

# Central Limit Theorem for dynamical systems using martingales

Vaughn Climenhaga

May 18, 2013

This post is based on notes from Matt Nicol's talk at the [UH summer school in dynamical systems](#). The goal is to present the ideas behind a proof of the central limit theorem for dynamical systems using martingale approximations.

## 1 Conditional expectation

Before we can define and use martingales, we must recall the definition of conditional expectation. Let  $(\Omega, \mathbb{P})$  be a probability space, with  $\mathbb{P}$  defined on a  $\sigma$ -algebra  $\mathcal{B}$ . Let  $\mathcal{F} \subset \mathcal{B}$  be a sub- $\sigma$ -algebra of  $\mathcal{B}$ .

**Example 1** Consider the doubling map  $T: [0, 1] \rightarrow [0, 1]$  given by  $T(x) = 2x \pmod{1}$ . Let  $\mathbb{P}$  be Lebesgue measure,  $\mathcal{B}$  the Borel  $\sigma$ -algebra, and  $\mathcal{F} = T^{-1}\mathcal{B} = \{T^{-1}(B) \mid B \in \mathcal{B}\}$ . Then  $\mathcal{F}$  is a sub- $\sigma$ -algebra of  $\mathcal{B}$ , consisting of precisely those sets in  $\mathcal{B}$  which are unions of preimage sets  $T^{-1}(x)$  – that is, those sets  $F \in \mathcal{B}$  for which a point  $y \in [0, \frac{1}{2}]$  is in  $F$  if and only if  $y + \frac{1}{2} \in F$ .

This example extends naturally to yield a decreasing sequence of  $\sigma$ -algebras

$$\mathcal{B} \supset T^{-1}\mathcal{B} \supset T^{-2}\mathcal{B} \supset \dots$$

Given a sub- $\sigma$ -algebra  $\mathcal{F} \subset \mathcal{B}$  and a random variable  $Y: \Omega \rightarrow \mathbb{R}$  that is measurable with respect to  $\mathcal{B}$ , the *conditional expectation* of  $Y$  given  $\mathcal{F}$  is any random variable  $Z$  such that

1.  $Z$  is  $\mathcal{F}$ -measurable (that is,  $Z^{-1}(I) \in \mathcal{F}$  for every interval  $I \subset \mathbb{R}$ ), and

$$2. \int_A Z d\mathbb{P} = \int_A Y d\mathbb{P} \text{ for every } A \in \mathcal{F}.$$

It is not hard to show that these conditions characterise  $Z$  for an almost-sure choice of  $\omega \in \Omega$ , and so the conditional expectation is uniquely defined as a random variable. We write it as  $\mathbb{E}[Y | \mathcal{F}]$ .

A key property is that conditional expectation is linear: for every  $\mathcal{F} \subset \mathcal{B}$ , every  $Y_1, Y_2: \Omega \rightarrow \mathbb{R}$ , and every  $a \in \mathbb{R}$ , we have

$$\mathbb{E}[aY_1 + Y_2 | \mathcal{F}] = a\mathbb{E}[Y_1 | \mathcal{F}] + \mathbb{E}[Y_2 | \mathcal{F}].$$

**Example 2** *If  $Y$  is already  $\mathcal{F}$ -measurable, then  $\mathbb{E}[Y | \mathcal{F}] = Y$ .*

**Example 3** *At the other extreme, if  $Y$  and  $\mathcal{F}$  are independent – that is, if  $\mathbb{E}[Y\mathbf{1}_A] = \mathbb{E}[Y]\mathbb{P}[A]$  for every  $A \in \mathcal{F}$  – then  $\mathbb{E}[Y | \mathcal{F}]$  is the constant function  $\mathbb{E}[Y]\mathbf{1}$ .*

**Example 4** *Suppose  $\{\Omega_i | i \in \mathbb{N}\}$  is a countable partition of  $\Omega$  such that  $\mathbb{P}[\Omega_i] > 0$  for every  $i$ . Let  $\mathcal{F} = \sigma(\Omega_1, \Omega_2, \dots)$  be the  $\sigma$ -algebra generated by the sets  $\Omega_i$ . Then*

$$\mathbb{E}[Y|\mathcal{F}] = \sum_{i \in \mathbb{N}} \frac{\mathbb{E}[Y\mathbf{1}_{\Omega_i}]}{\mathbb{P}[\Omega_i]} \mathbf{1}_{\Omega_i}.$$

## 2 Martingales

Now we can define martingales, which are a particular sort of stochastic process (sequence of random variables) with “enough independence” to generalise results from the IID case.

