Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements
Xavier Pennec

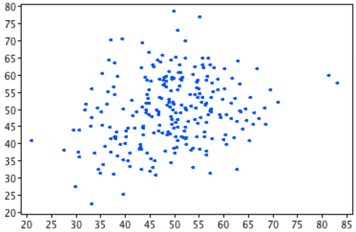
Daniel Levy [1 2] Mehran Shakerinava [1 2]

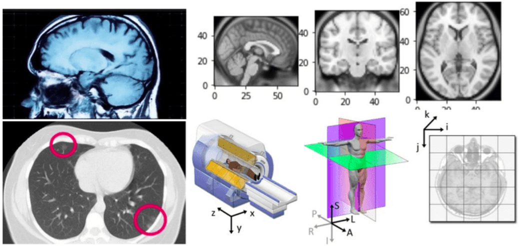
[1]McGill University

[2]Mila Quebec AI Institute

COMP 760 - September 30, 2022

- Statistics: In experiments, have set of samples, care about the central value, the dispersion around that value, and statistical tests
- Manifolds: Medical data and image analysis often have geometric objects that are better described as finite dimensional Riemannian manifolds instead of measurements in vectors spaces, such as surfaces, 3D rotations, or frames.





Motivation - Statistics on Manifolds

- Field of directional statistics
  - Tend to be extrinsic
  - Not a clear connection to manifold structure
- Approached from stochastic processes on Lie groups, stochastic differential geometry, stochastic calculus
  - Tend to focus on large sample sizes (CLT), and bounds on approximations
- This paper:
  - Intrinsic
  - Explicit about Riemannian structure
  - Approporiate for small sample sizes

Contributions

- Surveys different definitions of the expectation of a random point on a Riemannian manifold
- Characterizes these means, and provides algorithm to compute them
- Describes notions of variance, covariance, and χ^2 laws on Riemannian manifolds
- Defines generalized Gaussians on Riemannian manifolds, and describe their properties

Overview

- Background
- Expectation of a random point
- Variance and covariance
- Normal distribution
- χ^2 tests
- Applications and further extensions
- Conclude

Background - Riemannian metric, distance and geodesics I

- The structure of a manifold $\mathcal{M}$ is specified by a Riemannian metric.

Background - Riemannian metric, distance and geodesics I

- The structure of a manifold $\mathcal{M}$ is specified by a Riemannian metric.
- A ***Riemannian metric*** is a continuous collection of dot products $\langle \cdot, \cdot \rangle_x$ on the tangent space $T_x\mathcal{M}$ at each point x of the manifold.

Background - Riemannian metric, distance and geodesics I

- The structure of a manifold $\mathcal{M}$ is specified by a Riemannian metric.
- A ***Riemannian metric*** is a continuous collection of dot products $\langle \cdot, \cdot \rangle_x$ on the tangent space $T_x\mathcal{M}$ at each point x of the manifold.
- A local coordinate system $x = (x_1, ..., x_n)$ induces a basis $\frac{\partial}{\partial x} = (\frac{\partial}{\partial x_1}, ..., \frac{\partial}{\partial x_n})$ of the tangent spaces.

Background - Riemannian metric, distance and geodesics I

- The structure of a manifold $\mathcal{M}$ is specified by a Riemannian metric.
- A ***Riemannian metric*** is a continuous collection of dot products $\langle \cdot, \cdot \rangle_x$ on the tangent space $T_x\mathcal{M}$ at each point x of the manifold.
- A local coordinate system $x = (x_1, ..., x_n)$ induces a basis $\frac{\partial}{\partial x} = (\frac{\partial}{\partial x_1}, ..., \frac{\partial}{\partial x_n})$ of the tangent spaces.
- We can express the metric in this basis by a symmetric positive definite matrix $G(x) = [\langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \rangle]_{i,j}$ called the ***local representation of the Riemannian metric in the chart*** x.

