PROBABILITIES AND STATISTICS ON RIEMANNIAN MANIFOLDS: BASIC TOOLS FOR GEOMETRIC MEASUREMENTS

Xavier Pennec

INRIA - Projet Epidaure
2004 Route des Lucioles
BP93, 06902 Sophia Antipolis Cedex, France.
http://www-sop.inria.fr/epidaure/personnel/pennec/pennec.html
email: Xavier.Pennec@sophia.inria.fr

ABSTRACT

Measurements of geometric primitives, such as rotations or rigid transformations, are often noisy and we need to use statistics either to reduce the uncertainty or to compare measurements. Unfortunately, geometric primitives often belong to manifolds and not vector spaces. We have already shown [9] that generalizing too quickly even simple statistical notions could lead to paradoxes.

In this article, we develop some basic probabilistic tools to work on Riemannian manifolds: the notion of mean value, covariance matrix, normal law, Mahalanobis distance and χ^2 test. We also present an efficient algorithm to compute the mean value and tractable approximations of the normal and χ^2 laws for small variances.

1. INTRODUCTION

To represent the results of a random experiment, one usually constructs a probabilized space $(\Omega, \mathcal{A}, \Pr)$ and models the measurements by *random variables* or *observables* (maps $\boldsymbol{x} = x(\omega)$ from Ω to $\mathbb{R}$). This formalism allows to "forget" the original probabilized space and work directly in $\mathbb{R}$ by associating to each random variable $\boldsymbol{x}$ or $\boldsymbol{y}$ different probability density functions (pdfs). However, from a computational point of view, we have to restrict the measurements to a few numeric characteristics of a random variable. Thus, one usually approximates a unimodal pdf its *mean value* or *expectation* and its *variance*. As the expectation is a linear operator, it is easy to generalize that to the expectation and covariance matrix of a *random vector* (several simultaneous measurements of the same random experiment). To compare measurements, one can use a probabilistic distance between distributions such as the Mahalanobis distance or simple statistical tests such as the χ^2 test. This is justified by the fact that the Gaussian minimizes the information knowing the mean and the covariance.

The problem we investigate in this article is to generalize this framework to measurements in Riemannian manifolds instead of measurements in a vector space. We call them *random primitives*. Examples of manifolds we routinely use are 3D rotations and 3D rigid transformations as transformation groups and frames (a 3D point and an orthonormal trihedron) semi- or non-oriented frames (where 2 (resp. 3) of the trihedron unit vectors are given up to their sign), oriented or directed points. We have already shown in [10, 9] that this is not an easy problem and that some paradoxes can arise. In particular, we cannot generalize the expectation to give a mean value since it would be an integral with value in the manifold.

We review in the remainder of this section some basic notions of differential and Riemannian geometry (see [8, 1]). In Section 2, we develop the notion of pdf, mean value and covariance matrix of a random primitive. In Section 3, we investigate a generalization of the Normal law and show how it can be approximated for small variances. We finish in Section 4 by the generalization of the Mahalanobis distance and the χ^2 law.

1.1. Riemannian metric, distance and geodesics

In the geometric framework, one specifies the structure of a manifold $\mathcal{M}$ by a *Riemannian metric*. This is a continuous collection of dot products on the tangent space $T_{\mathrm{x}}\mathcal{M}$ at each point of the manifold. A local coordinate system induces a basis of the tangent spaces. Thus, we can express the metric in this basis by a symmetric positive definite matrix $Q(\mathrm{x}) = [\langle\, \partial_i \mid \partial_j \,\rangle_{\mathrm{x}}]$ where each element is given by the dot product of the tangent vector to the coordinate curves.

Considering a curve on the manifold, we can compute at each point its instantaneous speed vector and its norm, the instantaneous speed. The length is obtained as usual by integrating this value along the curve. To obtain a distance between two points of a connected Riemannian manifold, we simply have to take the minimum length among the smooth curves joining these points. The curves realizing this minimum for any two points of the manifold are called geodesics. The manifold is said to be *geodesically complete* if the definition domain of all geodesics can be extended to $\mathbb{R}$. As an important consequence, the Hopf-Rinow-De Rham theorem state that such a manifold is complete for the induced distance, and that there always exist at least one minimizing geodesic between any two points of the manifold (i.e. which length is the distance between the two points). From now on, we will assume that the manifold is geodesically complete.

