ON INEQUALITIES OF INFORMATION GEOMETRY

V.A. Voloshko
Belarusian State University
Minsk, BELARUS
e-mail: valeravoloshko@yandex.ru

Abstract

Some special cases of the triangle inequality for geometry of probability distributions are found. Examples of following numerical inequalities are given.

1 Introduction

The question about how to define a natural differentiable structure on manifolds of probability distributions appeared in the middle of last century. In particular, it was discussed by Kolmogorov in his unpublished lectures, being read in autumn 1955 in the institute of Henri Poincare. Later, when the information inequality was proved and fundamental role of Fisher information matrix was realized, it became clear that the Fisher matrix defines by itself a metric tensor, which induces a riemannian geometry on manifolds of probability distributions. The main results for this geometry were obtained by Chentsov [3] and Amari [1]. The last one coined the term *information geometry* for it.

Denote by $\mathcal{P}$ the set of probability distributions of the measurable space $(\Omega, \mathcal{F})$. Bhattacharya distance [1] between measures $P, Q \in \mathcal{P}$:

$$\rho(P,Q) = 2 \arccos \int_\Omega \sqrt{P(d\omega)Q(d\omega)}, \tag{1}$$

defines the spherical geometry on $\mathcal{P}$ with scalar product of increments dP_1, dP_2 in tangent space [1] $T(P)$ of dot $P \in \mathcal{P}$:

$$(dP_1, dP_2)_P = \int_\Omega \frac{dP_1(d\omega)dP_2(d\omega)}{P(d\omega)}.$$

A parametric family of measures $P(\theta), \theta \in \Theta \subset \mathbb{R}^d$, forms a finite-dimensional manifold $P(\Theta) \subset \mathcal{P}$. Subspace $T(\theta)$ of tangent space $T(P(\theta))$, formed by the increments of $P(\theta)$ along the manifold $P(\Theta)$, can be represented as a result of linear mapping from $\mathbb{R}^d$:

$$\mathbb{R}^d \ni d\theta \to \left(\frac{dP(\theta)}{d\theta}, d\theta\right) = \sum_{i=1}^{d} \frac{\partial P(\theta)}{\partial \theta_i} d\theta_i \in T(\theta),$$

which induces scalar product for the increments of θ:

$$(d\theta_1, d\theta_2)_\theta = (dP_1(\theta), dP_2(\theta))_{P(\theta)} = d\theta_1' I(\theta) d\theta_2, \tag{2}$$

where

$$I(\theta) = \left(I(\theta)_{i,j} = \left(\frac{\partial P(\theta)}{\partial \theta_i}, \frac{\partial P(\theta)}{\partial \theta_j} \right)_{P(\theta)} = \int_\Omega \frac{\partial P(\theta)}{\partial \theta_i} \cdot \frac{\partial P(\theta)}{\partial \theta_j} \cdot \frac{1}{P(\theta)} \bigg|_{d\omega} \right)_{i,j=1}^{d} -$$

is the Fisher information matrix. Then for $\Theta = [\theta_1, \theta_2] \subset \mathbb{R}$ the length of $P(\Theta)$ is.

$$|P(\Theta)| = \int_{\theta_1}^{\theta_2} \sqrt{I(\theta)} d\theta \geq 2 \arccos \int_\Omega \sqrt{P(\theta_1) P(\theta_2)} \Big|_{d\omega} = \rho(P(\theta_1), P(\theta_2)). \quad (3)$$

In this paper some special cases of the triangle inequality (3) was found.

2 Special cases of the triangle inequality

Consider Borel space $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and some probability measure P_0 on it with differentiable probability density function $p(x)$. Introduce the notation:

$$J_n = \int_{\mathbb{R}} \frac{(p'(x))^2}{p(x)} x^n dx, \ n = 0, 1, 2, \ldots \quad (4)$$

Theorem 1. *Under $\mu > 0$ the following inequality holds:*

$$2 \arccos \int_{\mathbb{R}} \sqrt{p(x) p(x-\mu)} dx \leq \mu \sqrt{J_0}. \quad (5)$$

