Riemannian centres of mass

and a new look at the classical central limit theorem

Wilfrid S. Kendall[1] Huiling Le[2]

[1]Department of Statistics, University of Warwick

[2]School of Mathematical Sciences, University of Nottingham

30th Leeds Annual Statistics Research Workshop
5–7 July 2011

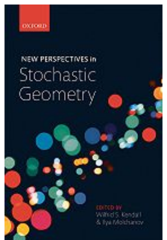
PLAN OF TALK

Riemannian barycentres

Riemannian barycentres have a history as old as that of measure-theoretic probability itself!

- Early years (Cartan 1929; Fréchet 1948);
- Careful description of properties in case of Riemannian manifolds (Karcher 1977);
- Appearing in applied statistics (Ziezold 1977, 1989, 1990, 1994, in the context of quasi-metric spaces).

Partly surveyed in WSK and Le (2010) (in W.S. Kendall and I.S. Molchanov, eds., *New Perspectives in Stochastic Geometry*, OUP, 2010, 608pp).
Follow-up: WSK and Le (2011) aimed to clarify the central limit theory for empirical barycentres.



Definitions

The definition of a Riemannian barycentre follows directly from mean-square-loss minimization definition.

- Conventional discrete sample mean / manifold-valued

$$\overline{y} = \operatorname*{argmin}_x \sum_y |x-y|^2\,; \quad \mathcal{E}_e[Y] = \operatorname*{argmin}_x \sum_y \operatorname{dist}(x,y)^2\,;$$

- Conventional expectation / manifold-valued

$$E[Y] = \operatorname*{argmin}_x \mathbb{E}\left[|x-Y|^2\right]\,; \quad \mathcal{E}[Y] = \operatorname*{argmin}_x \mathbb{E}\left[\operatorname{dist}(x,Y)^2\right]\,;$$

- It is convenient to have notation for the “energy” which is being minimized here, for example

$$\phi(x) \;=\; \frac{1}{2}\mathbb{E}\left[\operatorname{dist}(x,Y)^2\right]\,.$$

Uniqueness theorem(I)

A basic question:

- When are barycentres unique?
- It suffices to think about spheres; more specifically, hemispheres.

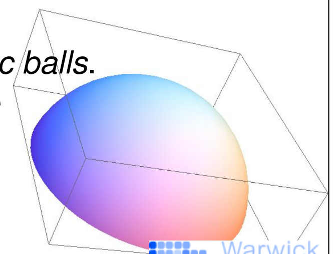
Uniqueness theorem(II)

Theorem

(WSK 1991) *Suppose that $S^{n-1}_{+,h}$ is a small hemisphere (intersection of sphere with half-space $x_1 \geq h$). Pick $\widetilde{h} \in (0, h)$. The following is convex if $2\nu\widetilde{h}^2(h^2 - \widetilde{h}^2) \geq 1$, and strictly convex outside the diagonal if strict inequality holds:*

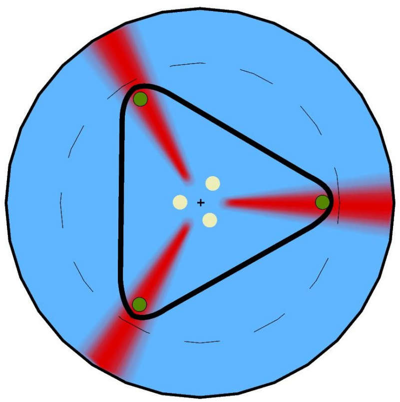
$$\Phi_{\nu,\widetilde{h}}(x, y) \quad = \quad \left(\frac{|x-y|}{2(x_1 y_1 - \widetilde{h}^2)}\right)^{\nu+1}$$

A direct modification holds for *regular geodesic balls*. In case of unit curvature upper bound, replace $p = \frac{1}{2}|x-y|^2$ by $1 - \cos \operatorname{dist}(x, y)$ and $q = x_1 y_1 - \widetilde{h}^2$ by $\cos \operatorname{dist}(x, \mathbf{o}) \cos(dist(y, \mathbf{o})) - \widetilde{h}^2$, for **o** the "north pole".



Failure of uniqueness

Uniqueness can fail in interesting ways if curvature varies.



(Negative curvature is blue, positive curvature is red.)

