Fractionalization and Polarization

Frazionalizzazione e polarizzazione

Giovanni Pistone and Maria Piera Rogantin

Abstract Fractionalization is a measure of concentration of a qualitative distribution, while Polarization is a measure of its bimodality. We study the steepest ascent lines of these indexes on the probability simplex using the notion of natural gradient of Amari. An index of variation is suggested.
Abstract *La frazionalizzazione di una distribuzione qualitativa è una misura di concentrazione; la polarizzazione è una misura di bimodalità. Studiamo le curve di massima pendenza di questi indici sul simplesso delle probabilità tramite la nozione di gradiente naturale secondo Amari. Si suggerisce un indice di variazione.*

Key words: Fractionalization Measure; Gini Index; Information Geometry; Polarization Measure.

1 Introduction

Given a discrete distribution π on $n+1$ classes, say $x = 0,1,\ldots,n$, we consider two indexes, the *fractionalization measure* $\mathrm{FRAC}(\pi) = 1 - \sum_x \pi_x^2$ and the *polarization measure* $\mathrm{POL}(\pi) = \sum_x \pi_x^2(1-\pi_x)$. If d is the discrete distance $d(x,y) = (x \neq y)$, $x,y = 0,1,\ldots,n$, and X, Y are independent samples of π, then $\mathrm{FRAC}(\pi) = \mathbb{E}[d(X,Y)]$, hence the fractionalization measure is a discrete version of the Gini index. It reaches its maximum value $n/(n+1)$ on the uniform distribution $\pi_x = 1/(n+1)$, $i = 0,1,\ldots,n$, and the minimum zero value if the distribution is concentrated on one class, see Fig. 1 (left).

G. Pistone
Collegio Carlo Alberto, Via Real Collegio 30, 10024 Moncalieri, Italy e-mail: `giovanni.pistone@carloalberto.org`

M.P. Rogantin
Dipartimento di Matematica, Via Dodecaneso 35, 16146 Genova, Italy e-mail: `rogantin@dima.unige.it`

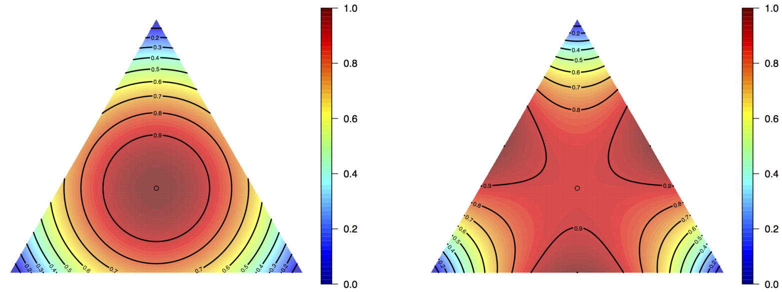

Fig. 1 Normalized fractionalization (left) and normalized polarization (right), discrete distance. Representation in barycentric coordinates. The usual gradient flow does not converge to the extremal point on the border of the simplex.

This index has been used in Economics by Alesina et. al [2]. The polarization measure has been introduced in full generality by Esterban and Ray [4]; the discrete version we consider here was used by Pino and Vidal-Robert [6, p. 10]. Let X, Y, Z be iid $\sim \pi$ and consider the indicator of *exactly two equal* $I_2 = (X = Y \neq Z) + (X = Z \neq Y) + (Y = Z \neq X)$. Then $\mathbb{E}[I_2] = 3\sum_x \pi_x^2(1-\pi_x) = 3\,\mathrm{POL}(\pi)$, see Fig. 1 (right). The polarization measure has an unstable critical point at the uniform distribution, it is zero in the case of concentration in one class, and reaches its maximum 1/4 on distributions on two classes with equal probabilities.