Background - Riemannian metric, distance and geodesics I

- The structure of a manifold $\mathcal{M}$ is specified by a Riemannian metric.
- A ***Riemannian metric*** is a continuous collection of dot products $\langle \cdot, \cdot \rangle_x$ on the tangent space $T_x\mathcal{M}$ at each point x of the manifold.
- A local coordinate system $x = (x_1, ..., x_n)$ induces a basis $\frac{\partial}{\partial x} = (\frac{\partial}{\partial x_1}, ..., \frac{\partial}{\partial x_n})$ of the tangent spaces.
- We can express the metric in this basis by a symmetric positive definite matrix $G(x) = [\langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \rangle]_{i,j}$ called the ***local representation of the Riemannian metric in the chart*** x.
- The dot product of two vectors v and w in $T_x\mathcal{M}$ is now $\langle v, w \rangle_x = v^T G(x) w$.

Background - Riemannian metric, distance and geodesics II

- The length of a curve $\gamma(t)$ on the manifold is equal to

$$\mathcal{L}_a^b(\gamma) = \int_a^b \|\dot{\gamma}(t)\| dt = \int_a^b \left(\langle \dot{\gamma}(t) \mid \dot{\gamma}(t) \rangle_{\gamma(t)} \right)^{\frac{1}{2}} dt \tag{1}$$

Background - Riemannian metric, distance and geodesics II

- The length of a curve $\gamma(t)$ on the manifold is equal to

$$\mathcal{L}_a^b(\gamma) = \int_a^b \|\dot{\gamma}(t)\| dt = \int_a^b \left(\langle \dot{\gamma}(t) \mid \dot{\gamma}(t) \rangle_{\gamma(t)}\right)^{\frac{1}{2}} dt \tag{1}$$

- The distance between two points of a manifold is the minimum length among the smooth curves joining these points:

$$\mathrm{dist}(\mathrm{x}, \mathrm{y}) = \min_{\gamma} \mathcal{L}(\gamma) \quad \text{with} \quad \gamma(0) = \mathrm{x} \quad \text{and} \quad \gamma(1) = \mathrm{y} \tag{2}$$

The curves realizing this minimum for any two points of the manifold are called ***geodesics***.

Background - Exponential map and cut locus I

- From the theory of second order differential equations, we know that there exists one and only one geodesic $\gamma(x, \partial_v)$ going through the point $x \in \mathcal{M}$ at $t = 0$ with tangent vector $\partial_v \in T_x\mathcal{M}$.

Background - Exponential map and cut locus I

- From the theory of second order differential equations, we know that there exists one and only one geodesic $\gamma(x, \partial_v)$ going through the point $x \in \mathcal{M}$ at $t = 0$ with tangent vector $\partial_v \in T_x\mathcal{M}$.
- The exponential map maps each vector ∂_v to the point of the manifold reached in a unit time:

$$\exp_x : T_x\mathcal{M} \to \mathcal{M}, \quad \exp_x(\partial_v) = \gamma_{(x,\partial_v)}(1) \tag{3}$$

Background - Exponential map and cut locus I

- From the theory of second order differential equations, we know that there exists one and only one geodesic $\gamma(x, \partial_v)$ going through the point $x \in \mathcal{M}$ at $t = 0$ with tangent vector $\partial_v \in T_x\mathcal{M}$.
- The exponential map maps each vector ∂_v to the point of the manifold reached in a unit time:

$$\exp_x : T_x\mathcal{M} \to \mathcal{M}, \quad \exp_x(\partial_v) = \gamma_{(x,\partial_v)}(1) \tag{3}$$

- We denote by $\log_x = \exp_x^{-1}$ the inverse map or simply $\overrightarrow{xy} = \log_x(y)$.

Background - Exponential map and cut locus I

- From the theory of second order differential equations, we know that there exists one and only one geodesic $\gamma(x, \partial_v)$ going through the point $x \in \mathcal{M}$ at $t = 0$ with tangent vector $\partial_v \in T_x\mathcal{M}$.
- The exponential map maps each vector ∂_v to the point of the manifold reached in a unit time:

$$\exp_x : T_x\mathcal{M} \to \mathcal{M}, \quad \exp_x(\partial_v) = \gamma_{(x,\partial_v)}(1) \tag{3}$$

- We denote by $\log_x = \exp_x^{-1}$ the inverse map or simply $\overrightarrow{xy} = \log_x(y)$.
- The exponential map is a local diffeomorphism from a sufficiently small neighborhood of 0 in $T_x\mathcal{M}$ into a neighborhood of the point $x \in \mathcal{M}$.

