Statistical properties of deterministic maps

Sébastien Gouëzel

27th November 2002

Abstract

In this note, we explain how the iterates of a typical point by a "good" map behave in a certain sense exactly like a random sequence.

1 Probabilities: law of large number, central limit theorem

1.1 The coin toss

The simplest example of a random event is certainly the coin toss, with a fair coin, that gives head or tail with probability 1/2. We toss the coin an infinite number of times, and at each time n we write $X_n = 0$ if we get head, and $X_n = 1$ if we get tail. Intuitively, we should get almost the same number of heads and tails, i.e. $\frac{1}{n}\sum_{i=0}^{n-1} X_i$ should go to 1/2.

This can be formalized in the following way. We introduce $\Omega = \{0,1\}^{\mathbb{N}}$, i.e. this is all the sequences of possible results of the experiment. The fact that the probability of getting tail at the first toss is 1/2 can be expressed as the fact that the probability of the set $\{(1, x_1, \dots) \mid x_i \in \{0,1\} \forall i \geqslant 1\}$ is 1/2. Thus, the function "probability" should be seen as a function P assigning to a set $A \subset \Omega$ a number $P(A) \in [0,1]$. For technical reasons, P should in fact be a measure.

In our case, to construct P, we use the fact that Ω is almost $[0,1]$. In fact, the application $\pi : \Omega \to [0,1]$, defined by assigning to a sequence the point whose dyadic expansion is this sequence, is surjective, and injective outside of a countable set. Pulling Lebesgue measure from $[0,1]$ to Ω through π, we obtain a measure on Ω which is reasonable: it assigns to the set $\{(x_0, \dots) \mid x_0 = a_0, \dots, x_k = a_k\}$ the measure $1/2^{k+1}$.

The observation X_n at time n can now be seen as a map $X_n : \Omega \to \{0,1\}$ that assigns to a sequence its n^{th} component. In this setting, we have the strong law of large numbers:

Theorem 1.1. *For almost all $\omega \in \Omega$,*

$$\frac{1}{n}\sum_{k=0}^{n-1} X_k(\omega) \to \frac{1}{2} \quad \textit{when } n \to \infty.$$

1.2 General probabilities

In this section, Ω will be a measurable space, endowed with a measure P of mass 1, called a probability measure. We consider functions $X_0, X_1, \dots$ from Ω to $\mathbb{R}$, that will correspond to the result at time n of the coin toss. They are called "random variables".

We want these random variables to describe the same experiment, at different times. For $A \subset \mathbb{R}$, we can consider $P(\{\omega \in \Omega \mid X_0(\omega) \in A\})$, also written $P(X_0 \in A)$. These numbers, for $A \subset \mathbb{R}$, describe the repartition of the random variable X_0 – they are called the *law* of X_0. This law can also be seen as a measure on $\mathbb{R}$, it is just $(X_0)_*(P)$. The random variables $X_0, X_1, \dots$ are *identically distributed* if their laws are equal.

We also want these random variables to be independent. For example, we do not want to have $X_0 = X_1 = X_2 = \dots$, which would mean that there is no randomness at all. The independence condition is the following: if $i \neq j$, $\forall A, B \subset \mathbb{R}$, $P(X_i \in A, X_j \in B) = P(X_i \in A)P(X_j \in B)$.

In this setting, the law of large numbers is:

Theorem 1.2. *Let $X_0, X_1, \dots$ be a sequence of independent identically distributed (i.i.d.) random variables on a probability space Ω. Assume that $X_0 \in L^1(P)$. Then, for almost all $\omega \in \Omega$,*

$$\frac{1}{n}\sum_{k=0}^{n-1} X_k(\omega) \to E(X_0)$$

where $E(X_0)$, the expectation of X_0, is simply $\int X_0(\omega)\,\mathrm{d}P(\omega)$.

1.3 The central limit theorem

We denote by $N = \mathcal{N}(0, \sigma^2)$ the law on $\mathbb{R}$ given by $N(-\infty, x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-t^2/2\sigma^2}\,\mathrm{d}t$. It is called the normal law (or Gaussian law) of 0 mean and σ^2 variance (the variance of the random variable X is $E(X^2) - E(X)^2$).

When $E(X_0) = 0$, the law of large numbers implies that $\frac{1}{n}\sum_{k=0}^{n-1} X_k$ goes almost surely to 0. Under stronger hypotheses on X_0 (namely $X_0 \in L^2$), it is possible to enhance this result and to get further informations on the convergence: this is the following central limit theorem.

Theorem 1.3. *If $X_0, X_1, \ldots$ are i.i.d. on Ω with $E(X_0) = 0$ and $E(X_0^2) = \sigma^2 \in (0, \infty)$, then*

$$\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} X_k \to \mathcal{N}(0, \sigma^2)$$

where the convergence is in distribution.

This means that $P(\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} X_k \in (-\infty, x)) \to \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-t^2/2\sigma^2} \, \mathrm{d}t$.

The surprising feature of this result is that the limit law is independent of the law of X_0: it is a universal law. This is why the normal law is so important in physics. For example, the error in a physical measurement can be seen as a sum of very small errors. Even though the law of these very small errors is unknown, the central limit theorem implies that the global error will be normal.

2 Iteration of maps

Let K be a space endowed with a measure μ of mass 1, and T an application from K to itself. If $f : K \to \mathbb{R}$, then the functions $X_i = f \circ T^i$ can be seen as random variables on the space K, but they are of course not independent, nor identically distributed. Nevertheless, it is sometimes possible to prove analogues of the strong law of large numbers or of the central limit theorem for these functions.

