Random Walk on an Arbitrary Set

By Simon C. Harris, Robin Sibson, and David Williams

School of Mathematical Sciences, University of Bath
Bath BA2 7AY, United Kingdom

1. The one-dimensional problem

Let I be a countably infinite set of points in $\mathbb{R}$, and suppose that I has no points of accumulation and that its convex hull is the whole of $\mathbb{R}$. It will be convenient to index I as $\{u_i : i \in \mathbb{Z}\}$, with $u_i < u_{i+1}$ for every i. Consider a continuous-time Markov chain $Y = \{Y(t) : t \geq 0\}$ on I, with the properties that:

Y *is driftless*; and
Y *jumps only between nearest neighbours*.

We propose this as the obvious analogue on an irregular support set of the simple symmetric random walk on the integers; it appears natural to change from a discrete-time to a continuous-time model to accommodate the irregularity, so it is not strictly speaking a generalisation. Suitably rescaled in time and space, the simple symmetric random walk converges in law to Brownian motion. In this paper we explore convergence properties for the irregular analogue.

In terms of elements of the Q-matrix of Y, the requirement of freedom from drift may be written in a general form as

$$\sum_{j \neq i} q_{i,j}\,(u_j - u_i) = 0 \text{ for every } i. \tag{1.1}$$

Write ℓ_i and r_i for the gaps to the left and to the right of u_i:

$$\ell_i := u_i - u_{i-1}, \quad r_i := u_{i+1} - u_i.$$

Then the limitation to nearest-neighbour jumps, namely

$$q_{i,j} = 0 \text{ for } j \neq i-1, i, i+1\ ,$$

allows the immediately off-diagonal elements of Q to be written as

$$q_{i,i+1} = q_i \frac{\ell_i}{\ell_i + r_i}\ , \quad q_{i,i-1} = q_i \frac{r_i}{\ell_i + r_i}\ ,$$

where $q_i = -q_{i,i}$ is the total jump-rate out of state i which we shall not yet fix. Thus even in the regularly spaced case, the change from a discrete-time to a continuous-time model permits some additional generality.

Let T_i be the closed interval extending from u_i half-way to each of its neighbours u_{i-1}, u_{i+1}, and denote by κ_i the length of this interval, namely

$$\kappa_i := \tfrac{1}{2}(\ell_i + r_i).$$

Define D_i as the 'variance' or 'diffusion' coefficient of Y at u_i:

$$D_i := \sum_{j \neq i} q_{i,j}\, (u_j - u_i)^2.$$

The sum has only two nonzero terms, those for $(i-1)$ and $(i+1)$, and D_i may thus be written in the form $\ell_i r_i q_i$. Finally, define

$$m_i := \kappa_i\, D_i^{-1} = \frac{\ell_i + r_i}{2\, \ell_i\, r_i\, q_i}.$$

Note that Q is m-symmetrizable because Q is nearest-neighbour and

$$m_i\, q_{i,i+1} = (2r_i)^{-1} = (2\ell_{i+1})^{-1} = m_{i+1}\, q_{i+1,i}.$$

The values m_i are all positive, and define a measure on I which will be denoted by m. The measure m is the (unique up to constant multiples) invariant measure for Y, and so, in various senses, describes how Y shares out its time amongst the states in I. If D_i is integrated with respect to m over the intersection of I with a large compact interval $K = [A, B]$ then

$$\sum_{u_i \in K} m_i\, D_i = \sum_{u_i \in K} \kappa_i = \lambda(K) + \text{`end effect'}, \tag{1.2}$$

where $\lambda(K)$ is the Lebesgue measure $|B - A|$ of K. The choice of normalisation in our definition of m is that required to give equality rather than just proportionality in (1.2). Let u_L, u_R be chosen so that $A \in T_L, B \in T_R$, choosing arbitrarily on a boundary. Then the end effect is bounded by $\kappa_L + \kappa_R$.

It follows that the end effect is of smaller order than the interval length as the latter tends to infinity, provided that $\kappa_i = o(i)$ as $i \to \pm\infty$, a mild condition to control the degree of irregularity with which the elements of I are positioned. This condition can of course be expressed in a variety of different ways; exploration of the multidimensional case, introduced in Section 2, may point towards the most natural form for it.

For convenience, we assume that $0 \in I$ and that $Y(0) = 0$. For $\theta > 0$, define the scaled version

$$Y_\theta(t) := \theta^{-1} Y(\theta^2 t)$$

of Y. For $\theta > 0$ let μ_θ be the measure on $\mathbb{R}$ which assigns mass m_i/θ to the point u_i/θ. Then μ_θ is the appropriately normalised invariant measure for Y_θ. Results (1.1) and (1.2) make it plausible that if μ_θ converges to Lebesgue measure, then Y, when suitably scaled, converges to Brownian motion. Theorem A makes this precise. In Theorem A, B denotes Brownian motion started at the origin and with unit variance coefficient (*standard* Brownian motion).

THEOREM A. *Suppose that, as* $\theta \to \infty$,

$$\mu_\theta([a, b]) \to b - a \quad \textit{whenever} \ -\infty < a \leq b < \infty. \tag{1.3}$$

Then

$$Y_\theta \Rightarrow B, \tag{1.4}$$

that is, the law of Y_θ *converges (in the usual 'weak' or 'narrow' topology) to the Wiener law of* B.

