Ergodicity in Science and Society

with Application to Information Geometry

Angela De Sanctis

Dipartimento di Metodi Quantitativi e Teoria Economica

Università “G. d’Annunzio” di Chieti-Pescara

a.desanctis@unich.it

Abstract

We discuss here the importance of Ergodic Theory in Science and Society. Then we introduce it in Information Geometry, beginning by the investigation of the ergodicity of the geodesic flow on the tangent bundle of the Normal Family, endowed with the Fisher information metric. We think that this method is able to have interesting applications in the study of Dynamic Econometrics.

1 Introduction: Ergodicity in Science and Society

Ergodic Theory studies the statistical properties of orbits, generated by maps, to investigate the predictability of them. Let $(M, \mathcal{A}, \mu, T)$ be a dynamical system, that is $(M, \mathcal{A}, \mu)$ is a measurable space and T is the law governing the time evolution of the system, the orbit by $x \in M$ is $\{x, Tx, T^2x, \dots\}$.

We wish to give an answer to the following question: How often does an orbit originating from a given point x visit a given region $A \in \mathcal{A}$?

If T is a transformation preserving μ , that is $\mu(T^{-1}(A)) = \mu(A)$, we define

$$V_n(x) \equiv \#\{k : 0 \leq k < n, T^k x \in A\}$$

$$\text{and } v_n(x) \equiv \frac{1}{n} V_n(x).$$

Therefore $V_n(x)$ denotes the number of visits of a set A after n iterations of T and $v_n(x)$ denotes the average number of visits. We would like to know whether the limit $\hat{v}(x) \equiv \lim_{n \rightarrow \infty} v_n(x)$ exists.

This can be established by means of a basic result, known as the Birkhoff-Khinchin Ergodic Theorem. In its generality, the Ergodic Theorem states that, if T preserves the probability measure μ and f is any integrable function on M then the limit $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) \equiv \hat{f}(x)$ exists for μ almost every point x of M and $\hat{f}(x)$ is T invariant i.e. $\hat{f}(Tx) = \hat{f}(x)$.

If we choose $f = \chi_A$, the characteristic function of A , then $\frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = v_n(x)$ and therefore the limit $\hat{v}(x)$ exists.

But in general it depends on x . This happens, for example, when the space is decomposable under the action of T and there exist two subspaces M_1 and M_2 , both invariant with respect to T , i.e. when T maps points of M_1 only to M_1 and points of M_2 only to M_2 .

But we are not interested in the properties of a single orbit starting from an arbitrary initial point but in the overall properties of ensembles of orbits originating from all possible initial conditions in a certain given region of the space. Thus it would be desirable that the average calculated along a particular “history” of the system should be equal to the averages evaluated over all possible histories. We call μ an ergodic measure if the limiting function $\hat{f}$ defined in the Ergodic Theorem is a constant, that is $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = \int_M f(x) d\mu(x)$ for μ almost every point x of M .

The ergodicity of μ implies that the average number of the visits to the region A of an orbit originating from almost every point is equal to the size that the ergodic measure assigns to that region (again using $f = \chi_A$).

Boltzman made the following:

Ergodic Hypotesis: A perfect gas (many molecules and elastic collisions) is ergodic.

This a difficult question which has been proved by Sinai [1976].

Contrast this with the 3-body problem (few particles and no collisions) which is not ergodic.

It is possible to prove that ergodicity is equivalent to indecomposability in the following sense: Whenever $T^{-1}(A) = A$ for some $A \in \mathcal{A}$ then either $\mu(A) = 1$ or $\mu(A) = 0$. This means roughly “any point goes everywhere”.

A classical example is the following: Consider the map $T_c : S^1 \longrightarrow S^1$, where S^1 denotes the unit circle, with $T_c(z) = e^{i\alpha 2\pi} z$, $\alpha \in [0, 1)$. A point on S^1 is identified by the angle formed by the abscissa and the line joining the point with the origin. The map T_c rotates points on the circle by an angle $\alpha 2\pi$.