1.2. Exponential map and cut locus

From the theory of second order differential equations, we know that there exits one and only one geodesic starting at a given point x with a given tangent vector. This allows to develop the manifold in the tangent space along the geodesics (think of rolling a sphere along its tangent plane at a given point). The geodesics going through this point are transformed into straight lines and the distance along these geodesics are conserved (at least in a neighborhood of x). The function that maps to each vector the corre-

sponding point on the manifold is called the *exponential map*.

This map is defined in the whole tangent space $T_{\mathrm{x}}\mathcal{M}$ (since the manifold is geodesically complete) but it is one-to-one only locally around the origin. We denote by $\overrightarrow{\mathrm{xy}}$ its the inverse: this is the smallest vector such that $\mathrm{y} = \exp_{\mathrm{x}}(\overrightarrow{\mathrm{xy}})$. If we look for the maximal definition domain, we find out that it is a star-shaped domain delimited by a continuous curve C_{x} called the *tangential cut-locus*. The image of C_{x} by the exponential map is the cut locus $\mathcal{C}_{\mathrm{x}}$ of point x. This is closure of the set of points where several minimizing geodesics starting from x meet. On the sphere $\mathcal{S}_2(1)$ for instance, the cut locus of a point x is its antipodal point and the tangential cut locus is the circle of radius π.

The exponential map within this domain realizes a chart called *the exponential chart*. It covers all the manifold except the cut locus of the development point, which has a null measure. In this chart, geodesics starting from x are straight lines. and the distance from the development point are conserved. This chart is somehow the "most linear" chart of the manifold with respect to the primitive x.

1.3. Riemannian measure or volume form

The Riemannian metric $Q(x)$ induces an infinitesimal volume element on each tangent space, and thus a measure on the manifold:

$$d\mathcal{M}(\mathrm{x}) = \sqrt{|Q(x)|}.dx \tag{1}$$

One can show that the cut locus has a null measure. This means that we can integrate indifferently in $\mathcal{M}$ or in any exponential chart.

2. PROBABILITIES ON A RIEMANNIAN MANIFOLD

In this section, we do not consider measurements that are real variables or vectors depending on the outcome of a random experiment, but rather random measurements of a manifold elements. We call such a measurement a *random primitive*.

Definition 1 (Random primitive) *Let $(\Omega, \mathcal{B}, \Pr)$ be a probabilized space. A random primitive in the Riemannian manifold $\mathcal{M}$ is a Borelian function $\mathbf{x} = \mathrm{x}(\omega)$ from Ω to $\mathcal{M}$.*

As in the real or vectorial case, we can now make abstraction of the original space Ω and directly work with the induced probability measure on $\mathcal{M}$.

2.1. Probability density function

Definition 2 *Let $\mathcal{A}$ be the Borelian tribe of $\mathcal{M}$. The random primitive $\mathbf{x}$ has a probability density function $p_{\mathbf{x}}$ (real, positive and integrable function) if $\forall \mathcal{X} \in \mathcal{A}$:*

$$\Pr(\mathbf{x} \in \mathcal{X}) = \int_{\mathcal{X}} p(\mathrm{y}).d\mathcal{M}(\mathrm{y}) \quad \textit{and} \quad \Pr(\mathcal{M}) = 1 \tag{2}$$

A simple example of a pdf is the *uniform pdf* in a bounded set $\mathcal{X}$: $p_{\mathcal{X}}(\mathrm{y}) = \mathbf{1}_{\mathcal{X}}(\mathrm{y})/\mathcal{V}(\mathcal{X})$, where $\mathcal{V}(\mathcal{X})$ is the "volume" of the set $\mathcal{X}$.

One must be careful that this pdf is uniform with respect to the measure $d\mathcal{M}$ and is not uniform for another measure on the manifold. This problem is the basis of the Bertrand paradox for geometrical probabilities [12, 6] and raise the problem of the measure to choose on the manifold. In our case, the measure is induced by the Riemannian metric but the problem is only lifted: which Riemannian metric do we have to choose ? We addressed this question in [9] for transformation groups and homogeneous manifolds by showing that an invariant metric is a good geometric choice.

2.2. Expectation and Mean value

Here, we focus on the notion of central value of a random primitive. We will preferably use the denomination *mean value or mean primitive* than *expected primitive* to stress the difference between this notion and the expectation of a real function.

