SCALING RANDOM WALKS ON ARBITRARY SETS

by

Simon C. Harris†, Robin Sibson* and David Williams†

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

* *The Registry, University of Kent at Canterbury*
Canterbury CT2 7NZ, United Kingdom

1. INTRODUCTION.

Let I be a countably infinite set of points in $\mathbb{R}$ which we can write as $I = \{u_i : i \in \mathbb{Z}\}$, with $u_i < u_{i+1}$ for every i and where $u_i \to \pm\infty$ if $i \to \pm\infty$. Consider a continuous-time Markov chain $Y = \{Y(t) : t \geq 0\}$ with state space I such that:

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

We remember that the simple symmetric random-walk, when repeatedly rescaled suitably in space and time, looks more and more like a Brownian motion. In this paper we explore the convergence properties of the Markov chain Y on the set I under suitable space-time scalings. Later, we consider some cases when the set I consists of the points of a renewal process and the jump rates assigned to each state in I are perhaps also randomly chosen.

This work sprang from a question asked by one of us (Sibson) about 'driftless nearest-neighbour' Markov chains on countable subsets I of $\mathbb{R}^d$, work of Sibson [7] and of Christ, Friedberg and Lee [2] having identified examples of such chains in terms of the Dirichlet tessellation associated with I. Amongst methods which can be brought to bear on this d-dimensional problem is the theory of Dirichlet forms. There are potential problems in doing this because we wish I to be random (for example, a realization of a Poisson point process), we do not wish to impose artificial boundedness conditions which would clearly make things work for certain deterministic sets I. In the 1-dimensional case discussed here and in the following paper by Harris, much simpler techniques (where we embed the Markov chain in a Brownian motion using local time) work very effectively; and it is these, rather than the theory of Dirichlet forms, that we use.

(1.1) The Q-matrix of the Markov chain. In terms of the elements of the Q-matrix of Y, the freedom-from-drift requirement may be written in a general form as

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

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

(1.3) $$\ell_i := u_i - u_{i-1}, \qquad r_i := u_{i+1} - u_i.$$

The limitation to nearest neighbour jumps is

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

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

$$q_{i,i+1} = q_i \frac{\ell_i}{\ell_i + r_i}, \qquad q_{i,i-1} = q_i \frac{r_i}{\ell_i + r_i}, \tag{1.4}$$

where $q_i = -q_{i,i}$ is the total jump rate out of state i which we shall not yet fix.

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 the length of this interval by κ_i, namely

$$\kappa_i := \tfrac{1}{2}(\ell_i + r_i). \tag{1.5}$$

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

$$D_i := \sum_{j \neq i} q_{i,j}(u_j - u_i)^2 = \ell_i r_i q_i. \tag{1.6}$$

Finally, we define

$$m_i := \kappa_i D_i^{-1} = \frac{\ell_i + r_i}{2\ell_i r_i q_i}, \tag{1.7}$$

and now notice that because Q is nearest neighbour we have

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

Q is therefore m-symmetrizable. The values m_i are all strictly positive, and define a measure on I which we denote by m. This is the (unique up to constant multiples) invariant measure for Y, and so, in various senses, descibes how Y shares out its time amongst the states in I.

Now we look at the 'average amount of diffusion' that takes place over intervals; integrate D with respect to m over the intersection of I with a large compact interval $K = [A, B]$, then

$$\sum_{u_i \in K} m_i D_i = \sum_{u_i \in K} \kappa_i = \lambda[K] + \text{'end effect'}, \tag{1.8}$$

where $\lambda[K]$ is the Lebesgue measure $|B-A|$ of K.
[*Note:* The choice of normalization in our definition of m was chosen so as to give equality rather than proportionality in (1.8). Let u_L and u_R be chosen so that $A \in T_L$ and $B \in T_R$ for the arbitrary set $[A, B]$, then the end effect is bounded by $\kappa_L + \kappa_R$. It will follow 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.]