Proof. Let the family $P(\mu), \mu \in \mathbb{R}$, be generated by the shifts of measure P_0. Then $P(\mu)$ has the density function $p(x-\mu)$ and:

$$I(\mu) = \int_{\mathbb{R}} \frac{(p'(x-\mu))^2}{p(x-\mu)} dx = J_0, \ \int_0^\mu \sqrt{I(x)} dx = \mu \sqrt{J_0},$$

whence by (3) under $\theta_1 = 0, \ \theta_2 = \mu > 0$ we get (5). □

Some examples are given below:

$$1) \ p(x) = \frac{1}{\sqrt{2\pi}} e^{-0.5x^2}, J_n = \int_{\mathbb{R}} x^{n+2} p(x) dx, J_0 = 1, \int_{\mathbb{R}} \sqrt{p(x) p(x-\mu)} dx =$$
$$= \int_{\mathbb{R}} \frac{dx}{\sqrt{2\pi}} e^{-0.5((x-\mu/2)^2 + (\mu/2)^2)} = e^{-\mu^2/8} \overset{(5)}{\Longrightarrow} \frac{\mu}{2} = u \geq \arccos\left(e^{-0.5u^2}\right), u \geq 0.$$

$$2) \ p(x) = \frac{1}{2} e^{-|x|}, J_n = \int_{\mathbb{R}} x^n p(x) dx, J_0 = 1, \int_{\mathbb{R}} \sqrt{p(x) p(x-\mu)} dx =$$
$$= \int_{\mathbb{R}} \frac{dx}{2} e^{-0.5(|x| + |x-\mu|)} = \left(\frac{\mu}{2} + 1\right) e^{-0.5\mu} \Rightarrow \frac{\mu}{2} = u \geq \arccos\left(\frac{u+1}{e^u}\right), u \geq 0.$$

Theorem 2. *Under $xp(x) \to 0, x \to \pm\infty$, and $\sigma > 0, \mu \in \mathbb{R}$, the following inequality holds:*

$$2 \arccos \int_{\mathbb{R}} \sqrt{p\left(\frac{x-\mu}{\sigma}\right)\frac{p(x)}{\sigma}} dx \leq \gamma\sqrt{J_0}\,\mathrm{arch}\left(1 + \frac{\mu^2 + 2\mu s\gamma_1 + s^2\gamma_2}{2\sigma\gamma^2}\right), \quad (6)$$

where $\mathrm{arch}(z) = \ln(z + \sqrt{z^2-1})$ *is the inverse hyperbolic cosine and:*

$$s = \sigma - 1, \ \gamma_1 = J_1/J_0, \ \gamma_2 = (J_2 - 1)/J_0, \ \gamma = \sqrt{\gamma_2 - \gamma_1^2}.$$

Proof. Let the family $P(\mu, \sigma), (\mu, \sigma) = (\theta_0, \theta_1) \in \Pi = \mathbb{R} \times \mathbb{R}_+$, be generated by the shifts and stretchings of measure P_0. Denote: $\widetilde{x} = (x - \mu)/\sigma$, then: $\partial\widetilde{x}/\partial\mu = -1/\sigma$, $\partial\widetilde{x}/\partial\sigma = -\widetilde{x}/\sigma \Rightarrow \partial\widetilde{x}/\partial\theta_i = -\widetilde{x}^i/\sigma|_{i=0,1}$. Measure $P(\mu, \sigma)$ has the density function $\sigma^{-1}p(\widetilde{x})$ and for $i, j \in \{0, 1\}$ $(l(x) = \ln p(x))$:

$$I_{i,j} = \int_{\mathbb{R}} \prod_{\gamma=i,j} \frac{\partial(l(\widetilde{x}) - \ln\sigma)}{\partial\theta_\gamma} p(\widetilde{x}) d\widetilde{x} = \int_{\mathbb{R}} \prod_{\gamma=i,j} \left(-l'(\widetilde{x})\frac{\widetilde{x}^\gamma}{\sigma} - \frac{\gamma}{\sigma}\right) p(\widetilde{x}) d\widetilde{x} =$$

$$= [t = \widetilde{x}] = \frac{1}{\sigma^2}\int_{\mathbb{R}} \left\{\frac{(p'(t))^2}{p(t)} t^{i+j} + (it^j + jt^i)p'(t) + ijp(t)\right\} dt =$$

$$= \left[\int_{\mathbb{R}} t^\gamma p'(t) dt = t^\gamma p(t)|_{-\infty}^{+\infty} - \int_{\mathbb{R}} \gamma t^{\gamma-1} p(t) dt = -\gamma\right]_{\gamma=0,1} = \frac{J_{i+j} - ij}{\sigma^2}.$$

