Convergence of symmetric Markov chains

Richard Bass
University of Connecticut
r.bass@uconn.edu
www.math.uconn.edu/∼bass

This is joint work with Takashi Kumagai and Toshihiro Uemura.

RB and T. Kumagai, TAMS 360 (2008) 2041–2075.

RB, T. Kumagai, and T. Uemura, PTRF 148 (2010) 107–140.

This problem has also been studied by Stroock-Zheng (1997) and DeMasi, Ferrari, Goldstein, Wick (1989). The latter paper is concerned with random walk in random environment.

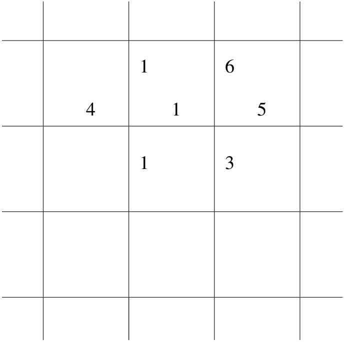
The model

To describe the model, first let's look at nearest neighbor random walks. We look at $\mathbb{Z}^2$, and on each edge we put a conductance. We let C_{xy} be the conductance between x and y.

The probability of going from a point x to a neighbor y is given by

$$\mathbb{P}^x(X_1 = y) = \frac{C_{xy}}{\sum_z C_{xz}},$$

where the sum in the denominator is over all neighbors z of the point x.


1
6
4
1
5
1
3

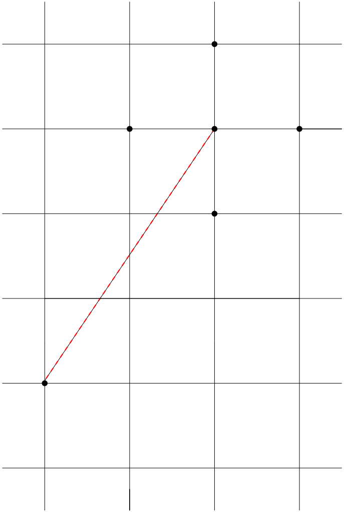
Such a Markov chain is called a symmetric Markov chain because the transition probabilities $\mathbb{P}^x(X_1 = y)$ have a Radon-Nikodym derivative $p(t, x, y)$ (called the transition densities) with respect to the measure $\mu(dx) = \sum_z C_{xz}$ and the transition densities are symmetric:

$$p(t, x, y) = p(t, y, x).$$

Note $\mathbb{P}^x(X_1 = y) \neq \mathbb{P}^y(X_1 = x)$ in general.

It is natural to generalize this model. First, we could allow jumps to nearby points, but so that the largest jump is of size K. This is called the bounded range model.

Second, we could allow arbitrarily large jumps (the unbounded range model).



By analogy with sums of independent random variables, one might impose a second moment condition.

Such a condition could be expressed by

$$C_{xy} \le \varphi(|x-y|)$$

and

$$\sum i^{d+1} \varphi(i) < \infty.$$

This implies

$$\sum_y |x-y|^2 C_{xy} \le C, \qquad x \in \mathbb{Z}^d.$$

(We convert to polar coordinates.)

Just as in the case of sums of independent random variables, one can look at the case where the limit is a generalization of the Gaussian distribution and also the case where the limit is a generalization of the stable law.

Therefore we want to impose some kind of decay property for the conductances, but want to get away with as weak a one as possible.

The second moment condition

Let us stick with the second moment condition for now.

Two other fairly technical assumptions we impose are:

(1)

$$c_1 \le \nu_x = \sum_y C_{xy} \le c_2 \qquad \text{for all } x$$

and

(2) $\exists M_0, \delta, N$ such that if $|x-y| = 1$, $\exists x = x_1, \dots, x_N = y \in B(x, M_0)$, $C_{x_i, x_{i+1}} \ge \delta$.

The second assumption says that one can get from any point to its nearest neighbor in no more than some fixed number of steps.

The main result we want in the second moment case is that under appropriate conditions on the conductances, the process X, normalized appropriately, converges to an elliptic operator in divergence form.

Continuous time vs. discrete time

First of all, it is easier to work with Y than the process X, where Y is defined as follows:

Wait at any point a length of time that is exponential with parameter 1 and independent of everything else, and then jump to another point with the same probabilities as X does.

The difference between X and Y is that X is a discrete time Markov chain while Y is a continuous time Markov chain. X and Y visit the same points in the same order, but one jumps at fixed times and the other at exponential times.

If T_n is the time of the n^{th} jump, then

$$\left|\frac{T_n}{n} - 1\right| \to 0, \qquad \text{a.s.}$$

by the strong law of large numbers. Note $Y_{T_n} = X_n$. Once we normalize Y, the difference between X and Y washes out, and we can thus work with either X or Y.