It is easy to see that T_c preserves the Lebesgue measure on the circle defined by $\hat{m} = m \circ \theta^{-1}$, where m is the Lebesgue measure and $\theta : [0, 1) \longrightarrow S^1$ is $\theta(x) = e^{i2\pi x}$. In words: when we apply the Lebesgue measure on the circle to a subset, first we map it to a corresponding subinterval of $[0, 1)$ and then we assign to that subinterval a value equal to its length.

If α is irrational, the measure $\hat{m}$ on S^1 is ergodic, as we deduce by Fourier analysis.

This map is the prototype of models generating (non chaotic, that is zero entropy) dynamics which arise in a number of models in Economics of which we mention here three main classes, all of them formulated in a discrete-time

setting. First, we have the models describing optimal growth; the second class comprises models of overlapping generations with production; finally we have models of Keynesian derivation describing the dynamics of a macroeconomic system characterized by nonlinear multiplier-accelerator mechanisms [Medio, 1999].

When dynamical systems are chaotic, the prototype is the well-known “logistic map” $T_L : [0, 1] \longrightarrow [0, 1]$, $T_L(x) = rx(1 - x)$ with suitable r .

2 Ergodic Theory

Let M be a measurable space endowed with a σ - algebra $\mathcal{A}$ of subsets. We suppose that a group of transformations G acts on M . That is for any $g \in G$ the transformation $T_g : M \longrightarrow M$ is defined so that:

- T_g is a measurable transformation;
- $T_g \circ T_h = T_{gh}$ (then T_g is invertible and $(T_g)^{-1} = T_{g^{-1}}$).

At first we will examine the case where G is countable.

Definition 2.1 *A finite measure μ (which can always be taken to satisfy $\mu(M) = 1$ by normalization) is called invariant under G if, for any set $A \in \mathcal{A}$ and for any element $g \in G$,*

$$\mu(A) = \mu(T_g^{-1}A) = \mu(T_g A)$$

i.e. the measure of any measurable set equals the measure of the image and the inverse image of that set.

Since M is a measurable space, it is possible to examine the measurable functions (random variables) $f(x)$.

The next theorem shows that the presence of an invariant measure can be related to the statistical properties of the action of the group G .

Theorem 2.1 (Ergodic Theorem of Birkhoff-Khinchin) *Let $f(x) \in L^1_\mu(M)$ and let μ be invariant with respect to G . Then the following limit exists with probability 1:*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T_g^k x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T_g^{-k} x)$$

In the Theory of Probability similar assertions are called laws of large numbers and since the convergence takes place almost everywhere the Ergodic Theorem of B.-K. is a theorem of the type of a strengthened law of large numbers [Billingsley, 1965].

Definition 2.2 *The transformation T_g is ergodic if, in the theorem of B.-K. for any function $f \in L^1_\mu(M)$ the limit function $\hat{f}$ is constant precisely:*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^n f(T_g^k x) = \hat{f}(x) = \int_M f(x) d\mu(x) \quad a.e.$$

In Probability Theory this is the form of the most common strengthening of the law of large numbers. The average converges almost everywhere to the mathematical expectation. It is possible to prove that the property of ergodicity of T_g is equivalent to the following assertion: for every invariant set A (that is $\mu(A) = \mu(T_g^{-1}A)$) we have that $\mu(A)$ is equal to 0 or 1. So the property of ergodicity is equivalent to the indecomposability of the system into nontrivial invariant subsets.

3 Geodesic Flow

In the previous section we talked about “abstract” dynamical systems as measure preserving transformations in Probability Theory. Now we will talk about “classical” systems as diffeomorphisms of smooth manifolds or smooth flows on smooth manifolds from the standpoint of Ergodic Theory.

Let (M, g) be a Riemannian manifold and let TM be its tangent bundle.

The geodesic flow on TM , denoted by $(S_t)_{t \in \mathbb{R}}$, assigns to every $(m, v) \in T_m M$ such that $\|v\| = \sqrt{\sum_{i,j} g_{ij} v^i v^j} = 1$ an (m_t, v_t) , obtained by marking the geodesic from m in the direction v , going along with constant speed 1 for a piece of length t in the metric g and stopping at m_t with speed v_t such that $\|v_t\| = 1$, as in Figure 1 below.

Figure 1: Geodesic flow

$(S_t)_{t \in \mathbb{R}}$ is a continuous group of smooth transformations.