Expectation of an observable Let $\varphi(\mathrm{x})$ be a Borelian real valued function defined on $\mathcal{M}$ and $\mathbf{x}$ a random primitive of pdf $p_{\mathbf{x}}$. Then, $\varphi(\mathbf{x})$ is a real random variable and we can compute its expectation:

$$\mathbf{E}\left[\,\varphi(\mathbf{x})\,\right] = \mathbf{E}_{\mathbf{x}}\left[\,\varphi\,\right] = \int_{\mathcal{M}} \varphi(\mathrm{y}).p_{\mathbf{x}}(\mathrm{y}).d\mathcal{M}(\mathrm{y}) \tag{3}$$

This notion of expectation corresponds to the one we defined on real random variables and vectors. However, we cannot directly extend it to define the mean value of the distribution since we have no way to generalize this integral in $\mathbb{R}$ into an integral with value in the manifold.

Fréchet expectation or mean value Let $\boldsymbol{x}$ be a random vector of $\mathbb{R}^n$. Fréchet observed in [3, 4] observed that the variance $\sigma^2_{\boldsymbol{x}}(y) = \mathbf{E}\left[\ \mathrm{dist}(\boldsymbol{x}, y)^2\ \right]$ is minimized for the mean vector $\overline{x} = \mathbf{E}\left[\,\boldsymbol{x}\,\right]$. The major point for the generalization is that the expectation of a real valued function is well defined for our connected and geodesically complete Riemannian manifold $\mathcal{M}$.

Definition 3 (Variance of a random primitive) *Let $\mathbf{x}$ be a random primitive of pdf $p_{\mathbf{x}}$. The variance $\sigma^2_{\mathbf{x}}(\mathrm{y})$ is the expectation of the squared distance between the random primitive and the fixed primitive y:*

$$\sigma^2_{\mathbf{x}}(\mathrm{y}) = \mathbf{E}\left[\ \mathrm{dist}(\mathrm{y}, \mathbf{x})^2\ \right] = \int_{\mathcal{M}} \mathrm{dist}(\mathrm{y}, \mathrm{z})^2.p_{\mathbf{x}}(\mathrm{z}).d\mathcal{M}(\mathrm{z}) \tag{4}$$

Definition 4 (Fréchet expectation of a random primitive) *If the variance $\sigma^2_{\mathbf{x}}(\mathrm{y})$ of a random primitive $\mathbf{x}$ is finite for all primitive $\mathrm{y} \in \mathcal{M}$, every primitive $\bar{\mathrm{x}}$ minimizing this variance is called an expected or mean primitive. Thus, the set of the mean primitives is:*

$$\mathbb{E}\left[\,\mathbf{x}\,\right] = \arg\min_{\mathrm{y} \in \mathcal{M}} \left(\mathbf{E}\left[\ \mathrm{dist}(\mathrm{y}, \mathbf{x})^2\ \right]\right) \tag{5}$$

One can define other types of central values based on the *mean deviation at order* α: $\sqrt[\alpha]{\mathbf{E}\left[\ \mathrm{dist}(\mathrm{y}, \mathbf{x})^\alpha\ \right]}$. For instance, the *modes* are obtained for $\alpha = 0$. Exactly like in a vector space, they are the primitives where the density is maximal on the manifold. The *median primitive* is obtained for $\alpha = 1$. For $\alpha \to \infty$, we obtain the "barycenter" of the distribution support (which has to be compact).

Existence and uniqueness: Karcher expectation As our mean primitive is the result of a minimization, its existence is not ensured and the result is a set and no longer a single element. This is to be compared with some central values in vector spaces, such as the modes. However, the Fréchet expectation does not define all the modes even in vector spaces: one only keeps the modes of maximal intensity.

To get rid of this constraint, [5] proposed to consider the local minima of the variance $\sigma^2_{\mathbf{x}}(\mathrm{y})$ instead of the global ones. Thus, the Fréchet mean primitives are a subset of the Karcher ones. Using this extended definition, [5] and [7] established that, for localized enough distributions, there exists one and only one Karcher mean.

The Karcher mean is perfectly adapted for our purpose, thanks to the good properties it has for optimization (see below). However, there are other works proposing different ways to generalize

the notion of mean value or barycenter of a distribution in a manifold. We review them in [11] but they do not seem to be practically applicable.