Y is easier to work with.

Dirichlet forms

To describe the possible limits, let me begin by describing Dirichlet forms.

Markov processes can be described by their infinitesimal generators:

$$\mathcal{L}f(x) = \lim_{t\to 0} \frac{P_t f(x) - f(x)}{t},$$

provided the limit exists.

One common class of infinitesimal generators is the class of non-divergence second order elliptic operators:

$$\mathcal{L}f(x) = \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x) + \sum_{i=1}^{d} b_i(x) \frac{\partial f}{\partial x_i}(x).$$

Another class is the class of divergence form operators:

$$\mathcal{L}f(x) = \sum_{i,j=1}^{d} \frac{\partial}{\partial x_i}\Big(a_{ij}(\cdot)\frac{\partial f}{\partial x_j}(\cdot)\Big)(x).$$

If the a_{ij} are differentiable, we use the product rule and this reduces to the case of non-divergence form operators. But people work with divergence form operators even when the a_{ij} are only bounded and measurable.

To interpret this, multiply $\mathcal{L}f(x)$ by $g(x)$ and integrate over $\mathbb{R}^d$. By integration by parts, we get

$$\int g(x)\mathcal{L}f(x)\,dx = -\mathcal{E}(f,g),$$

where

$$\mathcal{E}(f,g) = \sum_{i,j=1}^{d} \int a_{ij}(x)\frac{\partial f}{\partial x_i}(x)\frac{\partial g}{\partial x_j}(x)\,dx.$$

$\mathcal{E}$ is known as the Dirichlet form, and this makes sense if f and g have derivatives in L^2 and a_{ij} is bounded and measurable.

CLTs

If one looks up central limit theorem for Markov chains in the literature, one finds references to theorems along the lines of:

$\sum f(X_i)/\sqrt{n}$ converges weakly to the energy of f.

A central limit theorem for us is the version for symmetric Markov chains of what Stroock and Varadhan proved for a certain other type of Markov chain in Chapter 11 of their book.

Let $Z_t^{(n)} = Y_{nt}^{(n)}/\sqrt{n}$ and $W_t^{(n)} = X_{[nt]}^{(n)}/\sqrt{n}$. Then, under appropriate assumptions, $Z^{(n)}$ and $X^{(n)}$ converge in law (as processes) to an elliptic diffusion in divergence form. The convergence is with respect to the space $D[0, t_0]$, and this holds for every t_0.

Let me state a precise theorem.

We have $C_{xy}^{(n)}$ instead of just one set of conductances.

Let $R > 0$. (If the conclusion holds for one R, it holds for all.)

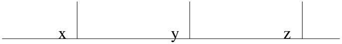
Let

$$C_{xy}^{n,R} = \begin{cases} C_{xy}^{(n)}, & |x - y| < nR; \\ 0, & \text{otherwise.} \end{cases}$$

Let us look at the case of dimension 1 first and set

$$a^{n,R}(x) = \sum C^{n,R}_{nx,n(y+k)} nk \operatorname{sgn} k,$$

where the sum is over those k such that $x \in [ny, n(y+k)]$.


x
y
z

Theorem. Suppose for each r

$$\lim_n \int_{|x|\leq r} |a^{n,R}(x) - a(x)|\, dx \to 0$$

where a is bounded above and below.

Then $Z^{(n)}$ converges to the diffusion corresponding to the the Dirichlet form $\mathcal{E}(f,f) = \int f'(x)a(x)f'(x)\, dx$.

When I say $Z^{(n)}$ converges, actually there is convergence in two different senses. For each starting point, the law of $Z^{(n)}$ converges weakly to the law of the elliptic diffusion with the same starting point with respect to the topology of $D[0, t_0]$ for each t_0. Also, the semigroup for $Z^{(n)}$ converges strongly to the semigroup of the limiting process.

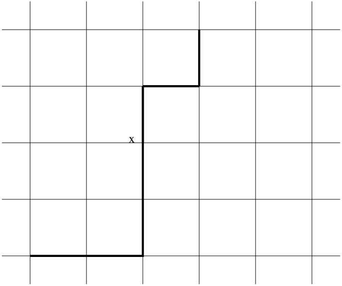
In the hypotheses, we can replace L^1 convergence by convergence a.e. or in measure, because the $a^{n,R}$ are bounded, but not by weak convergence.

An example of Stroock and Zheng shows this: On $\mathbb{Z}$, take $C_{j,j+1} = 1$ if j is odd and equal to 2 if j is even.

