Freie Universität Berlin
Fachbereich Mathematik und Informatik
Institut für Mathematik

The Diffusion Mean

Seminar Paper in the Module “Geometric Statistics”

Johannes Schade
5564109
15.03.2023

Contents

1 Introduction

This paper elaborates on the presentation held by me on 23/11/2022 in the seminar, covering the article “Diffusion Means in Geometric Spaces” from Benjamin Eltzner, Pernille Hansen, Stephan F. Huckemann and Stefan Sommer in 2021 [Elt+21].

The vast body of mathematical statistics rests on a Euclidean setting, i.e. only considers non-curved spaces. But there are many cases in which we can not assume data points to live in a Euclidean space. As an example, consider a study about the density of a species on our planet, and the researchers are interested in a very basic statistic like a mean, i.e. an “average point” on earth, where this species lives “on average”. An intuitive way of approaching this problems would be to consider the surface $\mathbb{S}^2$ (the 2-sphere, modelling the earth) embedded in the Euclidean space $\mathbb{R}^3$, i.e. every point on earth has three coordinates (x, y, z) with $\|(x, y, z)\| = 1$. For the sake of simplicity, assume that there only two points on earth, $(x, 0, 0)$ and $(-x, 0, 0)$ where the only populations of the species of exactly the same size live. One could be tempted to just apply the Euclidean mean, but would then be baffled to find out that this species “on average” lives in the earth’s core.

This example aims to underscore that we cannot just adapt Euclidean statistics to obtain a meaninful result - we wish the mean in this case to be again on the earth’s surface, i.e. on $\mathbb{S}^2$ in our model, and not in the core. This is of course not only applicable for $\mathbb{S}^2$, but for *manifolds* in general, i.e. spaces that locally resemble the well-known k-dimensional Euclidean space $\mathbb{R}^k$.

Several approaches have been made for the mean on curved spaces like $\mathbb{S}^2$. The *extrinsic mean* for spaces embedded in $\mathbb{R}^k$, studied e.g. in [BP03], where the Euclidean mean μ as in our example above is projected on the closest points the manifold.

This mean is in many cases hard to compute and by definitions can only exist for embeddings. A more generally applicable concept for *Riemannian manifolds* was introduced by in form of the *Fréchet means* [Fré48], which is defined by points closest to the sample means by *geodesic distance*, i.e. not the necessarily Euclidean distance, but the distance naturally associated with the manifold.

A more generalized form of the Fréchet mean was introduced by [Huc11], such that other loss functions than the geodesic distance can be used.

[Elt+21] eventually introduced the diffusion mean as a special case of the generalised Fréchet mean, the *diffusion mean* studied in this paper. They introduce a loss function with a deep connection to the normal distribution, and show that this allows, in contrast to the Fréchet mean, positive probability mass on the *cut locus*. Futhermore, they present estimators for sampled data.

In this paper, we first introduce basic notions from differential geometry in order to understand (Riemannian) manifolds, and also notions from mathematical statistics like estimators (sections 2 and 3). In section 4 the definition of the diffusion mean is presented, as well as its connection to the normal distribution and the Fréchet mean. In section 5, possible estimators and its properties are introduced. Finally, applications for the diffusion means are provided in section 6.

2 Notions from Differential Geometry

In this project the notions needed of Differential Geometry in general and Riemannian manifolds in particular are provided.

The most important concept for considering non-Euclidean (i.e. non-flat) spaces are manifolds:

1 Definition (Topological manifold)**.** A topological k-manifold M is a set M of points equipped with a topology $\mathcal{T} \subseteq \mathcal{P}(M)$ that provides the open sets of M, given that for every $p \in M$ there exists a neighbourhood $U_p \in \mathcal{T}$ such that U_p is homeomorphic to an open subset of $\mathbb{R}^k$.

In plain English, a space (that means a set equipped with a topology) is a manifold if it is locally like the $\mathbb{R}^k$, but not necessarily globally. Examples for such manifolds are the sphere, torus or just the $\mathbb{R}^k$ itself.