2.3. Characterizing a Karcher mean

Local minima are entirely characterized by a null gradient and a positive definite Hessian matrix. Let $f_{\mathrm{z}}(\mathrm{y}) = \mathrm{dist}(\mathrm{y},\mathrm{z})^2 = \|\overrightarrow{\mathrm{yz}}\|^2$. For a fixed z, the gradient of $f_{\mathrm{z}}(\mathrm{y})$ is defined everywhere on the manifold except on the cut locus $C(\mathrm{z})$. On can show that $\mathrm{grad}\, f_{\mathrm{z}} = 2.\overrightarrow{\mathrm{zy}} = -2.\overrightarrow{\mathrm{yz}}$.

Since the integration domain $\mathcal{M} - C(\mathrm{y})$ depends on the variable y, the conditions to differentiate under the sum are not fulfilled. However, when the manifold is compact (and of course when there is not cut locus), we were able to establish that:

$$\mathrm{grad}\,\left(\sigma^2_{\mathbf{x}}(\mathrm{y})\right) = \int_{\mathcal{M}-C(\mathrm{y})} -2\overrightarrow{\mathrm{yz}}.p_{\mathbf{x}}(\mathrm{z}).d\mathcal{M}(\mathrm{z}) = -2.\mathbf{E}\left[\,\overrightarrow{\mathrm{y\bar{x}}}\,\right]$$

We believe that this is still true for more general manifolds, with perhaps a few conditions on the density.

Going one step further is even more difficult: we do not have a general formula to compute the Hessian matrix, except when there is no cut-locus. In this case, we can differentiate once again under the sum: $\mathrm{Hess}\,\left(\sigma^2_{\mathbf{x}}(\mathrm{y})\right) = 2.\,\mathrm{Id}$

Theorem 1 (Characterization of Karcher means) *Let* $\mathbf{x}$ *be a random primitive in either a compact manifold or a manifold without cut-locus. A necessary (but not sufficient) condition for a point* $\bar{\mathrm{x}}$ *to be a Karcher mean of the random primitive* $\mathbf{x}$ *is that the random vector* $\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}$ *has a null expectation.*

$$\bar{\mathrm{x}} \in \mathbb{E}\left[\,\mathbf{x}\,\right] \quad \Longrightarrow \quad \mathbf{E}\left[\,\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}\,\right] = \mathbf{E}_{\mathbf{x}}\left[\,\exp_{\bar{\mathrm{x}}}\,\right] = 0 \tag{6}$$

If the manifold has no cut locus, the Hessian is constant and positive definite. Hence the condition is sufficient.

2.4. A gradient descent algorithm to obtain the mean

Let y be an estimation of the mean of the random primitive $\mathbf{x}$ and $f(\mathrm{y}) = \sigma^2_{\mathbf{x}}(\mathrm{y})$ the variance. The intrinsic second order Taylor expansion of f and y is:

$$f(\exp_{\mathrm{y}}(v)) = f(\mathrm{y}) + \mathrm{grad}\, f(v) + \frac{1}{2}\mathrm{Hess}\, f(v,v)$$

This is a function of the vector $v \in T_{\mathrm{y}}\mathcal{M}$. Using the approximation $\mathrm{Hess}\, f \simeq 2.\,\mathrm{Id}$, this function is concave and thus has a minimum. Let $H_f(v)$ denote the linear form verifying $\langle\, H_f(v) \mid w \,\rangle = \mathrm{Hess}\, f(v,w)$ for all w and $H_f^{(-1)}$ denote the inverse map. The minimum is characterized by

$$\mathrm{grad}\, f + H_f(v) = 0 \quad \Leftrightarrow \quad v = -H_f^{(-1)}(\mathrm{grad}\, f)$$

Since $\mathrm{grad}\, f = -2.\mathbf{E}\left[\,\overrightarrow{\mathrm{y\bar{x}}}\,\right]$, we obtain the following evolution equations:

$$\mathrm{y}_{t+1} = \exp_{\mathrm{y}_t}\left(\mathbf{E}\left[\,\overrightarrow{\mathrm{y_t\bar{x}}}\,\right]\right) \tag{7}$$

In the case of a vector space, these formula converge in a single step to the mean value and the barycenter: $y_{t+1} = \mathbf{E}\left[\,\boldsymbol{x}\,\right]$. In the case of the discrete or empirical mean, which is much more interesting from a statistical point of view, we have exactly the same algorithm, but with the empirical expectation $\mathrm{y}_{t+1} = \exp_{\mathrm{y}_t}\left(\frac{1}{n}\sum_i \overrightarrow{\mathrm{y_t x_i}}\right)$.

