THE HYPERBOLIC MODEL OF THE MEAN × STANDARD DEVIATION "PLANE"

Sueli I. R. Costa
and
Sandra A. Santos

RELATÓRIO TÉCNICO Nº 10/90

Abstract: C. Rao (1945) introduced a Riemannian metric based on the Fisher information matrix to measure the divergence between probability distributions. We consider here the normal distributions, giving geometrical approach to the univariated case and extending the analysis to the multivariated normal distributions through a product metric. In such an extension, the hyperbolic mean × standard deviation "plane" which is the model for the univariated case, is replaced by a constant negative mean curvature space.

Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Ciência da Computação
IMECC - UNICAMP
Caixa Postal 6065
13.081 - Campinas - SP
BRASIL

Fevereiro - 1990

THE HYPERBOLIC MODEL OF THE MEAN × STANDARD DEVIATION "PLANE"

Sueli I. R. Costa
Sandra A. Santos

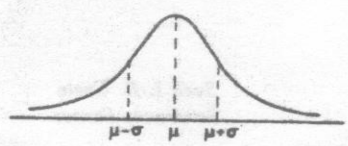
Introduction – Statisticians often assume a statistical model as a subset of the set of all the possible probability distributions. It was C. Rao who first pointed out the importance of the differential–geometrical approach in analysing statistical models. In an early paper (1945), he introduced a Riemannian metric over the space of a parametric family of probability distributions based on the Fisher information matrix. Since then, several authors have considered metrics arising out of a variety of divergence measures between probability distributions (references may be found in [1]). We consider here the univariated normal distribution with the Fisher metric by means of a geometrical approach. References dealing with this subject from a statistical view are [2] and [7]. Through a natural extension of our analysis, we could generalize this study to the multivariated normal distributions, using a product metric. In such an extension, the hyperbolic mean × standard deviation "plane" which is the model for the univariated case, is replaced by a constant negative mean curvature space.

Normally distributed random variables are associated exactly or in an approximated way with several physical and biological phenomena. For instance, Maxwell's law asserts that under appropriate conditions, the components of the velocity of a molecule of gas will be normally distributed with mean zero and standard deviation depending on the gas peculiarities. Other examples are the quantity of DNA in certain cells, measurement errors, variability of outputs from industrial production lines and biological variabilities such as height and weight.

A mean μ and a standard deviation σ define univocally a normal probability density function:

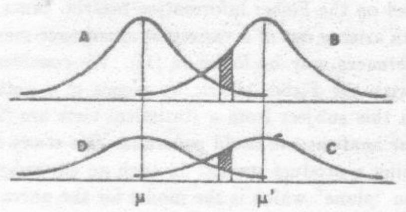
$$f(x) = p(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

Graphically, we have a Gaussian curve (fig. 1):



- fig. 1 -

A measure for the distance between two of these curves must reflect how much the probability (i.e. the integral of the density function) varies from one distribution to the other.

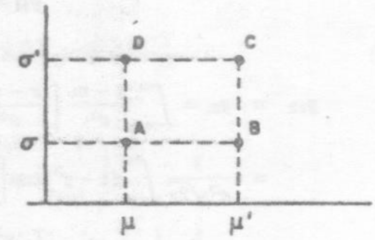


- fig. 2 -

Comparing two normal distributions with the same standard deviation σ and different means μ and μ', we can visualize, in an interval, the difference between them through the hatched areas. By fixing the means and increasing the standard deviation, we can see that the distance between the curves must be smaller because the curves become flatter. (fig. 2)

In the half-plane mean $\times$ standard deviation defined by the pairs (μ, σ), $\sigma > 0$, we identify each point with its associated normal distribution. In this way, we have a notion of distance in the half-plane that gives a measure of the "dissimilarity" between two distributions.

In Figure 3 we translate the situation illustrated by Figure 2 and we can observe that such a metric in the half-plane cannot be Euclidean, as it must vary with the inverse of the standard deviation σ . The points C and D are "nearer" than A and B, hence the distributions associated with A an B are more dissimilar: $d_*(C,D) < d_*(A,B)$.