Theorem A is proved in Section 5. The proof may be adapted to show that condition (1.3) is also *necessary* for property (1.4) to hold, but we regard that as primarily a technical point. If μ_θ converges in the weak* sense of property (1.3) to some measure μ, then, of course, Y_θ converges to the diffusion with identity scale function and μ as speed measure. We remark (without detailed proof) that since μ_θ is obtained by rescaling m, the limit μ must be a multiple of Lebesgue measure on each half-line $[0, \infty)$, $(-\infty, 0]$. Although the generalisation to different multiples on each half-line appears to be of little interest, it will often be natural to allow the speed measure to be an arbitrary multiple of Lebesgue measure on the whole line.

In the light of Theorem A, it is appropriate to return to the question of the choice of total jump-rates q_i. We identify three interesting cases.

First, consider the case where D_i is constant, say, $D_i \equiv D$. This implies $m_i \equiv \kappa_i D^{-1}$, so $\mu_\theta([a,b])$ differs from $D^{-1}(b-a)$ only through an end effect, and provided I satisfies the condition introduced above to control irregularity, convergence is assured. The limit has unit speed when $D = 1$. In many senses this is the most natural normalisation, and we call it the *well-normalised* case.

Secondly, consider the case where m_i is constant; this corresponds to

$$q_{i,i+1} = q_{i+1,i} \propto 1/[2(u_{i+1} - u_i)] \quad \text{for all } i \tag{1.5}$$

and is appropriately called the *symmetric* case. The measure μ_θ is in this case a scaled counting measure, and condition (1.3) can be thought of as requiring asymptotic uniformity of spacing of the points of I. The measure m actually *is* the counting measure on I when the constant of proportionality in (1.5) is 1, and the limit then has unit speed when the scaled counting measure converges to Lebesgue measure. Corollary A.1 develops this case further.

The third case of interest is where q_i is constant. In this case the total jump-rate from each state does not depend on the geometry, unlike the first two cases where it is to a greater or lesser extent adapted to it. We call this the *decoupled* case. Since in this case the jump times have a common exponential distribution, it is only a small step away from the discrete-time model. Corollaries A.2 and A.3, and Theorems B and C, provide some further development of this case.

In each of these three cases, some further generality is possible by allowing the quantity held constant (D_i, m_i, q_i respectively) to vary as independent realisations of a random variable; some of the results in Section 3 are expressed in this form.

2. A higher-dimensional analogue

Let I be a countable subset of $\mathbb{R}^d$ without accumulation points whose convex hull is the whole of $\mathbb{R}^d$. As in the one-dimensional case, it will be convenient to index the points of I, as $I = \{\mathbf{u}_i\}$ (where the $\mathbf{u}_i$ are regarded as column vectors), but there is no longer a natural order in which to do this. The *tile* T_i of $\mathbf{u}_i$ in the Dirichlet tessellation of $\mathbb{R}^d$ determined by I is the set

$$T_i := \{\mathbf{u} \in \mathbb{R}^d : \|\mathbf{u} - \mathbf{u}_i\| = \inf_{\mathbf{u}_j \in I} \|\mathbf{u} - \mathbf{u}_j\|\} \tag{2.1}$$

of those points $\mathbf{u} \in \mathbb{R}^d$ which are at least as close to $\mathbf{u}_i$ as to any other point of I. Write κ_i for $\lambda_d(T_i)$, where λ_d denotes d-dimensional Lebesgue measure. Define $S_{i,j} = T_i \cap T_j$, Call points $\mathbf{u}_i$ and $\mathbf{u}_j$ of I *neighbours* if $\lambda_{(d-1)}(S_{i,j})$ is positive, that is, if T_i and T_j have a common facet (face of codimension 1). Define the *boundary-over-distance weight* $\nu_{i,j}$ by

$$\nu_{i,j} := \frac{\lambda_{(d-1)}(S_{i,j})}{2\|\mathbf{u}_j - \mathbf{u}_i\|} \qquad (j \neq i). \tag{2.2}$$

Note that $\nu_{i,j}$ is positive if and only if $\mathbf{u}_i$ and $\mathbf{u}_j$ are neighbours, and that, for all $i \neq j$, $\nu_{i,j} = \nu_{j,i}$. Each point of I has finitely many neighbours, at least $(d+1)$. In one dimension, the concept of neighbour defined in this section is the familiar one, and, indexing as in Section 1,

$$\nu_{i,i+1} = \nu_{i+1,i} = 1/[2(u_{i+1} - u_i)].$$

These notations and definitions are consistent with those already introduced in Section 1 in the one-dimensional case.