**Definition 1** *A sequence  $Y_n: \Omega \rightarrow \mathbb{R}$  of random variables is a martingale if*

1.  $\mathbb{E}[|Y_n|] < \infty$  for all  $n$ ;
2. *there is an increasing sequence of  $\sigma$ -algebras (a filtration)  $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \subset \mathcal{B}$  such that  $Y_n$  is measurable with respect to  $\mathcal{F}_n$ ;*
3. *the conditional expectations satisfy  $\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = Y_n$ .*

The first condition guarantees that everything is in  $L^1$ . If  $\mathcal{F}_n$  is taken to be the  $\sigma$ -algebra of events that are determined by the first  $n$  outcomes of a sequence of experiments, then the second condition states that  $Y_n$  only depends on those first  $n$  outcomes, while the third condition requires that if the first  $n$  outcomes are known, then the expected value of  $Y_{n+1} - Y_n$  is 0.

**Example 5** Let  $Y_n$  be a sequence of fair coin flips – IID random variables taking the values  $\pm 1$  with equal probability. Let  $S_n = Y_1 + \dots + Y_n$ . As suggested in the previous paragraph, let  $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$  be the smallest  $\sigma$ -algebra with respect to which  $Y_1, \dots, Y_n$  are all measurable. (The sets in  $\mathcal{F}_n$  are precisely those sets in  $\mathcal{B}$  which are determined by knowing the values of  $Y_1, \dots, Y_n$ .)

It is easy to see that  $S_n$  satisfies the first two properties of a martingale, and for the third, we use linearity of expectation and the definition of  $\mathcal{F}_n$  to get

$$\mathbb{E}[S_{n+1}|\mathcal{F}_n] = \mathbb{E}[Y_1 + \dots + Y_n|\mathcal{F}_n] + \mathbb{E}[Y_{n+1}|\mathcal{F}_n] = S_n + 0 = S_n.$$

When  $Y_n$  is a sequence of random variables for which  $S_n = Y_1 + \dots + Y_n$  is a martingale, we say that the sequence  $Y_n$  is a *martingale difference*.

In the previous example the martingale property (the third condition) was a direct consequence of the fact that the random variables  $Y_n = S_{n+1} - S_n$  were IID. However, there are examples where the martingale differences are not IID.

**Example 6** *Polya's urn* is a stochastic process defined as follows. Consider an urn containing some number of red and blue balls. At each step, a single ball is drawn at random from the urn, and then returned to the urn, along with a new ball that matches the colour of the one drawn. Let  $Y_n$  be the fraction of the balls that are red after the  $n$ th iteration of this process.

Clearly the sequence of random variables  $Y_n$  is neither independent nor identically distributed. However, it is a martingale, as the following computation shows: suppose that at time  $n$  there are  $p$  red balls and  $q$  blue balls in the urn. (This knowledge represents knowing which element of  $\mathcal{F}_n$  we are in.) Then at time  $n + 1$ , there will be  $p + 1$  red balls with probability  $\frac{p}{p+q}$ , and  $p$  red balls with probability  $\frac{q}{p+q}$ . Either way, there will be  $p + q + 1$  total balls, and so the expected fraction of red balls is

$$\begin{aligned} \mathbb{E}[Y_{n+1}|\mathcal{F}_n] &= \frac{p}{p+q} \cdot \frac{p+1}{p+q+1} + \frac{q}{p+q} \cdot \frac{p}{p+q+1} \\ &= \frac{p(p+q+1)}{(p+q)(p+q+1)} = \frac{p}{p+q} = Y_n. \end{aligned}$$

If we assume that the martingale differences are stationary (that is, identically distributed) and ergodic, then we have the following central limit theorem for martingales, from a [1974 paper of McLeish](#) (we follow some [notes by S. Sethuraman](#) for the statement).

**Theorem 2** *Let  $Y_i$  be a stationary ergodic sequence such that  $\sigma^2 = \mathbb{E}[Y_i^2] < \infty$  and  $\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = 0$ , where  $\mathcal{F}_n$  is the  $\sigma$ -algebra generated by  $\{Y_i \mid 1 \leq i \leq n\}$ . Then  $S_n = \sum_{i=1}^n Y_i$  is a martingale, and  $\frac{S_n}{\sigma\sqrt{n}}$  converges in distribution to  $N(0, 1)$ .*

More sophisticated versions of this result are available, but this simple version will suffice for our needs.

