ANNALES DE L'I. H. P., SECTION B

MARTIN BAXTER

Discounted additive functionals of Markov processes

Annales de l'I. H. P., section B, tome 32, nº 5 (1996), p. 623-644

<http://www.numdam.org/item?id=AIHPB_1996__32_5_623_0>

Ann. Inst. Henri Poincaré,

Vol. 32, n° 5, 1996, p. 623-644 *Probabilités et Statistiques*

Discounted additive functionals of Markov processes

by

Martin BAXTER

Statistical Laboratory, University of Cambridge, Cambridge CB2 1SB.

ABSTRACT. – When taking the long term time average of the occupation measure of a Markov process, traditionally the uniform time average has been used. In this paper, we derive results for more general average shapes, such as the exponential discount. We obtain large-deviation results and even stronger density asymptotics for small rates of discount. A strong law and a central limit theorem are also proved.

RÉSUMÉ. – **Les fonctionnelles additives pondérées des processus de Markov**. La moyenne à long terme de la mesure d'occupation d'un processus Markovien est traditionnellement calculée en utilisant une moyenne uniforme. Nous étudions ici quelques formes plus générales de moyennes, telles que la pondération exponentielle. Nous obtenons un résultat de grandes déviations ainsi que le comportement asymptotique de la densité dans le cas d'un taux de pondération petit. Enfin, nous prouvons une loi forte et un théorème central limite.

1. INTRODUCTION

Much study has been made of the time averages of random processes. Most of this effort has been directed towards the Cesàro average which weights times uniformly up to a finite horizon. In this paper, we shall derive some results about more general averages. The initial motivation was the exponential discount (we will call this the Abel average), which appears frequently in the contexts of, among others, systems control and models of financial markets. The techniques developed, however, extend easily to other shapes of discount.

In this paper we will push forward two strands of inquiry that have been developed in proceeding work. The generality that we will obtain should strictly encompass the existing material and can be read on its own, though reference will be made to earlier papers when a proof is essentially the same as one before.

For general results about the large-deviations behaviour of discounted occupation times of processes, in Section 2 we will build on Section (a) of Baxter and Williams [2] and on Baxter [3]. Previously we knew that the large-deviation property held for the Abel discounted average of a general finite-state Markov chain, and we will fully extend this to completely general discounts of chains and partially extend further to a wide class of Markov processes. We discover that although the large-deviation rate function of the discounted average can be written in terms of that for the Cesàro, the rate is often different for a different discount. We also derive results about the smoothness and finiteness of the rate function which are used in the next Section to prove a central limit theorem.

Finally we shall go beyond the limited approximation precision of the large-deviation property and give an asymptotic expansion of the density of the distribution itself, following from the density studies of Section (b) of Baxter and Williams [2]. This again is now performed for general discounted averages of finite-state Markov chains. We also notice a pattern in the differential equations we worked with, and hypothesize about their full solutions and other generalizations.

The main objects studied are the Cesàro and Abel averages of a process X. Both are random measures of unit mass on the set S, the state space of X. The former is defined as

$$C_t(F) := \frac{1}{t}\int_0^t I_F(X_s)\,ds, \qquad \text{for } F \text{ measurable in } S, t > 0.$$

For any density m on $[0,\infty)$, that is for m in $L_1^+(\mathbb{R}^+)$, the Abel average A_λ is

$$A_\lambda(F) := \lambda\int_0^\infty m(\lambda t) I_F(X_t)\,dt, \qquad \text{for } F \text{ in } S, \lambda > 0.$$

The two main results of the paper are:

THEOREM A. – *Suppose that X is an irreducible Markov chain on a finite state-space S, with Q-matrix Q. Then both C_t and A_λ obey the large-deviation principle with rate functions I and K respectively. That is that*

$$\limsup_{t\to\infty} t^{-1}\log \mathbb{P}(C_t \in F) \le -\inf_{\nu\in F} I(\nu),$$
$$\liminf_{t\to\infty} t^{-1}\log \mathbb{P}(C_t \in G) \ge -\inf_{\nu\in G} I(\nu),$$
$$\limsup_{\lambda\downarrow 0} \lambda\log \mathbb{P}(A_\lambda \in F) \le -\inf_{\nu\in F} K(\nu),$$
$$\liminf_{\lambda\downarrow 0} \lambda\log \mathbb{P}(A_\lambda \in G) \ge -\inf_{\nu\in G} K(\nu),$$

for F and G respectively closed and open subsets of $M := M_1(S)$. Additionally, K is related to I by the equation

$$K^*(v) = \int_0^\infty I^*(m(t)v)\,dt, \qquad v \text{ in } \mathbb{R}^S,$$

where I^* and K^* are the Legendre transforms (convex conjugates) of I and K, satisfying

$$I^*(v) = \sup_{x\in M}(\langle v,x\rangle - I(x)),$$
$$I(x) = \sup_{v\in\mathbf{R}^S}(\langle v,x\rangle - I^*(v)),$$
$$K^*(v) = \sup_{x\in M}(\langle v,x\rangle - K(x)),$$
$$K(x) = \sup_{v\in\mathbf{R}^S}(\langle v,x\rangle - K^*(v)).$$

THEOREM B. – *Additionally, if m is of bounded variation, then the density of A_λ on M under the law starting X at i, f_i^λ can be written as*

$$f_i^\lambda(x) = e^{-K(x)/\lambda}(2\pi\lambda)^{-(n-1)/2}(\det H_K(x))^{1/2} z_i(x) r_i^\lambda(x),$$

where H_K is the Hessian of K taken with respect to M, $z(x)$ is the positive eigenvector of $Q + \mathrm{diag}(m_0\nabla K(x))$, and the residue term r^λ goes to 1 in the sense that $r_i^\lambda(x)\,dx$ converges weakly to dx on $\mathrm{Int}(M)$.

2. LARGE DEVIATIONS

We will work with general discount shapes and positive recurrent processes. Let X be a stochastic process with state-space E and invariant distribution π. In the Cesàro case we would expect some sort of ergodic theorem such as

$$C_t(F) := \frac{1}{t}\int_0^t I_F(X_s)\,ds \to \pi(F) \qquad \text{as } t \to \infty, \tag{1}$$

for all measurable subsets F of E. Then C_t takes values in $M_1(E)$, the space of probability measures on E. We might also have a large-deviation result, which we can think of for the moment as the slogan

$$\text{“}\mathbb{P}(C_t \in H) \approx \exp\big(-t \inf_{\nu \in H} I(\nu)\big) \ \text{ as } t \to \infty\text{”} \qquad H \subset M_1(E),$$

for some rate function I, with $I(\pi) = 0$. The space $M_1(E)$ and the continuous bounded functions on E, $C_b(E)$ are in duality via the bracket, $\langle v, \nu\rangle = \int_E v(x)\nu(dx)$. A related slogan is that of the Laplace transform

$$\text{“}\mathbb{E}\exp\big(\langle v, C_t\rangle t\big) \approx \exp\big(t\delta(v)\big) \ \text{ as } t \to \infty\text{”} \qquad v \in C_b(E),$$

where δ and I are related by Legendre transformation (convex conjugation), in that

$$I(\nu) = \delta^*(\nu) := \sup_{v \in C_b(E)} \big(\langle v, \nu\rangle - \delta(v)\big), \tag{2}$$

$$\delta(v) = I^*(v) := \sup_{\nu \in M_1(E)} \big(\langle v, \nu\rangle - I(\nu)\big). \tag{3}$$

Our program will be to study, for any discount density m in $L_1^+(\mathbb{R}^+)$, the average

$$A_\lambda(F) := \int_0^\infty \lambda m(\lambda t) I_F(X_t)\,dt. \tag{4}$$

We will show that $A_\lambda \to \pi$ as λ goes to 0, and that the large-deviation principle holds with rate K whose Legendre transform η is given by the equation

$$\eta(v) = \int_0^\infty \delta\big(m(t)v\big)\,dt. \tag{5}$$

This is actually the same as the η-equation at (1.16) in Baxter and Williams [2] (with discount $m_t = e^{-t}$) and at Theorem C in Baxter [3] (with discount $m_t = (1-t)^\gamma$), but (5) is a more natural formulation.