The weights $\nu_{i,j}$ satisfy the local balance property

$$\sum_{j \neq i} \nu_{i,j} \left(\mathbf{u}_j - \mathbf{u}_i\right) = \mathbf{0}_d \text{ for every } i, \tag{2.3}$$

where $\mathbf{0}_d$ is the d-dimensional zero vector. This property is trivial in one dimension, and indeed in Section 1 we took (1.1) as a starting point and only emerged later with (1.5). In higher dimensional cases, (2.3) does not determine the relative values of the weights once a point has more than $(d+1)$ neighbours, and it is natural to look for methods of choosing the weights which are canonical in some sense and lead to (2.3) as a property. The boundary-over-distance weights form one such system, but not the only one. Sibson (1980) gave the first example of such a system, with weights $\kappa_{i,j}$ defined as the d-dimensional measure of the subtile $T_{i,j}$ (the set of points with $\mathbf{u}_i$ as nearest and $\mathbf{u}_j$ as second-nearest elements of I). The proof of that result uses integral geometry arguments. The corresponding result for boundary-over-distance weights was proved by Christ, Friedberg, and Lee (1982) using the divergence theorem; note that we have normalised our boundary-over-distance weights differently, to avoid the appearance of an unwelcome constant in (2.4) and (2.5) below. For the two-dimensional case only, there is an easy alternative proof using vectors. The general result for boundary-over-distance weights can also be derived from the result for subtile weights.

What makes the boundary-over-distance weights particularly interesting in the present context is that they also have the symmetry property, which the subtile weights do not; it is not *a priori* obvious that it is possible to combine local balance and symmetry. The appearance of the divergence theorem in the proof of local balance is also suggestive of a link with Brownian motion. A further result (again see Christ, Friedberg, and Lee, 1982) is

$$\sum_{j\neq i} \nu_{i,j}\,(\mathbf{u}_j - \mathbf{u}_i)(\mathbf{u}_j - \mathbf{u}_i)^{\mathrm{T}} = \kappa_i\,\mathbf{I}_d + \text{`zero-trace matrix'} \tag{2.4}$$

where superscript T denotes transpose, $\mathbf{I}_d$ is the $d\times d$ identity matrix, and the zero-trace matrix is a sum over j of terms which are odd in i, j. Summing (2.4) over those $\mathbf{u}_i$ in a compact ball K in $\mathbb{R}^d$ accordingly yields a RHS whose principal term is $\left(\sum_{u_i\in K}\kappa_i\right)\mathbf{I}_d$ and whose other term is a zero-trace symmetric matrix arising from neighbour-pairs i, j where $\mathbf{u}_i \in K$ and $\mathbf{u}_j \notin K$, terms for neighbour-pairs i, j with $\mathbf{u}_i \in K$ and $\mathbf{u}_j \in K$ cancelling because of oddness. This zero-trace matrix is accordingly a boundary effect, and under suitable conditions may be expected to behave as $o\,(\lambda_d(K))$. The same is true of the difference between the principal term and $\lambda_d(K)\,\mathbf{I}_d$. On taking the trace of (2.4), we recover the trivial remark that

$$\sum_{j\neq i} \nu_{i,j}\,\|\mathbf{u}_i - \mathbf{u}_j\|^2 = \kappa_i\, d. \tag{2.5}$$

This simply reflects a decomposition of T_i into d-dimensional pyramids with common vertex $\mathbf{u}_i$ and the facets of T_i as bases.

These results suggest that a multidimensional generalisation of the Markov chains considered in Section 1 may usefully be obtained by considering Q-matrices of the form

$$q_{i,j} := q_i \frac{\nu_{i,j}}{\sum_{k\neq i}\nu_{i,k}} \quad \text{for } j\neq i \tag{2.6}$$

where, as before, q_i is the total jump-rate out of $\mathbf{u}_i$. Such a chain allows jumps only to neighbours, and, from (2.3), is driftless. If we write

$$\mathbf{D}_i := \sum_{j\neq i} q_{i,j}\,(\mathbf{u}_j - \mathbf{u}_i)(\mathbf{u}_j - \mathbf{u}_i)^{\mathrm{T}} \tag{2.7}$$

then, from (2.5),

$$\operatorname{trace}\mathbf{D}_i = q_i \frac{1}{\sum_{k\neq i}\nu_{i,k}}\kappa_i\, d.$$

If we define

$$D_i = d^{-1}\,\mathrm{trace}\,\mathbf{D}_i$$

and

$$m_i := \kappa_i\, D_i^{-1} = \frac{\sum_{k\neq i}\nu_{i,k}}{q_i} \quad \text{for all } i$$

then, as in the 1-dimensional case, m symmetrises Q:

$$m_i\, q_{i,j} = m_j\, q_{j,i} \quad \text{for all } i,j \text{ with } j\neq i.$$

It again follows that m is the invariant measure, unique up to a scale factor. Moreover, the remark about summing (2.4) shows that, with this choice of normalisation for m, we have

$$\begin{aligned}\sum_{\mathbf{u}_i\in K} m_i\,\mathbf{D}_i &= \Big(\sum_{\mathbf{u}_i\in K}\kappa_i\Big)\,\mathbf{I}_d + \text{`zero-trace matrix'}\\ &= \lambda_d(K)\,\mathbf{I}_d + \text{`edge effect'},\end{aligned} \tag{2.8}$$

where the edge effect is the sum of two terms: the discrepancy between $\sum\kappa_i$ and $\lambda_d(K)$, times the unit matrix; together with the zero-trace matrix. Under suitable conditions both of these terms can be expected to behave as $o\,(\lambda_d(K))$; we make no attempt here to explore what form those conditions should take. Hence, it is plausible that if m scales to Lebesgue measure, then a chain Y with Q-matrix Q will scale to Brownian motion. We hope to prove this in a sequel. As in the one-dimensional case, we can identify the well-normalised, symmetric, and decoupled cases as those in which D_i, m_i, q_i respectively are identically 1.