(1.9) The scaling to Brownian motion. For convenience, we assume that $0 \in I$ and that $Y(0) = 0$. For $\theta > 0$, define the scaled version of the Markov chain Y by

$$Y_\theta(t) := \theta^{-1} Y(\theta^2 t). \tag{1.10}$$

Also, let μ_θ be the measure on $\mathbb{R}$ which assigns a mass m_i/θ to the point u_i/θ, so that

$$\mu_\theta[K] = \theta^{-1} \sum_{u_i \in \theta K} m_i. \tag{1.11}$$

Then it is easily seen that μ_θ is the appropriately normalized invariant measure for Y_θ. Now results (1.2) and (1.8) make it plausible that if μ_θ converges to Lebesgue measure, then Y_θ converges to Brownian motion. This is made precise in the following theorem.

(1.12) THEOREM. *Suppose that, as $\theta \to \infty$,*

$$\mu_\theta[a, b] \to b - a, \quad \textit{whenever } -\infty < a \le b < \infty. \tag{1.13}$$

Then, if B is a standard Brownian motion started at the origin,

$$Y_\theta \Rightarrow B, \tag{1.14}$$

that is, the law of Y_θ converges weakly to the law of B.

Remarks. Our later proof of this theorem may be adapted to show that (1.13) is also *necessary* for property (1.14) to hold. If μ_θ converges weakly to some measure μ, then Y_θ converges to the diffusion with identity scale function and μ as speed measure. In fact, with the scaling of space and time in (1.10), any limit measure μ *must* be a multiple of Lebesgue measure on the positive and negative half-lines.

(1.15) Choice of jump rates. In the light of theorem 1.12, we return to the freedom of choice in the Q-matrix of the chain. Three special cases can be identified.
(i) *The Well-Normalized Case.* Consider the case where D_i is constant, say $D_i \equiv D$. This implies that $m_i = \kappa_i D^{-1}$ for every i, so that $\mu_\theta[a, b]$ differs from $D^{-1}(b-a)$ only through an end effect. Then provided I satisfies the condition introduced above

to control irregularity, convergence is assured. The limit has unit speed when $D = 1$ and we get a standard Brownian motion. In many senses this is the most natural normalization, and so we call it the *well-normalized* case.
(ii) *The Symmetric Case.* Consider the case where we hold $m_i \equiv 1$; this corresponds to

$$q_{i,i+1} = q_{i+1,i} = 1/\big(2(u_{i+1} - u_i)\big), \qquad \text{for all } i, \tag{1.16}$$

and is appropriately called the *symmetric* case. In this case, the measure m is the counting measure and the condition (1.13) can now be thought of as requiring asymptotic uniformity of spacing of the points of set I.
(iii) *The Decoupled Case.* If we set q_i equal to a constant then the total jump rate out of each state does not depend on the geometry, unlike the first two cases where it is adapted to it to some extent. Thus, we call it the *decoupled* case.

2. RANDOM WALKS ON RENEWAL PROCESSES.

Now that we have the 'basic' weak convergence theorem, our interest shifts to some cases where the state space and the jump rates of the Markov chain are generated in a random fashion.

Let $\{G_i : i \in \mathbb{Z}\}$ be a family of independent identically distributed random variables each distributed like some strictly positive random variable G. Define

$$u_i = \begin{cases} \sum_{j=1}^{i} G_j & \text{for } i > 0, \\ 0 & \text{for } i = 0, \\ -\sum_{j=1+i}^{0} G_j & \text{for } i < 0. \end{cases} \tag{2.1}$$

We now define the state space $I := \{u_i : i \in \mathbb{Z}\}$. The G_i random variables give the gaps between neighbours of I, that is $u_i - u_{i-1} = G_i$. As usual we start the chain from the origin, defining $Y(0) := u_0 = 0$.

We set the jump rates for the chain according to the well-normalized, symmetric or decoupled cases. At other times, we let the jump rate of the chain out of site u_i be $q_i := R_i$, where $\{R_i : i \in \mathbb{Z}\}$ is a family of independent identically distributed random variables each distributed like some strictly positive random variable R. We let $J := \{q_i : i \in \mathbb{Z}\}$.

The information about the state space and jump rates now specifies the Q matrix exactly. We have a nearest neighbour, driftless Markov chain which has a *random state space* and *random jump rates.*

Weak Convergence.

In looking at scalings of the 'randomized' Markov chains described above, randomness has entered at more than one stage of the problem, and appropriate types of convergence need to be defined with greater care. Background can be found in Billingsley [1], Ethier and Kurtz [3], Parthasarathy [4] and Rogers and Williams [5].