Therefore on twodimensional surface $P(\Pi) \subset \mathcal{P}$, formed by measures $P(\mu, \sigma)$, metric (1) induces riemannian geometry with metric tensor (2) $I(\mu, \sigma) = I(0,1)/\sigma^2$. As it known [2], the tensor of such form in the halfplane generates hyperbolic geometry. Using transformation: $\Pi \ni (\bar{\mu}, \bar{\sigma}) \to (\mu, \sigma) = (\bar{\mu} + a\bar{\sigma}, b\bar{\sigma}) \in \Pi$, $(a, b) \in \Pi$, let us bring the tensor to the canonical form $(I_{0,0} = I_{1,1}, I_{0,1} = 0)$:

$$J_0 \cdot d\mu^2 + 2J_1 \cdot d\mu d\sigma + (J_2 - 1) \cdot d\sigma^2 = J_0 \cdot d\bar{\mu}^2 + 2(J_0 a + J_1 b) \cdot d\bar{\mu} d\bar{\sigma} + $$
$$+ (J_0 a^2 + 2J_1 ab + (J_2 - 1)b^2) \cdot d\bar{\sigma}^2.$$

By equating factor $d\bar{\mu}d\bar{\sigma}$ to zero, and factors $d\bar{\mu}^2$ and $d\bar{\sigma}^2$ together, we get: $a = -\gamma_1/\gamma$, $b = 1/\gamma$, whence $(\bar{\mu}, \bar{\sigma}) = (\mu + \sigma\gamma_1, \sigma\gamma)$ and:

$$d\rho^2 = \rho^2(P(\mu, \sigma), P(\mu + d\mu, \sigma + d\sigma)) = \frac{J_0}{\sigma^2}(d\bar{\mu}^2 + d\bar{\sigma}^2) = \frac{\gamma^2 J_0}{\bar{\sigma}^2}(d\bar{\mu}^2 + d\bar{\sigma}^2).$$

This element of length $d\rho$ in coordinates $(\bar{\mu}, \bar{\sigma})$ corresponds to the model of Lobachevsky geometry, named *the Poincare model in halfplane* [2]. Therefore distance $\widetilde{\rho}(P_1, P_2)$ between measures $P_i = P(\mu_i, \sigma_i)|_{i=1,2}$ along the surface $P(\Pi)$, i.e. the length of the shortest curve, lying in $P(\Pi)$ and connecting P_1, P_2, satisfies the following relation [2]:

$$\mathrm{ch}\left(\frac{\widetilde{\rho}}{\gamma\sqrt{J_0}}\right) = 1 + \frac{(\bar{\mu}_1 - \bar{\mu}_2)^2 + (\bar{\sigma}_1 - \bar{\sigma}_2)^2}{2\bar{\sigma}_1\bar{\sigma}_2},$$

where $\mathrm{ch}(z) = (e^z + e^{-z})/2$ is the hyperbolic cosine. In particular, under $P_1 = P(\mu, \sigma)$, $P_2 = P(0, 1)$ we get:

$$\mathrm{ch}\left(\frac{\widetilde{\rho}}{\gamma\sqrt{J_0}}\right) = 1 + \frac{(\mu + s\gamma_1)^2 + (s\gamma)^2}{2\sigma\gamma^2} = 1 + \frac{\mu^2 + 2\mu s\gamma_1 + s^2\gamma_2}{2\sigma\gamma^2}.$$

Then inequality $\rho(P_1, P_2) \leq \widetilde{\rho}(P_1, P_2)$ takes the form (6). □