3. Random walks on renewal processes

In this section we explore some examples in which the support set I is generated as a realisation of a renewal process. In this case stochastic processes enter at more than one stage, and appropriate types of convergence need to be defined with care. Background can be found in Billingsley (1968), Ethier and Kurtz (1986), and Parthasarathy (1967). The paths of the scaled random walk Y_θ are in the space $D_{\mathbb{R}}[0,\infty)$ of right continuous functions on the nonnegative real line. Under the Skorohod metric d, the space $(D_{\mathbb{R}}[0,\infty), d)$ is a complete separable metric space. Y_θ is a $D_{\mathbb{R}}[0,\infty)$-valued random variable with probability measure $\mathbb{P}_\theta\in\mathcal{P}(D_{\mathbb{R}}[0,\infty))$ given by

$$\mathbb{P}_\theta(S) := \mathbb{P}(Y_\theta\in S) \quad \forall S\in\mathcal{S},$$

where $\mathcal{S}$ is the Borel σ-algebra of $(D_{\mathbb{R}}[0,\infty), d)$.

Interest lies in finding when Y_θ converges in distribution to a $D_{\mathbb{R}}[0,\infty)$-valued random variable Z, denoted by $Y_\theta \Rightarrow Z$, or equivalently, when $\mathbb{P}_\theta$ converges weakly to $\mathbb{P}_Z$, denoted by $\mathbb{P}_\theta \Rightarrow \mathbb{P}_Z$. In this situation, we have that weak convergence is equivalent to convergence in the Prohorov metric ρ, under which $(\mathcal{P}(D_{\mathbb{R}}[0,\infty)), \rho)$ is a complete separable metric space. So

$$\mathbb{P}_\theta \Rightarrow \mathbb{P}_Z \quad \text{iff} \quad \rho(\mathbb{P}_\theta, \mathbb{P}_Z) \to 0.$$

For our results we need to consider the convergence of *random* laws on the space $D_{\mathbb{R}}[0,\infty)$ of right continuous paths. Consider the law of the chain conditional on knowing the set of points on which it lives. Hence define the random probability law

$$\mathcal{L}_\theta(S) := \mathbb{P}(Y_\theta \in S \mid I).$$

$\mathcal{L}_\theta$ is a $\mathcal{P}(D_{\mathbb{R}}[0,\infty))$-valued random variable which is $\sigma(I)$-measurable. We define the *global law* $\mathbb{P}_{\mathcal{L}}$ associated with a random law $\mathcal{L}$ by

$$\mathbb{P}_{\mathcal{L}}(S) = \mathbb{E}(\mathcal{L}(S))$$

and avoid double-subscripting by writing $\mathbb{P}_\theta$ for this construct when $\mathcal{L}$ is $\mathcal{L}_\theta$.

Given a random probability law $\mathcal{L}$, *weak convergence in probability* (of $\mathcal{L}_\theta$ to $\mathcal{L}$) is denoted by

$$\mathcal{L}_\theta \overset{\mathbb{P}}{\Rightarrow} \mathcal{L}\,, \quad \text{defined via} \quad \forall \epsilon > 0, \quad \mathbb{P}(\rho(\mathcal{L}_\theta, \mathcal{L}) > \epsilon) \to 0 \quad \text{as} \quad \theta \to \infty.$$

and *weak convergence almost surely* is denoted by

$$\mathcal{L}_\theta \overset{\text{a.s.}}{\Rightarrow} \mathcal{L}\,, \quad \text{defined via} \quad \mathbb{P}(\rho(\mathcal{L}_\theta, \mathcal{L}) \to 0 \quad \text{as} \quad \theta \to \infty) = 1.$$

As usual, the latter condition implies the former. Also, weak convergence in probability of random laws to a limit implies weak convergence of the corresponding global laws. The relations

$$\mathcal{L}_\theta \overset{\text{a.s.}}{\Rightarrow} \mathcal{L}_Z \quad \text{implies that} \quad \mathcal{L}_\theta \overset{\mathbb{P}}{\Rightarrow} \mathcal{L}_Z \quad \text{implies that} \quad \mathbb{P}_\theta \Rightarrow \mathbb{P}_Z$$

are a special case, where the *deterministic law* $\mathcal{L}_Z$ is defined to have unit mass at $\mathbb{P}_Z$; the latter is in consequence its global law.

First, we present some corollaries of Theorem A.

COROLLARY A.1. *Let $\{G_j : j \in \mathbb{Z}\}$ be a family of independent identically distributed (IID) positive random variables, where each G_j has the distribution of a generic random variable G. Define*

$$u_0 := 0, \quad u_i := \sum_{j=0}^{i-1} G_j \quad (i > 0), \quad u_i := -\sum_{j=i}^{-1} G_j \quad (i < 0),$$

and let $I = \{u_i : i \in \mathbb{Z}\}$. If the q_i are chosen so that each m_i equals 1 (the symmetric case discussed in Section 1), and if $\mathbb{E}(G) = 1$, then $\mathcal{L}_\theta \overset{\text{a.s.}}{\Rightarrow} \mathcal{L}_B$, the Wiener law of B.

