Spectral methods – central limit theorem

Vaughn Climenhaga

March 7, 2013

    With the previous post on convergence of random variables, the law of large numbers, and Birkhoff's ergodic theorem as background, we return to the spectral methods discussed in the first two posts in this series. This post is based on Andrew Török's talk from March 4 and gives a proof of the central limit theorem using the spectral gap property.

1   Central limit theorem for IID

Now we recall the statement of the central limit theorem (CLT) and give a proof in the case of IID (independent identically distributed) random variables.

    The weak law of large numbers says that if X_n is a sequence of IID random variables with $\mathbb{E}[X_n] = 0$, then writing $S_n = \sum_{k=0}^{n-1} X_k$, the time averages $\frac{1}{n} S_n$ converge to 0 in probability, or equivalently (since the limit is a constant), in distribution. In the case when $\sigma^2 = \mathbb{E}[X^2] < \infty$, the central limit theorem strengthens this to the result that the sequence $\frac{1}{\sqrt{n}} S_n$ converges in distribution to $N(0, \sigma^2)$, the normal distribution with mean 0 and variance σ^2. That is, we have

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} S_n \leq c\right) \xrightarrow{n \to \infty} \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{c} e^{-t^2/2\sigma^2} \, dt \tag{1}$$

for every $c \in \mathbb{R}$.

    This can be established by the same method as we used last time for the proof of the weak law of large numbers, by studying the characteristic functions of $\frac{1}{\sqrt{n}} S_n$ and $N(0, \sigma^2)$. The characteristic function of $N(0, \sigma^2)$ is

$$\psi(t) = e^{-\frac{1}{2}\sigma^2 t^2}. \tag{2}$$

Arguing as in the proof of the weak law of large numbers in the previous post, we write φ_n for the characteristic function of $\frac{1}{\sqrt{n}}S_n$ and observe that

$$\varphi_n(t) = \mathbb{E}[e^{it\frac{1}{\sqrt{n}}(X_1+\cdots+X_n)}] = \prod_{j=1}^{n}\mathbb{E}[e^{\frac{it}{\sqrt{n}}X_j}] = \varphi\left(\frac{t}{\sqrt{n}}\right)^n, \qquad (3)$$

where φ is the characteristic function of the X_j (which are identically distributed), and the second equality uses the fact that the X_j are independent.

Now by Taylor's theorem, we have

$$\varphi(t/\sqrt{n}) = \mathbb{E}[e^{\frac{it}{\sqrt{n}}X_j}] = 1 + \frac{it}{\sqrt{n}}\mathbb{E}[X_j] - \frac{t^2}{2n}\mathbb{E}[X_j^2] + o(t^2)$$

$$= 1 - \frac{t^2}{2n}\sigma^2 + o(t^2),$$

using the fact that the X_j have mean 0 and variance σ^2. Thus we conclude from (3) that

$$\varphi_n(t) = \left(1 - \frac{t^2\sigma^2}{2n} + o(t^2)\right)^n \xrightarrow{n\to\infty} e^{-\frac{1}{2}t^2\sigma^2} = \psi(t),$$

which completes the proof of the CLT in the IID case.

2 CLT with spectral gap

To translate the CLT into the language of dynamical systems, we consider a space X and a map $T\colon X \to X$ with an invariant measure μ. In general there may be many T-invariant measures, and so it is important to choose a suitable measure μ. For example, when X is an interval and f is piecewise expanding, we are most interested in the case when μ is an acip.

Given a measurable function $f\colon X \to \mathbb{R}$, the sequence of functions f, $f\circ T$, $f\circ T^2, \ldots$ defines a sequence of identically distributed random variables on X. However, they are not independent, and so we need some information about the decay of correlations between them. In particular, we can replicate the proof from the previous section as long as the transfer operator has a spectral gap.

Let's make this precise in the case when T is a piecewise expanding interval map, so the Lasota–Yorke inequality we discussed in an earlier post

yields a spectral gap for the transfer operator $\mathcal{P}_T$ acting on BV, the space of functions of bounded variation, and in particular establishes the existence of an acip μ.