Standard set-up. – Let X be an ergodic Feller-Dynkin Markov process on a locally compact Polish space E, with generator L. We define the

Cesàro average C_t and the general average A_λ by (1) and (4) respectively. Then C_t and A_λ converge to π, the invariant distribution of X, with respect to the weak topology on $M_1(E)$, that is, in the sense of (1). A sufficient condition for the former limit is that, as in 8.11.2 of Bingham *et al.* [5], π is a limiting distribution of the transition semigroup (P_t). The latter limit follows from the former by a similar L_1-continuity argument to that which will be used in the proof of Theorem 1. Deuschel and Stroock [6] show that under an assumption of uniform ergodicity the large-deviation property holds for C_t with rate function I defined on $M_1(E)$. That is that

$$\begin{aligned} &\limsup_{t\to\infty} t^{-1}\log \mathbb{P}(C_t \in F) \le -\inf_{\nu\in F} I(\nu), \qquad (6)\\ \text{and}\qquad &\liminf_{t\to\infty} t^{-1}\log \mathbb{P}(C_t \in G) \ge -\inf_{\nu\in G} I(\nu), \end{aligned}$$

for F and G respectively closed and open subsets of $M_1(E)$. We learn from 4.2.17 of Deuschel and Stroock [6] that

$$\lim_{t\to\infty} t^{-1}\log \mathbb{E}\exp\int_0^t v(X_s)\,ds = \delta(v), \tag{7}$$

where δ and I are convex functions satisfying (2) and (3). Further there are, by 4.2.27 and 4.2.38 of Deuschel and Stroock [6], explicit expressions for δ and I as

$$\begin{aligned} \delta(v) &= \lim_{t\to\infty} t^{-1}\log \|P_t^v\|_{\mathrm{op}},\\ &\quad\text{where } P_t^v f(x) := \mathbb{E}_x\left(\exp\left(\left(\int_0^t v(X_s)\,ds\right) f(X_t)\right)\right), \qquad (8)\\ I(\nu) &= \sup\left\{-\int_E \frac{Lf(x)}{f(x)}\,\nu(dx) : f\ge 1,\ f\in \mathrm{Dom}(L)\right\}. \qquad (9) \end{aligned}$$

As in Baxter and Williams [2] we shall be particularly interested in the case where X is a Markov chain on a finite state-space S with Q-matrix Q. Then $\delta(v) = \sup\{\mathrm{Re}(z) : z \in \mathrm{spect}(Q+V)\}$, where V denotes the diagonal matrix $\mathrm{diag}(v)$ and $\mathrm{spect}(\cdot)$ denotes spectrum (here the set of eigenvalues). This expression for δ also holds in the general Markov process setting, if the generator L is π-symmetric.

We begin by proving a result whose first part is similar to one remarked by Kifer [9] in the context of the large-deviations of the averages of dynamical systems, but it is the second part which will be more useful in our further work. In earlier papers we derived a differential equation by the self-similarity of discount shapes such as e^{-t}, but it is enough to study the shifts of the discount along the time-axis, which provides a useful one-dimensional parameterisation.

THEOREM 1. – *Suppose that X is an FD Markov process on a space E, with generator L, and m is any density on $[0, \infty)$, and for x in E and v in $C_b(E)$ the limit $\delta(v) = \lim_\lambda \lambda \log \mathbb{E}_x \exp \int_0^{1/\lambda} v(X_s)\, ds$ exists uniformly in x on E. If we define φ by*

$$\varphi(x, t, \lambda, v) = \mathbb{E}_x \exp \int_0^\infty \theta_t m(\lambda s) v(X_s)\, ds, \tag{10}$$

where θ_t is the shift operator $\theta_t f(s) = f(t+s)$, then

$$\lim_{\lambda \downarrow 0} \lambda \log \varphi(x, 0, \lambda, v) = \eta(v) := \int_0^\infty \delta(m_t v)\, dt. \tag{11}$$

Further, $\varphi(\cdot, t, \lambda, v)$ is in the domain of L and $\varphi(x, \cdot, \lambda, v)$ is differentiable and

$$-\frac{\partial \varphi}{\partial t} = \lambda^{-1}(L + m_t V)\varphi. \tag{12}$$

Proof of Theorem 1. – We prove the first part using continuity arguments. If we define

$$H(\lambda, x, \alpha) := \lambda \log \mathbb{E}_x \exp \int_0^{1/\lambda} \alpha v(X_s)\, ds, \tag{13}$$

then $\sup_x |H(\lambda, x, \alpha) - \delta(\alpha v)|$ goes to 0 as λ does. We start by proving the general average limit for m of the form

$$m = \sum_{i=1}^n c_i I(a_i, b_i),$$

where $\{(a_i, b_i)\}$ are disjoint intervals in $\mathbb{R}^+$ and $c_i > 0$. Set

$$Y_i^\lambda := \exp\left(\int_{a_i/\lambda}^{b_i/\lambda} c_i v(X_s)\, ds - \lambda^{-1}(b_i - a_i)\delta(c_i v)\right), \tag{14}$$

and define $y_i^\lambda(x) := \mathbb{E}_x Y_i^\lambda$. Then for λ sufficiently small $|\lambda \log y_i^\lambda(x)| < \epsilon$ uniformly in x. Thus

$$\begin{aligned}\frac{\mathbb{E}_x \exp \int_0^\infty m(\lambda t) v(X_t)\, dt}{\exp \lambda^{-1} \int_0^\infty \delta(m_t v)\, dt} &= \mathbb{E}_x(Y_1^\lambda \dots Y_n^\lambda) = \mathbb{E}_x(Y_1^\lambda \dots Y_{n-1}^\lambda y_n^\lambda(X_{a_n/\lambda}))\\ &\le e^{\epsilon/\lambda} \mathbb{E}_x(Y_1^\lambda \dots Y_{n-1}^\lambda) \le \dots \le e^{n\epsilon/\lambda},\end{aligned}$$

and so we have the right upper bound. Similarly we have the lower bound.

Let us now define $I_\lambda(m) := \int_0^\infty m(\lambda t)v(X_t)\,dt$, $L_\lambda(m) := \lambda \log \mathbb{E} \exp I_\lambda(m)$, and $J(m) := \int_0^\infty \delta(m_t v)\,dt$. Then $\lambda|I_\lambda(m_1) - I_\lambda(m_2)|$ is bounded by $\|v\|_\infty \|m_1 - m_2\|_1$ uniformly in ω, and hence $|L_\lambda(m_1) - L_\lambda(m_2)|$ has the same bound. Because $|\delta(v_1) - \delta(v_2)| \leq |v_1 - v_2|$, the same bound also dominates $|J(m_1) - J(m_2)|$. This L_1-continuity and (careful) application of monotone class theorems let us generalize firstly to all bounded m of compact support and then to all m in $L_1^+(\mathbb{R}^+)$.