Theorem 3.1 (Liouville) *[Arnold & Avez, 1967]: The geodesic flow preserves the volume μ on TM as Riemannian manifold with the metric*

$$\left(\sum_{i,j} g_{ij}(m) dx^i dx^j + g_{ij}(m) dv^i dv^j \right)^{\frac{1}{2}}$$

that is

$$\mu(S_t A) = \mu(A) \quad , \quad A \subset TM \quad t \in \mathbb{R}.$$

4 The Geometry of the Normal Family

A famous example of a parametrised family of probability measures is the Normal Family [Murray & Rice, 1993].

The sample space is $\mathbb{R}$ and the densities are with respect to Lebesgue measure dx on $\mathbb{R}$, so that as a set of probability measures the Normal Family

is

$$\mathcal{N} = \{p(\mu, \sigma)dx : \mu \in \mathbb{R}, \sigma > 0\}$$

with densities

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

The parameter of this family is the pair (μ, σ) which varies over the open subset of $\mathbb{R}^2$ determined by $\sigma > 0$.

The Normal Family is a particular type of Exponential Family. In fact if V is the span of x^2 and x , let U be the following open set

$$U = \{ax^2 + bx : a < 0\}$$

Every density as well as its usual parametrization can be written

$$p(\theta^1, \theta^2)(x) = \exp(x^2\theta^1 + x\theta^2 - k(\theta))$$

where

$$\theta^1 = -\frac{1}{2\sigma^2} \quad \theta^2 = \frac{\mu}{\sigma^2} \quad \text{and} \quad k(\theta) = \frac{1}{2} \ln\left(-\frac{\pi}{\theta^1}\right) - \frac{(\theta^2)^2}{4\theta^1}$$

The parameter $\theta = (\theta^1, \theta^2)$ is called the canonical parameter and lies in the open subset of $\mathbb{R}^2$ defined by $\theta^1 < 0$.

The log-likelihood function for $\mathcal{N}$ is

$$l(x, p) = x^2 \left(\frac{-1}{2\sigma^2}\right) + x \left(\frac{\mu}{\sigma^2}\right) - \ln(\sqrt{2\pi}\sigma) - \frac{\mu^2}{2\sigma^2}$$

As

$$\frac{\partial l}{\partial \mu} = \frac{x - \mu}{\sigma} \quad \text{and} \quad \frac{\partial l}{\partial \sigma} = \frac{(x - \mu)^2}{\sigma^3} - \frac{1}{\sigma}$$

the Fisher information (Riemannian) metric corresponds to the matrix with elements

$$g\left(\frac{\partial}{\partial \mu}, \frac{\partial}{\partial \sigma}\right) = 0 \quad , \quad g\left(\frac{\partial}{\partial \mu}, \frac{\partial}{\partial \mu}\right) = \frac{1}{\sigma^2} \quad , \quad g\left(\frac{\partial}{\partial \sigma}, \frac{\partial}{\partial \sigma}\right) = \frac{2}{\sigma^2}.$$

This means that the Normal Family can be identified with the upper half-plane $H = \{z = (\mu, \sigma) : \sigma > 0\}$ and the geometry of this Riemannian manifold is the hyperbolic geometry of Lobachevskian plane. The geodesics are either vertical lines or half-circles intersecting the μ axis perpendicularly. This geometry satisfies all the axioms of Euclidean plane geometry except the parallelism postulate. Indeed, through each point not on a geodesic γ there pass infinitely many geodesics that do not meet γ . The (gaussian) curvature is constant and equal to -1 . The isometries are the fractional linear transformations $f(z) = \frac{az+b}{cz+d}$ that take the upper half-plane into itself, so the group of these isometries is identifiable with $PSL(2, \mathbb{R})$ by the correspondence $f \longrightarrow \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

The hyperbolic plane H has a much richer symmetry structure than the euclidean plane. This has to do with the fact that its group of isometries $PSL(2, \mathbb{R})$ is much subtler than the group of euclidean traslations plus rotations in $\mathbb{R}^2$. This observation is emphasized by Poincarè model for the hyperbolic plane in which H is the interior of the unit disk, the geodesics are circles orthogonal to the unit circle and non- euclidean angles are equal to euclidean angles (see Figure 2).