Note that homeomorphy requires only continuity between $\mathcal{M}$ and $\mathbb{R}^k$, but not necessarily smoothness. Requiring smoothness leads to *differentiable manifolds*:

2 Definition (Differentiable manifold)**.** A differentiable k-manifold is a set M with a family $(M_i)_{i \in I} \subseteq \mathcal{P}(M))$ such that

1. $M = \bigcup_{i \in I} M_i = M$,
2. for every $i \in I$ exists a injective $\varphi_i : M_i \to \mathbb{R}^k$ such that $\varphi_i(M_i)$ is an open set in $\mathbb{R}^k$,
3. if $M_i \cap M_j \neq \varnothing$, then $\varphi_i(M_i \cap M_j)$ is open in $\mathbb{R}^k$ and
$$\varphi_j \circ \varphi_i^{-1} : \varphi_i(M_i \cap M_j) \to \varphi_j(M_i \cap M_j)$$
is differentiable for any i, j [Küh99].

Note that 1. and 2. provide the definition of a manifold, and 3. secures differentiability.

We continue with an informal definition of *tangent spaces*:

3 Definition (Tangent space)**.** T_pM denotes the *tangent space* of M at p, which can be thought of the set of all direction one can travel along in M from p). Note that $T_pM \cong \mathbb{R}^k$ as a neighbourhood of p is homeomorphic to $\mathbb{R}^k$.

Now we can define a *Riemannian manifold*:

4 Definition (Riemannian manifold)**.** A Riemannian manifold is a tuple $M = (M, g)$ where M is a differentiable manifold and $g_p : T_pM \times T_pM \to \mathbb{R}$ is smooth function for every $p \in M$ such that

1. $g_p(x, y) = g_p(y, x)$ (symmetry),
2. $g_p(x, x) > 0$ if $x \neq 0$ (positive definiteness),
3. the coefficient functions $g_{ij}(p)$ in

$$g_p = \sum_{i,j} g_{ij}(p)\,\mathrm{d}x^i \otimes \mathrm{d}x^j$$

is differentiable (i.e. shifting to another $p' \in M$ is smooth), where $\mathrm{d}x^i \otimes \mathrm{d}$ is a basis vector for every i, j.

Note that this implies that $g_p(\cdot, \cdot)$ defines a scalar product at every p. This induces a norm

$$\|x\| := \sqrt{g_p(x, y)}$$

and an angle

$$\cos \angle(x, y) := \frac{g_p(x, y)}{\|x\|\|y\|}$$

for $x, y \in T_pM$ [Küh99] and eventually a length of a curve $\gamma : [a, b] \to M$ defined by

$$L(\gamma) := \int_a^b \|\dot{\gamma}(t)\|_{\gamma(t)}\,\mathrm{d}t$$

.

Furthermore, we call shortest curves γ that connect two points $x, y \in M$ *geodesics*. This gives rise to a metric $d : M \times M \to \mathbb{R}$, where $d(x, y)$ is the length of the geodesic connecting x and y.

3 Notions from Mathematical Statistics

This section introduces some basic notions of parametric mathematical statistics from [Geo15]. Generally, in Mathematical Statistics we try to characterize the unknown distribution of a random variable by computing estimations from certain statistics like mean and variance from observed realizations of it.

The main tool for doing so is approximating (or *estimating*) these statistics:

5 Definition (Statistic, Estimator)**.** Let $(\Omega, \mathcal{A}, \{\mathbb{P}_\vartheta : \vartheta \in \Theta\})$ be a statistical Experiment with parameter space Θ and let $(\Omega', \mathcal{B})$ be an event space. Then we call

1. a random variable $X : \Omega \to \Omega'$ a *statistic*.

2. Let $\tau : \Theta \to \Omega'$ be a map that associates to each $\theta \in \Theta$ a $\omega \in \Omega$. Then a statistic $\hat{\tau} : \Omega \to \Omega'$ is called an *estimator* of τ.

The definition(s) above require some context. The definition of a statistic just appears to just rename random variables. This is technically true, but we equip a statistic with a different interpretation: while a random variable models randomness, a statistic extracts a "meaningful " quantity from observed data. To asses whether a statistic is "meaningful", one later investigates whether a statistic satisfies certain conditions like sufficiency (e.g. a vector $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$ interpreted as mean and squared variance is a sufficient statistic for the one dimensional normal distribution as it characterizes it fully). Furthermore, τ can be thought of an function that e.g. projects a parameters vector $\vartheta \in \mathbb{R}^k$ on one of its components, e.g. $\tau(\vartheta = [\mu, \sigma^2]) = \mu$.