Theorem 1 *Let T be a piecewise expanding interval map and μ the acip constructed before. Suppose that μ is mixing. Then μ satisfies the central limit theorem as follows: given any $f \in BV$ with $\int f \, d\mu = 0$ and writing $S_n f(x) = \sum_{k=0}^{n-1} f \circ T^k$, we have*

$$\mu \left\{ x \mid \frac{1}{\sqrt{n}} S_n f(x) \leq c \right\} \xrightarrow{n \to \infty} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{c} e^{-x^2/2\sigma^2} \, dx \qquad (4)$$

for all $c \in \mathbb{R}$, where σ is given by the Green–Kubo formula

$$\sigma^2 = \sum_{n \in \mathbb{Z}} \int_X f \cdot (f \circ T^n) \, d\mu, \qquad (5)$$

and $\sigma = 0$ if and only if there exists $g \in BV$ and $c \in \mathbb{R}$ such that $f = c + g \circ T - g$.

Before proving the theorem, we make some remarks concerning the Green–Kubo formula (5). First, note that the sum converges as soon as we establish exponential decay of correlations for functions in BV, since each integral in the sum is just the correlation function at time n. Second, note that if we replace the functions $f \circ T^n$ with independent random variables, then all the terms with $n \neq 0$ vanish, and the $n = 0$ term is just the variance $\mathbb{E}[X^2]$, as in the previous section.

Note also that using (5), σ^2 can be written as

$$\sigma^2 = \lim_{n \to \infty} \frac{1}{n} \int (S_n f)^2 \, d\mu.$$

Now we prove the central limit theorem (4). As in the IID case, we use the characteristic functions

$$\psi(t) = e^{-\sigma^2 t^2/2},$$

$$\varphi_n(t) = \mathbb{E}_\mu[e^{it(S_n f)/\sqrt{n}}] = \int e^{\frac{it}{\sqrt{n}} S_n f} \, d\mu,$$

where $\psi(t)$ is the characteristic function of the normal distribution and φ_n is the characteristic function of $\frac{1}{\sqrt{n}} S_n f$, so it suffices to show that $\varphi_n(t) \to \psi(t)$ for all t.

3

To prove this convergence of the characteristic functions, we use the following procedure.

1. Write the characteristic functions φ_n in terms of a *twisted transfer operator* $\mathcal{P}_{f,t}$, where f is the function we are investigating in the CLT, and $t \approx 0$ is a small real parameter. The operator $\mathcal{P}_{f,t}$ is a small perturbation of the transfer operator $\mathcal{P}_T$.

2. Use perturbation theory of operators to show that $\mathcal{P}_{f,t}$ has a spectral gap and to derive asymptotics for the leading eigenvalue $\lambda(t)$. In particular, relate $\lambda'(0)$ and $\lambda''(0)$ to the mean and variance of the limiting distribution.

First we define the transfer operator itself by the implicit equation

$$\int (\mathcal{P}_T g) \cdot h \, d\mu = \int g \cdot (h \circ T) \, d\mu \qquad (6)$$

for all $g \in L^1(\mu)$ and $h \in L^\infty$. Note that this is different from the transfer operator defined by integrating with respect to Lebesgue measure in (6) – it is a worthwhile exercise to determine the precise relationship between the two.

More directly, the transfer operator can be defined by

$$\mathcal{P}_T g(x) = \sum_{y \in T^{-1}(x)} \frac{g(y) h(y)}{|T'(y)|}, \qquad (7)$$

where h is the density of μ with respect to Lebesgue measure.

Now given $f \in BV$ and $t \in \mathbb{R}$, we define the *twisted* transfer operator by

$$\mathcal{P}_{f,t} g = \mathcal{P}_T(e^{itf} g). \qquad (8)$$

To see the utility of this definition, we first note that

$$\int \mathcal{P}_{f,t}(g) \, d\mu = \int \mathcal{P}_T(e^{itf} g) \mathbf{1} \, d\mu = \int e^{itf} g \, d\mu,$$

and so by induction we have

$$\int \mathcal{P}_{f,t}^n(g) \, d\mu = \int e^{itS_n f} g \, d\mu.$$

4

In particular, considering the characteristic function φ_n, we have

$$\varphi_n(t) = \int e^{\frac{it}{\sqrt{n}} S_n f} \, d\mu = \int \mathcal{P}^n_{f, \frac{t}{\sqrt{n}}}(\mathbf{1}) \, d\mu, \tag{9}$$

which accomplishes the first stage of the proof – writing the characteristic function in terms of the twisted transfer operator.