- fig. 3 -

The mean × standard deviation "plane" - One of the metrics which is often used in statistical models is the one induced by the Fisher[1] information matrix. Its coefficients are calculated as the expectation of a product involving the partial derivatives of the logarithm of the probability density function:

$$\begin{aligned} g_{ij} &= E\left[-\frac{\partial^2 \log p(x,\theta)}{\partial\theta_i\partial\theta_j}\right] \\ &= E\left[\frac{\partial \log p(x,\theta)}{\partial\theta_i}\ \frac{\partial \log p(x,\theta)}{\partial\theta_j}\right] \\ &= \int_{-\infty}^{\infty} \frac{\partial \log p(x,\theta)}{\partial\theta_i}\ \frac{\partial \log p(x,\theta)}{\partial\theta_j}\ p(x,\theta)dx. \end{aligned}$$

In the univariated normally distributed case we have $\theta = (\theta_1, \theta_2) = (\mu, \sigma)$ and we obtain the 2×2 Fisher information matrix

$$\begin{aligned} g_{11} &= \int_{-\infty}^{\infty} \frac{(x-m)^2}{\sigma^4}\ \frac{1}{\sqrt{2\pi}\sigma}\ \exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]dx \\ &= \frac{1}{\sigma^2\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 \exp\left[-\frac{y^2}{2}\right]dy \\ &= \frac{1}{\sigma^2\sqrt{2\pi}}\left[y \exp\left[-\frac{y^2}{2}\right]\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \exp\left[-\frac{y^2}{2}\right]dy\right] \end{aligned}$$

since

$$\lim_{y\to\pm\infty} y \exp\left[-\frac{y^2}{2}\right] = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} \exp\left[-\frac{y^2}{2}\right]dy = \sqrt{2\pi},$$

[1]Ronald A. Fisher (1890 – 1962) made an enormous contribution to the development of statistical techniques and their application in biology. He introduced the theory of statistical inference and his concept of information appeared in one of his articles in 1921.

it follows that

$$g_{11} = \frac{1}{\sigma^2} \tag{I}$$

$$\begin{aligned} g_{12} &= g_{21} = \int_{-\infty}^{\infty} \frac{x-m}{\sigma^2}\left[\frac{(x-m)^2}{\sigma^3} - \frac{1}{\sigma}\right]\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]dx \\ &= \frac{1}{\sigma^2\sqrt{2\pi}}\int_{-\infty}^{\infty}(1-y^2)\exp\left[-\frac{y^2}{2}\right](-y\,dy) \\ &= \frac{1}{\sigma^2\sqrt{2\pi}}\left[(1-y^2)\exp\left[-\frac{y^2}{2}\right]\right]_{-\infty}^{\infty} + 2\int_{-\infty}^{\infty} y\exp\left[-\frac{y^2}{2}\right]dy \Bigg] \end{aligned}$$

and since

$$\lim_{y\to\pm\infty}(1-y^2)\exp\left[-\frac{y^2}{2}\right] = 0$$

it follows that

$$g_{12} = g_{21} = 0 \tag{II}$$

$$\begin{aligned} g_{22} &= \int_{-\infty}^{\infty}\left[\frac{(x-m)^2}{\sigma^3} - \frac{1}{\sigma}\right]^2 \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-m)^2}{2\sigma^2}\right]dx \\ &= \frac{1}{\sigma^2\sqrt{2\pi}}\int_{-\infty}^{\infty}(y^4 - 2y^2 + 1)\exp\left[-\frac{y^2}{2}\right]dy. \end{aligned}$$

But

$$\int_{-\infty}^{\infty} y^2 \exp\left[-\frac{y^2}{2}\right]dy = \sqrt{2\pi}.$$

Let

$$I = \int_{-\infty}^{\infty} y^3 \exp\left[-\frac{y^2}{2}\right](y\,dy) = -y^3\exp\left[-\frac{y^2}{2}\right]\Bigg|_{-\infty}^{\infty} + 3\sqrt{2\pi}$$

since

$$\lim_{y\to\pm\infty} -y^3\exp\left[-\frac{y^2}{2}\right] = 0, \quad I = 3\sqrt{2\pi}.$$

Therefore,

$$g_{22} = \frac{1}{\sigma^2\sqrt{2\pi}}[3\sqrt{2\pi} - 2\sqrt{2\pi} + \sqrt{2\pi}] = \frac{2}{\sigma^2} \tag{III}$$

$$(g_{ij}) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{pmatrix}.$$

That is, the expression for the metric is:

$$ds_*^2 = \frac{d\mu^2 + 2d\sigma^2}{\sigma^2}.$$

As we are going to see, the half plane mean × standard deviation ($I\!H_*^2$) provided with this metric, is a model for hyperbolic geometry which can be related to the Poincaré half-plane ($I\!H^2$), the metric of which is given by the matrix

$$(\bar{g}_{ij}) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 1/\sigma^2 \end{pmatrix}.$$