In Fig. 1 the simplex is represented as an equilateral triangle, the Cartesian coordinate being related with the probabilities by $\pi_1 = -(u-1/2) - (\frac{1}{\sqrt{3}})(v - \sqrt{3}/2)$, $\pi_2 = (u-1/2) - (\frac{1}{\sqrt{3}})(v-\sqrt{3}/2)$, while the probabilities are the barycentric coordinates:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \pi_0 \begin{bmatrix} 1/2 \\ \sqrt{3}/2 \end{bmatrix} + \pi_1 \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \pi_2 \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

We want to study this indexes in a dynamical context, i.e. to characterize evolutions that maximize or minimize the relevant index. This study requires tools from Information Geometry, see e.g. [3, 8, 5, 7].

Note that the curve of steepest variation of FRAC from π is the mixture model of π and $(1/3, 1/3, 1/3)$ but in general the minimum on this line is not reached at a vertex. A similar situation arises for POL. As the ordinary gradient flow does not lead to the extrema of interest on the border of the simplex, we turn to a different way to compute the gradient.

2 Natural gradient

Let π be a strictly positive multinomial on $0,1,\ldots,n$, that is

$$\pi \in \Delta_n^\circ = \left\{ \pi \in \mathbb{R}^{n+1} \left| \sum_{x=0}^{n} \pi_x = 1, \pi_x > 0, x = 0,1,\ldots,n \right. \right\}.$$

and define B_π to be the vector space of random variables U that are π-centered, $\mathbb{E}_\pi[U] = 0$.

The set $T\Delta_n^\circ = \{(\pi,u) | \pi \in \Delta_n^\circ, u \in B_\pi\}$ is the *tangent bundle* of the multinomial model. If $t \mapsto \pi(t) \in \Delta_n^\circ$ is a curve, its *score* $\mathrm{dl}\,\pi(t) = \dot{\pi}(t)/\pi(t) = d\log\pi(t)/dt$ belongs to $B_{\pi(t)}$, hence $t \mapsto (\pi(t), \mathrm{dl}\,\pi(t))$ is a curve in the tangent bundle.

For example, consider the immigration process $\pi_x(t) \propto \pi_x(0) + \alpha t(x=0)$, $x = 0,1,\ldots,n$. Then $\sum_{x=0}^n \pi_x(t) = 1+\alpha t$ and $\log(\pi_x(t)) = \log(\pi_x(0) + \alpha t(x=0)) - \log(1+\alpha t)$ and

$$\mathrm{dl}\,\pi_x(t) = \frac{\alpha(x=0)}{\pi_x(0)+\alpha t(x=0)} - \frac{\alpha}{1+\alpha t} = \frac{\alpha}{\pi_0(0)+\alpha t}(x=0) - \frac{\alpha}{1+\alpha t},$$

so that it is easy to check $\mathbb{E}_{\pi(t)}[\mathrm{dl}\,\pi(t)] = 0$.

A *vector field* F is a mapping on the multinomial model such that $F(\pi) \in B_\pi$, i.e. such that $(\pi, F(\pi)) \in T\Delta_n^\circ$ for all π. A *differential equation* is an equation of the form $\mathrm{dl}\,\pi(t) = F(\pi(t))$. Given a real function $\phi\colon \Delta_n^\circ \to \mathbb{R}$, its *gradient* is the vector field $\nabla\phi$ such that for all curves $\pi(\cdot)$ we have

$$\frac{d}{dt}\phi(\pi(t)) = \langle \nabla\phi(\pi(t)), \mathrm{dl}\,\pi(t) \rangle_{\pi(t)}, \qquad \langle u,v \rangle_\pi = \mathbb{E}_\pi[uv], \quad u,v \in B_\pi. \tag{1}$$

The *gradient flow equation* is the differential equation $\mathrm{dl}\,\pi(t) = \pm\nabla\phi(\pi(t))$, which implies that $d\phi(\pi(t))/dt = \pm\langle \nabla\phi(\pi(t)), \nabla\phi(\pi(t)) \rangle_{\pi(t)}$ is of definite sign.