Figure 2: Poincarè disk

If we denote by $z = (\mu, \sigma)$ with $\sigma > 0$ and by (z, w) an element of the tangent bundle of $\mathcal{N}$ with unit speed, we can find only one $g \in PSL(2, \mathbb{R})$ which transforms (z, w) into $(i = (0, 1), v = (0, 1))$. We deduce that the tangent bundle with elements of unit speed denoted by $T^1\mathcal{N}$ is also identifiable with $PSL(2, \mathbb{R})$.

5 Ergodicity and Normal Family

Two directed (oriented) geodesics of the Normal Family are called “positively asymptotic” if they end in the same point of the abscissa (real μ -axis). Let that point be denoted u , as in Figure 3a below.

Figure 3a: Positively asymptotic geodesics

We will determine the speed of contraction of such asymptotic geodesics as they tend to u .

Proposition 5.1. *Asymptotic geodesics approach each other with exponential speed.*

Proof: We perform a fractional linear transformation ρ taking the upper half-plane into itself and the point u to infinity. Then $\rho(\gamma^1)$ and $\rho(\gamma^2)$ will be taken to two parallel lines intersecting the abscissa at the points μ_1 and μ_2 , as in Figure 3b.

Figure 3b: Asymptotic geodesics approach each other with exponential speed.

The distance between the points of $\rho(\gamma^1)$ and $\rho(\gamma^2)$ located on the level σ equals

$$s = \int_{\mu_1}^{\mu_2} \frac{d\mu}{\sigma} = \frac{\mu_2 - \mu_1}{\sigma}$$

and in turn

$$r = \sqrt{2} \int_1^{\sigma} \frac{d\sigma}{\sigma} = \sqrt{2} \ln \sigma$$

i.e. the relative distance is

$$s = \frac{\mu_2 - \mu_1}{e^{r/\sqrt{2}}}$$

■

This property of the asymptotic geodesics is the basic for the proof of Ergodicity.

Fixing a point u of the abscissa, we will examine the family of mutually asymptotic directed geodesics ending at the point u . It is clear that through

every point of the plane there passes one and only one geodesic of the family. We will examine the trajectories orthogonal to the family of asymptotic geodesics. Because non euclidean angles coincide with the euclidean angles it follows that the orthogonal trajectories form circles that are tangent to the abscissa at a point. These circles are called “horocycles”. In this way, to the family of asymptotic directed geodesics corresponds a family of orthogonal trajectories.

A calculation similar to the above shows an important relation that holds between these two families. We take the interval $\tilde{\gamma}_0$ of the horocycle from a to b . Through each of its points c we construct a geodesic interval of length t in the direction of the point u . Then the union of the ends of these intervals is again an interval $\tilde{\gamma}_t$ of the horocycle. Their lengths are related by: $l(\tilde{\gamma}_t) = e^{-t}l(\tilde{\gamma}_0)$, see Figure 4 below.

Figure 4: Contracting length

Now we consider the tangent bundle with elements of unit speed $T^1\mathcal{N}$ of the Normal Family $\mathcal{N}$. Every directed geodesic in $\mathcal{N}$ generates a curve in $T^1\mathcal{N}$, as we saw in Section 3. In an analogous way every horocycle in $\mathcal{N}$ can be lifted to a curve in $T^1\mathcal{N}$. Namely, if $(z, v) \in T^1\mathcal{N}$, then, through z , pass two circles $\tilde{\Gamma}^{(c)}$ and $\tilde{\Gamma}^{(e)}$ tangent to the abscissa and orthogonal to v (see Figure 5).

Figure 5: Contracting and expanding horocycles

We endow these with unit normal vectors directed as v , obtaining $\Gamma^{(c)}$ and $\Gamma^{(e)}$. We will call the curves obtained in $T^1\mathcal{N}$ horocycles as before.