For the differential equation, we need only apply the Feynman-Kač formula to the space-time process $Y_t := (X_t, \tau_t)$, where $\tau_t = \tau_0 + t$, which has generator $L + \partial_t$. Taking $\lambda = 1$ for simplicity, for v in $C_b(E)$, we define v_Y on $C_b(E \times \mathbb{R}^+)$ by $v_Y(x,t) := m_t v(x)$, and

$$A_t := \int_0^t v_Y(Y_s)\,ds = \int_0^t \theta_{\tau_0} m(s) v(X_s)\,ds. \tag{15}$$

Without loss of generality we can assume that v is non-negative, because if (11) and (12) hold for some v, then they hold for all vectors of the form $v + \alpha\mathbf{1}$, where $\mathbf{1}$ is the constant vector $(1, 1, \ldots, 1)$. This shifting identity follows from the fact that $\delta(v + \alpha\mathbf{1}) = \delta(v) + \alpha$, a property which η inherits. Then the semigroup P^v defined by

$$P_t^v f(y) = \mathbb{E}_{Y_0 = y}\big(e^{A_t} f(Y_t)\big)$$

has generator $L^v := L + \partial_t + m_t V$, as seen in, for example, III.39 of Williams [12]. Then if we set $\varphi((x,t)) := \varphi(x,t,1,v)$, which is continuous in t, we have that

$$P_t^v \varphi(Y_0) = \mathbb{E}_{Y_0}\big(e^{A_t}\varphi(Y_t)\big) = \mathbb{E}_{Y_0}(\mathbb{E}(\exp A_\infty | \mathcal{F}_t)) = \varphi(Y_0). \tag{16}$$

Thus $t^{-1}(P_t^v - I)\varphi = 0$, implying that φ is in the domain of L^v and is annihilated by it. The equation $L^v \varphi = 0$ is exactly (12). □

We note that in the case of X a standard Brownian motion and the exponential discount $m_t = e^{-t}$ and $v(x) = I(x > 0)$, then (12) is equation (3.5) of Baxter and Williams [1].

COROLLARY 2. – *Suppose that X is an irreducible Markov chain on a finite state-space S, with Q-matrix Q, and m is any density on $[0, \infty)$, and A_λ is defined by (4), then the large-deviation property analogue of (6) holds for A_λ with rate function K,*

$$\limsup_{\lambda \to 0} \lambda \log \mathbb{P}(A_\lambda \in F) \leq - \inf_{\nu \in F} K(\nu),$$

and

$$\liminf_{\lambda \to 0} \lambda \log \mathbb{P}(A_\lambda \in G) \geq - \inf_{\nu \in G} K(\nu), \tag{17}$$

for F and G respectively closed and open subsets of M. The rate function K relates to the η of (11) through the following equations:

$$K(x) = \sup_{v \in \mathbb{R}^S} \langle v, x \rangle - \eta(v), \tag{18}$$

$$\eta(v) = \sup_{x \in M} \langle v, x \rangle - K(x), \tag{19}$$

where $M := M_1(S) = \{(x_i)_{i=1}^n : \sum_i x_i = 1,\ x \geq 0\}$.

Proof of Corollary 2. – We are in the context of Theorem 1 because X will satisfy condition ($\tilde{\text{U}}$) of 4.2.7 of Deuschel and Stroock [6], which is sufficient for the limit δ to exist as required by the theorem. The large-deviation property and (18) come from theorem II.2 of Ellis [7]. In his language, t is our v, Y_n is our A_λ, $c_n(\cdot)$ is our $\lambda \log \varphi(x, 0, \lambda, \cdot)$, and $c(\cdot)$ is our $\eta(\cdot)$. As η is defined and differentiable on the whole of $\mathbb{R}^S$, it meets Ellis' "steep" hypothesis. From (11), η inherits the (strict) convexity and differentiability of δ, which gives (19). □

We complete this Section with a pair of results about the large-deviation rate function K. The former of these is in the spirit of Proposition D of Baxter and Williams [2] and identifies the points where the various suprema in Legendre transforms (18) and (19) are achieved. This leads to central limit results and the major result of the next Section.

PROPOSITION 3. – *Under the conditions of Corollary 2, K is finite, twice differentiable and strictly convex on* $\text{Int}(M)$, *and the supremum of (18) is attained uniquely (up to multiples of* $\mathbf{1}$*) at $v = \nabla K(x)$, and the supremum of (19) is attained uniquely at $x = \nabla \eta(v)$.*

Proof of Proposition 3. – It is immediate from its definition that $\delta(v)/\|v\|_\infty \to 1$ as $\|v\|_\infty \to \infty$ with $v \geq 0$. But as also $|\delta(v)| \leq \|v\|_\infty$, the Dominated Convergence theorem gives us that $\eta(v)/\|v\|_\infty$ goes to 1 as well. Take $x \in \text{Int}(M)$ and suppose there exists a sequence of vectors (v_n) such that

$$\langle v_n, x \rangle - \eta(v_n) \to \infty.$$

Without loss of generality we can replace (v_n) by $(v_n - (\min_i v_n(i))\mathbf{1})$, because $\eta(v + \alpha \mathbf{1}) = \eta(v) + \alpha$, and thus assume that the (v_n) are positive, with at least one zero co-ordinate. The sequence must still get infinitely large, but

$$\langle v_n, x \rangle - \eta(v_n) \leq \|v\|_\infty \Big((1 - \min_i x_i) - \eta(v_n)/\|v_n\|_\infty \Big), \tag{20}$$

which is large and negative for large n, contradicting our supposition of $K(x) = \infty$.

As remarked in Corollary 2, η inherits the smoothness and the strict convexity on $\mathbf{1}^{\perp}$ of δ. Its continuity means that the supremum must be attained at some finite point $\hat{v}(x)$, and the convexity gives the uniqueness. The differentiability shows that the maximizing $\hat{v}$ will be the solution of $\nabla\eta(v) = x$. We can expand $\hat{v}$ around an x as $\hat{v}(x+\epsilon) = \hat{v}(x) + H_\eta^{-1}\epsilon$ to see that

$$K(x+\epsilon) = K(x) + \langle \hat{v}(x), \epsilon\rangle + \frac{1}{2}\epsilon^\top H_\eta^{-1}\epsilon + o(\epsilon^2). \tag{21}$$

Thus K is twice differentiable, $\nabla K(x) - \hat{v}(x)$ is a multiple of $\mathbf{1}$, and K is locally (and hence globally) strictly convex. (Technical note: we are regarding H_η, the Hessian of η, as an automorphism of $\mathbf{1}^{\perp}$.) By the above $x = \nabla\eta(v)$ is a solution of (19), and the strict convexity of K shows it is unique. □

In the simple example studied in Section (d) of Baxter and Williams [2], the rate function was calculated exactly as $K(x) = \sum \pi_i \log(\pi_i/x_i)$, which is infinite on the boundary of M, whilst the Cesàro rate function I is finite everywhere. Note that we can see that I is finite in the general Corollary 2 situation by considering equation (9). We shall have further remarks about this example in the next Section, but for the moment we derive a necessary and sufficient condition for K to be everywhere finite or infinite.