Background - Exponential map and cut locus II

- What is the maximal domain where the exponential map is a diffeomorphism?

Background - Exponential map and cut locus II

- What is the maximal domain where the exponential map is a diffeomorphism?
- If we follow a geodesic it is either always minimizing or it is minimizing only up to a time $t_0 < \infty$. In this last case, the point $z = \gamma_{(x,\partial_v)}(t_0)$ is called a ***cut point*** and the corresponding tangent vector $t_0\partial_v$ a ***tangential cut point***.

Background - Exponential map and cut locus II

- What is the maximal domain where the exponential map is a diffeomorphism?
- If we follow a geodesic it is either always minimizing or it is minimizing only up to a time $t_0 < \infty$. In this last case, the point $z = \gamma_{(x,\partial_v)}(t_0)$ is called a ***cut point*** and the corresponding tangent vector $t_0\partial_v$ a ***tangential cut point***.
- The set of all cut points of all geodesics starting from x is the ***cut locus*** $C(x) \in \mathcal{M}$ and the set of corresponding vectors the ***tangential cut locus*** $\mathcal{C}(x) \in T_x\mathcal{M}$.

Background - Exponential map and cut locus II

- What is the maximal domain where the exponential map is a diffeomorphism?
- If we follow a geodesic it is either always minimizing or it is minimizing only up to a time $t_0 < \infty$. In this last case, the point $z = \gamma_{(x,\partial_v)}(t_0)$ is called a ***cut point*** and the corresponding tangent vector $t_0\partial_v$ a ***tangential cut point***.
- The set of all cut points of all geodesics starting from x is the ***cut locus*** $C(x) \in \mathcal{M}$ and the set of corresponding vectors the ***tangential cut locus*** $\mathcal{C}(x) \in T_x\mathcal{M}$.
- The maximal definition domain for the exponential chart is the domain $D(x)$ containing 0 and delimited by the tangential cut locus.

Example on the Sphere $\mathcal{S}_n$

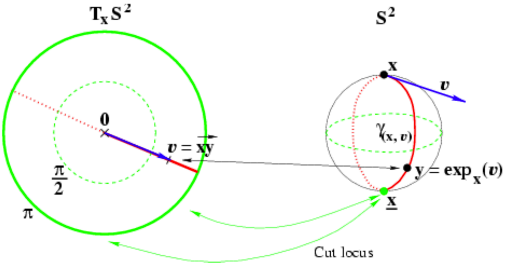


Figure: The geodesics are the great circles and the cut locus of a points x is its antipodal point −x. The exponential chart is obtained by rolling the sphere onto its tangent space so that the great circles going through x become lines. The maximal definition domain is thus the open ball $\mathcal{D} = \mathcal{B}_n(\pi)$.

Background - Random points on a Riemannian Manifold

Definition (Random point on a Riemannian manifold)

Let $(\Omega, \mathcal{B}(\Omega), Pr)$ be a probability space, $\mathcal{B}(\Omega)$ being the Borel σ-algebra of Ω (i.e. the smallest σ-algebra containing all the open subsets of Ω) and Pr a measure on that σ-algebra such that $Pr(\Omega) = 1$. A random point in the Riemannian manifold $\mathcal{M}$ is a Borel measurable function $\mathbf{x} = x(\omega)$ from Ω to $\mathcal{M}$.

Background - Riemannian Measure or volume form I

- **Vector space**
  In a vector space with basis $\mathcal{A} = (a_1, ..., a_n)$, the local representation of the metric is given by $G = A^T A$ where $A = [a_1, ...a_n]$ is the matrix of coordinates change from $\mathcal{A}$ to an orthonormal basis.
  The measure (or the infinitesimal volume element) is given by the volume of the parallelepipedon spanned by the basis vectors: $d\mathcal{V} = |\det(A)|dx = \sqrt{|\det(G)|}dx$.