### 3 Koopman operator and transfer operator

Now we want to apply 2 to a dynamical system  $T: X \rightarrow X$  with an ergodic measure  $\mu$  by taking  $Y_i = \varphi \circ T^i$  for some observable  $\varphi: X \rightarrow \mathbb{R}$ .

To carry this out, we consider two operators on  $L^2(\mu)$ . First we consider the Koopman operator  $U: \varphi \mapsto \varphi \circ T$ . Then we define the transfer operator  $\mathcal{P}$  to be its  $L^2$  adjoint – that is,

$$\int (\mathcal{P}\varphi)\psi \, d\mu = \int \varphi(\psi \circ T) \, d\mu \tag{1}$$

for all  $\varphi, \psi \in L^2$ . The key result for our purposes is that the operators  $\mathcal{P}$  and  $U$  are one-sided inverses of each other.

**Proposition 3** *Given  $\varphi \in L^2$ , we have*

1.  $\mathcal{P}U\varphi = \varphi$ ;
2.  $U\mathcal{P}\varphi = \mathbb{E}[\varphi|T^{-1}\mathcal{B}]$ , where  $\mathcal{B}$  is the  $\sigma$ -algebra on which  $\mu$  is defined.

PROOF: For the first claim, we see that for all  $\psi \in L^2$  we have

$$\int \psi \cdot (\mathcal{P}U\varphi) \, d\mu = \int (\psi \circ T)(\varphi \circ T) \, d\mu = \int \psi\varphi \, d\mu,$$

where the first equality uses the definition of  $\mathcal{P}$  and the second uses the fact that  $\mu$  is invariant. To prove the second claim, we first observe that given an interval  $I \subset \mathbb{R}$ , we have

$$(\mathcal{P}U\varphi)^{-1}(I) = T^{-1}((\mathcal{P}\varphi)^{-1}(I)) \in T^{-1}\mathcal{B},$$

since  $\mathcal{P}$  maps  $\mathcal{B}$ -measurable functions to  $\mathcal{B}$ -measurable functions. This shows that  $\mathcal{P}U\varphi$  is  $T^{-1}\mathcal{B}$ -measurable, and it remains to show that

$$\int_{T^{-1}B} U\mathcal{P}\varphi d\mu = \int_{T^{-1}B} \varphi d\mu \text{ for all } B \in \mathcal{B}. \quad (2)$$

This follows from a similar computation to the one above: given  $B \in \mathcal{B}$  we have

$$\begin{aligned} \int_{T^{-1}B} U\mathcal{P}\varphi d\mu &= \int ((\mathcal{P}\varphi) \circ T) \cdot \mathbf{1}_{T^{-1}B} d\mu = \int ((\mathcal{P}\varphi) \circ T)(\mathbf{1}_B \circ T) d\mu \\ &= \int (\mathcal{P}\varphi)\mathbf{1}_B d\mu = \int \varphi \cdot (\mathbf{1}_B \circ T) d\mu = \int_{T^{-1}B} \varphi d\mu, \end{aligned}$$

which establishes (2) and completes the proof.  $\square$

We see from Proposition 3 that a function has zero conditional expectation with respect to  $T^{-1}\mathcal{B}$  if and only if it is in the kernel of  $\mathcal{P}$ . In particular, if  $\mathcal{P}h = 0$  then  $\sum_{j=1}^n h \circ T^j$  is a martingale; this will be a key tool in the next section.

**Example 7** Let  $X = [0, 1]$  and  $T$  be the doubling map. Let  $\mu$  be Lebesgue measure. For convenience of notation we consider the  $L^2$  space of complex-valued functions on  $X$ ; the functions  $\varphi_n: x \mapsto e^{2\pi inx}$  form an orthonormal basis for this space. A simple calculation shows that

$$U\varphi_n(x) = \varphi_n(2x) = e^{2\pi in(2x)} = \varphi_{2n}(x),$$

so  $U: \varphi_n \rightarrow \varphi_{2n}$ . For the transfer operator we obtain  $\mathcal{P}: \varphi_{2n} \rightarrow \varphi_n$ , while for odd values of  $n$  we have

$$\begin{aligned} \mathcal{P}\varphi_n(x) &= \frac{1}{2} \left( \varphi_n\left(\frac{x}{2}\right) + \varphi_n\left(\frac{1+x}{2}\right) \right) \\ &= \frac{1}{2} (e^{\pi inx} + e^{\pi inx} e^{\pi in}) = 0. \end{aligned}$$

## 4 Martingale approximation and CLT

The machinery of the Koopman and transfer operators from the previous section can be used to apply the martingale central limit theorem to observations of dynamical systems via the technique of martingale approximation, which was introduced by [M. Gordin in 1969](#).