PROPOSITION 4. – *Under the conditions of Corollary 2, the rate function K is either everywhere finite or everywhere infinite on the boundary of M according as to whether the support of the discount function m is of finite or infinite (Lebesgue) length.*

Proof of Proposition 4. – Firstly let us define V_+ to be the space of elements of $(\mathbb{R}^+)^n$ which have at least one zero component. We note that $\lambda V_+ = V_+$ for any positive λ, which is a feature we shall use later. For x in $\mathrm{Int}(M)$, we take v_x to be the unique choice in V_+ of the $v = \nabla K(x)$ in Proposition 3. In fact the pair $(\nabla K, \nabla\eta)$ represents a homeomorphism between $\mathrm{Int}(M)$ and V_+. Then $x = \nabla\eta(v_x)$, so by taking the gradient of (11), we can write x as

$$x = \int_0^\infty m_t \nabla\delta(m_t v_x)\, dt.$$

Then because $\langle v, \nabla\delta(v)\rangle = \delta(v) + I(\nabla\delta(v))$ for all v in $\mathbb{R}^n$,

$$\langle v_x, x\rangle = \int_0^\infty \Big(\delta(m_t v_x) + I\big(\nabla\delta(m_t v_x)\big)\Big)\, dt = \eta(v_x) + \int_0^\infty I\big(\nabla\delta(m_t v_x)\big)\, dt.$$

As v_x is the optimal v in (18), we can express $K(x)$ as

$$K(x) = \int_0^\infty I\big(\nabla\delta(m_t v_x)\big)\,dt. \tag{22}$$

Thus, for an upper bound,

$$K(x) \le \sup_{y\in M} I(y) \int_0^\infty I_{\{m_t>0\}}\,dt = \sup_{y\in M} I(y)\ \mathrm{Leb\,supp}(m),$$

and so K is bounded on all of M if the support of m, $\mathrm{supp}(m)$, is compact. The rate function I is only 0 in M at π, and $\nabla\delta$ only takes the value π in V_+ at 0. Thus from (22) we have the lower bound

$$K(x) \ge \mathrm{Leb}\big\{t : m_t \ge \|v_x\|_\infty^{-1}\big\}\ \inf\big\{I\big(\nabla\delta(v)\big) : v \in V_+,\ \|v\|_\infty \ge 1\big\} > 0.$$

Now as x tends towards ∂M, the boundary of M, the vector v_x tends to infinity in V_+. So if m has unbounded support, then $K(x)$ tends to infinity as x tends to ∂M. The intuition, of course, is that X can with positive probability avoid hitting a certain state for all times in a finite length set but not for all times in an infinite length set. □

3. MORE EXACT RESULTS FOR MARKOV CHAINS

Our aim is to obtain a sharper version of (11) for finite Markov chains, and then to derive more terms of the asymptotic expansion of the density of A_λ.

The initial case studied in Baxter and Williams [2] was of a symmetrizable (reversible) Markov chain and a smooth discount density m. It turns out that m need only be of bounded variation (*see* below), but for technical ease we shall give the proof first in the case where m is also absolutely continuous.

More interestingly, the symmetrizability is seen now to have only been needed to make one of the eigenvalues of Q real and its corresponding eigenvector orthogonal to the others. This in fact happens automatically because every (non-diagonal) element of Q is non-negative (we say that Q is *essentially non-negative*). The following theorem collects all the facts about non-negative matrices that we will need.

THEOREM 5. – *Let R be an essentially non-negative $n \times n$ matrix. Let δ be its principal eigenvalue (the one with greatest real part). Then δ is itself real, and its corresponding eigenvector is non-negative and no other is positive.*

If, in addition, R is irreducible (in the stochastic sense), then δ is simple, its eigenvector is strictly positive and no other is non-negative, and there exists a real diagonal matrix F with positive elements such that $S := \delta I - F^{-1}RF$ has a simple eigenvalue zero, with an orthogonal eigenprojection P, and that, for some positive σ

$$\langle Sx, x\rangle \geq \sigma\|(I-P)x\|^2, \qquad \textit{for all } x \in \mathbb{R}^n. \tag{23}$$

(Where $\langle\ ,\ \rangle$ and $\|\ \|$ are the standard inner product and its norm on $\mathbb{R}^n$, and an orthogonal projection P satisfies $P^\top = P^2 = P$.)

Proof of Theorem 5. – For the first parts *see* the Perron-Frobenius theorem in, for example, theorem 1.5 of Seneta [11] or theorems I.7.5 and I.7.10 of Kato [8]. For the existence of F *see* Theorem B of Baxter [3], which itself is adapted from theorem I.7.13 of Kato [8]. □

We recall that the *variation* of a measurable function $x : [0,\infty) \to \mathbb{R}$ on $[a,b]$ is defined as

$$V_x(a,b) := \sup \sum_{i=1}^{n} |x(t_i) - x(t_{i-1})|, \tag{24}$$

where the supremum is taken over all partitions: $a = t_0 < t_1 < \ldots < t_n = b$ of $[a,b]$. We say that x is of finite variation (FV) if $V_x(0,t)$ is finite for all t, and that x is of bounded variation (BV) if $V_x(0,\infty) := \lim_{t\to\infty} V_x(0,t)$ is finite. An absolutely continuous BV function is the partial integral of a function in $L_1(0,\infty)$. (My thanks to James Norris for correcting a previous mis-statement here.)

We can now begin by strengthening (11):

THEOREM 6. – *Let X be an irreducible continuous-time Markov chain on a finite set S, with Q-matrix Q. Let m be a non-negative absolutely continuous density on $[0,\infty)$ of bounded variation. Then, if X starts in state i and v is in $\mathbb{R}^S$,*

$$\mathbb{E}_i \exp \int_0^\infty m(\lambda t) v(X_t)\, dt = e^{\eta(v)/\lambda}\big(w_i(m_0 v) + o(1)\big), \tag{25}$$

where $\eta(v)$ is as in (11) and $w(v)$ is the positive eigenvector of $Q+V$ and $o(1)$ tends to 0 locally uniformly in v as λ goes to 0.

Proof of Theorem 6. – The chain has an invariant distribution π, but we do not need to assume that Q is π-symmetric. We will aim to get a uniform bound for all v in some compact subset V_K of $\mathbb{R}^S$, and for a fixed m such that $V_m(0,\infty) \leq K_V$. Since Theorem 1 gives us the asymptotic exponential size of φ, it is sensible to discount it by the same, by defining

$$\psi_i(t,\lambda,v) := \exp\left(-\lambda^{-1}\int_0^\infty \delta(\theta_t m_s v)\, ds\right)\varphi(i,t,\lambda,v). \tag{26}$$

Then ψ satisfies the vector differential equation transformed from (12)

$$\partial_t\psi = \lambda^{-1}R(m_t v)\psi, \qquad \psi(\infty,\lambda) = \mathbf{1}, \tag{27}$$

where $R(v)$ is $\delta(v)I - (Q+V)$ which has a simple eigenvalue at 0, and all its other eigenvalues have positive real part. From Theorem 5, there exists a real diagonal matrix $F(v)$ with positive elements, such that $S(v) := F^{-1}(v)R(v)F(v)$ has an orthogonal eigenprojection $P(v)$ onto the space spanned by the strictly positive eigenvector $y(v)$ corresponding to the eigenvalue zero. Further there exists a positive $\sigma(v)$ such that (23) holds, that is

$$\langle S(v)x, x\rangle \geq \sigma(v)\|(I-P(v))x\|^2 \qquad \text{for all } x, \tag{28}$$

where $\langle\ ,\ \rangle$ and $\|\ \|$ are the standard inner product and its norm on $\mathbb{R}^S$. Kato [8], or otherwise, tells us that R, S, F, P, y and σ are smooth in v with bounded derivatives on V_K. Let $\sigma_0 := \inf_{v\in V_K}\sigma(v)$, which is positive. We now fix v, although our bounds will still be uniform, and write R_α for $R(\alpha v)$, and so on. We can choose the normalisation of F uniquely such that $F_0 = \operatorname{diag}(\pi_i^{-1/2})$ and $P_\alpha F_\alpha^{-1}F'_\alpha y_\alpha = 0$, and by choosing $\|y(v)\| = 1$ we ensure that $P_\alpha y'_\alpha = 0$ and $y_0 = \sqrt{\pi}$.

