Mean and variance

In order to perform statistics on shapes we must first try to define central statistical concepts on manifolds. In this section we focus on a geodsically complete Riemannian manifold, (M, g), of dimension n, and present ways of defining the mean, variance and covariance of M-valued random variables. Given an underlying probability space, $(\Omega, \mathcal{F}, P)$, a M-valued random variable is a $\mathcal{F}/\mathcal{B}(M)$ measurable map, $X : \Omega \to M$, and we denote by $x = X(\omega)$ a realization of X on M.

In order to perform statistics on M we need to construct a measure on M. This measure is induced by the metric g in the following way. Let $x = (x^1, \dots, x^n)$ be representation of $x \in M$ in local coordinates, and let $\frac{\partial}{\partial x} = (\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n})$ be the corresponding basis of $T_x M$. The metric g is then expressed in this basis by the matrix $G = [g_{ij}(x)]$ where $g_{ij}(x) = \langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \rangle = g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)$. The measure on M is then defined by $dM(x) = \sqrt{|\det G(x)|}dx$. X is said to have density p_X w.r.t. dM if

$$P(X \in \mathcal{A}) = \int_{\mathcal{A}} p_X(y) dM(y),$$

holds for all $\mathcal{A} \in \mathcal{B}(M)$ and if the integral over M is equal to 1. Here p_X is a density in the usual sense. It is a real-valued, positive and integrable function. If π is a chart of the manifold, then $r := \pi(X(\omega))$ defines a random vector with density, ρ_r, w.r.t to the Lebesgue measure given by $\rho_r(y) = p_X(y)\sqrt{|\det G(y)|}$. If $\varphi : M \to \mathbb{R}$ is a $\mathcal{B}(M)/\mathcal{B}(\mathbb{R})$-measurable map, then $\varphi(X)$ deifnes a real-valued random variable for which the expection is

$$\mathbb{E}(\varphi(X)) = \int_M \varphi(y) p_X(y) dM(y).$$

Unfortunately, we cannot define the expectation of M-valued random variables in a smiliar manner, since the real-valued integral does not generalize to an integral with values on M. Instead we generalize the notion of mean value by first defining the variance of a M-valued random variable and then defining the so-called *Frchet means* as minimizers of the variance.