An important point for this algorithm is to determine a good starting point. In the case on a set of measurements $\{\mathrm{x}_i\}$, on can choose at random one of the measurements as the starting point. Another solution is to map to each point x_i is mean distance with respect to other points (or the median distance to be robust) and choose as the starting point the minimizing point.

2.5. Covariance matrix

With the mean value, we have a dispersion value: the variance. To go one step further, we observe that the covariance matrix of a random vector $\boldsymbol{x}$ with respect to a point y is the *directional* dispersion of the "difference" vector $\overrightarrow{y\boldsymbol{x}} = \boldsymbol{x} - y$:

$$\mathrm{Cov}_y(\boldsymbol{x}) = \mathbf{E}\left[\,\overrightarrow{y\boldsymbol{x}}.\overrightarrow{y\boldsymbol{x}}^{\mathrm{T}}\,\right] = \int_{\mathbb{R}^n} (\overrightarrow{y\boldsymbol{x}}).(\overrightarrow{y\boldsymbol{x}})^{\mathrm{T}}.p_{\boldsymbol{x}}(x).dx$$

This definition is easily extendible to a complete Riemannian manifold using the random vector $\overrightarrow{\mathrm{y\bar{x}}}$ in $T_{\mathrm{y}}\mathcal{M}$ and the Riemannian measure. As in the vector case, the covariance is related to the variance by $\mathrm{Tr}(\mathrm{Cov}_{\mathrm{y}}(\mathbf{x})) = \sigma^2_{\mathbf{x}}(\mathrm{y})$. In fact, we are usually interested in the covariance relative to the mean value:

Definition 5 (Covariance) *Let* $\bar{\mathrm{x}} \in \mathbb{E}\left[\,\mathbf{x}\,\right]$ *be the unique mean value of the random primitive* $\mathbf{x}$*. We note* $\Sigma_{\mathbf{xx}}$ *and we call covariance the expression:*

$$\Sigma_{\mathbf{xx}} = \mathrm{Cov}_{\bar{\mathrm{x}}}(\mathbf{x}) = \mathbf{E}\left[\,\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}.\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}^{T}\,\right] \tag{8}$$

The covariance depends on the basis used for the exponential chart if we see it as a matrix, but it does not depend on it if we consider it as a bilinear form over the tangent plane. In fact, as soon as we have found the mean value, everything appears to be similar to the case of a centered random vector by developing the manifold onto the tangent space at the mean value. Indeed, $\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}$ is a random vector in $T_{\bar{\mathrm{x}}}\mathcal{M}$ whose expectation is null and whose covariance matrix is $\mathbf{E}\left[\,\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}.\overrightarrow{\bar{\mathrm{x}}\mathbf{x}}^{\mathrm{T}}\,\right]$. Thus, we could define higher order moments of the distribution by tensors on this tangent space, just as we have done for the covariance.

3. GENERALIZING THE NORMAL LAW

We now present an approach based on the the information minimization to generalize the Normal distribution to a manifold. In this section the symbols $\log$ and $\exp$ denote the standard logarithmic and exponential functions in $\mathbb{R}$.

3.1. Information and uniform law

As we can integrate a real valued function, the extension of the *entropy* $\mathbf{H}\left[\,\mathbf{x}\,\right]$ (or its opposite, the *information* $\mathbf{I}\left[\,\mathbf{x}\,\right]$) of a random primitive is straightforward:

$$\mathbf{I}\left[\,\mathbf{x}\,\right] = -\mathbf{H}\left[\,\mathbf{x}\,\right] = \mathbf{E}\left[\,\log(p_{\mathbf{x}}(\mathbf{x}))\,\right] \tag{9}$$

This definition is consistent since the pdf $p_{\mathcal{U}}$ that minimize the information when we only know that the measure is in a compact set $\mathcal{U}$ is the uniform density in this set.