Background - Riemannian Measure or volume form I

- **Vector space**
  In a vector space with basis $\mathcal{A} = (a_1, ..., a_n)$, the local representation of the metric is given by $G = A^T A$ where $A = [a_1, ... a_n]$ is the matrix of coordinates change from $\mathcal{A}$ to an orthonormal basis.
  The measure (or the infinitesimal volume element) is given by the volume of the parallelepipedon spanned by the basis vectors: $d\mathcal{V} = |\det(A)| dx = \sqrt{|\det(G)|} dx$.
- **Riemannian manifold**
  The Riemannian metric $G(x)$ induces an infinitesimal volume element on each tangent space, and thus a measure on the manifold:

$$d\mathcal{M}(\mathrm{x}) = \sqrt{|G(x)|} dx \tag{4}$$

Background - Riemannian Measure or volume form II

- The cut locus has a null measure. This means that to integrate in $\mathcal{M}$ we can integrate in any exponential chart.

Background - Riemannian Measure or volume form II

- The cut locus has a null measure. This means that to integrate in $\mathcal{M}$ we can integrate in any exponential chart.
- If f is an integrable function of the manifold and $f_x(\overrightarrow{xy}) = f(\exp_x(\overrightarrow{xy}))$ is its image in the exponential chart at x, we have:

$$\int_{\mathcal{M}} f(\mathrm{x})d\mathcal{M} = \int_{\mathcal{D}(\mathrm{x})} f_{\mathrm{x}}(\overrightarrow{\mathrm{z}})\sqrt{G_{\overrightarrow{\mathrm{x}}}(\overrightarrow{\mathrm{z}})}d\overrightarrow{\mathrm{z}} \tag{5}$$

Probability Density Function

Definition

Let $\mathcal{B}(\mathcal{M})$ be the Borel σ-algebra of $\mathcal{M}$. The random point $\mathbf{x}$ has a probability density function $p_{\mathbf{x}}$ (real, positive and integrable function) if:

$$\forall \mathcal{X} \in \mathcal{B}(\mathcal{M}), \quad \Pr(\mathbf{x} \in \mathcal{X}) = \int_{\mathcal{X}} p(\mathrm{y}) d\mathcal{M}(\mathrm{y}) \quad \text{and} \quad \Pr(\mathcal{M}) = \int_{\mathcal{M}} p(\mathrm{y}) d\mathcal{M}(\mathrm{y}) = 1$$

Probability Density Function

Definition

Let $\mathcal{B}(\mathcal{M})$ be the Borel σ-algebra of $\mathcal{M}$. The random point $\mathbf{x}$ has a probability density function $p_{\mathbf{x}}$ (real, positive and integrable function) if:

$$\forall \mathcal{X} \in \mathcal{B}(\mathcal{M}), \quad \Pr(\mathbf{x} \in \mathcal{X}) = \int_{\mathcal{X}} p(\mathrm{y}) d\mathcal{M}(\mathrm{y}) \quad \text{and} \quad \Pr(\mathcal{M}) = \int_{\mathcal{M}} p(\mathrm{y}) d\mathcal{M}(\mathrm{y}) = 1$$

Example Uniform pdf in a bounded set $\mathcal{X}$:

$$p_{\mathcal{X}}(\mathrm{y}) = \frac{1}{\int_{\mathcal{X}} d\mathcal{M}} \mathbf{1}_{\mathcal{X}}(\mathrm{y}) = \frac{\mathbf{1}_{\mathcal{X}}(\mathrm{y})}{\mathcal{V}ol(\mathcal{X})} \tag{6}$$

Expectation of an observable

- $\varphi(x)$ Borelian real valued function on $\mathcal{M}$,
- $\mathbf{x}$ random point of pdf $p_{\mathbf{x}}$
- $\varphi(\mathbf{x})$ real random variable with expectation:

$$\mathbf{E}[\varphi(\mathbf{x})] = \mathbf{E}_{\mathbf{x}}[\varphi] = \int_{\mathcal{M}} \varphi(\mathrm{y}) p_{\mathbf{x}}(\mathrm{y}) d\mathcal{M}(\mathrm{y})$$

- Can do this integration since we're integrating over a real variable!
- However, can't define a mean like this

Fréchet expectation of a random point

- Idea: in $\mathbb{R}^n$, the mean is the point that minimizes the variance of a distribution
- Variance of a random point:

$$\sigma_{\mathbf{x}}^2(\mathrm{y}) = \mathbf{E}\left[\mathrm{dist}(\mathrm{y}, \mathbf{x})^2\right] = \int_{\mathcal{M}} \mathrm{dist}(\mathrm{y}, \mathrm{z})^2 p_{\mathbf{x}}(\mathrm{z}) d\mathcal{M}(\mathrm{z})$$