We are dealing with paths of a scaled random walk which are in the space of right continuous functions, that is $Y_\theta = \{Y_\theta(t) : t \geq 0\} \in D_{\mathbb{R}}[0,\infty)$. When comparing paths of the chain we use the *Skorokhod metric*, d, and then $(D_{\mathbb{R}}[0,\infty), d)$ is a complete, separable metric space. Heuristically, in saying the Skorokhod distance between two right-continuous paths is small, we are allowing not only small variations between the path heights at *fixed* times, but we also permit some flexibility in the time scale to allow for (possibly large) jumps occurring in each path at times *near* each other.

The path of the Markov chain, Y_θ, is a $D_{\mathbb{R}}[0,\infty)$-valued random variable with probability measure $\mathbb{P}_\theta \in \mathcal{P}\big(D_{\mathbb{R}}[0,\infty)\big)$ given by

$$\mathbb{P}_\theta(S) := \mathbb{P} \circ Y_\theta^{-1}(S) = \mathbb{P}(Y_\theta \in S) \quad \forall S \in \mathcal{S},$$

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,

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

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 state space, I, on which it lives and the jump rates, J, out of each state. Hence define the 'random probability law' $\mathcal{L}_\theta$ to be the *regular conditional probability of* $\mathbb{P} \circ Y_\theta^{-1}$ *given* (I, J), which we think of as

$$\mathcal{L}_\theta(S) := \mathbb{P}\big(Y_\theta \in S \,\big|\, \sigma(I, J)\big).$$

$\mathcal{L}_\theta$ is a $\mathcal{P}(D_{\mathbb{R}}[0,\infty))$-valued random variable which is $\sigma(I, J)$-measurable.

We define the *global law* $\mathbb{P}_{\mathcal{L}}$ associated with the random law $\mathcal{L}$ by

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

and avoid double-subscripting by writing $\mathbb{P}_\theta$ for this construct when $\mathcal{L}$ is $\mathcal{L}_\theta$. Then, as is natural in our case of interest, $\mathbb{P}_\theta$ is the global law for the random law $\mathcal{L}_\theta$.

(2.2) Definitions. *Suppose we have a sequence of random probability laws $\mathcal{L}_\theta$ and a random probability law $\mathcal{L}$, then we say*

(i) '$\mathcal{L}_\theta$ weakly converges in probability to $\mathcal{L}$', *denoted by*

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

(ii) '$\mathcal{L}_\theta$ weakly converges almost surely to $\mathcal{L}$', *denoted by*

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

As usual, the latter condition implies the former. Also, weak convergence in probability of random laws to a (possibly random) limit law implies weak convergence of the global laws:

$$\mathcal{L}_\theta \overset{a.s.}{\Rightarrow} \mathcal{L} \textit{ implies that } \mathcal{L}_\theta \overset{p}{\Rightarrow} \mathcal{L} \textit{ implies that } \mathbb{P}_\theta \Rightarrow \mathbb{P}_{\mathcal{L}}.$$

Some Further Results.

We first present some elementary corollaries of the main theorem 1.12 which cover many of the randomized state space and jump rate cases. However, we also state a theorem which demonstrates how the scaling procedure and convergence can be strongly affected by the gap distribution generating the state space. In particular, even a driftless, nearest-neighbour decoupled Markov chain on the points of a Poisson process *fails* to fit into the 'standard' convergence framework.

(2.3) COROLLARY. *Let G be a strictly positive random variable with $\mathbb{E}G < \infty$ that specifies the distribution of I. Suppose we choose the jump rates, q_i, according to one of the following situations:*

(i) Well normalized case. *Let $D_i \equiv \sigma$ for some constant $\sigma > 0$.*

(ii) Symmetric case. *Let $m_i \equiv 1$ and define $\sigma^2 := \mathbb{E}G$.*

(iii) Decoupled case. *Let $q_i \equiv 1$ and define*

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

(iv) Independent case. *Let $q_i := R_i$ where the R_i are independent, each distributed like some positive random variable R. Define*

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

Then, for each separate case, we find that $\mathcal{L}_\theta \overset{a.s.}{\Rightarrow} \mathcal{L}_{\sigma B}$, the law of Brownian motion with variance coefficient σ.