As in Baxter [3], we change bases appropriately by defining

$$\chi(t,\lambda) := F_{m(t)}^{-1}\psi(t,\lambda). \tag{29}$$

The differential equation (27) now becomes

$$\partial_t\chi = \lambda^{-1}S_{m(t)}\chi + J_{m(t)}\chi m', \qquad \chi(\infty,\lambda) = \sqrt{\pi}, \tag{30}$$

where $J_\alpha := -F_\alpha^{-1}F'_\alpha$. Then by taking the inner product of (30) with χ we can produce a differential inequality in the norm of χ,

$$\frac{1}{2}\partial_t\|\chi\|_t^2 \geq 0 - K_1\|\chi\|_t^2|m'_t|, \quad \text{so that} \quad \|\chi\|_t \leq \exp\left(K_1\int_t^\infty |m'_s|\, ds\right), \tag{31}$$

where $K_1 := \sup\|J_\alpha\|$, the supremum taken over the range $\alpha\in[0,K_V]$ and $v\in V_K$. Whence we deduce that χ is uniformly bounded in t and λ

by $K_\chi := \exp(K_1 K_V)$. Now we split χ up according to the decomposition $I = P_\alpha \oplus (I - P_\alpha)$, and define $\chi_-(t) := (I - P_{m(t)})\chi_t$. We differentiate χ_-, using (30), to get

$$\partial_t \chi_- = \lambda^{-1} S_m \chi_- + (I - P_m) J_m \chi m' - P'_m \chi m', \qquad \chi_-(\infty, \lambda) = 0. \quad (32)$$

Taking the inner product of this with χ_- itself, we derive the inequality

$$\frac{1}{2}\partial_t \|\chi_-\|_t^2 \geq \lambda^{-1}\sigma_0 \|\chi_-\|_t^2 - K_\chi(K_1 + K_2)\|\chi_-\|_t |m'_t|, \quad (33)$$

where $K_2 := \sup \|P'_\alpha\|$, with the supremum taken over the same range as K_1. Which we can integrate to get the upper bound

$$\|\chi_-\|_t \leq (K_1 + K_2)K_\chi \int_t^\infty e^{-\sigma_0(s-t)/\lambda} |m'_s|\, ds. \quad (34)$$

And so we see that $\chi_-(t, \lambda)$ tends to 0 as λ tends to 0 for all finite t, though note that the convergence is not necessarily uniform in t. Finally we consider the component of χ in the y_m direction, $\xi(t, \lambda) := \langle \chi(t, \lambda), y_{m(t)} \rangle$, which is governed by the differential equation obtained from (30)

$$\partial_t \xi_t = \langle \chi_-, Jy + y' \rangle m'_t, \qquad \xi(\infty, \lambda) = 1, \quad (35)$$

where we used the fact that $PJy = Py' = 0$. Then

$$|\xi(t, \lambda) - 1| \leq (K_1 + K_2) \int_t^\infty \|\chi_-(s, \lambda)\|\, |m'_s|\, ds, \quad (36)$$

which, by the Dominated Convergence theorem, tends to 0 uniformly in t as λ goes to 0. So

$$\varphi(i, 0, \lambda, v) = \exp(\eta(v)/\lambda)\big((Fy)_i(m_0 v) + o(1)\big), \quad (37)$$

and $F(v)y(v) = w(v)$, where $w(v)$ is the positive eigenvector of $Q + V$, with the normalisation that $w(0) = 1$ and $\partial_\alpha w(\alpha v)$ is orthogonal to the positive eigenvector of $Q^\top + \alpha V$. In the case where Q is π-symmetrizable ($\pi_i q_{ij} = \pi_j q_{ji}$), then the normalisation condition becomes $\|w\|_\pi = 1$, where $\|v\|_\pi^2 = \sum \pi_i v_i^2$. □

The next theorem removes the restriction that m need be continuous, but takes us into the technicalities of FV functions. The casual reader can pass this by without disadvantage.

An FV function can be written as the difference of two increasing functions, that is

$$x_t = x_0 + x_t^+ - x_t^-,$$

where

$$x_t^+ := \frac{1}{2}(x_t + V_x(0,t) - x_0) \quad \text{and} \quad x_t^- := \frac{1}{2}(x_0 + V_x(0,t) - x_t).$$

And so x has only countably many discontinuities (though they may even be dense), and thus can be taken to be an R-function (right-continuous with left limits). We shall take all our functions to be R-functions. We adapt the calculus from the left-continuous integrands of V.18 of Rogers and Williams [10] (changing some signs) to give the formulae

(Decomposition)

$$x = x_0 + x^c + x^a$$

(Integration by parts)

$$d(xy) = x\, dy + y\, dx - \triangle x \triangle y$$

(Itô's formula)

$$d(f(x)) = f'(x)\, dx^c + \triangle(f(x))$$

where x and y are FV and f is C^1, $\triangle x_t$ is $x_t - x_{t-}$, and x^c and x^a denote the continuous and purely discontinuous parts of x respectively. There is an expression for x^a as $\sum_{0<s\leq t} \triangle x_s$. As x^+ and x^- are increasing they induce positive σ-finite Lebesgue-Stieltjes measures on $(0, \infty)$, via $x^+(a,b] = x_b^+ - x_a^+$. So we can associate x with the (signed) measure of their difference. We write $dx_t = dx_t^+ - dx_t^-$. We will also use the notation $|dx_t|$ for $dx_t^+ + dx_t^- = dV_x$. The differential expressions above are symbolic, being merely shorthand for integral expressions.

We will also use an FV exponential result in that if x is BV and

$$\text{if} \quad dx_t \geq -x_t|dy_t|, \quad \text{then} \quad x_t \leq x_\infty \prod_{s>t}(1 + |\triangle y_s|) \exp V_{y^c}(t, \infty). \tag{38}$$

Another useful result follows from integration by parts, in that

$$\frac{1}{2}d||x_t||^2 = \langle x_t, dx_t \rangle - \frac{1}{2}||\triangle x_t||^2. \tag{39}$$

THEOREM 7. – *Theorem 6 remains true if m is a discontinuous non-negative density of bounded variation.*

Proof of Theorem 7. – We follow the proof of Theorem 6 exactly down to (29), except that we take R, F, S and P to be functions of t rather than v. We write $G := F^{-1}$, $w := Fy$ and $w^* := Gy$, and take the normalisation that $\|y\| = 1$ ($\iff$ $\langle w, w^* \rangle = 1$) and $\langle w, dw^* \rangle = 0$ ($\iff$ $\langle dw, w^*_- \rangle = 0$). Note that all these functions are BV. Then (30) becomes

$$d\chi = \lambda^{-1} S\chi \, dt + dG\, F\chi. \tag{40}$$

So

$$\frac{1}{2} d\|\chi_t\|^2 \geq \langle dG\, F\chi, \chi \rangle_t - \frac{1}{2}\|\triangle GF\chi\|_t^2 \geq K_1 \|\chi_t\|^2 \|dG\|, \tag{41}$$

by (39) and (38), for some constant K_1. Hence $\|\chi_t\|$ is uniformly bounded in t by some constant K_χ. Now using (40) and (39) we again work with the components of χ orthogonal to y,

$$\begin{aligned} \frac{1}{2} d\|\chi_-\|_t^2 &\geq \frac{\sigma_0}{\lambda}\|\chi_-\|_t^2\, dt + \langle dG\, F\chi - dP\,\chi + \triangle P \triangle \chi, \chi_- \rangle_t - \frac{1}{2}\|\triangle(P\chi)\|_t^2 \\ &\geq \frac{\sigma_0}{\lambda}\|\chi_-\|_t^2\, dt - K_2(\|dG_t\| + \|dP_t\|), \end{aligned}$$

for some constant K_2. So a result of the same form as (34) holds. Finally we find that

$$d\xi = \xi\langle w, dw^* \rangle + \langle F\chi_-, dw^* \rangle,$$

and we have a bound similar to that of (36), because $\langle w, dw^* \rangle = 0$. Explicitly, w_t is the positive eigenvector of $Q + m_t V$ with the normalisation that $w_\infty = 1$ and dw_t is orthogonal to the positive eigenvector of $Q^\top + m_{t-} V$. □

We can calculate an exact expression for w.