Computations are usually performed in a parameterization $\pi\colon \Theta \ni \boldsymbol{\theta} \mapsto \pi(\boldsymbol{\theta}) \in \Delta_n^\circ$, Θ being an open set in $\mathbb{R}^n$. The random variables $\mathrm{dl}_j\,\pi(\boldsymbol{\theta}) = \partial\log\pi(\boldsymbol{\theta})/\partial\theta_j$, $j = 1,\ldots,n$, form a vector basis of $B_{\pi(\boldsymbol{\theta})}$ and the representation of the scalar product in the basis is $\left\langle \sum_{i=1}^n \alpha_i\,\mathrm{dl}_i\,\pi(\boldsymbol{\theta}), \sum_{j=1}^n \beta_j\,\mathrm{dl}_j\,\pi(\boldsymbol{\theta}) \right\rangle_{\pi(\boldsymbol{\theta})} = \sum_{i,j=1}^n \alpha_i\beta_j I_{ij}(\boldsymbol{\theta})$, where the *Fisher information* matrix is:

$$I(\boldsymbol{\theta}) = [\mathbb{E}_\pi[\mathrm{dl}_i\,\pi(\boldsymbol{\theta})\,\mathrm{dl}_j\,\pi(\boldsymbol{\theta})]]_{i,j=1}^n = \left[\langle \mathrm{dl}_i\,\pi(\boldsymbol{\theta}), \mathrm{dl}_j\,\pi(\boldsymbol{\theta}) \rangle_{\pi(\boldsymbol{\theta})} \right]_{i,j=1}^n.$$

If $\boldsymbol{\theta} \mapsto \tilde{\phi}(\boldsymbol{\theta})$ is the representation of a function ϕ in the parameters and $t \mapsto \boldsymbol{\theta}(t)$ is the expression of a generic curve in the parameters, then the components of the gradient in (1) are expressed in terms of the ordinary gradient by observing that

$$\frac{d}{dt}\phi(p(t)) = \frac{d}{dt}\tilde{\phi}(\boldsymbol{\theta}(t)) = \sum_{j=1}^n \frac{\partial}{\partial\theta_j}\tilde{\phi}(\boldsymbol{\theta}(t))\dot{\theta}_j(t),$$

and $\operatorname{dl}\pi(t) = \sum_j \operatorname{dl}_j \pi(\boldsymbol{\theta}(t))\dot{\theta}_j(t)$. The vector $\widetilde{\nabla}\tilde{\phi}(\boldsymbol{\theta}) = \nabla\tilde{\phi}(\boldsymbol{\theta})I^{-1}(\boldsymbol{\theta})$ is the *natural gradient*.

3 Computation of the natural gradient in a parameterization

The simplest possible example where the notion of parameterization is meaningful happens for example in case of an electoral competition with 3 parties. In this scenario, te maximum polarizion corresponds to the outcome where one of the parties wins zero members of Parliament (MP), while the other two parties win the same number of MPs.

The most common parameterization of the flat simplex Δ_n° is the projection on the solid simplex, that is $\pi\colon \boldsymbol{\theta} \mapsto (1-\sum_{j=1}^n \theta_j, \theta_1, \dots, \theta_n)$, in which case $\partial_j \pi(\boldsymbol{\theta})$ is the random variable with values -1 at $x=0$, 1 at $x=j$, 0 otherwise, or $\operatorname{dl}_j \pi(\boldsymbol{\theta}) = ((x=j)-(x=0))/\pi_x(\boldsymbol{\theta})$. The elements of the Fisher information matrix are $I_{jh}(\boldsymbol{\theta}) = \theta_j^{-1}(j=h) + (1-\sum_k \theta_k)^{-1}$, hence

$$I(\boldsymbol{\theta}) = (1-\sum_{j=1}^n \theta_j)^{-1}\left(\operatorname{diag}\left(\theta_j^{-1}\colon j=1,\dots,n\right) + (1-\sum_{j=1}^n \theta_j)^{-1}[1]_{i,j=1}^n\right).$$

The inverse $I^{-1}(\boldsymbol{\theta})$ is computable in closed form. For example, for $n=2$,

$$\begin{aligned} I(\theta_1,\theta_2) &= \begin{bmatrix} \theta_1^{-1} + (1-\theta_1-\theta_2)^{-1} & (1-\theta_1-\theta_2)^{-1} \\ (1-\theta_1-\theta_2)^{-1} & \theta_2^{-1} + (1-\theta_1-\theta_2)^{-1} \end{bmatrix}, \\ I^{-1}(\theta_1,\theta_2) &= \begin{bmatrix} (1-\theta_1)\theta_1 & -\theta_1\theta_2 \\ -\theta_1\theta_2 & (1-\theta_2)\theta_2 \end{bmatrix}. \end{aligned}$$