The idea is that if  $\mathcal{P}^n \varphi \rightarrow 0$  quickly enough for functions  $\varphi$  with  $\int \varphi d\mu = 0$ , then we can approximate the sequence  $\varphi \circ T^n$  with a martingale sequence  $h \circ T^n$ .

More precisely, suppose that the sequence  $\|\mathcal{P}^n \varphi\|_2$  is summable; then we can define an  $L^2$  function  $g$  by

$$g = \sum_{n=1}^{\infty} \mathcal{P}^n \varphi. \quad (3)$$

Let  $h = \varphi + g - g \circ T \in L^2$ . We claim that  $S_n h = \sum_{j=1}^n h \circ T^j$  is a martingale. Indeed,

$$\mathcal{P}h = \mathcal{P}\varphi + \mathcal{P} \sum_{n \geq 1} \mathcal{P}^n \varphi - \mathcal{P}U \sum_{n \geq 1} \mathcal{P}^n \varphi,$$

and since  $\mathcal{P}U$  is the identity we see that the last term is just  $\sum_n \mathcal{P}^n \varphi$ , so that  $\mathcal{P}h = 0$ .

Proposition 3 now implies that  $\mathbb{E}[h|T^{-1}\mathcal{B}] = 0$ , and we conclude that  $S_n h$  is a martingale, so by the martingale CLT  $\frac{S_n h}{\sigma\sqrt{n}}$  converges in distribution to  $N(0, 1)$ , where  $\sigma^2 = \mathbb{E}[h^2] = \int h^2 d\mu$ .

Now we want to apply this result to obtain information about  $\varphi$  itself, and in particular about  $S_n \varphi = \sum_{j=1}^n \varphi \circ T^j$ . We have  $\varphi = h + g \circ T - g$ , and so

$$S_n \varphi = S_n h + \sum_{j=1}^n (g \circ T^{j+1} - g \circ T^j) = S_n h + g \circ T^{n+1} - g.$$

This yields

$$\frac{S_n \varphi}{\sigma\sqrt{n}} = \frac{S_n h}{\sigma\sqrt{n}} + \frac{g \circ T^{n+1} - g}{\sigma\sqrt{n}},$$

and the last term goes to 0 in probability, which yields the central limit theorem for  $S_n \varphi$ .

**Remark 4** *There is a technical problem we have glossed over, which is that the sequence of  $\sigma$ -algebras  $T^{-n}\mathcal{B}$  is decreasing, not increasing as is required by the definition of a martingale. One solution to this is to pass to the natural extension  $\hat{T}$  and to consider the functions  $h \circ \hat{T}^{-j}$  and the  $\sigma$ -algebras  $\hat{\mathcal{F}}_j = \hat{T}^j \mathcal{B}$ . Another solution is to use reverse martingales, but we do not discuss this here.*

**Example 8** Let  $X = [0, 1]$  and  $T: X \rightarrow X$  be an intermittent type (Manneville–Pomeau) map given by

$$T(x) = \begin{cases} x(1 - (2x)^\alpha) & x \in [0, \frac{1}{2}), \\ 2x - 1 & x \in [\frac{1}{2}, 1], \end{cases}$$

where  $0 < \alpha < 1$  is a fixed parameter. It can be shown that  $T$  has a unique absolutely continuous invariant probability measure  $\mu$ , and that the transfer operator  $\mathcal{P}$  has the following contraction property: for every  $\varphi \in L^2$  with  $\int \varphi d\mu = 0$ , there is  $C \in \mathbb{R}$  such that  $\|\mathcal{P}^n \varphi\|_2 \leq Cn^{-\gamma}$ , where  $\gamma = \frac{1}{2}(\frac{1}{\alpha} - 1)$ .

For small values of  $\alpha$ , this shows that  $\|\mathcal{P}^n \varphi\|_2$  is summable, and consequently  $\mu$  satisfies the CLT by the above discussion.