LEMMA 8. – *Let $y(v)$ be the positive eigenvector of $Q + V$ of constant norm, with $y(0) = \mathbf{1}$. If $w_t^0 := y(m_t v)$ and w_t^* is the positive eigenvector of $Q^\top + m_t V$ satisfying $\langle w_t^0, w_t^* \rangle = 1$, then*

$$w_t = w_t^0 \prod_{s>t} (1 + \langle \triangle w_s^0, w_{s-}^* \rangle) \exp \int_t^\infty \langle dw_s^{0,c}, w_s^* \rangle.$$

Further $w_t = w_t(v, m)$ is continuous in v.

Proof of Lemma 8. – If we set $w_t = r_t w_t^0$, then

$$dw_t = dr_t^c\, w_t^0 + r_t\, dw_t^0 + \triangle r_t w_{t-}^0 \quad \text{so} \quad \langle dw_t, w_{t-}^* \rangle = dr_t + r_t \langle dw_t^0, w_{t-}^* \rangle.$$

An application of (38) gives the expression for w. Elementary perturbation results, in for example Kato [7], tell us that y is smooth in v and so

$$dw_t^0(v) = \langle v, \nabla\rangle y(m_t v)\, dm_t^c + \Delta(w_t^0(v)).$$

Thus the difference $dw_t^0(v) - dw_t^0(u)$ can be written as

$$\Big(\langle v-u, \nabla\rangle y(m_t v) + \langle u, \nabla\rangle \big(y(m_t v) - y(m_t u)\big)\Big)\, dm_t^c + \Delta\big(y(m_t v) - y(m_t u)\big).$$

Hence

$$|dw_t^0(v) - dw_t^0(u)| \le K_1 \|v-u\|\, |dm_t^c| + K_2 \|v-u\|\, |\Delta m_t|,$$

where $K_1 := \sup_{K_V} \|\nabla y(v)\| + V_K \sup_{K_V} \|v\|\, \|\nabla y(v) - \nabla y(u)\| / \|v-u\|$ and some constant K_2. Hence w_t is (Lipschitz) continuous in v. □

THEOREM 9. – *Let X be an irreducible continuous-time Markov chain on a finite set S, with Q-matrix Q. Let m be a non-negative density on $[0, \infty)$ of bounded variation. Then where f_i^λ is the density of A_λ on M under the law starting X at i, the (f_i^λ) can be written as*

$$f_i^\lambda(x) = e^{-K(x)/\lambda} (2\pi\lambda)^{-(n-1)/2} \big(\det H_K(x)\big)^{1/2} z_i(x) r_i^\lambda(x), \qquad (42)$$

where K is as defined by (18), H_K denotes its Hessian taken with respect to M, $z(x)$ is the positive eigenvector of $Q + \operatorname{diag}(m_0 \nabla K(x))$, and the residue term r^λ goes to 1 as λ goes to 0, in the sense that

$$\limsup_{\lambda \downarrow 0} \int_F r_i^\lambda(x + \sqrt{\lambda} y)\, dy \le |F|, \quad \text{and} \quad \liminf_{\lambda \downarrow 0} \int_G r_i^\lambda(x + \sqrt{\lambda} y)\, dy \ge |G|, \qquad (43)$$

for all x in $\operatorname{Int}(M)$, and for F and G respectively closed and open bounded subsets of $\mathbf{1}^\perp$.

Notes. – (1) We take the Hessian regarding K as a function on an open subset of $\mathbb{R}^{n-1}$, that is $K(x_1, \ldots, x_{n-1}, 1 - \sum_i^{n-1} x_i)$. *See* the example at the end of this paper.

(2) Unfortunately we would really like to prove the result that

$$\mathbb{P}_i(A_\lambda \in H) \sim \int_H e^{-K(x)/\lambda} (2\pi\lambda)^{-(n-1)/2} \big(\det H_K(x)\big)^{1/2} z_i(x)\, dx, \quad (44)$$

for suitable H, as λ goes to 0. This could be proved if the integrand in our control of r^λ was $r_i^\lambda(x + \lambda y)$ rather than $r_i^\lambda(x + \sqrt{\lambda} y)$.

Proof of Theorem 9. – Theorem 6 can be taken as saying that for v in $\mathbb{R}^S$,

$$f_i^{v,\lambda}(x) := \exp\Big(\lambda^{-1}\big(\langle v, x\rangle - \eta(v)\big)\Big) f_i^{\lambda}(x)/w_i(v) \tag{45}$$

is (asymptotically) a density on M, where $w_i(v)$ is the $w_0(v, m)(i)$ of Lemma 8, and our $z_i(x)$ will be $w_i(\nabla K(x))$. If A_λ^v under $\mathbb{P}_i$ has the law $f_i^{v,\lambda}$, then we can derive a central limit result by considering $Z_\lambda^v := (A_\lambda^v - \nabla\eta(v))/\sqrt{\lambda}$. We see that for u in $\mathbb{R}^S$,

$$\begin{aligned}\mathbb{E}_i\big(\exp\langle u, Z_\lambda^v\rangle\big) &= \exp\Big(\lambda^{-1}\big(\eta(v+\sqrt{\lambda}u) - \eta(v) - \langle\sqrt{\lambda}u, \nabla\eta(v)\rangle\big)\Big)\\ &\quad\ldots\Big(w_i(v+\sqrt{\lambda}u) + o(1)\Big)/w_i(v),\end{aligned} \tag{46}$$

using the local uniformity in v of the convergence of $o(1)$. As η inherits the smoothness of δ, we can expand it about v as

$$\eta(v+\sqrt{\lambda}u) = \eta(v) + \langle\sqrt{\lambda}u, \nabla\eta(v)\rangle + \frac{1}{2}\lambda u^\top H_\eta(v)u + o(\lambda), \tag{47}$$

and hence deduce that

$$\lim_{\lambda\downarrow 0}\mathbb{E}_i\big(\exp\langle u, Z_\lambda^v\rangle\big) = \exp\big(\frac{1}{2}u^\top H_\eta(v)u\big). \tag{48}$$

In other words

$$Z_\lambda^v \xrightarrow{\mathcal{D}} N(0, H_\eta(v)).$$

We can think of $f_i^{v,\lambda}$ as the distribution of A_λ conditioned in some way to converge to $\nabla\eta(v)$, but we do not make this formal. Proposition 3 provides the interpretation of $\nabla\eta(v)$ as the maximizing x in the Legendre transform.