Associated to a metric matrix $G = (g_{ij})$, we have an inner product for vectors in the plane:

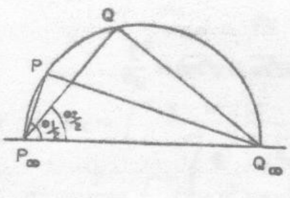
$$\langle u, v\rangle_G = u^t(g_{ij})v; \quad \|u\|_G = \langle u, u\rangle_G^{1/2}.$$

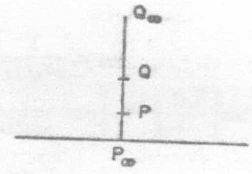
The distance between two points P, Q in $I\!H_*^2$ is given by the number which is the minimum of the lengths of all the piecewise smooth paths γ_P^Q joining these two points (see [5], §3). The length of a path $\gamma(t)$ is calculated by using the inner product $\langle\ ,\ \rangle_G$: length of $\gamma = \int_\gamma ds = \int_\gamma \|\dot{\gamma}\|_G dt \quad d_G(P,Q) = \min_{\gamma_P^Q}$ {length of γ_P^Q}

The *Fisher distance* is the one associated to the information matrix. In order to express such a notion of distance and to characterize the geometry in the mean × standard deviation plane, we analyse its analogies with the Poincaré half-plane.

It can be shown (see [3]–chapter 7) that in $I\!H^2$, the curves which minimize length – geodesics – are vertical half-lines and half-circles centered at $\sigma = 0$. Furthermore, the distance between two points is given by the logarithm of the cross-ratio between these two points and the points at the infinite: $d_H(PQ) = \log(P\,Q\,P_\infty\,Q_\infty)$. It can be expressed by the following formulae, considering P and Q as vertical lined or not (fig. 4).

(fig. 4).





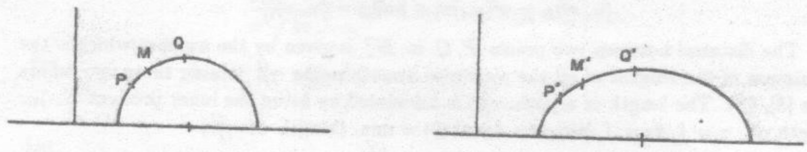
– *fig. 4* –

$$d_H(PQ) = \log\left(\frac{\frac{PP_\infty}{PQ_\infty}}{\frac{QP_\infty}{QB_\infty}}\right) = \log\left(\frac{tg\theta_{1/2}}{tg\theta_{2/2}}\right) \qquad d_H(PQ) = \log\left(\frac{\sigma_Q}{\sigma_P}\right).$$

We remark that the Poincaré and the Fisher distances are related by

$$d_*\Big((\mu_1,\sigma_1),(\mu_2,\sigma_2)\Big) = \sqrt{2}d_H\left(\left(\frac{\mu_1}{\sqrt{2}},\sigma_1\right),\left(\frac{\mu_2}{\sqrt{2}},\sigma_2\right)\right).$$

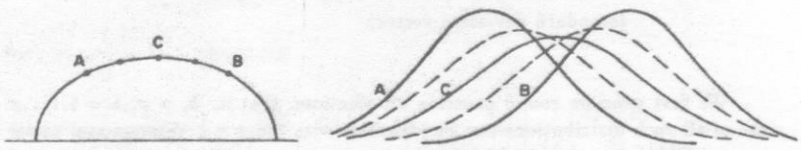
Therefore, the geodesics of $I\!H^2_*$ are the image of the geodesics of $I\!H^2$ by means of the transformation $(m,\sigma) \longrightarrow \left(\frac{m}{\sqrt{2}},\sigma\right)$. In fact, they are the vertical half-straight lines and half-ellipses centered in $\sigma = 0$ and having excentricity $1/\sqrt{2}$.



– *fig. 5* –

Such a metric can be used to establish the concept of "average distribution" between two given distributions P and Q. This is determined by the point M on the geodesic segment joining P and Q and which is equidistant to these points. (fig. 5)

We observe that by adopting the Fisher distance, the shortest path between two normal distributions A and B with the same standard deviation σ is a one-parameter family which obviously does not preserve σ (fig. 6).



– fig. 6 – Shortest path between the normal distributions A and B with the same standard deviation $\sigma = 0.5$ and means $\mu_A = 0$ and $\mu_B = 1$. C is the mean distribution between A and B.