Laws of large numbers (I)

- Consider empirical barycentres of independent samples $X_1, \ldots, X_n$ from metric space $\mathcal{X}$.
- Ziezold (1977): strong law of large numbers for i.i.d. samples from separable finite quasi-metric spaces $\mathcal{X}$:

$$\bigcap_{k=1}^{\infty} \overline{\bigcup_{n=k}^{\infty} \arg\min_x \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2} \operatorname{dist}(X_i, x)^2} \subseteq \arg\min_x \mathbb{E}\left[\frac{1}{2} \operatorname{dist}(X_1, x)^2\right].$$

- Significant links with set-valued expectations in stochastic geometry (Stoyan and Molchanov 1997).
- Consistency under local compactness (Bhattacharya and Patrangenaru 2003).

Laws of large numbers (II)

Two equivalent variant Lindeberg conditions for finite-energy independent random variables $X_1, X_2, \ldots$ from $\mathcal{X}$. Let $\phi_n(x)$ be the aggregate energy at $x \in \mathcal{X}$. The following are equivalent:

Local: As $n \to \infty$, for each $\varepsilon > 0$,

$$\frac{1}{\phi_n(x)} \sum_{m=1}^{n} \mathbb{E}\left[\frac{1}{2}\operatorname{dist}(X_m, x)^2 \;;\; \frac{1}{2}\operatorname{dist}(X_m, x)^2 > \varepsilon\phi_n(x)\right] \to 0$$

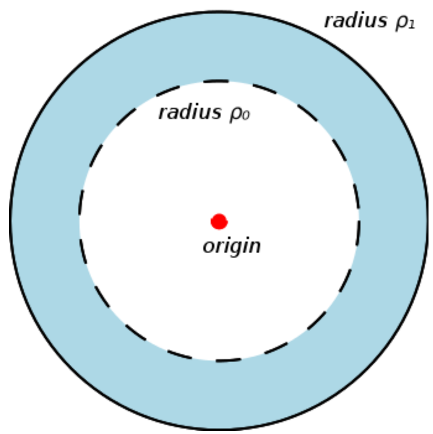
Semi-global: As $n \to \infty$, for each $\varepsilon > 0$,

$$\frac{1}{n\phi_n(x)} \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\left[\frac{1}{2}\operatorname{dist}(X_i, X_j)^2 \;;\; \frac{1}{2}\operatorname{dist}(X_i, X_j)^2 > \varepsilon\phi_n(x)\right] \to 0\,.$$

Second form is *semi*-global, as $\phi_n(x)$ still depends on $x \in \mathcal{X}$.

Laws of large numbers

visualizing the result



We want the mean aggregate energies to be uniformly strictly higher in rings away from the origin.
Then we can prove a weak law of large numbers for empirical barycentres.

Laws of large numbers (III)

- Let $\mathcal{X}$ be a locally compact separable metric space. Consider independent non-identically distributed X_1, X_2, $\ldots$ from $\mathcal{X}$ with $\mathbb{E}\left[\operatorname{dist}(X, \mathbf{o})^2\right] < \infty$.
- Suppose for some $\rho_1 > 0$, for every positive $\rho_0 \leq \rho_1$, we have $(1+\kappa)\phi_n(\mathbf{o}) < \inf\{\phi_n(y) : \rho_0 \leq \operatorname{dist}(y, \mathbf{o}) \leq \rho_1\}$.
- Suppose a local variant Lindeberg condition holds.
- (Theorem) For any sequence of local minimizers[1] of empirical energy lying within ball$(\mathbf{o}, \rho_1)$, there is a subsequence which converges to $\mathbf{o}$ in probability[2].

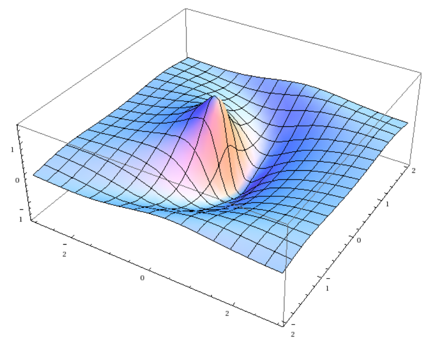
[1] Minimizers of empirical energy restricted to ball$(\mathbf{o}, \rho_1)$

[2] So local minimizers also local minima of unrestricted empirical energy.

Optimizing the empirical energy

Best way to locate point where empirical energy is at minimum:

- Seek out where gradient vanishes;
- Do this by pretending the energy surface is quadratic;



- Newton's method for cumulative empirical energy $\frac{1}{2}\sum_{i=1}^{n}\operatorname{dist}(x, X_i)^2$ (cf: Groisser 2004).