In the solid simplex parameterization, the fractionalization measure is

$$\widetilde{\mathrm{FRAC}}(\boldsymbol{\theta}) = 1 - (1-\sum_{j=1}^n \theta_j)^2 - \sum_{j=1}^n \theta_j^2,$$

with partial derivatives $\partial_j \widetilde{\mathrm{FRAC}}(\boldsymbol{\theta}) = 2(1-\sum_{h=1}^n \theta_h - \theta_j)$. If $n=2$, the gradient is $\nabla\widetilde{\mathrm{FRAC}}(\boldsymbol{\theta}) = (2(1-2\theta_1-\theta_2), 2(1-\theta_1-2\theta_2))$ and the natural gradient is

$$\begin{aligned} \widetilde{\nabla}\widetilde{\mathrm{FRAC}}(\boldsymbol{\theta}) = 2\big(&2\theta_1^3 + 2\theta_1^2\theta_2 + 2\theta_1\theta_2^2 - 3\theta_1^2 - 2\theta_1\theta_2 + \theta_1, \\ &2\theta_1^2\theta_2 + 2\theta_1\theta_2^2 + 2\theta_2^3 - 2\theta_1\theta_2 - 3\theta_2^2 + \theta_2\big). \end{aligned}$$

On the vertices of Δ_n, i.e. $\pi = \delta_x$, the polarization is zero. On 1-faces of Δ_n, i.e. $\pi = (1-\theta)\delta_x + \theta\delta_y$, $x \neq y$, $\theta \in [0,1]$, we have $\mathrm{POL}(\pi) = (1-\theta)^2\theta + \theta^2(1-\theta) = \theta(1-\theta)$, then values outside the vertices are strictly larger than 0, with maximum value $1/4$ obtained with equal probabilities on the two classes x, y. It follows that

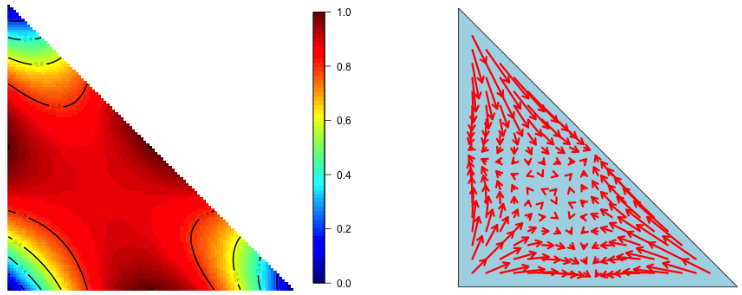

Fig. 2 Polarization measure in the solid simplex parameterization (left) and its natural gradient field (right)

this polarization measure is larger in 2 clusters than in 1 cluster, and the maximum in the case of 2 clusters is obtained with equal (probability of) classes. Interpretation is less clear in 2 faces (3 clusters). If $\pi = \theta_1 \delta_x + \theta_2 \delta_y + (1-\theta_1-\theta_2)\delta_z$, $x \neq y \neq x$, then $\operatorname{POL}(\pi) = \theta_1^2(1-\theta_1) + \theta_2^2(1-\theta_2) + (1-\theta_1-\theta_2)^2(\theta_1+\theta_2)$, hence it is not true that all 2 faces values are larger that the values in the 1 faces of the border. Among all 3 clusters, the maximum is obtained a equal probabilities.