For the second stage of the proof, we consider the twisted transfer operator as a perturbation of $\mathcal{P}_T$. From the Lasota–Yorke inequality and the fact that μ is mixing, we know that the spectrum of $\mathcal{P}_T$ has the form $\{1\} \cup Z$, where Z is contained in a disc of radius $r < 1$ centred at the origin.

By the perturbation theory of linear operators, the spectrum of $\mathcal{P}_{f,t}$ has the same form for small enough $|t|$: there is a leading eigenvalue $\lambda(t)$ that is close to 1, and the rest of the spectrum is contained in the disc of radius r. Moreover, the leading eigenvalue satisfies *(Edit: see the end of the post for a proof)*

$$\lambda'(0) = \int (if) \, d\mu = 0$$

and

$$\lambda''(0) = \lim_{n \to \infty} \frac{1}{n} \int (S_n(if))^2 \, d\mu = -\sigma^2,$$

which is the origin of the expression in the Green–Kubo formula.

Now we use the Riesz functional calculus, whose general ideas we briefly recall here. Let X be a Banach space and $\mathcal{B}(X)$ the space of bounded linear operators on X. Given $S \in \mathcal{B}(X)$, let $\sigma \subset \mathbb{C}$ be the spectrum of S. Then there is a unique way to associate to each analytic function $g \colon \sigma \to \mathbb{C}$ an operator $g(S)$ such that the map $g \mapsto g(S)$ is a homomorphism mapping the constant function to the identity operator and the identity function to S.

This mapping can be defined by integrating around a curve γ surrounding the spectrum σ (this is similar to the Cauchy formula from complex analysis):

$$g(S) = \frac{1}{2\pi i} \int_\gamma (S - zI)^{-1} g(z) \, d\gamma,$$

where we recall that $S - zI$ is invertible for all z in the resolvent $\mathbb{C} \setminus \sigma$. If we take g to be the characteristic function of part of the spectrum, we obtain a projection to the eigenspace associated with that part.

In particular, considering the operator $\mathcal{P}_{f,t}$, we may set $g = \mathbf{1}_{\lambda(t)}$ and obtain a projection Π_t onto the eigenspace of $\lambda(t)$. Similarly, setting $g(z) =$

5

$z\mathbf{1}_{Z(t)}(z)$, where $Z(t)$ is the part of the spectrum contained in a disc of radius $r < 1$, we get an operator R_t such that

$$\Pi_t R_t = R_t \Pi_t = 0, \qquad \|R_t\|_{BV} < r.$$

Moreover, we have

$$\mathcal{P}_{f,t} = \lambda(t)\Pi_t + R_t,$$

which allows us to write the operator in (9) as

$$
\begin{aligned}
\mathcal{P}^n_{f,\frac{t}{\sqrt{n}}} &= \lambda\left(\frac{t}{\sqrt{n}}\right)^n \Pi_{\frac{t}{\sqrt{n}}} + R^n_{\frac{t}{\sqrt{n}}} \\
&= \left(1 + \lambda'(0)\frac{t}{\sqrt{n}} + \frac{\lambda''(0)}{2}\frac{t^2}{n} + o\left(\frac{t^2}{n}\right)\right)^n \Pi_{\frac{t}{\sqrt{n}}} + R^n_{\frac{t}{\sqrt{n}}} \\
&= \left(1 - \frac{\sigma^2 t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \left(\Pi_0 + O\left(\frac{t}{\sqrt{n}}\right)\right) + R^n_{\frac{t}{\sqrt{n}}} \\
&\xrightarrow{n\to\infty} e^{-t^2\sigma^2/2}\Pi_0,
\end{aligned}
$$

using the fact that $\|R_t\|_{BV} < r < 1$. Now (9) yields

$$\varphi_n(t) \to e^{-t^2\sigma^2/2} \int \Pi_0(\mathbf{1})\, d\mu = e^{-t^2\sigma^2/2} = \psi(t),$$

which completes the proof of the CLT.