- The distance is a real valued function, so we can integrate it and use it to get a variance
- The set of mean points:

$$\mathbb{E}[\mathbf{x}] = \arg\min_{\mathbf{y} \in \mathcal{M}} \left(\mathbf{E}\left[\mathrm{dist}(\mathrm{y}, \mathbf{x})^2\right]\right)$$

Fréchet expectation [cont]

- Discrete or empirical mean:

$$\mathbb{E}[\{x_i\}] = \arg\min_{\mathbf{y}\in\mathcal{M}} \left(\frac{1}{n} \sum_i \mathrm{dist}(\mathrm{y}, x_i)^2 \right)$$

- Mean deviation at order α:

$$\sigma_{\mathbf{x},\alpha}(y) = (\mathbf{E}[\mathrm{dist}(\mathrm{y}, x_i)^{\alpha}])^{1/\alpha}$$

- $\alpha = 0$ yields ***modes***, $\alpha = 1$ yields ***median***, and $\alpha \to \infty$ yields ***barycenter*** of distribution support

Uniqueness of the mean

- In general, no guarantee of a global minimum, so no guaranteed Fréchet mean
- Instead consider all local minima, called the ***Riemannian centers of mass***
- A ball of radius r centered on y $\mathcal{B}(y, r)$ is ***geodesic*** if it does not contain the cut locus of its center, y. It is ***regular*** if it has radius $r < \frac{\pi}{2\sqrt{\kappa}}$ where κ is the maximum Riemannian curvature in the ball.
  - Example: sphere $\mathcal{S}_2$ with unit radius, has curvature equal to 1. A ball on the sphere is regular if $r < \pi/2$.

Theorem

If $\mathbf{x}$ is a random point of pdf $p_{\mathbf{x}}$ and the support of $p_{\mathbf{x}}$ is contained in a regular geodesic ball $\mathcal{B}(y, r)$, then $\exists!$ Riemannian center of mass on the ball.

Alternative definitions of the mean

- Instead of minimizing variance, we could choose other ways to define a mean as the $\bar{x}$ satisfying:

$$\forall \mathrm{y} \in \mathcal{M} \quad \mathsf{dist}(\mathrm{y}, \bar{\mathrm{x}}) \leq \mathbf{E}[\mathsf{dist}(\mathbf{x}, y)]$$

- Instead of looking at distances, could use general convex functions – the ***convex barycenter*** is the set $\mathbb{B}(\mathbf{x})$ of points $y \in \mathcal{M}$ satisfying:

$$\alpha(y) \leq \mathbf{E}[\alpha(\mathbf{x})]$$

  where α is any convex function

- These definitions are not useful for compact groups

Gradient of the Variance Function

Theorem

- *The variance $\sigma^2(y)$ for a probability P on the manifold $\mathcal{M}$ is differentiable at any point $y \in \mathcal{M}$ where it is finite and where the cut locus $C(y)$ has a null probability measure, $P(C(y)) = 0$*
- *At such a point, it has the gradient:*

$$\left(\operatorname{grad}\sigma^2\right)(\mathrm{y}) = -2\int_{\mathcal{M}/C(\mathrm{y})} \overrightarrow{\mathrm{yz}}\, dP(\mathrm{z}) = -2\mathbf{E}[\overrightarrow{\mathrm{yx}}]$$

Characterization of Riemannian centers of mass

- Define $\mathbf{x}$ a random point with finite variance everywhere, and $\mathcal{A}$ the set of points where the cut locus has a non-zero probability measure.
- A necessary condition for $\bar{x}$ to be a Riemannian center of mass is that $\mathbf{E}[\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}] = 0$ if $\bar{x} \notin \mathcal{A}$, or $\bar{x} \in \mathcal{A}$
- If $\mathcal{M}$ is simply connected, complete, and has a non-positive curvature that is bounded from below, then there is a unique center of mass with $\mathbf{E}[\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}] = 0$.