Proof of Corollary A.1. This is an immediate consequence of Theorem A and the Strong Law. □

COROLLARY A.2. *Let I be as in Corollary A.1. Let the jump rates $\{q_i : i \in \mathbb{Z}\}$ be a family of IID positive random variables, where each q_i has the distribution of a generic random variable Q, and where the family $\{q_i\}$ is independent of the family $\{G_j\}$. Then, as $\theta \to \infty$, $\mathcal{L}_\theta \overset{\text{a.s.}}{\Rightarrow} \mathcal{L}_{\sigma B}$, where the variance coefficient of the Brownian motion on the right-hand side is given by*

$$\sigma^2 = \frac{\mathbb{E}(G)}{\mathbb{E}(1/G)\mathbb{E}(1/Q)}\,.$$

Proof of Corollary A.2. For any interval $[a, b]$, the measure $\mu_\theta[a, b]$ is θ^{-1} times the sum of the m_i-values over points u_i in $[\theta a, \theta b]$. Since the average gap size is $\mathbb{E}(G)$, the number of such points is, more or less, $\theta(b-a)/\mathbb{E}(G)$. Since, moreover,

$$r^{-1}\sum_{i=1}^{r} m_i = (2r)^{-1}\sum_{i=1}^{r}\left(G_i^{-1} + G_{i+1}^{-1}\right) q_i^{-1} \to \mathbb{E}(G^{-1}Q^{-1}),$$

we see that μ_θ assigns mass approximately $(b-a)/\sigma^2$ to $[a, b]$. □

For real $k > 0$, say that a positive random variable G has a $\Gamma(k, 1)$ distribution if it has probability density function

$$g^{k-1}e^{-g}/\Gamma(k) \qquad (g > 0).$$

COROLLARY A.3. *Let I be as in Corollary A.1, and let G be distributed as $\Gamma(k,1)$ with $k \geq 1$. Let $q_i \equiv 1$ (the decoupled case of Section 1). Then $\mathcal{L}_\theta \overset{\text{a.s.}}{\Rightarrow} \mathcal{L}_{\sigma B}$, where $\sigma^2 = k(k-1)$.*

Proof of Corollary A.3. This is an application of Corollary A.2. □

If k is 1, so that I consists of all points of a Poisson process, then the occurrence of many clusters of close points between which Y will make many time-consuming oscillations will result in weak convergence of Y_θ to the zero process. Such 'small-gap' considerations lead one to ask in general about other normalizations of the Y process. For the particular case where the renewal distribution is a $\Gamma(k,1)$ distribution, this question is answered completely by Theorems B and C below, which complement Corollary A.3 and show some of the strange things that can happen.

THEOREM B. *Let I be as in Corollary A.1, and let G be distributed as $\Gamma(1,1)$; that is, G has an exponential distribution with parameter 1, and the renewal process is a unit-rate Poisson process. Redefine Y_θ for this case using the normalisation*

$$Y_\theta(t) := \theta^{-1} Y((\theta^2 \log \theta)t)$$

and let $\mathcal{L}_\theta$ be the corresponding random law. Then

$$\mathcal{L}_\theta \overset{\text{P}}{\Rightarrow} \mathcal{L}_B \quad \textit{but} \quad \mathcal{L}_\theta \overset{\text{a.s.}}{\not\Rightarrow} \mathcal{L}_B.$$

THEOREM C. *Let I be as in Corollary A.1, and let G be distributed as $\Gamma(k,1)$ with $k < 1$. Redefine Y_θ for this case using the normalisation*

$$Y_\theta(t) := \theta^{-1} Y(\theta^{(k+1)/k} t)$$

and let $\mathcal{L}_\theta$ be the corresponding random law. Let $\mathbb{P}^k_{FM}$ be the probability measure of a randomised Feller-McKean chain driven by the jumps of a stable process of characteristic exponent k. Then $\mathbb{P}_\theta \Rightarrow \mathbb{P}^k_{FM}$, but the random law $\mathcal{L}_\theta$ is not even weakly convergent in probability.

Proofs of Theorems B and C will be given by SCH in a further paper. The probability measure $\mathbb{P}^k_{FM}$ referred to in Theorem C is defined as follows.

Let $J \subset \mathbb{R}$ be a countable dense subset of the line, and let $m := \{m_y : y \in J\}$ be a set of positive real numbers indexed by the elements of J. By associating the mass m_y with the point y for each $y \in J$ we define an atomic measure on the Borel subsets of $\mathbb{R}$. The *Feller-McKean chain* with respect to (J, m) is the process $Z = \{Z(t) : t \geq 0\}$ where

$$Z(t) := B(\tau_t), \quad \tau_t := \inf\{u : A_u > t\}, \quad A_u := \sum_{y \in J} m_y \, L(u, y)$$

and B is a Brownian motion with local time process L. The Feller-McKean chain is a continuous process which spends almost all of its time on the countable set J and leaves each state instantly (its Q-matrix has $-\infty$ down the diagonal and zeroes elsewhere). It is a time change of a Brownian motion (see Williams, 1979).