Central Limit Theorem (I)

We are now in a position to consider a CLT for non-identically distributed random variables on a Riemannian manifold $\mathbb{M}$. Consider independent $X_1, X_2, \ldots$ all with barycentre at $\mathbf{o}$.

- Gradient $\operatorname{grad}_x \frac{1}{2}\operatorname{dist}(x,y)^2$:
  $\operatorname{dist}(x,y)\operatorname{grad}_x \operatorname{dist}(x,y) = -\operatorname{Exp}_x \chi_x^{(y)}$,
  where $\chi_x^{(y)}$ is vector pointing from x to y.
- Hessian:
  define $\operatorname{Hess}_x \frac{1}{2}\operatorname{dist}(x,y)^2$ using covariant derivatives.
- Euclidean counterparts:
  $\operatorname{grad}_x \frac{1}{2}\operatorname{dist}(x,y)^2 \approx x - y$,
  $\operatorname{Hess}_x \frac{1}{2}\operatorname{dist}(x,y)^2 \approx$ identity matrix.

Central Limit Theorem (II) – geometric technicalities

- Taylor expansion of gradient of cumulative empirical energy $\frac{1}{2}\sum_{i=1}^n \operatorname{dist}(x, X_i)^2$:

$$\Pi_{x,\mathbf{o}} \sum_{i=1}^n Y_i(x) = \sum_{i=1}^n Y_i(\mathbf{o}) - \sum_{i=1}^n H_i(\mathbf{o})\gamma_x'(0) + \Delta_n(x)\gamma_x'(0)\,.$$

- Here $\Pi_{x,\mathbf{o}}$ is parallel transport, Y_i is the individual gradient field, H_i is the individual Hessian field, γ_x is minimal geodesic from $\mathbf{o}$ to x, and Δ_n is a correction term.
- If x is empirical barycentre then

$$0 \approx \sum_{i=1}^n Y_i(\mathbf{o}) - \sum_{i=1}^n H_i(\mathbf{o})\gamma_x'(0)$$

Central Limit Theorem (III)

Hence the following vector points towards the empirical barycentre:

$$\left(\frac{1}{\sqrt{2\phi_n(\mathbf{o})}}\sum_{i=1}^n H_i(\mathbf{o})\right)^{-1} \cdot \frac{1}{\sqrt{2\phi_n(\mathbf{o})}}\sum_{i=1}^n Y_i(\mathbf{o})\,.$$

Requirements:

- Cut-locus of $\mathbf{o}$ must be avoided;
- Control of $\Delta_n(x)$ (control local geometry *via* local mean variation of Hessian);
- $\phi_n(\mathbf{o})$ grows at least linearly, and bounds the sum of mean squared Frobenius norms of $H_1, \ldots H_n$, so that inverse matrix multiplier converges to constant in probability;
- Limit inverse matrix multiplier has an inverse;
- There is a vector-valued CLT for the normalized sum of gradients.

Classical innovations(I)

- We need a classical CLT for a sequence of independent non-identically distributed vector-valued random variables.
- A Lindeberg condition is needed. However the classical, coordinate-wise, Lindeberg condition is unnatural!
- We prefer a variant Lindeberg condition on the lengths of the random vectors: as $n \to \infty$ for each $\varepsilon > 0$,

$$\frac{1}{\phi_n}\sum_{m=1}^n \mathbb{E}\left[\frac{1}{2}\|Y_m\|^2\,;\,\frac{1}{2}\|Y_m\|^2 > \varepsilon\phi_n\right] \to 0$$

  where $\phi_n = \sum_{i=1}^n \mathbb{E}\left[\frac{1}{2}\|Y_m\|^2\right]$.
- But this is not sufficient to ensure convergence to a normal distribution.
  (Simple counterexample: consider sequences of bivariate normal distributions!)

A visual clue

Two problems which amount to the same thing:

- Measuring the distance between two probability distributions;
- Moving sand in the most efficient way possible

$$W_1(\mu,\nu) = \inf\{\mathbb{E}[\|U-V\|] : \mathcal{L}(U)=\mu, \mathcal{L}(V)=\nu\}$$
$$\widetilde{W}_1(\mu,\nu) = \inf\{\mathbb{E}[1\wedge\|U-V\|] : \mathcal{L}(U)=\mu, \mathcal{L}(V)=\nu\}$$

Classical innovations(II)