Similarly, in the second definition the connection between τ and $\hat{\tau}$ is somewhat unclear. As for the statistic, one usually looks for certain conditions to asses whether an estimator is "good". This is usually done by demanding a form of *consistency*, i.e.

$$\hat{\tau}_n \xrightarrow{m} \tau$$

for some mode of convergence m in the number of realizations n.

A very popular estimator also employed in [Elt+21] is the *Log-Likelihood-Estimator*:

6 Definition. A estimator $\hat{\vartheta}$ is a Log-Likelihood estimator of ϑ if it is of the form

$$\hat{\vartheta}(x) = \arg\max_{\vartheta} \left(\log f_\vartheta(x)\right)$$

where $f_\vartheta(x)$ is the density, interpreted as a function in both x and ϑ.

Applying the logarithm does not alter the optimization problem, but since $f_\vartheta(x)$ is usually a product density (i.e. the product of densities), it yields an easier to handle sum instead of a product.

4 The Diffusion Mean and its Derivation

In this section we introduce the definition of the diffusion mean and show its connection with the Fréchet mean. We start with the definition of the diffusion mean:

7 Definition (Diffusion mean)**.** Let $X : \Omega \to M$ be a random variable on the manifold M. Then we call

$$E_t(X) := \operatorname*{arg\,min}_{y \in M} \mathbb{E}_X[-\ln u(X, y, t)] \subseteq M,$$

where u denotes the heat kernel in M, the diffusion mean.

The heat kernel is defined as follows:

8 Definition (Heat kernel)**.** The heat kernel is defined as the solution of the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t} u(x, y, t) = \Delta_x u(x, y, t)$$

where Δ_x denotes the Laplace operator with respect to $x \in M$.

We want to consider only manifolds that satsisfly the following condition:

9 Definition (Stochastical completeness)**.** We call a manifold M stochastically complete if there exists a minimal heat kernel p such that

$$\int_M p(x, y, t)\,\mathrm{d}t = 1 \qquad \forall x \in M, t \in \mathbb{R}$$

It can be shown that if M is geodesically complete (i.e. for any point $p \in M$, T_pM is defined everyhwere), stochastically complete and has bounded sectional curvature, then the heat kernel is positive definite and symmetric with respect to x and y. The diffusion mean provides a possible definition of the normal distribution on non-Euclidean manifolds, since in $\mathbb{R}^k$ it coincides with the density function of the normal distribution [Hsu02]: The heat kernel in $\mathbb{R}^k$ is given by

$$u(x, y, t) = \frac{1}{\sqrt{2\pi t}^d} \exp\left(-\frac{\|x - y\|}{4t}\right)$$

and substituting $y = \mu$ and $2t = \sigma^2$ results in the density of the multinomial normal distribution with covariance matrix $\Sigma = tI$, where I denotes the unit matrix. In particular, t can be seen as a variance parameter.

The diffusion mean has a strong relationship with the *Fréchet mean* [FLE87], which is defined as follows:

10 Definition (Fréchet mean)**.** The set

$$E(X) := \underset{y \in M}{\arg\min}\, \mathbb{E}_X[d^2(x,y)] \subseteq M$$

where $d : M^2 \to \mathbb{R}; (x,y) \mapsto d(x,y)$ is the geodesic distance, i.e. the length of the shortes path between x and y, is called the Fréchet mean.

This mean can be generalized to the *generalized Fréchet mean* [Huc11]:

11 Definition (Generalized Fréchet mean)**.** The set

$$E_\rho(X) := \underset{y \in M}{\arg\min}\, \mathbb{E}_X[\rho(X,y)] \subseteq M,$$

where $\rho : M^2 \to \mathbb{R}$ is a loss function, is called the generalized Fréchet mean with respect to ρ.

Thus the diffusion mean is $E_{-\ln u}$. Additionally, we also have by

$$\lim_{t \to 0} -2t \ln u(x,y,t) = d^2(x,y)$$

a connection to the geodesic distance [Wei78].

Last but not least we cite an important property of $E_{t,n}$: For $k \geq 2$, it allows positive probability mass on the cut locus, i.e. if $\mu^* \in \mathbb{S}^k$ is a unique diffusion means, then the *cut locus* $-\mu$ can have positive probability mass.

5 Estimators

In this section we will discuss possible ways to estimate the parameters y (the mean of the heat kernel) and t individually and jointly.