Proof. 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,

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

we see that μ_θ assigns mass approximately $(b-a)/\sigma^2$ to $[a, b]$. We can now specialize this to each of the above situations easily. □

We notice from corollary 2.3 that if the *gap distribution has a finite mean*, but the *reciprocal gap has infinite mean*, then we will get weak convergence to the zero process in the decoupled case. In these situations, the occurrence of many clusters of close points between which Y makes many time-consuming oscillations results in the limit process getting 'trapped'. We may hope that in some of these cases, a suitable adjustment of the time scaling may yield some other non-trivial convergence. The following example is typical of what can happen here.

(2.4) Decoupled chain with Gamma gap distribution. Let the common gap distribution of the state space, G, have a Gamma distribution with parameter $k > 0$, where

$$\mathbb{P}(G \in dx) = \Gamma(k)^{-1} x^{k-1} e^{-x}\, dx$$

and $\Gamma(k)$ is the gamma function. Some calculations reveal that

$$\mathbb{E}G = k, \qquad \mathbb{E}(1/G) = \begin{cases} 1/(k-1) & \text{for } k > 1, \\ +\infty & \text{for } k \leq 1. \end{cases}$$

Consider the *decoupled* case, so that $q_i \equiv 1$. With this setup, we have the following result.

(2.5) THEOREM.

(a) Case $k > 1$. Finite mean reciprocal gap. *Define*

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

which has a random law $\mathcal{L}_\theta$. Let $\mathcal{L}_{\sigma B}$ be the (deterministic) law of a Brownian motion with variance coefficient $\sigma^2 := k(k-1)$. Then

$$\mathcal{L}_\theta \overset{a.s.}{\Rightarrow} \mathcal{L}_{\sigma B}.$$

(b) Case $k = 1$. Poisson gaps. *Define*

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

which has a random law $\mathcal{L}_\theta$. Let $\mathcal{L}_B$ be the (deterministic) law of a Brownian motion with unit variance coefficient. Then

$$\mathcal{L}_\theta \overset{p}{\Rightarrow} \mathcal{L}_B\ .$$

We do not have almost sure convergence of the random laws, that is $\mathcal{L}_\theta \overset{a.s.}{\not\Rightarrow} \mathcal{L}_B$.

(c) Case $k < 1$. Infinite mean reciprocal gap. *Define*

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

which has a random law $\mathcal{L}_\theta$. Let $\mathbb{P}^k_{FM}$ be the probability measure of a Feller-McKean chain driven by the jumps of a stable process of characteristic exponent k (see below). Now we find

$$\mathbb{P}_\theta \Rightarrow \mathbb{P}^k_{FM}.$$

However, the random laws $\mathcal{L}_\theta$ are not even weakly convergent in probability.

The first part of this theorem is simply a special case of corollary 2.3(iii). The proof of the other parts require special treatment and will be given in a paper by Harris. The last case is when the many small gaps cause ‘clustering’ in the resulting scaled state space, which now becomes a countable, dense subset of the line. However, we still end up with the nearest process to Brownian motion that we can given the circumstances, that is, a Brownian motion viewed only at times when it visits the ‘clustered’ state space. This is a randomised Feller-McKean chain, as now described.

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

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

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 I and leaves each state instantly - its Q matrix has $-\infty$ down the diagonal and zeroes elsewhere, hence they are also referred to as *instantaneous Markov chains.* As is apparent from our construction, it is a time change of a Brownian motion. See Rogers and Williams [5] for more details.

A *randomized Feller-McKean chain* can be constructed by choosing (I, m) randomly in the following manner. For $k < 1$ we take the increasing 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, I, with only finitely many large jumps. Thus it induces a random measure, $dX(y)$, which consists of a dense set I of atoms with masses $m_y := X(y) - X(y-)$ at $y \in I$. Hence we can define a Feller McKean chain as above, but with respect to the random pair (I, m) generated from X. That is, define the process Z by

$$Z(t) := B(\tau_t), \qquad \tau_t := \inf\{u : A_u > t\}, \qquad A_u := \int_{y \in \mathbb{R}} L(u, y) \, dX(y).$$

The Brownian motion is taken to be independent of the stable process. Finally, $\mathbb{P}^k_{FM}$ is defined to be the probability measure for Z.