A *randomised* Feller-McKean chain is constructed by choosing (J, m) randomly in the following manner. For $k < 1$ we take the stable process, $X = \{X(t) : t \in (-\infty, \infty)\}$ with $X(0) = 0$, of characteristic exponent k. This is an increasing random process which in any time interval has a random countable dense set of jumps, J, with only finitely many large jumps. Thus it induces a random measure $\mathrm{d}X(y)$ consisting of a dense set J of atoms with masses $m_y := X(y) - X(y-)$ at $y \in J$. Hence we can define a Feller-McKean chain as above, but with respect to the random pair (J, m) generated from X. The Brownian motion (determining the local time) is taken to be independent of the stable process. Finally, $\mathbb{P}^k_{FM}$ is defined to be the probability measure for Z.

4. Heuristics

We now want to *explain* Theorem A via an argument which is not intrinsically tied to the one-dimensional situation. In the next section, we *prove* Theorem A by a local-time method, much more powerful for one-dimensional problems, but restricted to them. Underpinning both approaches is Lévy's theorem that *if* $U = \{U(t) : t \geq 0\}$ *is a path-continuous martingale starting at* 0 *such that* $(U(t)^2 - t)$ *is also a martingale, then* U *is standard Brownian motion.* Purists should note that we shall blur the distinction between martingale and local martingale in this section.

For the one-dimensional case, define the *density* ρ on the interior of the tile T_i by

$$\rho(u) := \rho_i := m_i/\kappa_i = D_i^{-1};$$

the definition of ρ on the tile boundaries, a set of measure zero, is irrelevant. Then the process $M = \{M(t)\}$ defined by

$$M(t) := \int_0^t \rho(Y(s))^{\frac{1}{2}}\,\mathrm{d}Y(s) \tag{4.1}$$

(this 'stochastic integral' being a sum of jumps) is a martingale such that $(M(t)^2 - t)$ is also a martingale. Multiplying the dY increment by $\rho^{\frac{1}{2}} = D^{-\frac{1}{2}}$ corresponds exactly to dividing a random variable by its standard deviation to make the variance 1.

The θ-renormalised version of (4.1) reads:

$$M_\theta(t) := \int_0^t \rho_\theta(Y_\theta(s))^{\frac{1}{2}}\,\mathrm{d}Y_\theta(s)$$

where

$$\rho_\theta(u) := \rho(u/\theta), \quad (u \in \mathbb{R}).$$

Like M, M_θ is a martingale such that $(M_\theta(t)^2 - t)$ is also a martingale. Under appropriate conditions, the jumps in M_θ will become small as $\theta \to \infty$, and Lévy's theorem suggests that as a consequence $M_\theta \Rightarrow B$. If the function ρ_θ converges to the constant function 1 on $\mathbb{R}$ in some appropriate sense, then for large θ the difference between Y_θ and M_θ should be small, and these two remarks taken together make it plausible that the result of interest, namely $Y_\theta \Rightarrow B$, will hold. Theorem A identifies the appropriate mode of convergence of ρ_θ to 1. We plan to address higher-dimensional cases by making this argument rigorous and by identifying the appropriate higher-dimensional analogues to plug into it. There may also be a possibility of extending it to deal with transitions more general than just to neighbours, and perhaps also to weaken the requirement of exact freedom from drift.

Consider any irreducible standard Markov chain on $I \subset \mathbb{R}^d$, with Q-matrix Q, not necessarily of the form we have been considering, and let m be some invariant measure for it. Write

$$\mathbf{h}_i = \sum_{j \neq i} q_{i,j}(\mathbf{u}_j - \mathbf{u}_i)$$

and

$$\mathbf{D}_i = \sum_{j \neq i} q_{i,j}(\mathbf{u}_j - \mathbf{u}_i)(\mathbf{u}_j - \mathbf{u}_i)^{\mathrm{T}}$$

Define μ_θ as before. Define the vector-valued measure η_θ to have density $\theta\,\mathbf{h}_i$ at $\mathbf{u}_i/\theta$, and the matrix-valued measure Δ_θ to have density $\mathbf{D}_i$ at that point, both densities being with respect to μ_θ. The condition for convergence to Brownian motion could take the form that there exists a normalisation for m such that

$$\mu_\theta \Rightarrow \lambda_d, \qquad \text{(Condition 0)}$$
$$\eta_\theta \Rightarrow \mathbf{0}_d\lambda_d, \qquad \text{(Condition 1)}$$
$$\Delta_\theta \Rightarrow \mathbf{I}_d\lambda_d \qquad \text{(Condition 2)}$$

We have made life easy for ourselves by our choice of Q. First, symmetrisability of Q makes it possible to write down m explicitly; without this there would be great difficulty in general in finding m to allow these conditions to be testable. Secondly, drift-freedom of Q guarantees that Condition 1 is satisfied as an equality. Because of the form of η_θ, if Condition 1 is not satisfied as an equality then it requires a strong condition to enforce it in the limit. Thirdly, the specific choice of boundary-over-distance weights allows identity (2.4) and the remark following it to be used to identify the appropriate normalisation for m, and reduces Condition 2 to a question of controlling edge-effects. This concentrates attention on Condition 0, the condition introduced in the one-dimensional case in Theorem A. In the well-normalised case $m_i = \kappa_i$, this again reduces to a question of edge effects, the appropriate conditions probably being a subset of those required for Condition 2. In other cases, for example the symmetric case and the decoupled case, Condition 0 serves as a genuine test condition; as demonstrated in Section 3, it can be checked in interesting examples.