5.1 Estimating y

Since we use the logarithmic heat kernel u as generalized normal distribution density, we can use definition 6 to yield the estimator

$$\hat{y}_{t.n} := E_{t,n} := \underset{y \in M}{\arg\min} -\frac{1}{n} \ln \left(\prod_{i=1}^{n} u(x_i, y, t) \right) = \underset{y \in M}{\arg\min} -\frac{1}{n} \ln \sum_{i=1}^{n} u(x_i, y, t)$$

for a fixed $t \in \mathbb{R}_+$, where x_i is the i^{th} realization of the observed random variable, $i = 1, \ldots, n$.

This estimator satisfies certain consistency conditions, making it “meaningful” in the sense that it actually converges to the desired set of values:

12 Definition (Ziezold Consistency)**.** An estimator $\hat{E}_n$ of a set E is Ziezold consistent if for almost all $\omega \in \Omega$

$$\bigcap_{n=1}^{\infty} \mathrm{cl}\left(\bigcup_{k=n}^{\infty} \hat{E}_k(\omega)\right) \subset E$$

where $\mathrm{cl}\, \hat{E}_k$ denotes the closure of the set $\hat{E}_k$.

Note that the Ziezold consistency condition is a variation of the $\liminf$.

13 Definition (Bhattacharya-Patrangenaru Consistency)**.** An estimator $\hat{E}_n$ of a set E is Bhattacharya-Patrangenaru consistent (BP consistent)if $E \neq \varnothing$, for almost all $\omega in \Omega$ and for every $\varepsilon > 0$ there exists an $n = n(\varepsilon, \omega) \in \mathbb{N}$

$$\bigcup_{k=n}^{\infty} E_k(\omega) \subset B(E, \varepsilon)$$

where $B(E, \varepsilon)$ is the ε-ball around E.

[Elt+21] in combination with [Huc11] show that the estimator $E_{t,n}$ is both Ziezold and BP consistent, assuming reasonable conditions.

We remark at this point that [Elt+21] also prove a variant of the Central Limit theorem under certain conditions for this estimator (see Theorem 4.5).

5.2 Estimating t

In this subsection a possible estimation of t for a fixed y will be presented. We define the set

$$T(y) := \arg\min_{t \in \mathbb{R}_+} \mathbb{E}_X[-\ln(u(X, y, t))]$$

with u denoting the heat kernel. An approach to solve this optimization problem is by differentiation, which can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_M -\ln(u(x, \mu_s, t))\, \mathrm{d}\mathbb{P}(x) = -\int_M \frac{\frac{\mathrm{d}}{\mathrm{d}t} u(x, \mu_s, t)}{(x, \mu_s, t)}\, \mathrm{d}\mathbb{P}(x) = -\int_M \frac{\Delta u(x, \mu_s, t)}{u(x, \mu_s, t)}\, \mathrm{d}\mathbb{P}(x)$$

where μ_s is a diffusion mean. This can be approximated with the gradient descent algorithm given by the recursion

$$t_n + 1 = t_n + \beta \int_M \frac{\Delta u(x, \mu_s, t)}{u(x, \mu_s, t)}\, \mathrm{d}\mathbb{P}(x)$$

for a $\beta > 0$.

5.3 Joint Estimation of $E_t(X)$ and t

A joint estimation can be done by solving the optimization problem

$$\hat{\vartheta} := \operatorname*{arg\,min}_{(\mu,t)\in M\times\mathbb{R}_+} -\frac{1}{n}\ln\sum_{i=1}^{n} u(x_i, y, t)$$

6 Applications

A remarkable feature of a unique diffusion mean μ is that it admits, in contrast to the Fréchet mean, positive probability on the cut locus $-\mu$ on the k-sphere $\mathbb{S}^k$ for $k \geq 2$ (see [Elt+21], section 3.1). This makes the diffusion mean a useful statistic on spherical data. Spherical data routinely occur in physiscs, e.g. the arrival directions of showers of cosmic rays, as well in geology in the facing directions of conically folded planes and the measurements of magnetic remanence in rocks. In biology, the orientations of the dendritic fields at different sites in the retina of a cat's eye are spherical data of interest [FLE87], or the dispersion of birds or turtles from a certain location [SA01]. In medical sciences and public health, arrival times in the emergency room are studied [JM09]. In such *directional* data (also known as *orientational* data) only the angles between two basis vectors is of interest, and not, let say, the magnitude of the vectors in the Euclidean space. Thus one can think of these data of points on sphere, uniquely determined by the angles.