In the solid simplex parameterization we have

$$\widetilde{\operatorname{POL}}(\boldsymbol{\theta}) = \left(1 - \sum_{j=1}^{n} \theta_j\right)^2 \left(\sum_{j=1}^{n} \theta_j\right) + \sum_{j=1}^{n} \theta_j^2 (1-\theta_j),$$

with gradient

$$\nabla \widetilde{\operatorname{POL}}(\boldsymbol{\theta}) = \left(6\theta_1\theta_2 + 3\theta_2^2 - 2\theta_1 - 4\theta_2 + 1, 3\theta_1^2 + 6\theta_1\theta_2 - 4\theta_1 - 2\theta_2 + 1\right)$$

and natural gradient

$$\begin{aligned}\widetilde{\nabla}\, \widetilde{\operatorname{POL}}(\boldsymbol{\theta}) = \big(&-9\theta_1^3\theta_2 - 9\theta_1^2\theta_2^2 + 2\theta_1^3 + 14\theta_1^2\theta_2 + 5\theta_1\theta_2^2 - 3\theta_1^2 - 5\theta_1\theta_2 + \theta_1, \\ &-9\theta_1^2\theta_2^2 - 9\theta_1\theta_2^3 + 5\theta_1^2\theta_2 + 14\theta_1\theta_2^2 + 2\theta_2^3 - 5\theta_1\theta_2 - 3\theta_2^2 + \theta_2\big)\end{aligned}$$

See in Fig. 2 the natural gradient field. Note that the increase near the maxima is not very sharp with this measure.

4 Applications

Consider a study of the evolution in time of the polarization measure, see [6]. In the time series $\pi(1), \pi(2), \ldots$, the evolution of the polarization, $\operatorname{POL}(\pi(1)), \operatorname{POL}(\pi(2)), \ldots$, could be misleading, because an increase in the index could be associated to a shift from a basin of attraction to a different basin of attraction. We suggest, given a movement from $\pi(t)$ to $\pi(t+1)$ to compare an estimate $\overrightarrow{\pi(t)\pi(t+1)}$ of the velocity vector to the gradient field of the polarization measure, that is compute $\left\langle \overrightarrow{\pi(t)\pi(t+1)}, \nabla \operatorname{POL}(\pi(t)) \right\rangle_{\pi(t)}$. An estimate of the velocity of change is a mapping from a couple of densities π_{initial}, π_{final} to the tangent space at the initial density $T_{\pi_{\text{initial}}} \Delta_n^\circ$. Such a mapping goes under the name of *inverse retraction* in the literature about optimzation on manifold e.g., [1], the simplest example here being $\overrightarrow{\pi_t \pi_{t+1}} = (\pi_{t+1} - \pi_t)/\pi_t = \pi_{t+1}/\pi_t - 1$. Other authors suggest the use the initial velocity of the Riemannian geodesic connecting π_t to π_{t+1}, but this requires the computation of the geodesic itself.

Acknowledgements G. Pistone is supported by the de Castro Statistics Initiative, Collegio Carlo Alberto, Moncalieri, Italy, and is a member of GNAMPA-IndAM.

References

1. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization algorithms on matrix manifolds. Princeton University Press, Princeton, NJ (2008). With a foreword by Paul Van Dooren
2. Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., Wacziarg, R.: Fractionalization. Journal of Economic Growth **8**(2), 155–194 (2003).
3. Amari, S., Nagaoka, H.: Methods of information geometry. American Mathematical Society, Providence, RI (2000).
4. Esteban, J., Ray, D.: On the measurement of polarization. Econometrica **62**(4), 819–851 (1994).
5. Gibilisco, P., Riccomagno, E., Rogantin, M.P., Wynn, H.P. (eds.): Algebraic and geometric methods in statistics. Cambridge University Press, Cambridge (2010)
6. Pino, F.J., Vidal-Robert, J.: Habemus papam? polarization and conflict in the papal states (2013). `http://www.afse-lagv.com/lagv/submissions/index.php/LAGV2013/LAGV12/paper/viewFile/362/152`
7. Pistone, G.: Nonparametric information geometry. In: F. Nielsen, F. Barbaresco (eds.) Geometric Science of Information, no. 8085 in LNCS, pp. 5–36. Springer-Verlag, Berlin Heidelberg (2013).
8. Pistone, G., Rogantin, M.: The exponential statistical manifold: mean parameters, orthogonality and space transformations. Bernoulli **5**(4), 721–760 (1999)