3.2. Constrained information minimization

Now assume that we know the mean (that we suppose to be unique) and the covariance of a random primitive: we denote it by $\mathbf{x} \sim$

$(\bar{\mathrm{x}}, \Sigma)$. The least informative pdf minimizes the conditional information:

$$\mathbf{I}\left[\,\mathbf{x} \mid \bar{\mathrm{x}} \in \mathbb{E}\left[\,\mathrm{x}\,\right]\ ,\ \Sigma_{\mathbf{xx}} = \Sigma\,\right]$$

Expressing the normalization, fixed mean value and fixed covariance constraints in the exponential chart at the mean value and neglecting any continuity or differentiability constraint on the cut locus, we can write the Lagrangian in the vector space $T_{\mathrm{x}}\mathcal{M}$ and we obtain:

Theorem 2 (Normal law) *We call Normal law on the manifold $\mathcal{M}$ the pdf minimizing the information with a fixed mean value and covariance. Assuming no continuity nor differentiability constraint on the cut locus $C(\bar{\mathrm{x}})$ and a symmetric domain $\mathcal{D}(\bar{\mathrm{x}})$, the Normal law of mean $\bar{\mathrm{x}}$ and concentration matrix Γ is given by:*

$$N_{(\bar{\mathrm{x}},\Gamma)}(\mathrm{y}) = k.\exp\left(\Leftrightarrow\frac{\overleftrightarrow{\mathrm{xy}}^{T}.\Gamma.\overleftrightarrow{\mathrm{xy}}}{2}\right) \tag{10}$$

where the normalization constant is

$$k^{(-1)} = \int_{\mathcal{M}} \exp\left(\Leftrightarrow\frac{\overleftrightarrow{\mathrm{xy}}^{T}.\Gamma.\overleftrightarrow{\mathrm{xy}}}{2}\right).d\mathcal{M}(\mathrm{y}) \tag{11}$$

The covariance Σ and concentration Γ are related by:

$$\Sigma = k.\int_{\mathcal{M}} \overleftrightarrow{\mathrm{xy}}.\overleftrightarrow{\mathrm{xy}}^{T}.\exp\left(\Leftrightarrow\frac{\overleftrightarrow{\mathrm{xy}}^{T}.\Gamma.\overleftrightarrow{\mathrm{xy}}}{2}\right).d\mathcal{M}(\mathrm{y}) \tag{12}$$

From the concentration matrix, we can compute the covariance of the random primitive, at least numerically, but the reverse is more difficult. Of course, in a vector space, the integrals can be entirely computed, and we find the usual Gaussian density.

3.3. Example on a simple manifold: the circle

The exponential chart for the circle of radius 1 with the canonical metric is the angle with respect to the development point $\theta \in \mathcal{D} =]\Leftrightarrow\pi;\pi[$, and the measure is simply $d\theta$. For a circle of radius r, the exponential chart becomes $x = r.\theta$. The domain is $\mathcal{D} =]\Leftrightarrow a; a[$ (with $a = \pi.r$) and the measure is $dx = r.d\theta$. Thus, the normalization factor of the Normal density is:

$$k^{(-1)} = \int_{-a}^{a} \exp\left(\Leftrightarrow\frac{\gamma.x^2}{2}\right).dx = \sqrt{\frac{2\pi}{\gamma}}.\mathrm{erf}\left(\sqrt{\frac{\gamma}{2}}.a\right)$$

The density is the truncated Gaussian $N_{(0,\gamma)}(x)$ for $x \in]\Leftrightarrow a; a[$. It is continuous but not differentiable on the cut locus $\pi \equiv \Leftrightarrow\pi$. The truncation introduces a bias in the relation between the variance and the concentration parameter:

$$\sigma^2 = \frac{1}{\gamma}\left(1 \Leftrightarrow 2.a.k.\exp\left(\Leftrightarrow\frac{\gamma.a^2}{2}\right)\right)$$

As the circle is compact, the variance cannot becomes infinite as in the real case when γ goes to zero: a Taylor expansion give $\sigma^2 = a^2/3 + O(\gamma)$. Thus, the maximal variance on the circle is $a^2/3$ with the density $N_{(0,0)}(x) = 1/(2a)$. As expected, the Normal density of concentration 0 is the uniform density. On the other hand, if γ goes to infinity, the variance goes to zero and the density tends to a Dirac (see figure 1).

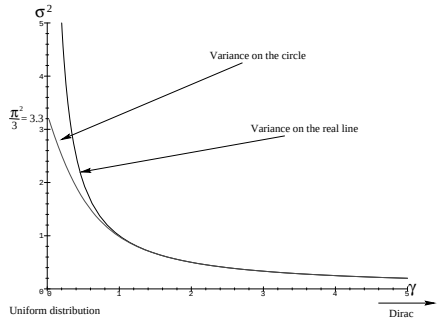

Figure 1: Variance σ^2 with respect to the concentration parameter γ on the circle of radius 1 and the real line.