Recall that a sequence of laws (ν_n) on a Polish space E converges to a law ν with respect to the weak topology on $M_1(E)$ if $\langle v, \nu_n\rangle \to \langle v, \nu\rangle$ for all v in $C_b(E)$. Billingsley [4], 2.1, shows that this is equivalent to each of the following

$$\limsup_{n\to\infty}\nu_n(F) \le \nu(F) \qquad \text{F closed in } E,$$

$$\liminf_{n\to\infty}\nu_n(G) \ge \nu(G) \qquad G \text{ open in } E,$$

and

$$\lim_{n\to\infty}\nu_n(H) = \nu(H) \qquad H \text{ in } E \text{ with } \nu(\partial H) = 0.$$

Setting $x = \nabla\eta(v)$, we recall from Proposition 3 that $\nabla K(x)$ is v, up to a multiple of $\mathbf{1}$. The asymptotics of the density of Z^v_λ are given by

$$f_i^{Z,\lambda}(y) := \lambda^{(n-1)/2} f_i^{v,\lambda}(x + \sqrt{\lambda}y) \sim (2\pi)^{-(n-1)/2} \det H_K(x)^{\frac{1}{2}} \times e^{-\frac{1}{2}y^\top H_K(x) y}\, r_i^\lambda(x + \sqrt{\lambda}y),$$

because

$$\begin{aligned}&\langle v, x + \sqrt{\lambda}y\rangle - \eta(v) - K(x + \sqrt{\lambda}y)\\ &= \langle v, x\rangle - \eta(v) - K(x) + \sqrt{\lambda}\langle v - \nabla K(x), y\rangle - \frac{1}{2}\lambda y^\top H_K(x) y + o(\lambda)\\ &= \lambda(-\frac{1}{2}y^\top H_K(x) y + o(1)).\end{aligned}$$

The normal distribution $N(0, H_\eta(v))$ itself has density

$$f(y) = (2\pi)^{-(n-1)/2} \det H_K(x)^{\frac{1}{2}} e^{-\frac{1}{2}y^\top H_K(x) y},$$

as Proposition 3 tells us that $H_K = H_\eta^{-1}$ on M. By Lemma I.45.1 of Williams [12], if H is bounded and $|\partial H| = 0$ then

$$\int I_H(y)(f(y))^{-1} f_i^{Z,\lambda}(y)\, dy \to |H|, \quad \text{or} \quad \int_H r_i^\lambda(x + \sqrt{\lambda}y)\, dy \to |H|. \tag{49}$$

Hence by the equivalence of the above expressions for weak convergence, the result is proved. □

The following Corollary is intended in the way of a remark, and was the original statement of Theorem 9, but is now seen to be weaker, although perhaps a more natural formulation.

COROLLARY 10. – *Under the conditions of Theorem 9,*

$$\limsup_{\lambda\downarrow 0} \int_F r_i^\lambda(x)\, dx \le |F| \quad \textit{and} \quad \liminf_{\lambda\downarrow 0} \int_G r_i^\lambda(x)\, dx \ge |G|, \tag{50}$$

for F closed in $\mathrm{Int}(M)$ *and G open in* $\mathrm{Int}(M)$. *In other words, $r_i^\lambda(x)\, dx$ converges weakly to dx on* $\mathrm{Int}(M)$.

Proof of Corollary 10. – Take G open in $\mathrm{Int}(M)$, δ small and positive with $G_\delta := \{y \in G : B(y, \delta) \subseteq G\}$, and B a ball around 0, then by Fatou's lemma and Fubini's theorem

$$\begin{aligned}|G_\delta|\,|B| &= \int_{G_\delta} \left(\liminf_{\lambda\downarrow 0} \int_B r_i^\lambda(x + \sqrt{\lambda}y)\, dy\right) dx\\ &\le \liminf_{\lambda\downarrow 0} \int_{G_\delta}\int_B r_i^\lambda(x + \sqrt{\lambda}y)\, dy\, dx \le \left(\liminf_{\lambda\downarrow 0} \int_G r_i^\lambda(x)\, dx\right) |B|.\end{aligned}$$

Letting δ tend to 0, we have one of our bounds. For some F closed in $\mathrm{Int}(M)$, we need $\int_B r_i^\lambda(x+\sqrt{\lambda}y)\,dy$ to be uniformly bounded on F and for λ near 0. It is, and the bound is

$$\sup_{x\in F}(2\pi)^{(n-1)/2}\det H_K(x)^{-\frac{1}{2}}\sup_{y\in B} e^{\frac{1}{2}y^\top H_K(x)y}<\infty.$$

Working with this F and with $F^\delta := \{y\in M : d(y,F)\le\delta\}$, we can show in a similar way that

$$\limsup_{\lambda\downarrow 0}\int_F r_i^\lambda(x)\,dx\le|F^\delta|,$$

and hence we are home. □

Some remarks.

(a) Were the r_i^λ to be equicontinuous (or some such condition) we would have that $r_i^\lambda(x)\to 1$ for all x and hence that $f_i^\lambda(x)/f_j^\lambda(x)\to z_i(x)/z_j(x)$ and $\lim -(Qf^\lambda)_i/f_i^\lambda$ differs from $m_0\nabla K(x)$ only by a multiple of $\mathbf{1}$, as in Section (c) of Baxter and Williams [2], where the choice of $\nabla K(x)$ in $\ker(\delta)$ was called $g(x)$.

(b) Note that the proof of Theorem 9 gives us a central limit theorem for A_λ as

$$Z_\lambda := (A_\lambda-\pi)/\sqrt{\lambda}\xrightarrow{\mathcal{D}}N(0,H_\eta(0)). \tag{51}$$

Taking a Taylor expansion of δ about 0 and integrating we discover that $H_\eta(0)=\sigma^2 H_\delta(0)$, where $\sigma=\|m\|_{L_2}$ which is finite because m is in both L_1 and L_∞.

Example. – (This case was first studied in Section (d) of Baxter and Williams [2].) Suppose we have a Markov chain which is symmetric and space-homogeneous, with Q-matrix $q_{ij}:=\pi_j-\delta_{ij}$, where π is a distribution on a finite set S. The Cesàro large-deviation rate function is $I(x)=1-(\sum\sqrt{\pi_i x_i})^2$, and the exponentially discounted large-deviation rate is $K(x)=\sum\pi_i\log(\pi_i/x_i)$. We found then that $\delta(v)$ is the unique root δ in $(\max_i(v_i-1),\infty)$ of

$$\sum_{i\in S}\frac{\pi_i}{\delta+1-v_i}=1,$$

and that η is given by $\eta(v) = \delta(v) - \sum_i \pi_i \log(\delta(v) + 1 - v_i)$. We find now that

$$\begin{aligned}
\nabla_i I(x) &= 1 - \sqrt{\frac{\pi_i}{x_i}}\Big(\sum_{j\in S}\sqrt{\pi_j x_j}\Big),\\
\nabla_i \delta(v) &= \frac{\pi_i}{(\delta(v)+1-v_i)^2}\Big/\sum_{j\in S}\frac{\pi_j}{(\delta(v)+1-v_j)^2},\\
\nabla_i K(x) &= 1 - \frac{\pi_i}{x_i},\\
\text{and}\quad \nabla_i \eta(v) &= \frac{\pi_i}{\delta(v)+1-v_i}.
\end{aligned}$$

Here we chose ∇I and ∇K to be in the kernel of δ. The distribution of A_λ can be calculated explicitly to be a multidimensional β-distribution with density

$$f_i^\lambda(x) = \frac{x_i}{\pi_i}\Gamma(\lambda^{-1})\prod_{j\in S}\frac{x_j^{(\pi_j/\lambda)-1}}{\Gamma(\pi_j/\lambda)}.$$

