a brief summary of known results, which thus provide a theoretical underpinning to considerations of sampling variation.

We begin by considering laws of large numbers for sequences of empirical barycentres, going beyond the set-theoretic perspective of Ziezold. The i.i.d. case is considered by Bhattacharya & Patrangenaru (2003): more recently Kendall & Le (2011) proved a more general weak law of large numbers based on a general weak law for independent non-identically distributed non-negative scalar random variables $Z_1, ..., Z_n$. Such a generalization is appealing in examples such as the hurricane tracks dataset, where an assumption of identical distribution over a period of 160 years would be foolhardy. In the non-negative scalar case, existence of a weak law is *equivalent* to a condition on the tails of the expectations of the $Z_1, ..., Z_n$, which can be viewed as an $L^1$ version of the celebrated Lindeberg condition for the central limit theorem. A similar result holds for the metric-space-valued random variables $X_1, ..., X_n$ so long as (a) an appropriate modification of the $L^2$ version of the Lindeberg condition holds (for bounded metric spaces there is less difference between $L^2$ and $L^1$ conditions), (b) a subsequence of local minimizers of the empirical energy function converge to the target limit **o**, (c) the cumulative energy functional $\mathbf{E}[\sum_{i=1,...,n}(X_i - \xi)^2]$ behaves well for $\xi$ close to the target limit **o**. The formulation is phrased carefully to allow for the possibility of non-unique barycentres.

The next step beyond weak laws of large numbers is the central limit theorem, specifically in the case of Riemannian manifolds. The i.i.d. case is treated in Bhattacharya & Patrangenaru (2005) using extrinsic means; Kendall & Le (2011) supply an intrinsic approach which again covers the independent non-identically distributed case. Proof techniques are related to Newton's method for root finding. The full formulation is rather technical, since conditions must be given to control the effects of curvature, though these conditions can be simplified in the important case of constant curvature. As would be expected from the classic Lindeberg-Feller central limit theorem, some kind of $L^2$ Lindeberg condition is required. However an analytic argument, involving the Riemannian exponential map, allows this condition to be referred to the Euclidean multivariate case. Most unexpectedly, it transpires that an economical formulation of the correct condition requires a re-working of the classic Lindeberg-Feller result to encompass the multivariate case; it turns out that the natural formulation of the multivariate Lindeberg-Feller result is actually an *approximation* result rather than a limit result, involving the celebrated Wasserstein metric for probability distributions.

The Kendall & Le (2011) central limit theorem concerns a sequence of Riemannian-manifold-valued random variables: Kumar (2012) discusses the case of triangular arrays. In a further development, Hotz *et al.* (2012) prove a central limit theorem for random variables taking values in a certain stratified metric space which is *not* a Riemannian manifold; when considering the hurricane track dataset, this may offer an alternative approach to the penalty-based method described above.

## 5. Conclusion

This short paper provides a brief survey of Riemannian barycentres, motivated by discussion of examples in particular the intriguing hurricane track dataset. There is now available a body of results which provide a framework within which an initial data analysis may be conducted. Surprisingly, the fluctuation theory (in particular the central limit theory) provides new perspectives on classical scalar limit theory.

It is hoped in due course to report elsewhere on the details of such a data-analytic approach to the hurricane tracks problem. Note that there are also significant ways in which the current theory could be developed; in particular it is an interesting question to what extent the central limit theory for empirical barycentres could be developed to parallel results in the scalar theory concerning stable laws (see for example Johnson & Samworth, 2005).

## References

Afsari, B., 2011. Riemannian $L^p$ center of mass: Existence, uniqueness, and convexity. *Proc. AMS*, 139(2):655–674.

Arnaudon, M., Dombry, C., Phan, A. & Yang, L., 2012. Stochastic algorithms for computing means of probability measures. *Stochastic Processes & Applications*, 122:1437–1455.

Arsigny, V., Fillard, P., Pennec, X., & Ayache. N., 2006. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic resonance in medicine*, 56(2):411–421.

Bhattacharya, R.N., & Patrangenaru, V., 2003, 2005. Large sample theory of intrinsic and extrinsic sample means on manifolds. *Annals of Statistics* 31(1):1–29 (Part I); 33(3):1225–1259 (Part II).