The local-time argument for the one-dimensional case is presented in Section 5; we now outline it at the heuristic level. The local-time approach concentrates on escape times rather than variances. For neighbour-limited transitions, these can be thought of as 'mutually inverse' concepts: the greater the variance, the shorter the time to escape. If Y is started at u_i, then the expected time before Y first hits the set $N_i := \{u_{i-1}, u_{i+1}\}$ is, of course, q_i^{-1}. If a Brownian motion B is started at u_i, then the expected time before it hits N_i is $\ell_i r_i$, as proved in the Note below. Thus $\rho_i = D_i^{-1}$ is the ratio of these times: the Brownian motion is providing the correct space-time reference frame. But we need something much more precise. The 'local time' spent by B at u_i before B hits N_i is exponentially distributed with mean $2\ell_i r_i/(\ell_i + r_i)$, so that if we multiply this by m_i we obtain an *exponentially* *distributed variable of mean* q_i^{-1} exactly like the holding time of Y in state u_i. Thus we can find the Y process in the Brownian motion.

Note. It may be helpful to recall an elementary argument. Let B be standard Brownian motion. Let ℓ and r be positive numbers. Since $B(t)$ and $(B(t)^2 - t)$ are martingales, so is $(\{\ell + B(t)\}\{r - B(t)\} + t)$, and on applying the stopping-time result at T, we have $\mathbb{E}(T) = \ell r$.

5. Proof of Theorem A

We present this argument in a relatively heuristic fashion. All results used may be found (with full rigour) in Williams (1979) or in Rogers & Williams (1987).

Step 1: *Brownian local time.* Let $B = \{B(t) : t \geq 0\}$ be a standard Brownian motion started at 0. Trotter's great theorem guarantees the existence of a local time

$$\{L(t,x) : t \geq 0; x \in \mathbb{R}\},$$

jointly continuous in t and x, and such that, for any nice function f on $\mathbb{R}$,

$$\int_0^t f(B(s))\,\mathrm{d}s = \int_{\mathbb{R}} f(x)\,L(t,x)\,\mathrm{d}x.$$

(Here, 'nice' can mean anything from 'C^∞ of compact support' to 'non-negative Borel-measurable' without changing the sense.) The normalization $L(t,x)$ is called ℓ_t^x in Rogers & Williams (1987).

For each fixed x, the function $t \mapsto L(t,x)$ is a continuous non-decreasing 'Cantor-like' function which grows only when B is at x. A good way to think about $L(t,x)$ is as

$$\int_0^t \delta_x(B(s))\,\mathrm{d}s,$$

where δ_x is the Dirac delta function concentrated at x.

Let $[a,b]$ be an interval, and let x and y belong to (a,b). Let $G(x,y)$ denote the expected local time spent at y by a Brownian motion started at x before that Brownian motion first hits a or b. The delta-function formula helps explain why G is the Green's function for the operator $\frac{1}{2}\mathrm{d}^2/\mathrm{d}x^2$ on $[a,b]$, namely

$$G(x,y) = G(y,x) = \frac{2(x-a)(b-y)}{b-a} \text{ when } 0 \leq x \leq y \leq b. \qquad \square$$

Moreover, a lack-of-memory property guarantees that the local time spent by this Brownian motion at its starting position x before it first hits either a or b has an exponential distribution, the mean of which is, of course, $G(x, x)$.

Step 2: *a representation for Y.* The following lemma is more or less an immediate consequence of the remarks just made.

LEMMA 5.1. *We can explicitly construct a Markov chain with the same law as Y, and we henceforth think of Y as this chain, by writing*

$$Y(t) := B\left(\alpha(t)\right), \tag{5.1}$$

where, for $t \geq 0$, $\alpha(t)$ is the largest solution of

$$\int_{x\in\mathbb{R}} L\left(\alpha(t), x\right)\, m(\mathrm{d}x) = t. \tag{5.2}$$

Proof. The construction forces Y to live on the set I. Moreover, Y inherits the 'driftless' property from B. It only remains to note that, by the Green's function formula, the (real) time spent by Y at u_i before its first jump from u_i is exponentially distributed with mean

$$m_i \, \frac{2\ell_i r_i}{\ell_i + r_i}\,,$$

and this has been 'cooked' to agree with q_i^{-1}. □

Step 3: *rescaling our representation.* For the moment, fix $\theta > 0$. We have

$$Y_\theta(t) = \theta^{-1} B(\alpha(\theta^2 t)). \tag{5.3}$$

We exploit the standard fact that $\tilde{B}(t) := \theta^{-1} B(\theta^2 t)$ defines a new Brownian motion started at 0.