In the case $d \geq 2$, let

$$(a^{n,R}(x))_{ij} = \sum C^{n,R}_{ny,n(y+k)} nk_j \operatorname{sgn} k_i,$$

where the sum is over those rectilinear paths that touch the cube containing x. To say a bit more about this, one chooses a class of shortest rectilinear paths that touch the cube containing x and then takes an appropriate average.


x

If a is symmetric and uniformly elliptic and $a^{n,R}$ converges to a in measure, then we get weak convergence.

An example: if $C_{xy}^{n,R}$ is constant, we get a linear transformation of BM.

If we make some sort of "continuity condition" on the $C^{n,R}_{xy}$, we get a more natural theorem. In this case, we take

$$(b^{n,R}(x))_{ij} = \sum_{k \in n^{-1}\mathbb{Z}^d} C^{n,R}_{nx,n(x+k)} n^2 k_i k_j.$$

If the $b^{n,R}$ converge uniformly to a, we get the same conclusion about convergence of the Markov chains.

Without the second moment condition

Now let us consider more general Markov chains, namely, those where we do not have a second moment condition. The case where the Markov chains look like stable processes was considered by Husseini and Kassmann.

In general we have a family of conductances C^n_{xy} defined for $x, y \in n^{-1}\mathbb{Z}^d$.

We impose a condition reminiscent of the second moment condition:

$$C^n(x,y) \le n^{-(d+2)}\varphi(|x-y|),$$

where

$$\int_0^\infty \varphi(t)t^{d-1}(1\wedge t^2)\,dt < \infty.$$

This actually is the analogue of the condition

$$\int (1\wedge |x|^2)\,n(dx) < \infty$$

that appears in the study of Lévy processes.

We choose $\varepsilon_n \to 0$ and define $C^{n,C}_{x,y}$ to be equal to $C^n_{x,y}$ if $|x-y| \le \varepsilon_n$ and $C^{n,J} = C^n - C^{n,C}$. We define $a^n_{ij}(x)$ similarly to the second moment case and require that the $a^n_{ij}(x)$ be uniformly strictly elliptic and converge to a symmetric $a_{ij}(x)$.

We define

$$j^n(x,y) = n^{d+2} C^{n,J}_{[x],[y]}.$$

Here $[x]$ is the lower left hand corner of the cube with vertices in $n^{-1}\mathbb{Z}^d$ in which x lies.

We then require that for each δ, N,

$$j^n([x],[y])\, 1_{[\delta,N]}(|x-y|)\, dx\, dy$$

converges to

$$j(x,y)\, 1_{[\delta,N]}(|x-y|)\, dx\, dy$$

for some symmetric function $j(x,y)$.

Here convergence is weak convergence in the sense of probability measures, extended to allow finite measures.

Thus finite measures μ_n converge to a finite measure μ if

$$\int f(x)\,\mu_n(dx) \to \int f(x)\,\mu(dx)$$

for every bounded continuous function f.

We then have the following theorem.

Theorem. For every starting point x, the laws of X_t^n and Y_t^n, $t \in [0,1]$, under $\mathbb{P}^{[x]}$ converge weakly with respect to the topology of $D[0,1]$. If $\mathbb{P}^x$ denotes the limit law, then the limit $(X_t, \mathbb{P}^x)$ is a symmetric strong Markov process with Dirichlet form

$$\mathcal{E}(f,g) = \int \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial f}{\partial x_i}(x) \frac{\partial g}{\partial x_j}(x)\,dx + \int\int (f(x) - f(y))(g(x) - g(y)) j(x,y)\,dx\,dy.$$

For every continuous $f \in L^2$, $P_t^n f$ converges in L^2 to $P_t f$, where $P_t^n f(x) = \int p^n(t,[x],[y]) f([y])\,dy$.

Idea of proof

There are two main steps. The first is to show that the law of Y^n, a normalized version of Y, is tight with respect to the topology on $D[0, 1]$.

The second step is to show that every weak subsequential limit point is the same.

The first step is fairly complicated, but it follows what is now a well understood procedure, so it actually is easier.

We can show tightness by the following procedure:

1. Get an upper bound on $p(t, x, y)$ by a method due to Nash.

2. Get an upper bound on $pt, x, y)$ that decays when $|x - y|$ gets large. This is based on work by Carlen, Kusuoka, and Stroock, by Bass and Levin, and by Barlow, Grigor'yan, and Kumagai.

