



Tat Dat Tran*, Jürgen Jost*# and Julian Hofrichter*
 Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany*
 The Santa Fe Institute, Santa Fe, New Mexico, USA#
 (trandat@mis.mpg.de, jost@mis.mpg.de, julian.hofrichter@mis.mpg.de)

Motivation

- Information geometry studies the geometry of a family of probability distributions with concept like the Fisher information metric or the Amari-Chentsov connection.
- Population genetics studies the change in frequency of alleles/genotypes within populations due to evolutionary forces such as random genetic drift, mutation, selection and recombination.
- When population size is infinite, the change in frequency of genotypes follows a deterministic dynamical system on the state space (considered as the space of probability distributions on the discrete space of genotypes).
- When population size is finite, this change is random and follows a stochastic dynamical system on the state space which may arise dramatically different evolutionary behaviours (metastability, stochastic tunnelling, etc.). Another approach is to consider its probability distribution which satisfies an infinite dimensional deterministic dynamical system on the space of probability distributions on the state space.

Aim: Using techniques/results in information geometry allow us to geometrically explain many evolutionary phenomenon in population genetics.

References

- [1] Akin, E. (1979), *The geometry of population genetics*, Lecture Notes in Biomathematics, Springer-Verlag, Berlin-New York, Volume 31. [Book link](#)
- [2] Amari, S. (1985), *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statistics, Springer, Berlin., Volume 28. [Book Link](#)
- [3] Amari S. (2001), Information geometry on hierarchy of probability distributions, *IEEE Transactions on Information Theory*, Vol. 47, No. 5, 1701-1711. [Journal Link](#)
- [4] Antonelli P. L., Strobeck C. (1977), The Geometry of random drift I. Stochastic Distance and Diffusion, *Adv. Appl. Probab.*, 9(2):238-249. [Journal link](#)
- [5] Antonelli P. L., Morgan K., Lathrop G. M. (1977), The geometry of random drift III. Recombination and diffusion, *Adv. Appl. Probab.*, 9(2):260-267. [Journal link](#)
- [6] Shahshahani, S. (1979), A new mathematical framework for the study of linkage and selection, *Mem. Amer. Math. Soc.*, 17. [Journal link](#)
- [7] Tran T.D. (2012) *Information Geometry and the Wright-Fisher model of mathematical population genetics*, PhD Thesis.
- [8] Hofrichter J., Jost J., Tran T.D. (2016) *Information geometry and population genetics*, submitted.

The state space Ω_n

In the population of 1-locus: The interior part of the state space is an open simplex:

$$\Omega_n^\circ = \Delta_n^1 = \left\{ \theta = (\theta^1, \dots, \theta^n) : \theta^i > 0 \text{ for } i = 1, \dots, n, \text{ and } \sum_{i=1}^n \theta^i < 1 \right\},$$

where θ^i is the frequency of allele A_i . The Fisher information metric for a family of probability distributions $p(w, \theta) = \theta^w$ on the space of alleles $A = \{1, \dots, n+1\}$ parameterized in $\theta \in \Delta_n^1$ is

$$(g_F)_{ij}(\theta) = \sum_{w \in A} \frac{1}{p(w, \theta)} \frac{\partial p(w, \theta)}{\partial \theta^i} \frac{\partial p(w, \theta)}{\partial \theta^j} = \frac{\delta_{ij}}{\theta^i} + \frac{1}{\theta^{n+1}}, \quad \forall i, j = \overline{1, n},$$

where $\theta^{n+1} = 1 - \theta^1 - \dots - \theta^n$. This metric g_F is also called Shahshahani metric and makes Δ_n^1 become an n -dimensional smooth statistical manifold. Moreover, (Δ_n^1, g_F) is a Riemannian manifold with the constant sectional curvature of $\frac{1}{4}$; an Einstein manifold with the Ricci tensor $R_{ik} = \frac{n-1}{4} g_{ik}$; and its the scalar curvature equals to $\frac{n(n-1)}{4}$.