Buss, S.R. & Fillmore, J.P., 2001. Spherical averages and applications to spherical splines and interpolation. *ACM Transactions on Graphics*, 20(2):95–126.

Cartan, E., 1929. Groupes simples clos et ouverts et géométrie riemannienne. *Journal de mathématiques pures et appliqués*, 8:1–43.

Charlier, B., 2011. Necessary and sufficient condition for the existence of a Frechet mean on the circle. *arXiv:*1109.1986, 21 pp.

Émery, M. & Mokobodzki, G., 1991. Sur le barycentre d'une probabilité dans une variété. *Séminaire de probabilités XXV:*220–233.

Fournel, A.P., Reynaud, E., Brammer, M.J., Simmons, A. & Ginestet, C.E., 2012. Group Analysis of Self-organizing Maps based on Functional MRI using Restricted Frechet Means. *arXiv:*1205.6158, 23 pp.

Fréchet, M., 1948. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré*, 10(4):215–310.

Ginestet, C.E., 2013. Strong Consistency of the Frechet Sample Mean in Bounded Metric Spaces. Submitted. 25 pp.

Hill, B.J., Kendall, W.S. & Thönnes, E., 2012. Fibre-generated Point Processes and Fields of Orientations. *Annals of Applied Statistics*, 6(3):994–1020.

Hotz, T., Huckemann, S., Le, H., Marron, J.S., Mattingly, J.C., Miller, E., Nolen, J., Owen, M., Patrangenaru, V. & Skwerer, S., 2012. Sticky central limit theorems on open books. *arXiv:*1202.4267, 20 pp.

Johnson, O., & Samworth, R., 2005. Central limit theorem and convergence to stable laws in Mallows distance. *Bernoulli* 11(5):829-845.

Karcher, H., 1977. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541.

Kendall, D.G. *et al*, 1991. Andrei Nikolaevich Kolmogorov. 25 April 1903 – 20 October 1987. *Biographical Memoirs of Fellows of the Royal Society*, 37(300-319):1748–8494.

Kendall, D.G., Barden, D., Carne, T., & Le, H., 1999. *Shape and Shape Theory*, Chichester: Wiley.

Kendall, D.G. & Kendall, W.S., 1980. Alignments in two-dimensional random sets of points. *Advances in Applied Probability*, 12(2):380–424.

Kendall, W.S., 1990. Probability, convexity, and harmonic maps with small image I: Uniqueness and fine existence. *Proc. LMS (Third Series)*, 61:371–406.

Kendall, W.S., 1991. Convexity and the hemisphere. *Journal LMS (Second Series)*, 43:567–576.

Kendall, W.S., 1992. The Propeller: A counterexample to a conjectured criterion for the existence of certain convex functions. *Journal LMS (Second Series)*, 46:364–374.

Kendall, W.S., & Le, H., 2010. Statistical Shape Theory. In *New Perspectives in Stochastic Geometry*, Clarendon Press, pp.348–373.

Kendall, W.S., and Le, H., 2011. Limit theorems for empirical Fréchet means of independent and non-identically distributed manifold-valued random variables. *Brazilian Journal Prob. & Stat.* 25(3):323-352.

Kumar, A., 2012. Riemannian Barycenters and CLT for the vector valued random variables. *University of Warwick Dissertation for degree of M.Res*.

MacManus, L., 2011. Smith Institute: Modelling hurricane track memory. Available at: http://www.smithinst.ac.uk/Projects/ESGI73/ESGI73-Lloyds/index_html.

Mardia, K. V & Jupp, P.E., 2009. *Directional Statistics*, Chichester: Wiley.

Su, J., Dryden, I.L., E Klassen, E., Le, H., & Srivastava, A., 2012. Fitting smoothing splines to time-indexed, noisy points on nonlinear manifolds. Image and Vision Computing 30(6-7):428–442.

Ziezold, H., 1977. On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. *Trans. 7th Prague Conf. Info. Theory, Stat. Dec. Functions, Rand. Proc.*, A, pp.591–602.