3.4. Approximated Normal law

If the pdf is sufficiently concentrated (a high concentration matrix Γ or a small covariance matrix Σ), then we can use the Taylor expansion of the metric in the exponential chart at the mean value given in [2, p84]. We easily deduce the Taylor expansion of the measure around the origin (Ric is the Ricci (or scalar) curvature matrix in the considered normal coordinate system):

$$d\mathcal{M}(y) = \sqrt{\det(Q(y))} = 1 \Leftrightarrow \frac{y^{\mathrm{T}}.\mathrm{Ric}\,.y}{6} + O(\|y\|^3)$$

Reporting this Taylor expansion in the integrals and manipulating the formulas leads to the following theorem.

Theorem 3 (Approximate normal density) *Let r be the injection radius at the mean point. The normalization constant and the concentration matrices are approximated by the following expressions for a small variance $\sigma^2 = \mathrm{Tr}(\Sigma)$:*

$$k = \frac{1 + O(\sigma^3) + \varepsilon\left(\frac{\sigma}{r}\right)}{\sqrt{(2\pi)^n \det(\Sigma)}} \qquad \Gamma = \Sigma^{(-1)} \Leftrightarrow \frac{\mathrm{Ric}}{3} + O(\sigma) + \varepsilon\left(\frac{\sigma}{r}\right)$$

Here, $\varepsilon(x)$ is a function that is a $O(x^k)$ for any positive k, i.e. such that $\forall k \in \mathbb{R}^+,\ \lim_{0+} x^{-k}\varepsilon(x) = 0$.

4. MAHALANOBIS DISTANCE AND χ^2 LAW

The problem we are now confronted with is to determine if a measure $\hat{\mathrm{y}}$ was drawn from a random primitive $\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma_{\mathbf{xx}})$. In the vectorial case and assuming a Gaussian distribution, the χ^2 test is well adapted to do that. This test measures the probability of the Mahalanobis distance $\chi^2 = (\hat{x} \Leftrightarrow \bar{x})^{\mathrm{T}}.\Sigma^{(-1)}_{\boldsymbol{xx}}.(\hat{x} \Leftrightarrow \bar{x})$ assuming that $\hat{x}$ is drawn from $\boldsymbol{x}$. If the probability is too small (i.e. χ^2 is too large), the hypothesis is rejected.

Definition 6 (Mahalanobis distance) *We call Mahalanobis distance between a random primitive $\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma_{\mathbf{xx}})$ and a (deterministic) point y on the manifold the value*

$$\mu^2_{\mathbf{x}}(\mathrm{y}) = \overleftrightarrow{\mathrm{xy}}^{T}.\Sigma^{(-1)}_{\mathbf{xx}}.\overleftrightarrow{\mathrm{xy}} \tag{13}$$

In fact, the Mahalanobis distance measures the distance between y and the mean value $\bar{\mathrm{x}}$ according to the "metric" $\Sigma^{(-1)}_{\mathbf{xx}}$.

4.1. Properties

Since $\mu_{\mathbf{x}}^2$ is a function from $\mathcal{M}$ to $\mathbb{R}$, $\mu_{\mathbf{x}}^2(\mathbf{y})$ is a real random variable. The expectation of this random variable is well defined and turns out to be quite simple:

$$\mathbf{E}\left[\ \mu_{\mathbf{x}}^2(\mathbf{y})\ \right] = \mathrm{Tr}\left(\Sigma_{\mathbf{xx}}^{(-1)}.\mathrm{Cov}_{\bar{\mathrm{x}}}(\boldsymbol{y})\right)$$

The expectation of the Mahalanobis distance of a random primitive with itself is even simpler:

Theorem 4 (Mean Mahalanobis distance) *The expected Mahalanobis distance of a random primitive with itself is independent of the distribution and does only depend on the dimension of the manifold:* $\mathbf{E}\left[\ \mu_{\mathbf{x}}^2(\mathbf{x})\ \right] = n$

This identity can be used to verify with a posteriori measurements that the covariance matrix has been correctly estimated. It can be compared with the expectation of the "normalized" squared distance, which is by definition: $\mathbf{E}\left[\ \mathrm{dist}(\mathbf{x},\bar{\mathrm{x}})^2/\sigma_{\mathbf{x}}^2\ \right] = 1$.