This might be awkward for some purposes. In [7] and [2] this situation is considered separately, taking for the distance between these two points, the measure of the horizontal segment $(\frac{|\mu_A - \mu_B|}{\sigma})$, which is not a geodesic arc.

The mapping $(\mu, \sigma) \mapsto (\frac{1}{\sqrt{2}}\mu, \sigma)$, which is a homothety between $\mathbb{H}^2_*$ and $\mathbb{H}^2$, is also an isometry from $\mathbb{H}^2_*$ to the half plane $\mathbb{H}^2_2$, where the metric is defined by $ds^2 = \frac{2}{\sigma^2}(d\mu^2 + d\sigma^2)$. This allows us to conclude that the curvature of $\mathbb{H}^2_*$ is $-\frac{1}{2}$. (See [4] for further references).

Multivariated Normal Distributions – A normal probability density function of p variables is given by

$$f(x, \mu, \bar{\sigma}) \quad = \quad (2\pi)^{-p/2}(\det\Sigma)^{-1/2}\exp[-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)]$$

$$\text{where} \quad x = (x_1, \ldots, x_p)^t, \quad x_i \in \mathbb{R} \quad i = 1, \ldots, p;$$

$$\mu = (\mu_1, \ldots \mu_p)^t, \quad \mu_i \in \mathbb{R}, \quad i = 1, \ldots, p \qquad \text{(mean vector)}$$

and $\Sigma = (\sigma_{ij})$ is a symmetric positive definite $p \times p$ matrix.

We analyse here the independently distributed case, where Σ is diagonal,

$$\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2); \quad \tilde{\sigma} = (\sigma_1, \ldots, \sigma_p)^t \quad \sigma_i > 0, i = 1, \ldots, p$$

(standard deviation vector).

We first consider round gaussian distributions, that is, $\sigma_i = \sigma$, $i = 1, \ldots, p$. The set of all such distributions can be identified with the $p + 1$ dimensional upper half-space ($\mathbb{H}_*^{p+1}$), parametrized by $\theta = (\mu_1, \ldots, \mu_p, \sigma)$, $\mu_i \in \mathbb{R}$, $i = 1, \ldots, p$, $\sigma > 0$. The result obtained in the univariated case can be extended very naturally to this situation: the Fisher information metric will provide a hyperbolic geometry to this half-space with constant (sectional) curvature equal to $-\frac{1}{2}$. The Fisher matrix for p-variated distributions is given by

$$g_{ij} = \int_{\mathbb{R}^p} \frac{\partial \log f}{\partial \theta_i} f(\mathbf{x}, \theta) \cdot d\mathbf{x}. \qquad (*)$$

Thus, for $\theta = (\mu_1, \mu_2, \ldots, \mu_p, \sigma)$ we calculate

$$(g_{ij}) = \begin{bmatrix} 1/\sigma^2 & & & & 0 \\ & 1/\sigma^2 & & & \\ & & \ddots & & \\ & & & 1/\sigma^2 & \\ 0 & & & & 2/\sigma^2 \end{bmatrix}.$$

As in the bidimensional situation, we have a homothety between the half space $\mathbb{H}_*^{p+1}$ and the Poincaré half space $\mathbb{H}^{p+1}$. Through this homothety, we can conclude that $\mathbb{H}_*^{p+1}$ *has constant sectional curvature equal to* $-\frac{1}{2}$ and its geodesics are "vertical" half-lines and half ellipses centered at $\sigma = 0$ with excentricity $1/\sqrt{2}$.

Considering the general standard deviation vector, we have elliptic Gaussian distributions which can be identified with the $2p$-dimensional polyhedron $\mathcal{P}$ parametrized by

$$\theta = (\mu_1, \sigma_1, \mu_2, \sigma_2, \ldots, \mu_p, \sigma_p), \quad \sigma_i > 0, \ \forall i.$$

We make the calculations of the Fisher information matrix coefficients (*) explicit here for $p = 2$, without loss of generality. For $\theta = (\mu_1, \sigma_1, \mu_2, \sigma_2)$, we have:

$$g_{11} = \int_{-\infty}^{\infty} \frac{(x_1-\mu_2)^2}{\sigma_1^4} \frac{1}{\sqrt{2\pi}\sigma_1} \exp -\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 dx_1$$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_2} \exp -\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^d dx_2$$

$$g_{11} = \frac{1}{\sigma_1^2} \cdot 1 \qquad \text{(from (I))}$$

$$g_{12} = \int_{-\infty}^{\infty} \frac{x_1-\mu_1}{\sigma_1^2} \frac{1}{\sigma_1}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 1\right] \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right] dx_1$$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_2} \exp -\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 dx_2$$

$$g_{12} = 0 \qquad \text{(from (II))}$$

$$g_{13} = \left[\int_{-\infty}^{\infty} \frac{x_1-\mu_1}{\sigma_1^2} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 dx_1\right]\right]^2$$