3. Get an estimate on

$$\mathbb{P}^x(\sup_{s \leq t} |Y_s - x| > \lambda).$$

4. Use a criterion of Aldous to get tightness in the space $D[0, 1]$.

To give an idea of how this goes, let me consider the continuous state space case rather than the discrete. Fix x_0 and let

$$E(t) = \int p(t, x_0, y)^2 \, dy.$$

Then

$$\begin{aligned} E'(t) &= 2 \int p(t, x_0, y) \frac{\partial p}{\partial t}(t, x_0, y) \, dy \\ &= 2 \int p(t, x_0, y) \mathcal{L} p(t, x_0, y) \, dy \\ &= -2 \mathcal{E}(p(t, x_0, \cdot), p(t, x_0, \cdot)). \end{aligned}$$

Here $\mathcal{L}$ is the infinitesimal generator of the process.

An inequality due to Nash (which can be proved via Fourier transforms) says that

$$\|f\|_2^{1+2/\nu} \le c\mathcal{E}(f,f)\|f\|_1^{4/\nu}.$$

We take $f(y) = p(t, x_0, y)$, note $\|f\|_1 = 1$, and obtain

$$E'(t) \ge -cE(t)^{1+2/\nu}.$$

We solve this differential inequality to obtain

$$p(2t, x_0, x_0) = E(t) \le ct^{-\nu/2}.$$

A variation of this gives better estimates when $|x - y|$ is large, and a similar technique works in the discrete case.

Once one has bounds on the transition densities, integrating gives bounds on

$$\mathbb{P}^x(|X_t - x| > r),$$

and standard Markov process techniques then leads to bounds on

$$\mathbb{P}^x(\sup_{s \le t} |X_s - x|).$$

In order to identify the limit, we also need to get equicontinuity on the $P_t f$, where

$$P_t f(x) = \int p(t,x,y) f(y)\, dy.$$

Again, although complicated, the argument is a modification of what is known:

1. Prove a weighted Poincaré inequality. This is the analogue to

$$\int_{\mathbb{R}^d} |f(x) - \overline{f}|^2 \varphi_A(x)\, dx \le c(A) \int_{\mathbb{R}^d} |\nabla f(x)|^2 \varphi_A(x)\, dx$$

where $\overline{f} = \int_{\mathbb{R}^d} f(x)\varphi_A(x)\, dx$ and $\varphi_A(x) = e^{-A|x|}$ for a parameter A.

2. Prove lower bounds on $p(t, x, y)$. This is basically an argument of Fabes and Stroock.

3. Obtain Hölder regularity of $p(t, x, y)$.

One starts with

$$G(t) = \int p(t, x_0, y) \log p(t, x_0, y) \varphi(y) \, dy,$$

where $\varphi(y) = c_1 e^{-c_2|y|}$, takes the derivative in t, and ends up with a differential inequality.

Once we have upper and lower bounds on $p(t, \cdot, \cdot)$, the regularity is not hard to prove.

Since $x, y \in \mathbb{Z}^d$, by Hölder regularity we mean

$$|p(t,x,y) - p(t,x',y')| \le c(t)(|x-x'|^\beta + |y-y'|^\beta)$$

provided either $x = x'$ or $|x - x'| \ge 1$ and similarly with y and y'.

To identify the limit, it suffices to show that

$$U_\lambda^n f \longrightarrow U_\lambda f$$

when f is a smooth function with compact support and U_λ^n and U_λ are the resolvents.

This is actually quite tricky and complicated.

Let me indicate the argument in an easier framework, to give a hint of the technique.

Let us consider diffusions on the real line with infinitesimal generator

$$\mathcal{L}f = (af')'.$$

The corresponding Dirichlet form is

$$\mathcal{E}_a(f,f) = \int a(f')^2.$$

Suppose a_n converges boundedly and a.e. to a.

There is a technique, called Mosco convergence, that can be used here, but this is not useful in the case of discrete Markov chains converging to continuous ones because the state spaces change.

Since the $U^n_\lambda f$ are equicontinuous, they converge to something, say, H. Then

$$\int (U^n_\lambda f)' g = -\int (U^n_\lambda f) g' \to -\int H g' = \int H' g,$$

or $(U^n_\lambda f)'$ converges weakly to H'.

It is easy to see that consequently $a_n (U^n_\lambda f)'$ converges weakly to aH', and then

$$\mathcal{E}_{a_n}(U^n_\lambda f, g) = \int a_n (U^n_\lambda f)' g' \to \int a H' g' = \mathcal{E}_a(H, g).$$

On the other hand,

$$\mathcal{E}_{a_n}(U^n_\lambda f, g) = (f, g) - \lambda(U^n_\lambda f, g) \to (f, g) - \lambda(H, g).$$

Therefore

$$\mathcal{E}_a(H, g) = (f, g) - \lambda(H, g),$$

which implies that $H = U_\lambda f$. Since $U^n_\lambda f \to H$, that does it.