4.2. A generalized χ^2 law

Assuming that the random primitive $\mathbf{x} \sim (\bar{\mathrm{x}}, \Sigma_{\mathbf{xx}})$ is normal, we can go one step further and compute the probability that $\chi^2 = \mu_{\mathbf{x}}^2 < \alpha^2$. This generalization of the χ^2 law turns out to be still independent of the mean value and the covariance matrix of the random primitive (at least up to the order $O(\sigma^3)$):

Theorem 5 (Approximate χ^2 law) *With the same hypotheses as for the approximate normal law, the χ^2 density probability function is*

$$p_{\chi^2}(u) = \frac{1 + O(\sigma^3) + \varepsilon\left(\frac{\sigma}{r}\right)}{2.\mathbf{\Gamma}\left(\frac{n}{2}\right)} \left(\frac{u}{2}\right)^{\frac{n}{2}-1} \exp\left(\Leftrightarrow\frac{u}{2}\right) \qquad (14)$$

Thus, up to the third order, the χ^2 law on a Riemannian manifold can be approximated by the standard χ^2 law.

5. DISCUSSION

We presented in this article the bases of a probability and statistical theory on geodesically complete Riemannian manifolds. Basically, we show that almost everything could be done almost as usual by using the the exponential chart at the mean point. This chart can be considered as the development of the manifold along the geodesics on the tangent space at a given point. It is the "most linear" representation of the manifold with respect to this development point.

From a theoretical point of view, we were able to obtain a very simple necessary conditions to characterize the Karcher mean value in the case of a manifold with no cut locus or compact manifolds (e.g. rotations). We believe that the same result could be obtain for other manifolds (such as rigid transformations), but the proof is still to be done. The information minimization approach to generalize the normal distribution to Riemannian manifolds is interesting since we obtain a whole family of densities going from the Dirac to the uniform distribution (or the uniform measure if the manifold in only locally compact). Unfortunately, this distribution is generally not differentiable at the cut locus, and often even not continuous. However, if the relation between the parameters and the moments of the distribution are not as elegant as in the vector case (but can we expect something simpler in the general case of Riemannian manifolds ?), the approximation for small covariances turns out to be rather simple. Thus, this approximate distribution can be handled quite easily for statistical purposes.

From a practical point of view, we were able to implement this framework for several kind of geometric primitives including rotations and rigid transformations by providing the manifolds with an invariant metric [11]. Thanks to the invariance properties, we could use only one exponential chart (at a point called the origin) and "translate" it at any other point of the manifold using the action of a well chosen transformation. The gradient descent algorithm to obtain the mean primitive is very efficient as it usually require between 5 and 10 iterations to converge at numerical precision of the machine.

The application field is wide. We have developed applications in computer vision, medical imaging and molecular biology [11]. Other important applications could take place in robotics, artificial intelligence, etc.

6. REFERENCES

[1] M. do Carmo. *Riemannian Geometry*. Mathematics. Birkhäuser, Boston, Basel, Berlin, 1992.

[2] I. Chavel. *Riemannian geometry - A modern introduction*, volume 108 of *Cambridge tracts in mathematics*. Cambridge university press, 1993.

[3] M. Fréchet. L'intégrale abstraite d'une fonction abstraite d'une variable abstraite et son application à la moyenne d'un élément aléatoire de nature quelconque. *Revue Scientifique*, pages 483–512, 1944.

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

[5] H. Karcher. Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math*, 30:509–541, 1977.

[6] M.G. Kendall and P.A.P. Moran. *Geometrical probability*. Number 10 in Griffin's statistical monographs and courses. Charles Griffin & Co. Ltd., 1963.

[7] W.S. Kendall. Probability, convexity, and harmonic maps with small image i: uniqueness and fine existence. *Proc. London Math. Soc.*, 61(2):371–406, 1990.

[8] W. Klingenberg. *Riemannian Geometry*. Walter de Gruyter, Berlin, New York, 1982.

[9] X. Pennec and N. Ayache. Uniform distribution, distance and expectation problems for geometric features processing. *Journal of Mathematical Imaging and Vision*, 9(1):49–67, July 1998.

[10] X. Pennec and J.P. Thirion. A framework for uncertainty and validation of 3D registration methods based on points and frames. *Int. Journal of Computer Vision*, 25(3):203–229, 1997.

[11] Xavier Pennec. *L'Incertitude dans les Problèmes de Reconnaissance et de Recalage – Applications en Imagerie Médicale et Biologie Moléculaire*. PhD thesis, Ecole Polytechnique, Palaiseau (France), december 1996.

[12] H. Poincaré. *Calcul des probabilités*. 2nd edition, Paris, 1912.