3. PROOF OF THEOREM 1.12.

The idea of our proof is to first construct the Markov chain Y by embedding it within a Brownian motion. This brings in the use of *local time* for a Brownian motion and the invariant measure for Y. Although this is a method specific to the one-dimensional picture, it is also the most powerful and gives the clearest intuitive picture of the scaling process. We proceed with a fairly heuristic fashion, but note that all steps can be tightened to a rigorous level, full details of which can be found in Rogers and Williams [5]&[6].

(3.1) Brownian Local time. Let $B = \{B(t) : t \geq 0\}$ be a standard Brownian motion started at the origin. Trotter's theorem tells us of the existence of a local time $\{L(t,x) : t \geq 0, x \in \mathbb{R}\}$ which is jointly continuous in t and x, and such that, for any 'nice' function f on $\mathbb{R}$,

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

(In this context, 'nice' can be regarded as anything from 'C^∞ of compact support' to 'non-negative Borel-measurable' without changing the sense.) Local time can therefore be considered as the occupation density of Brownian motion.

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. Heuristically, we can think of $L(t,x)$ as

$$\int_0^t \delta\big(B(s) - x\big)\, ds,$$

where δ is the Dirac delta function concentrated at the origin.

Now it is well known that if we start a Brownian motion B at the origin, the amount of local time (at zero) that it clocks up before first hitting either r or $-\ell$ (where $r, \ell > 0$) is exponentially distributed with mean $2r\ell/(r+\ell)$. Further, the probability of exiting the interval at ℓ is given by $r/(r+\ell)$.

If Y is started in state u_i, then the time until the chain leaves this point to enter into $N_i := \{u_{i-1}, u_{i+1}\}$ is exponentially distributed (independently of everything else) with parameter q_i, thus the mean holding time is q_i^{-1}. So if we start a Brownian motion at u_i, then m_i multiplied by the local time spent at u_i before hitting N_i is also an exponentially distributed random variable of mean q_i^{-1}. Thus, we can find the Y process in the Brownian motion if we view it only when it visits I and have as our natural clock the sum of local times at points of I weighted by the invariant measure. This is expressed more precisely in the following lemma.

(3.2) LEMMA. 'Construction of the Markov chain using a Brownian motion.' *Consider the process Y constructed from the Brownian motion B, with associated local time process $L(t,x)$, as follows:*

$$Y(t) := B\big(\alpha(t)\big), \tag{3.3}$$

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

$$\int_{x\in\mathbb{R}} L\big(\alpha(t), x\big)\, m(dx) = \sum_{i\in\mathbb{Z}} m_i L\big(\alpha(t), u_i\big) = t. \tag{3.4}$$

Then Y is a driftless Markov chain which has state space I and moves only between nearest neighbours of I.

Henceforth, we can (and do) think of the chain Y as being constructed in this way.

We can now consider pathwise rescalings of Y and utilise the natural rescaling of Brownian motion, where $\widetilde{B}(t) := \theta^{-1}B(\theta^2 t)$ is also a Brownian motion started at the origin. Let $\widetilde{L}$ be the local time process for $\widetilde{B}$. For the moment consider θ as fixed, then

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

Now we call upon the following:

(3.5) Lemma. *We have*

$$Y_\theta(t) = \widetilde{B}\big(\widetilde{\alpha}_\theta(t)\big), \tag{3.6}$$

where $\widetilde{\alpha}(t)$ is the largest solution of

$$\int_{x\in\mathbb{R}} \widetilde{L}\big(\widetilde{\alpha}(t), x\big)\, \mu_\theta(dx) = t. \tag{3.7}$$