The following properties hold:

- $S_t\Gamma^{(c)}(m) = \Gamma^{(c)}(S_tm)$
 $S_t\Gamma^{(e)}(m) = \Gamma^{(e)}(S_tm)$
- $l(S_t\Gamma^{(c)}(m)) = e^{-t}l(\Gamma^{(c)}(m))$
 $l(S_t\Gamma^{(e)}(m)) = e^tl(\Gamma^{(e)}(m))$

where $\{S_t\}_{t \in \mathbb{R}}$ is the group of traslations along geodesics. So it is natural to call the horocycles $\Gamma^{(c)}$ and $\Gamma^{(e)}$ “contracting” and “expanding” respectively.

We denote by $R_\tau^{(c)}$ the transformation of the space $T^1\mathcal{N}$ consisting of the traslation of every element (m, v) along the contracting horocycle determined by it through a distance τ . The transformation $R_\tau^{(c)}$ gives rise to a one-parameter group of smooth diffeomorphisms of $T^1\mathcal{N}$ and hence that it is generated by some smooth vector field which we denote $a^{(c)}$. The vector field $a^{(e)}$ for the horocycles $\Gamma^{(e)}$ is constructed analogously. We also denote by $a^{(0)}$ the smooth vector field corresponding to the geodesic flow. It is easy to see

that at each point $(m, v) \in T^1\mathcal{N}$ the vector fields $a^{(c)}$, $a^{(e)}$, $a^{(0)}$ are linearly independent among themselves.

It is known [Anosov, 1967] that the most general compact surface Σ of constant negative curvature is obtained as a quotient space of the Lobachevskian plane $H = \{z = (\mu, \sigma) : \sigma > 0\}$ with an hyperbolic fuchsian group Γ that is $\Sigma = H/\Gamma$ endowed with the $PSL(2, \mathbb{R})$ - invariant metric $ds^2 = (d\mu^2 + d\sigma^2)/\sigma^2$. Geometrically the surface Σ can be identified with a fundamental domain Σ_0 of Poincarè disk with “opposite sides” identified modulo Γ , as in Figure 6 below.

Figure 6: Quotient space of the Poincarè disk

If we consider Σ and $T^1\Sigma$ the following holds

Theorem 5.1 (Hedlund-Hopf) [Sinai, 1976]: *The geodesic flow is ergodic that is, if f is a continuous function on $T^1\Sigma$, we have that the limit*

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(S_t x) dt = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(S_{-t} x) dt$$

is constant almost everywhere.

We do not know the statistical interpretation of Γ and Σ , when we identify H with the Normal Family $\mathcal{N}$.

6 Conclusions

We introduce Ergodic Theory in Information Geometry. Precisely we investigate the ergodicity of the geodesic flow on the tangent bundle of the Normal Family, endowed with the Fisher information metric. This method promises interesting applications in Dynamic Econometrics.

References

- [1] Anosov, D. [1967], “Geodesic Flows on Closed Riemannian Manifolds with Negative Curvature”, *Proc. of Steklov. Inst. N.90*.
- [2] Arnold, V.I. and Avez, A. [1967], *Problèmes ergodiques de la mécanique classique*, Gauthier-Villars.
- [3] Billingsley, P. [1965], *Ergodic Theory and Information*, New York , John Wiley and Sons.
- [4] Gallavotti, G. [1982], “Aspetti della Teoria Ergodica Qualitativa e Statistica del moto”, *Quaderno U.M.I., no. 21, Ed. Pitagora Bologna*, pp. 1-300.
- [5] Hendry, D.F. [1997], *Dynamic Econometrics*, Oxford University Press.
- [6] Liverani, C. and Fidaleo, F. [1999], “Ergodic properties for a quantum non linear dynamics” *Journal of Statistical Mechanics*, no. 97, 5/6, pp. 957-1009.
- [7] Medio, A. [1999], “Nonlinear dynamics and caos. Part II: Ergodic approach” *Macroeconomic Dynamics*, 3, Cambridge Univ. Press, pp. 84-114.
- [8] Murray, M.K. and Rice, J.W. [1993], *Differential Geometry and Statistics*, Chapman and Hall.
- [9] O’Neill, B. [1983], *Semi-Riemannian Geometry with Applications to Relativity*, Academic press, INC.
- [10] Sinai, Y. [1976], *Introduction to Ergodic Theory*, Princeton Univ. Press.

Received: June 15, 2005