$g_{13} = 0$ **(The integrand is skew-symmetric with respect to the line $x_1 = \mu_1$)**

$$g_{14} = \int_{\mathbb{R}^2}\int \frac{x_1-\mu_1}{\sigma_1^2} \frac{1}{\sigma_2}\left[\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 1\right] \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\} dx_1 dx_2$$

$$g_{14} = 0 \qquad \text{(Skew-symmetry again)}$$

$$g_{22} = \int_{-\infty}^{\infty} \frac{1}{\sigma_1^2}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 1\right)^2 \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right] dx_1$$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right] dx_2$$

$$g_{22} = \frac{2}{\sigma_1^2} \qquad \text{(from (III))}$$

$$g_{24} = \left(\int_{-\infty}^{\infty} \frac{1}{\sigma_1}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 1\right]\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left[-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right]dx_1\right)^2$$

$g_{24} = 0$ (The integrand is the difference between two probability density functions).

Using the analogies and symmetry, we can fill the Fisher information matrix associated to the p-variated independent distributed normal:

$$(g_{ij}) = \begin{bmatrix} 1/\sigma_1^2 & 0 & & & \\ 0 & 2/\sigma_1^2 & & 0 & \\ & & \ddots & & \\ 0 & & & 1/\sigma_p^2 & 0 \\ & & & 0 & 2/\sigma_p^2 \end{bmatrix}.$$

We remark that associated with this matrix we have a product metric. That is, in the polyhedron $\mathcal{P} = I\!H^2_* \times I\!H^2_* \times \cdots \times I\!H^2_*$, the inner product is given by the sum of the p inner products of each projection in the mean $\times$ standard deviation plane. This means we can describe the Fisher metric as $g = \sum_{i=1}^{p} \pi_i^*(g_{I\!H^2_*})$ (tensorial notation, see [6], chapter 3), where π_i is the projection on the ith couple $\pi_i(\mu_1, \sigma_1, \ldots, \mu_p, \sigma_p) = (\mu_i, \sigma_i)$, and $g_{I\!H^2_*}$ is the Fisher tensor metric on the $\mu_i \times \sigma_1$ half plane.

Using results on Riemannian product manifolds (see [6], chapter 7), we remark that the sectional curvature of a $\mu_i \times \sigma_i$ plane is $-1/2$ and the curvature associated to any other canonical 2-space is zero. Therefore we can state that $\mathcal{P}$, the Fisher model for independent p-variated normal distributions, has *constant scalar (mean) curvature equal to* $-\frac{1}{2(2p-1)}$. Moreover, a curve

$$\alpha(t) = (\alpha_1(t), \ldots, \alpha_p(t)) \quad \text{in} \quad \mathcal{P} = I\!H^2_* \times I\!H^2_* \times \cdots \times I\!H^2_*$$

is a geodesic if, and only if, $\alpha_i(t)$ is a geodesic in $I\!H^2_*$. This means that the shortest path between two p-variated normal distributions is a curve (one-parameter family), the projections of which are half ellipses and vertical half lines.

References

[1] Amari S., - *Differential - Geometrical Methods in Statistics*. Lecture Notes in Statistics 28 - Spring-Verlag - 290 p.p., Heidelberg, 1986.

[2] Atkinson, C. and Mitchell, A.F.S. - "Rao's Distance Measure", *Samkhyā*, 43, 345 - 365 (1981).

[3] Beardon, A. F., *The Geometry of Discrete Groups*, Spring-Verlag, New York, 1982.

[4] Klingenberg, W., *A Course in Differential Geometry*, Spring-Verlag, New York, 1978.

[5] Millman, R. S. and Stehney, A. K., "The Geometry of Connections", The American Mathematical Monthly, 80 (1973), 475-500.

[6] O'Neill, N., *Semi-Riemannian Geometry*, Academic Press, New York, 1983.

[7] Rao, C.R. - "Information and Accuracy Attainable in the Estimation of Statistical Parameters", *Bull. Calcutta Math. Soc.* 37, 81 - 91 (1945).

Universidade Estadual de Campinas
Instituto de Matemática
13081 - Campinas - SP - Brazil

RELATÓRIOS TÉCNICOS — 1990