3 Proof of formulas for derivatives of λ

The formulas given above for $\lambda'(0)$ and $\lambda''(0)$ were not explained. Here is a derivation of these formulas.

Let g_t be the eigenfunction of $\mathcal{P}_{f,t}$ corresponding to the eigenvalue $\lambda(t)$. That is, g_t satisfies

$$\mathcal{P}_T(e^{itf}g_t) = \lambda(t)g_t.$$

Multiplying by a test function h and integrating against μ gives

$$\int \mathcal{P}_T(e^{itf}g_t)h\, d\mu = \lambda(t)\int g_t h\, d\mu.$$

Recalling the definition of $\mathcal{P}_T$, this gives

$$\int (e^{itf}g_t)(h\circ T)\, d\mu = \lambda(t)\int g_t h\, d\mu. \tag{10}$$

Let $g_t' = \frac{d}{dt} g_t$ and $g_t'' = \frac{d^2}{dt^2} g_t$. Then differentiating (10) with respect to t gives

$$\int (if)(e^{itf} g_t)(h \circ T) \, d\mu + \int (e^{itf} g_t')(h \circ T) \, d\mu = \lambda'(t) \int g_t h \, d\mu + \lambda(t) \int g_t' h \, d\mu. \tag{11}$$

Setting $t = 0$ and using the fact that $\lambda(0) = 1$ and $g_0 \equiv 1$, we get

$$\int (if)(h \circ T) \, d\mu + \int (q)(h \circ T) \, d\mu = \lambda'(0) \int h \, d\mu + \int (q)(h) \, d\mu,$$

where we write $q = \frac{d}{dt} g_t|_{t=0}$.

Putting $h \equiv 1$ gives the expression for $\lambda'(0)$. Before finding $\lambda''(0)$, we observe that the above equation can also be used to find $\int qh \, d\mu$, which will be important later on. Indeed, using the assumption that $\int f \, d\mu = 0$, we have $\lambda'(0) = 0$, and so the above equation becomes

$$\int (if)(h \circ T) \, d\mu + \int (q)(h \circ T) \, d\mu = \int (q)(h) \, d\mu. \tag{12}$$

Similarly, replacing h with the test functions $h \circ T^k$ for $k \geq 1$, (12) gives

$$\int (if)(h \circ T^2) \, d\mu + \int (q)(h \circ T^2) \, d\mu = \int (q)(h \circ T) \, d\mu,$$

$$\int (if)(h \circ T^3) \, d\mu + \int (q)(h \circ T^3) \, d\mu = \int (q)(h \circ T^2) \, d\mu,$$

and so on. Observe that $\sum_{k \geq 1} (q)(h \circ T^k) \, d\mu$ converges because we have exponential decay of correlations. Thus we may add the above equations (infinitely many of them) and subtract this sum from both sides to obtain

$$\int qh \, d\mu = \sum_{k \geq 1} \int (if)(h \circ T^k) \, d\mu. \tag{13}$$

Now we can find the expression for $\lambda''(0)$. Set $h \equiv 1$ in (11) and differentiate to get

$$\int \left((if)^2 (e^{itf} g_t) + 2(if)(e^{itf} g_t') + e^{itf} g_t'' \right) \, d\mu = \int (\lambda''(t) g_t + 2\lambda'(t) g_t' + \lambda(t) g_t'') \, d\mu. \tag{14}$$

7

At $t = 0$ we see that the terms containing g''_t are equal, while $\lambda'(0) = 0$ by the assumption that $\int f\, d\mu = 0$, and so (14) gives

$$\int (-f^2)\, d\mu + 2 \int (if)(q)\, d\mu = \lambda''(t). \tag{15}$$

From (13), we have

$$\int (if)(q)\, d\mu = \sum_{k \geq 1} \int (if)(if \circ T^k)\, d\mu,$$

which together with (15) suffices to complete the proof of the expression for $\lambda''(0)$.