From such directional data one wishes the mean to be again to be directional, (imagine e.g. one wants to find a mean location for swarms of fish on the earth surface) and here this can be achieved by the diffusion mean, with less restrictions than the Fréchet mean.

Of course the use of the diffusion mean is not restricted to spheres, but it can also used for other manifolds. It might be particularly useful for manifolds for which the analytic expression of the heat kernel is known. Besides spheres, this is in particularly the case for the hyperbolic spaces $\mathbb{H}^k$.

Moreover, by swapping in the density of Brownian motion, one can obtain a diffusion mean for graphs and Lie groups as well. A Lie group M is a manifold equipped with a smooth group operation $\odot$, such that $(M, \odot)$ forms an algebraic group. For example, on $\mathbb{S}^3$ the operation $x \odot y$ can be defined as rotating the sphere along the geodesic between x and y such that y is rotated to the spot of x, called the translation from x to y. There always exists a symmetric bi-invariant connection on Lie groups that result in (possibly non-metric) affine structures, thus it is possible to define a mean by using exponential barycenters. This allows the definition of random walks, and using its densities as transition density as "heat kernel", one can obtain a version of the diffusion mean on Lie groups[Pen16] [MP15]. This is important in field like robotics and anatomy, since $SE(3)$, the group of rigid motions in $\mathbb{R}^3$, is a Lie group.

For a graph $G = (V, E)$, a random walk can also be defined by using the transition probability for moving from vertex v_i to vertex v_j given by $\mathbb{P}_{ij} = \frac{\deg v_i}{c_{ij}}$, where $c_{ij} := |\{e \in E : e \text{ edge between } v_i \text{ and } v_j\}|$. Using the resulting counting density (i.e. encoded as $|E| \times |E|$-matrix) as "heat kernel", where t is the (discrete) point of time, on which we visit a certain vertex. Doing statistics on graphs is useful in network analyisis, e.g. neural networks in the brain or rail infrastructure.

References

[BP03] Rabi Bhattacharya and Vic Patrangenaru. "Large sample theory of intrinsic and extrinsic sample means on manifolds". In: *The Annals of Statistics* 31.1 (2003), pp. 1–29.

[Elt+21] Benjamin Eltzner et al. *Diffusion Means in Geometric Spaces*. 2021. DOI: 10.48550/ARXIV.2105.12061. URL: https://arxiv.org/abs/2105.12061.

[FLE87] N. I. Fisher, T. Lewis, and B. J. J. Embleton. *Statistical Analysis of Spherical Data*. Cambridge University Press, 1987. DOI: 10.1017/CBO9780511623059.

[Fré48] Maurice Fréchet. "Les éléments aléatoires de nature quelconque dans un espace distancié". In: *Annales de l'institut Henri Poincaré*. Vol. 10. 4. 1948, pp. 215–310.

[Geo15] Hans-Otto Georgii. *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Walter de Gruyter GmbH & Co KG, 2015.

[Hsu02] Elton P Hsu. *Stochastic analysis on manifolds*. 38. American Mathematical Soc., 2002.

[Huc11] Stephan F. Huckemann. "Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth". eng. In: *The Annals of statistics* 39.2 (2011), pp. 1098–1124. ISSN: 0090-5364.

[JM09] Peter E Jupp and Kanti V Mardia. *Directional statistics*. John Wiley & Sons, 2009.

[Küh99] Wolfgang Kühnel. *Differentialgeometrie*. Vol. 2003. Springer, 1999.

[MP15] Nina Miolane and Xavier Pennec. “Statistics on Lie groups: a need to go beyond the pseudo-Riemannian framework”. In: *AIP Conference Proceedings*. Vol. 1641. 1. American Institute of Physics. 2015, pp. 59–66.

[Pen16] Xavier Pennec. *Barycentric Subspace Analysis on Manifolds*. 2016. DOI: 10.48550/ARXIV.1607.02833. URL: https://arxiv.org/abs/1607.02833.

[SA01] Jammalamadaka S Rao and Sengupta Ashis. *Topics In Circular Statistics-vol 5.* Series on Multivariate Analysis v. 5. World Scientific, 2001. ISBN: 9789810237783. URL: https://search.ebscohost.com/login.aspx?direct=true&db=e000xww&AN=235938&site=ehost-live.

[Wei78] Joel L Weiner. “On a problem of Chen, Willmore, et al”. In: *Indiana University mathematics journal* 27.1 (1978), pp. 19–35.