Application:

- In a large 1-locus multi-allele diploid population with only selection, $m_{ij} = m_{ji}$ is the fitness of genotype $A_i A_j$:

$$\dot{\theta}^i = \theta^i (m_i(\theta) - \bar{m}(\theta)), \quad \forall i = \overline{1, n},$$

where $m_i(\theta) = \sum_{j=1}^n m_{ij} \theta^j$ and $\bar{m}(\theta) = \sum_{i=1}^n m_i \theta^i$. This system is difficult to solve explicitly but we can have the qualitative behaviour because of the flow can be seen as an Fisher-gradient flow, i.e.

$$\dot{\theta} = \nabla_{g_F} \left(\frac{1}{2} \bar{m} \right).$$

This explains the Fisher's fundamental theorem and Kimura's maximum principle in population genetics.

The space of probability distributions on the state space $\mathcal{P}(\Omega_n)$

We define the **Fisher-Wasserstein distance** on $\mathcal{P}(\Omega_n)$ by

$$G_{FW}(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \left\{ \int_{\Omega_n \times \Omega_n} d_{g_F}^2(\mathbf{x}, \mathbf{y}) \pi(d\mathbf{x}, d\mathbf{y}) \right\}, \quad \forall \mu, \nu \in \mathcal{P}(\Omega_n),$$

where d_{g_F} is the Fisher distance derived from Fisher information metric, $\Pi(\mu, \nu)$ is the space of probability measures on $\Omega_n \times \Omega_n$ with marginals μ and ν .

Applications:

- In 1-locus Wright-Fisher population: The probability density satisfy the Kolmogorov backward equation:

$$\frac{\partial u}{\partial t} = \frac{1}{2} g_F^{ij}(p) \frac{\partial^2 u}{\partial p^i \partial p^j}.$$

where $(g_F^{ij}(p))$ is the covariance matrix of the multinomial distribution, also inverse of Fisher information metric. We can represent it in the standard form

$$\frac{\partial u}{\partial t} = \frac{1}{2} \Delta_{g_F} u + ch_{g_F}^i(p) \frac{\partial u}{\partial p^i},$$

where $ch_{g_F}^k(p) = \frac{1}{2} g_F^{ij}(p) \Gamma_{ij}^k(p)$ is the Christoffel force and Δ_{g_F} is the Laplace-Beltrami operator with respect to the Fisher information metrics. The Christoffel forces can be calculated easily to have $ch_{g_F}^k(p) = -\frac{1}{4}(1 - (n+1)p^k)$. This explains many biological behaviours as follows:

- The Christoffel velocities vanish at the centroid of the frequency space, so that near the centroid, the diffusion equation for the general WF model is very well approximated by that of spherical Brownian motion.
- The Christoffel velocities drive the populations toward the vertices of the frequency space. It explains why there is fixation of one allele when there is no other evolutionary forces.

- In 1-locus Wright-Fisher population with uniform mutations: The probability density satisfy the Kolmogorov forward equation (Fokker-Planck equation):

$$\frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial^2}{\partial x^i \partial x^j} \left(g_F^{ij}(x) u \right) - \frac{\partial}{\partial x^i} \left((\mu_i - |\mu| x^i) u \right).$$

The free energy functional is

$$\mathcal{F}(f) = \int_{\Omega_n} f(\mathbf{x}) (\ln f(\mathbf{x}) + \psi(\mathbf{x})) d\mathbf{x}, \quad \forall f \in \text{Density}(\Omega_n, d\mathbf{x})$$

where the potential energy $\psi(\mathbf{x}) = \sum_{i=1}^{n+1} (1 - \mu_i) \ln x_i$. We then can rewrite the flow as

$$\partial_t u(t, \cdot) = -\nabla_{G_{FW}} \mathcal{F}(u(t, \cdot)).$$

This explains asymptotical behaviour of the flow of probability density.

Acknowledgements



The research leading to these results has received funding from the European Research Council under the European Union Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 267087.