LEMMA 5.2. *Let $\tilde{L}$ be local time for $\tilde{B}$. Then*

$$Y_\theta(t) = \tilde{B}(\tilde{\alpha}_\theta(t)), \tag{5.4}$$

where $\tilde{\alpha}_\theta(t)$ is the largest solution of

$$\int_{x\in\mathbb{R}} \tilde{L}(\tilde{\alpha}_\theta(t), x)\, \mu_\theta(\mathrm{d}x) = t. \tag{5.5}$$

Proof of Lemma 5.2.. For a nice function f on $\mathbb{R}$, we have

$$\begin{aligned}\int_{z\in\mathbb{R}} f(z)\,\tilde{L}(s,z)\,\mathrm{d}z &= \int_{r=0}^{s} f(\tilde{B}(r))\,\mathrm{d}r = \int_{r=0}^{s} f(\theta^{-1}B(\theta^2 r))\,\mathrm{d}r \\ &= \theta^{-2}\int_{u=0}^{\theta^2 s} f(\theta^{-1}B(u))\,\mathrm{d}u = \theta^{-2}\int_{y\in\mathbb{R}} f(\theta^{-1}y)\,L(\theta^2 s, y)\,\mathrm{d}y \\ &= \theta^{-1}\int_{z\in\mathbb{R}} f(z)\,L(\theta^2 s, \theta z)\,\mathrm{d}z,\end{aligned}$$

so that $\tilde{L}(s,z) = \theta^{-1}L(\theta^2 s, \theta z)$ for all s and z. Result (5.5) is now a rephrasing of equation (5.2) with $\tilde{\alpha}_\theta(t) = \theta^{-2}\alpha(\theta^2 t)$; and then, result (5.4) merely rephrases equation (5.1). □

Step 4: *completion of proof of Theorem A.* From Lemma 5.2, the process Y_θ is identical in law to the process Z_θ, where

$$Z_\theta(t) := B(\alpha_\theta(t)), \text{ where } \int_{x\in\mathbb{R}} L(\alpha_\theta(t), x)\,\mu_\theta(\mathrm{d}x) = t.$$

Suppose now that property (1.3) holds. For the moment, fix t. Then, for $u > t$, we have, since $x \mapsto L(u,x)$ is continuous of compact support and since property (1.3) implies weak* convergence of μ_θ to Lebesgue measure,

$$\int L(u,x)\,\mu_\theta(\mathrm{d}x) \to \int L(u,x)\,\mathrm{d}x = u > t,$$

whence $\alpha_\theta(t) \le u$ for all large θ. Hence,

$$\limsup \alpha_\theta(t) \le u, \qquad \text{for all } u > t,$$

so that $\limsup \alpha_\theta(t) \le t$. A similar argument shows that $\liminf \alpha_\theta(t) \ge t$, so that $\alpha_\theta(t) \to t$, and $Z_\theta(t) \to B(t)$.

For the weak-convergence result, we need to prove that $Z_\theta(t)$ converges to $B(t)$ uniformly on compact t-intervals. However, because of the uniform continuity of $B(\cdot)$ on compact intervals, it is only necessary to prove uniform convergence on compact intervals of $\alpha_\theta(t)$ to t. However, it is elementary that a sequence of monotone functions which converges pointwise to a continuous function converges uniformly on compact intervals. □

Why is the method successful? Lévy's theorem characterizes Brownian motion as the unique path-continuous process $B(t)$ such that both $B(t)$ and

$(B(t)^2 - t)$ are martingales. We embedded the original Markov chain Y is a richer 'Brownian' structure for which we can assert that $Z_\theta(t)$ and $(Z_\theta(t)^2 - \alpha_\theta(t))$ are (local) martingales, although not path-continuous ones. We then showed that $\alpha_\theta(t)$ converges uniformly to t on compact intervals. This makes the convergence of the law of Z_θ (equivalently, that of Y_θ) to Wiener measure very plausible. *The nice process α_θ does not exist on the original impoverished sample space of Y*; and we cannot study the 'quadratic' properties of Y and Y_θ as neatly on their own sample spaces. This difficulty is amongst those which must be confronted in more general situations.

References

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Christ, N. H., Friedberg, R., and Lee, T. D. (1982). Weights of links and plaquettes in a random lattice. *Nuclear Physics B, Field Theory and Statistical Systems*, **B210**, 337-346.

Ethier, S. N., and Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence.* Wiley, New York.

Parthasarathy, K. R. (1967). *Probability Measures on Metric Spaces.* Academic Press, New York.

Rogers, L. C. G., and Williams, D. (1987). *Diffusions, Markov Processes, and Martingales. Volume 2: Itô Calculus.* Wiley, Chichester and New York.

Sibson, R. (1980). A vector identity for the Dirichlet tessellation. *Mathematical Proceedings of the Cambridge Philosophical Society*, **87**, 151-155.

Williams, D. (1979). *Diffusions, Markov Processes, and Martingales. Volume 1: Foundations.* Wiley, Chichester and New York. (New Edition, joint with L. C. G. Rogers, to appear 1994).