2.1 Birkhoff's Theorem

We would like to obtain a law of large numbers, that is $\frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k(x) \to \int f \, \mathrm{d}\mu$ for μ-almost every x, and for all function $f \in L^1$.

Changing variables in the sum and letting n go to infinity proves that, necessarily, $\int f \circ T = \int f$. This means that $\forall A$, $\mu(T^{-1}A) = \mu(A)$: we say that the measure μ is *invariant* by T.

Moreover, if $T(A) = A$, then either $\mu(A) = 0$, or $\mu(A) > 0$. In the latter case, for $f = 1_A$, there exists $x \in A$ such that $\frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k(x) \to \int f \, \mathrm{d}\mu = \mu(A)$. But, on the left-hand side, every term is equal to 1, whence $\mu(A) = 1$. We have proved that an invariant set necessarily has measure 0 or 1. Such an application T is called *ergodic*.

Invariance and ergodicity are necessary conditions to obtain a law of large numbers. Surprisingly, they are also sufficient:

Theorem 2.1 (Birkhoff's Theorem, 1934). *Let T be a measure-preserving, ergodic* *map of a probability space (K,μ). If $f : K \to \mathbb{R}$ is integrable, then for μ-almost every x,*

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) \to \int f \,\mathrm{d}\mu.$$

Taking $f = 1_A$, we get that the proportion of time spent in A by the iterates of x is equal to $\mu(A)$. Thus, the iterates of x are very well equi-distributed in the space K, as if they were points picked at random.

For example, take $K = S^1 \subset \mathbb{C}$ and $T : z \mapsto z^2$. It preserves Lebesgue measure, since the preimage of a set A is two copies of A, but twice smaller. Moreover, it is ergodic: assume A is invariant and $\mathrm{Leb}(A) > 0$. Then A has points of density, i.e. there exists x such that $\frac{\mathrm{Leb}(A\cap[x-\varepsilon,x+\varepsilon])}{\mathrm{Leb}([x-\varepsilon,x+\varepsilon])} \to 1$ when $\varepsilon \to 0$. In particular, for any small δ, there is an interval of size $1/2^n$ such that the proportion of this interval covered by A is at least $1-\delta$. Iterating n times, and since A is invariant, we get that A covers a proportion at least $1-\delta$ of S^1, i.e. $\mu(A) \geqslant 1-\delta$. Thus, $\mu(A) = 1$. Birkhoff's Theorem gives then a law of large numbers for the iterates of almost all points of S^1.

Another example is given by the irrational rotations of the circle, i.e. rotations by an irrational angle. They preserve Lebesgue measure and, using Fourier series, it is not hard to show that they are ergodic.

However, this example seems to exhibit less randomness than the previous one: points starting at a distance d will remain at a distance d after any number of iterations, while the map $z \mapsto z^2$ mixes everything. While the law of large numbers is true in both examples, the central limit theorem will hold only in the latter.

2.2 Central limit theorem

The analogue of the central limit theorem in our dynamical setting is the fact that $\mu\left\{x \mid \frac{1}{\sqrt{n}}\sum_{i=0}^{n-1} f(T^i x) \in (a,b)\right\} \to N(a,b)$ for any interval (a,b), where $N = \mathcal{N}(0,\sigma^2)$ is a normal law for some parameter $\sigma^2 \geqslant 0$.

This kind of result is much harder to prove than laws of large numbers: while Birkhoff's Theorem dates back to the 30's, the first deterministic central limit theorems were proved in the 80's! The required hypotheses are also much stronger, some kind of expansion is needed. For example, a rotation on the circle never satisfies the central limit theorem.

Theorem 2.2. *For $T : z \mapsto z^2$ from S^1 to S^1, the central limit theorem is satisfied for any C^1-function f with $\int f = 0$.*

The variance σ^2 is then $\int f^2 + 2\sum_{n\geqslant 1}\int f \cdot f\circ T^n$, and an important part of the proof is to prove that this sum is indeed finite.

The central limit theorem is in fact valid in a wider setting, that we precise in the following theorem.

Theorem 2.3. *Let* $T : S^1 \to S^1$ *be* C^2 *with* $T' \geqslant \lambda > 1$*. Then there exists a function* $\varphi \in L^1(\mathrm{dLeb})$ *such that* $\mathrm{d}\mu = \varphi \, \mathrm{dLeb}$ *is an invariant probability.*

In fact, $\mathrm{d}\mu$ *is ergodic, whence* T *satisfies a law of large numbers for any* $f \in L^1(\mathrm{d}\mu)$*.*

Moreover, any C^1 f *with* $\int f \, \mathrm{d}\mu = 0$ *also satisfies a central limit theorem, with* $\sigma^2 = \int f^2 \, \mathrm{d}\mu + 2 \sum_{n=1}^{\infty} \int f \cdot f \circ T^n \, \mathrm{d}\mu$*.*

There are mainly two methods of proof for the central limit theorem: the first one is to prove first that $\mathrm{Cor}(f, f \circ T^n) := \int f \cdot f \circ T^n$ is summable, and then to use this information to construct a *reverse martingale*: we bring back independence and use probabilistic results.

The second method is a spectral method: instead of iterating forward, we iterate backward, which corresponds to looking at an operator, called the *transfer operator*. This operator has a regularizing effect, and nice spectral properties. The study of perturbations of this operator gives the central limit theorem.

References

[Fel67] William Feller. *An Introduction to Probability Theory and its Applications, volume 1*. Wiley Series in Probability and Mathematical Statistics. J. Wiley, 1967.

[Via97] Marcelo Viana. Stochastic dynamics of deterministic systems. Book on his web page, 1997.

[Wal82] Peter Walters. *An introduction to ergodic theory*, volume 79 of *Graduate texts in mathematics*. Springer, 1982.