Gradient descent algorithm to obtain the mean

- Using the exponential map between $\mathcal{M}$ and its tangent space, we can use an iterative algorithm for gradient descent.
- Using a Taylor expansion and neglecting any issues from the cut locus, we get:

$$\mathrm{y}_{t+1} = \exp_{\mathrm{y}_t}\left(\mathbf{E}\left[\overrightarrow{\mathrm{y_t}\mathbf{x}}\right]\right)$$

- Can use a random observation as the starting point
- Can repeat with several starting points to verify uniqueness

Covariance Matrix

- The covariance matrix of a random vector $\mathbf{x}$ with respect to a point y can be thought of as the directional dispersion of the difference vector $\overrightarrow{y\mathbf{x}} = \mathbf{x} - y$
- Can extend this to Riemannian manifold:

$$\Sigma_{\mathbf{xx}} = \mathrm{Cov}_{\mathbf{x}}(\bar{x}) = \mathbf{E}\left[\overrightarrow{\bar{x}\mathbf{x}}\,\overrightarrow{\bar{x}\mathbf{x}}^T\right] = \int_{\mathcal{D}(\bar{x})} (\overrightarrow{\bar{x}x})(\overrightarrow{\bar{x}x})^T p_{\mathbf{x}}(\mathrm{x}) d\mathcal{M}(\mathrm{x})$$

- The trace is related to the variance:

$$\mathrm{Tr}(\Sigma_{\mathbf{xx}}) = \sigma_{\mathbf{x}}^2$$

Covariance

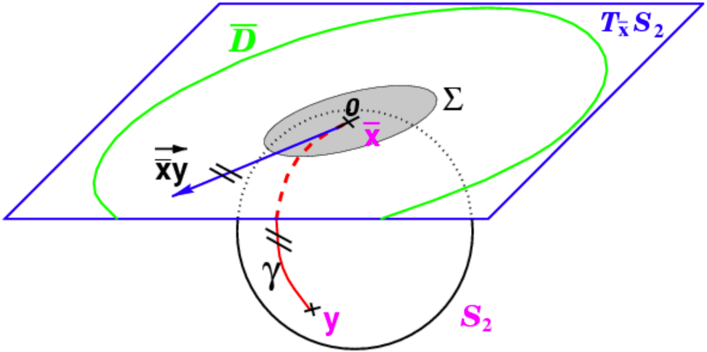


Figure: Covariance on sphere $\mathcal{S}_2$ at the mean point $\bar{x}$

Generalization of the Normal distribution

- Stochastic calculus: generalize solutions to the heat equation. This is impossible to explicitly calculate in most cases
- Directional statistics: use a wrapped Gaussian distribution. This can be very complex to calculate the properties for in less-simple cases.
- This paper: taking the mean and covariance matrix for granted, what distribution maximizes entropy?
- Entropy:

$$\mathbf{H}[\mathbf{x}] = \mathbf{E}[-\log(p_x(\mathbf{x}))] = -\int_{\mathcal{M}} \log(p_x(x))p_x(x)d\mathcal{M}(x)$$

Normal Law

Theorem

The Normal law on the manifold $\mathcal{M}$ is the maximum entropy distribution knowing the mean value and covariance. Assuming no continuity or differentiability constraint on the cut locus $C(\bar{x}$ and sa symmetric domain $\mathcal{D}(\bar{x})$, pdf of Normal law of mean $\bar{x}$ and concentration matrix Γ is:

$$N_{(\bar{x},\Gamma)}(y) = k \exp\left(- \frac{\overrightarrow{\bar{x}y}^T \Gamma \overrightarrow{\bar{x}y}}{2} \right)$$

Normal Law (cont)

Theorem

$$N_{(\bar{x},\Gamma)}(y) = k \exp\left(-\frac{\overrightarrow{\bar{x}y}^T \Gamma \overrightarrow{\bar{x}y}}{2}\right)$$

Normalization constant and covariance are related to the concentration matrix by:

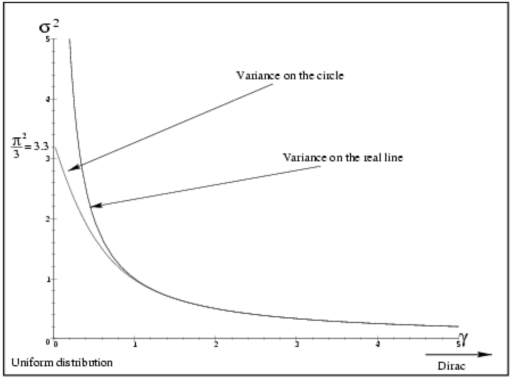
$$k^{(-1)} = \int_{\mathcal{M}} \exp\left(-\frac{\overrightarrow{\bar{x}y}^T \Gamma \overrightarrow{\bar{x}y}}{2}\right) d\mathcal{M}(y)$$

and

$$\Sigma = k \int_{\mathcal{M}} \overrightarrow{\bar{x}y}\,\overrightarrow{\bar{x}y}^T \exp\left(-\frac{\overrightarrow{\bar{x}y}^T \Gamma \overrightarrow{\bar{x}y}}{2}\right)$$

Example: Circle

On a circle with radius r, the concentration is related to the variance by:
$\sigma^2 = \frac{1}{\gamma}(1 - 2\pi r k \exp(-\frac{\gamma(\pi r)^2}{2}))$



Approximations

- Small concentration matrix on a compact manifold:
  - We recover the uniform distribution:

$$N(y) = 1/Vol(\mathcal{M}) + O(\mathrm{Tr}(\Gamma))$$

  - Moreover, Σ stays finite

Approximations

- Small concentration matrix on a compact manifold:
  - We recover the uniform distribution:

$$N(y) = 1/Vol(\mathcal{M}) + O(\mathrm{Tr}(\Gamma))$$

  - Moreover, Σ stays finite
- Large concentration matrix, using a Taylor expansion around mean:

$$k = \frac{1 + O(\sigma^3) + \epsilon(\frac{\sigma}{r})}{\sqrt{(2\pi)^n \det(\Sigma)}}$$

$$\Gamma = \Sigma^{(-1)} - \frac{1}{3}\,\mathrm{Ric} + O(\sigma) + \epsilon(\frac{\sigma}{r})$$

Mahalanobis distance and χ^2 law

- What if we want to find out if an observation $\hat{x}$ came from a given distribution $\mathbf{x}$ on $\mathcal{M}$, or if it is an outlier?
- For $\mathbb{R}^n$, use Mahalnobis distance $\mu^2 = (\hat{x} - \bar{x})^T \Sigma_{\mathbf{xx}}^{(-1)} (\hat{x} - \bar{x})$
- D^2 test: compare μ^2 to tail of its expected distribution.
- When we assume that $\mathbf{x}$ is Gaussian, then μ^2 should be χ^2 distributed if observation is not an outlier.
- Note: we don't need to know the mean or variance of the Gaussian to use this test
- Mahalnobis distance for Riemannian manifold:

$$\mu_{\mathbf{x}}^2(y) = \overrightarrow{\bar{x}y}^T \Sigma_{\mathbf{xx}}^{(-1)} \overrightarrow{\bar{x}y}$$

Generalized χ^2 law

- If we assume a random point $\mathbf{x} \sim (\bar{x}, \Sigma_{\mathbf{xx}})$ is normal, we can compute the probability that $\chi^2 = \mu_{\mathbf{x}}^2 < \alpha^2$
- This χ^2 law is independent of the mean value and covariance of the random point, up to order $O(\sigma^3)$.
- In practice, pick confidence level γ, then find $\alpha(\gamma)$ such that $\gamma = Pr\{\chi^2 \leq \alpha^2\}$

Applications

- These techniques have found significant use in computer vision on medical images: for example, predicting 3D rotations given a set of 2D images
- Generalizing k-means to 3D shapes and rotations
- Applications to generative modelling:
- If the latent space of a VAE is taken to be a manifold, this theory can be used to model a Gaussian distribution in that space.
- The same goes for normalizing flows which learn to map a Gaussian distribution, one-to-one, onto a desired distribution. We may wish to learn a distribution on a manifold, in which case we will need a Gaussian distribution on a manifold.

Summary

We have seen how to extend the following notions from statistics in vector spaces to Riemannian manifolds.

- Mean
- Variance
- Covariance
- Gaussian
- χ^2