Proof of Lemma 3.5. For a 'nice' function f on $\mathbb{R}$ and $s > 0$, we have

$$\int_{z\in\mathbb{R}} f(z)\widetilde{L}(s,z)\, dz = \int_{r=0}^{s} f\big(\widetilde{B}(r)\big)\, dr = \int_{r=0}^{s} f\big(\theta^{-1}B(\theta^2 r)\big)\, dr$$

$$= \theta^{-2}\int_{u=0}^{\theta^2 s} f\big(\theta^{-1}B(u)\big)\, du = \theta^{-2}\int_{y\in\mathbb{R}} f(\theta^{-1})L(\theta^2 s, y)\, dy$$

$$= \theta^{-1}\int_{z\in\mathbb{R}} f(z)L(\theta^2 s, \theta z)\, dz,$$

so that we have

$$\widetilde{L}(s,z) = \theta^{-1} L(\theta^2 s, \theta z) \qquad \text{for all } s \text{ and } z. \tag{3.8}$$

The result is now a rephrasing of Lemma 3.2 with

$$\widetilde{\alpha}_\theta(t) = \theta^{-2}\alpha(\theta^2 t). \tag{3.9}$$

□

We have done a rescaling for a *fixed* θ where it should be noted that the new Brownian motion used, $\widetilde{B}$, is constructed pathwise in a θ *dependent* fashion from the original B. Then given a *particular* path of B, as we run through the scaling procedure, our pathwise construction of Y_θ process from B naturally leads to an ever changing path picture. However, to simplify completion of the proof, we note that we are allowed to alter the construction of each Y_θ as long as we maintain their laws - it is only the sequence of laws that we are interested in. Observe that we can use one fixed Brownian path that we do *not* scale and find each chain by 'watching' the Brownian path on the required lattice.

Then the process Y_θ is equal in law to the process Z_θ, where

$$Z_\theta(t) := B\big(\alpha_\theta(t)\big), \ \textit{and } \alpha_\theta(t) \textit{ is the largest solution of } \int_{x\in\mathbb{R}} L\big(\alpha_\theta(t)\big)\mu_\theta(dx) = t.$$

Suppose now that property (1.13) holds. Consider t as fixed for the moment. Let $u > t$ be fixed. Since $x \mapsto L(u,x)$ is continuous of compact support and since property (1.13) implies the weak convergence of μ_θ to Lebesgue measure,

$$\int_{x\in\mathbb{R}} L(u,x)\,\mu_\theta(dx) \to \int_{x\in\mathbb{R}} L(u,x)\,dx = u > t,$$

whence

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

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 $Z_\theta(t) \to B(t)$.

Now, proving that $Z_\theta(t)$ converges to $B(t)$ uniformly on compact t-intervals will certainly give us the weak convergence result. However, because of the uniform continuity of $B(\cdot)$ on compact time intervals, it is only necessary to prove uniform convergence on compact time intervals of $\alpha_\theta(t)$ to t. This is clear, since a sequence of monotone functions converging pointwise to a continuous function converges uniformly on compact intervals. □

Remark. Lévy's theorem characterizes Brownian motion by the unique path-continuous process $B(t)$ such that both $B(t)$ and $B(t)^2 - t$ are martingales. We embedded the Markov chain Y in a richer 'Brownian' structure where we can assert that $Z_\theta(t)$ and $Z_\theta(t)^2 - \alpha_\theta(t)$ are (local) martingales, if not path-continuous ones. Then showing that $\alpha_\theta(t)$ converges uniformly to t on compact time intervals and that Z_θ tends to have smaller and smaller jumps makes the convergence of the law of Z_θ to Wiener measure very plausible. Note, *the process $\alpha_\theta(t)$ 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.

REFERENCES

[1] P. BILLINGSLEY. *Convergence of Probability Measures.* Wiley, New York, (1968).

[2] N. H. CHRIST, R. FRIEDBERG and T. D. LEE. Weights of links and plaquettes in a random lattice. *Nuclear Physics B, Field Theory and Statistical Systems*, **B210** (1982), 337-346.

[3] S. N. ETHIER and T. G. KURTZ. *Markov Processes: Characterization and Convergence.* Wiley, New York, (1986).

[4] K. R. PARTHASARATHY. *Probability Measures on Metric Spaces.* Academic Press, New York, (1967).

[5] L. C. G. ROGERS and D. WILLIAMS. *Diffusions, Markov Processes, and Martingales. Volume 1: Foundations.* (New Edition). Wiley, Chichester and New York, (1994).

[6] L. C. G. ROGERS and D. WILLIAMS. *Diffusions, Markov Processes, and Martingales. Volume 2: Itô Calculus.* Wiley, Chichester and New York, (1987).

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