Note that the Hessian of K on M is not the same as that derived from the extension of K to $\mathbb{R}^S$, but by using any of the following co-ordinate schemes:

$$\begin{aligned}
&K_i : \mathbb{R}^{S\setminus\{i\}} \longrightarrow \mathbb{R} \qquad \text{for each } i \in S\\
\text{where}\quad &(x_j)_{j\neq i} \longmapsto K(x_1,\ldots,x_{i-1},1-\sum_{j\neq i}x_j,x_{i+1},\ldots,x_n)\\
\text{or}\quad &K_0 : \mathbb{R}^n \longrightarrow \mathbb{R}\\
\text{where}\quad &x \longmapsto K(x + (1-\mathbf{1}^\top x)\mathbf{1}/n) + \frac{1}{2}(\mathbf{1}^\top x)^2.
\end{aligned}$$

What is happening here is that our choice of basis for evaluating the Hessian corresponds to our choice of basis for integrating which was made back at the start of Section (b) of Baxter and Williams [2]. The K_0 representation projects onto M and adds a strictly convex term which is perpendicular to M. This representation is more natural, though cumbersome to calculate with, and can be shown equivalent to any of the others by verifying that the change of basis matrix has determinant one. Thus the Hessian (in the K_n realisation) and its determinant are given by

$$H_{K_n}(x)_{ij} = \frac{\pi}{x_i^2}\delta_{ij} + \frac{\pi_n}{x_n^2} \qquad \text{and} \qquad \det(H_K(x)) = \Big(\prod_{i\in S}\frac{\pi_i}{x_i^2}\Big)\sum_{j\in S}\frac{x_j^2}{\pi_j}.$$

The normalisation of the eigenvector $z_i(x)$ is that $\|z\|_\pi = 1$, so it is given by

$$z_i(x) = \frac{x_i}{\pi_i}\Big(\sum_{j\in S}\frac{x_j^2}{\pi_j}\Big)^{-1/2}.$$

(The corresponding vector for the Cesàro case is $\sqrt{x_i/\pi_i}(\sum_j\sqrt{\pi_j x_j})$.) We can now calculate the residual functions using Stirling's formula

$$\Gamma(x) = \sqrt{2\pi}x^{x-\frac{1}{2}}e^{-x}(1+O(x^{-1})),$$

where $|O(x^{-1})| \leq K/x$ as $x \to \infty$ for some constant K. It is thus discovered that the residual functions $r_i^\lambda(x)$ can be calculated and are found to be independent of both i and x, and are of size $1 + O(\lambda)$.

Hypothesis 11. – We recall from Theorem C of Baxter and Williams [2] that in the set-up of Theorem 9 with the exponential discount ($m_t = e^{-t}$), the density f^λ satisfies the vector differential equation

$$\mathcal{L}f^\lambda = -\lambda^{-1}Qf^\lambda, \tag{52}$$

where $\mathcal{L}$ is the matrix differential operator $\mathcal{L} = \operatorname{diag}(\sum_{j\neq i}(\partial_j - \partial_i)x_j)_{i\in S}$. Here we have changed the domain of f^λ from a subset of $\mathbb{R}^{n-1}$ equivalent to M, to a neighbourhood of M in $\mathbb{R}^n$ by extension. The operator $\mathcal{L}$ is invariant to the extension chosen. If we discount f^λ by the known large-deviation rate function K, that is by defining g^λ by

$$f^\lambda(x) = e^{-K(x)/\lambda}(2\pi\lambda)^{-(n-1)/2}g^\lambda(x)$$

then

$$\mathcal{L}g^\lambda = \lambda^{-1}R\big(\nabla K(x)\big)g^\lambda.$$

This compares with equation (27) which said that

$$\partial_t\psi = \lambda^{-1}R(m_t v)\psi,$$

where ψ is the discount of φ as defined by (26). The matrix $R(v)$ has a simple eigenvalue 0 and all other eigenvalues have positive real part. We saw that ψ tended to a multiple of the 0-eigenvector of $R(m_t v)$ as λ went to 0, and also that g^λ tended (in some sense) to $z(x)$, which was the 0-eigenvector of $R(\nabla K(x))$. We can formulate an analogue of (52) for the general discount case as follows.

Let us write $A_{\lambda,t}$ for $\lambda M_t^{-1}\int_0^\infty \theta_t m(\lambda s)\delta_{X_s}\,ds$, where $M_t := \int_0^\infty \theta_t m(s)\,ds$, and $f_i^{\lambda,t}$ for the density of $A_{\lambda,t}$ if X starts in state i. Then $A_{\lambda,t}$ will satisfy the large-deviation property with rate function K_t, where

$$K_t = \eta_t^* \quad \text{where} \quad \eta_t(v) := \int_0^\infty \delta(M_t^{-1}\theta_t m(s)v)\,ds,$$

and we write $f^{\lambda,t}$ as

$$f^{\lambda,t}(x) = e^{-K_t(x)/\lambda}(2\pi\lambda)^{-(n-1)/2}g^{\lambda,t}(x).$$

Then

$$M_t^{-1}m_t\mathcal{L}g^{\lambda,t} + \partial_t g^{\lambda,t} = \lambda^{-1}R\big(M_t^{-1}m_t\nabla K_t(x)\big)g^{\lambda,t}.$$

Again Theorem 9 tells us that $g^{\lambda,t}$ tends (in some sense) to the 0-eigenvector, z, of the matrix R. We hypothesize that the convergence is in fact pointwise.

REFERENCES

[1] M. W. BAXTER and D. WILLIAMS, Symmetry characterizations of certain distributions, 1, *Math. Proc. Cambridge Philos. Soc.*, Vol. **111**, 1992, pp. 387-397.

[2] M. W. BAXTER and D. WILLIAMS, Symmetry characterizations of certain distributions, 2: Discounted additive functionals and large deviations, *Math. Proc. Cambridge Philos. Soc.*, Vol. **112**, 1992, pp. 599-611.

[3] M. W. BAXTER, Symmetry characterizations of certain distributions, 3: Discounted additive functionals and large deviations for a general finite-state Markov chain, *Math. Proc. Cambridge Philos. Soc.*, Vol. **113**, 1993, pp. 381-386.

[4] P. BILLINGSLEY, *Convergence of Probability Measures*, Wiley, 1968.

[5] N. H. BINGHAM, C. M. GOLDIE and J. L. TEUGELS, Regular Variation, *Encyclopedia of Mathematics and its Applications*, Vol. **27**, Cambridge University Press, 1987.

[6] J.-D. DEUSCHEL and D. W. STROOCK, *Large deviations*, Academic Press, 1989.

[7] R. S. ELLIS, Large deviations for a general class of random vectors, *Ann. Probab.*, Vol. **12**, 1984, pp. 1-12.

[8] T. KATO, *A Short Introduction to Perturbation Theory for Linear Operators*, Springer-Verlag, 1982.

[9] Y. KIFER, Averaging in Dynamical Systems and Large Deviations, *Inventiones Math.*, Vol. **110**, 1992, pp. 337-370.

[10] L. C. G. ROGERS and D. WILLIAMS, *Diffusions, Markov Processes and Martingales*, Vol. **2**: Itô calculus, Wiley, 1987.

[11] E. SENETA, *Non-negative Matrices, an Introduction to Theory and Applications*, Allen and Unwin, 1973.

[12] D. WILLIAMS, *Diffusions, Markov Processes and Martingales*, Vol. **1**: Foundations, Wiley, 1979.

(Manuscript accepted November 7, 1995.)