- Counterexample begs the question: perhaps we can show the variant Lindeberg condition forces normalized sum to become more normal even if it doesn't actually converge?
- "Central approximation" rather than "Central limit".
- Key idea: Use "truncated Wasserstein distance" (or "Dudley distance"):

$$\widetilde{W}_1(\mu,\nu) = \inf\{\mathbb{E}[1\wedge\|U-V\|] : \mathcal{L}(U)=\mu, \mathcal{L}(V)=\nu\}$$

- Metrizes weak convergence (Müller 1997; Villani 2003).
- Using Kantorovich-Rubinstein duality (Villani 2003, Remark 7.5(i))

$$\widetilde{W}_1(\mu,\nu) = \sup\{\int f\,\mathrm{d}(\mu-\nu) : f \text{ Lip}(1) \text{ for distance } 1\wedge\|x-y\|\},$$

  we can generalize the classical characteristic functions proof of the Lindeberg-Feller CLT.
- We can even prove the Feller converse!

Classical innovations(III)

We therefore have a clear picture of requirements for barycentre CLT:

$$\left(\frac{1}{\sqrt{2\phi_n(\mathbf{o})}}\sum_{i=1}^{n} H_i(\mathbf{o})\right)^{-1}\cdot\frac{1}{\sqrt{2\phi_n(\mathbf{o})}}\sum_{i=1}^{n} Y_i(\mathbf{o}).$$

converges to normality in sense of truncated Wasserstein distance when:

- Cut-locus of **o** is avoided;
- Control of $\Delta_n(x)$ *via* local mean variation of Hessian;
- Linearly growing $\phi_n(\mathbf{o})$ bounds the sum of mean square Frobenius norms of $H_1, \ldots H_n$, so that inverse matrix multiplier converges to constant in probability;
- Limit inverse matrix multiplier has an inverse;
- The variant Lindeberg condition holds.

Conclusion

- Clarification of requirements for barycentre CLT;
- A surprising new development for the classical Feller-Lindeberg theorem;
- My student Ashish Kumar has checked the results carry through for triangular arrays;
- Question: Desirable to find a simple example to show arbitrarily slow convergence for Lindeberg CLT;
- Question: Infinite dimensions?
- Question: Metric spaces instead of manifolds? (cf: hurricanes!)
- Question: Stable laws?

Bhattacharya, R. N. and V. Patrangenaru (2003).
Large sample theory of intrinsic and extrinsic sample means on manifolds – I.
The Annals of Statistics 31(1), 1–29.

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

Doss, S. (1949).
Sur la moyenne d'un élément aléatoire dans un espace distancié.
Bull. Sci. Math. (2) 73, 48–72.

Fréchet, M. (1948).
Les éléments aléatoires de nature quelconque dans un espace distancié.
Annales de l'institut Henri Poincaré 10(4), 215—-310.

Groisser, D. (2004, July).
Newton's method, zeroes of vector fields, and the Riemannian center of mass.
Advances in Applied Mathematics 33(1), 95–135.

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

MacManus, L. (2011).
Smith Institute: Modelling hurricane track memory.

Müller, A. (1997).
Integral probability metrics and their generating classes of functions.
Advances in Applied Probability 29(2), 429–443.

Stoyan, D. and I. S. Molchanov (1997).
Set-valued means of random particles.
J. Math. Imaging Vision 7(2), 111–121.

Villani, C. (2003).
Topics in optimal transportation, Volume 58 of *Graduate Studies in Mathematics*.
Providence, RI: American Mathematical Society.

WSK (1991).
Convexity and the hemisphere.
The Journal of the London Mathematical Society (Second Series) 43, 567–576.

WSK and H. Le (2010, January).
Statistical Shape Theory.
In W. S. Kendall and I. S. Molchanov (Eds.), *New Perspectives in Stochastic Geometry*, Chapter 10, pp. 348–373. The Clarendon Press Oxford University Press.

WSK and H. Le (2011).
Limit theorems for empirical Fréchet means of independent and non-identically distributed manifold-valued random variables.
Brazilian Journal of Probability and Statistics 25(3), to appear.

Ziezold, H. (1977).
On expected figures and a strong law of large numbers for random elements in quasi-metric spaces.
Transactions of Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes A, 591–602.

Ziezold, H. (1989).
On expected figures in the plane.
In *Geobild '89 (Georgenthal, 1989)*, Volume 51 of *Math. Res.*, Berlin, pp. 105–110. Akademie-Verlag.

Ziezold, H. (1990).
Mean Figures and Mean Shapes in the Plane.

Ziezold, H. (1994).
Mean Figures and Mean Shapes Applied to Biological Figure and Shape Distributions in the Plane.
Biometrical Journal 36(4), 491–510.