Note, that under $\gamma_1 = \mu = 0$ the right-hand side of (6):

$$\gamma\sqrt{J_0}\cdot \mathrm{arch}(1 + (\sigma - 1)^2/2\sigma) = \gamma\sqrt{J_0}\cdot \mathrm{arch}(\mathrm{ch}(\ln\sigma)) = \gamma\sqrt{J_0}\cdot |\ln\sigma|.$$

Under conditions $\sigma = 1$ and $\mu > 0$ the left-hand parts of inequalities (5) and (6) are match, while parts of (6) are closer to each other, than parts of (5):

$$2\arccos\int_{\mathbb{R}}\sqrt{p(x-\mu)p(x)}dx \leq \sqrt{J_0}\gamma\,\mathrm{arch}\left(1 + \frac{\mu^2}{2\gamma^2}\right) \leq \sqrt{J_0}\mu. \tag{7}$$

Examples of formula (6) for the density functions $p(x)$, examined after Theorem 1, are given below ($u = \mu/2 \geq 0$):

1) $J_0 = 1,\ J_1 = 0,\ J_2 = 3,\ \gamma_1 = 0,\ \gamma_2 = 2,\ \gamma = \sqrt{2},\ \cos\dfrac{\rho(P_0, P(\mu,\sigma))}{2} =$

$$= \int_{\mathbb{R}}\frac{dx}{\sqrt{2\pi\sigma}}\exp\left\{-\frac{(x\sigma)^2 + (x-\mu)^2}{4\sigma^2}\right\} = \sqrt{\frac{2\sigma}{\sigma^2+1}}\exp\left\{-\frac{\mu^2}{4(\sigma^2+1)}\right\} \Rightarrow$$

$$\Rightarrow \arccos\left(\sqrt{\frac{2\sigma}{\sigma^2+1}}\exp\left\{-\frac{\mu^2}{4(\sigma^2+1)}\right\}\right) \leq \frac{1}{\sqrt{2}}\mathrm{arch}\left(1 + \frac{\mu^2 + 2(\sigma-1)^2}{4\sigma}\right),$$

$$\arccos\left(\sqrt{\frac{2\sigma}{\sigma^2+1}}\right) \leq \frac{|\ln\sigma|}{\sqrt{2}}\bigg|_{\mu=0},\quad \arccos\left(e^{-0.5u^2}\right) \leq \frac{\mathrm{arch}(1+u^2)}{\sqrt{2}}\bigg|_{\sigma=1}.$$

2) $J_0 = 1,\ J_1 = 0,\ J_2 = 2,\ \gamma_1 = 0,\ \gamma_2 = 1,\ \gamma = 1,\ \cos\dfrac{\rho(P_0, P(\mu,\sigma))}{2} =$

$$= \int_{\mathbb{R}}\frac{dx}{2\sqrt{\sigma}}\exp\left\{-\frac{1}{2}\left(|x| + \left|\frac{x-\mu}{\sigma}\right|\right)\right\} = \frac{2\sqrt{\sigma}}{\sigma^2-1}\left(\sigma e^{-\mu/2\sigma} - e^{-\mu/2}\right) \Rightarrow$$

$$\Rightarrow \arccos\left(\frac{2\sqrt{\sigma}}{\sigma^2-1}\left(\sigma e^{-\mu/2\sigma} - e^{-\mu/2}\right)\right) \leq \frac{1}{2}\mathrm{arch}\left(1 + \frac{\mu^2 + (\sigma-1)^2}{2\sigma}\right),$$

$$\arccos\left(\frac{2\sqrt{\sigma}}{\sigma+1}\right) \leq \frac{|\ln\sigma|}{2}\bigg|_{\mu=0},\quad \arccos\left(\frac{u+1}{e^u}\right) \leq \frac{\mathrm{arch}(1+2u^2)}{2}\bigg|_{\sigma=1}.$$

References

[1] Amari S. (2000). *Methods of Information Geometry*. Oxford University Press, New York.

[2] Anderson J.W. (2005). *Hyperbolic Geometry*. Springer, New York.

[3] Morozova E.A., Chentsov N.N. (1991). Natural geometry of families of probability laws. *Probability theory*. Vol. **8**, pp. 133-265.