Tunneling from Random Walks to Markov Chains

Neil O'Connell, Adam Shwartz*
Basic Research Institute in the Mathematical Sciences
HP Laboratories Bristol
HPL-BRIMS-98-09
May, 1998

large deviations,
Markov chains,
contraction principle

As a philosophically motivated exercise, we obtain the large deviation principle for a Markov chain by viewing it as a functional i.i.d random variables.

*Electrical Engineering, Technion Institute of Technology, Haifa, 32000 Israel

1 what and why

A sequence of random variables Z_n, taking values in a Hausdorff topological space $\mathcal{Z}$ is said to satisfy the *large deviation principle* (LDP) with good rate function $J : \mathcal{Z} \to \mathbb{R}_+$ if J has compact level sets and, for all Borel sets B,

$$-\inf_{B^\circ} J \leq \liminf_n \frac{1}{n} \log P(Z_n \in B) \leq \limsup_n \frac{1}{n} \log P(Z_n \in B) \leq -\inf_{\bar{B}} J.$$

In any given example, this can be proved directly or, if $\mathcal{Z}$ is a topological vector space with dual $\mathcal{Z}^*$, by considering the asymptotic behaviour of the scaled cumulant generating functions $H_n : \mathcal{Z}^* \to \mathbb{R}$ defined by

$$H_n(t) = \frac{1}{n} \log E \exp(n\langle t, Z_n \rangle).$$

Alternatively, one can appeal to the *contraction principle* of large deviation theory, which states the following. Suppose we have a sequence of random variables X_n satisfying the LDP in a Hausdorff topological space $\mathcal{X}$ with good rate function I, and $Z_n = f(X_n)$, for some continuous $f : \mathcal{X} \to \mathcal{Z}$. Then the sequence Z_n satisfies the LDP with good rate function given by

$$J(z) = \inf\{I(x) : \ f(x) = z\}.$$

The law of large numbers states, in its most general form, that a 'smooth' function of many independent random variables is roughly constant. There is a rapidly developing branch of probability theory concerned with making this statement precise, and does so mainly in the form of concentration inequalities. The contraction principle of large deviation theory provides an alternative. To see this, think of X_n as some (one to one) encoding of n i.i.d.

random variables, $\xi_1, \ldots, \xi_n$. For example, we could let

$$X_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{(\xi_i, i/n)} \in M_1(\mathbb{R} \times [0,1])$$

where by $M_1(S)$ we denote the space of probability measures on S. If we use a sufficiently course topology on $M_1(\mathbb{R} \times [0,1])$, such as the weak topology, then the sequence X_n satisfies the LDP with a good rate function I [1]. Moreover, $I(\nu) = 0$ if, and only if, $\nu = \mu \times \lambda$, where μ is the law of ξ_1 and λ is Lebesgue measure on $[0,1]$. (This implies that the law of large numbers holds for the sequence of measures X_n.) By the contraction principle, if $f : M_1(\mathbb{R} \times [0,1]) \to \mathcal{Z}$ is continuous, then the sequence $f(X_n)$ satisfies the LDP in $\mathcal{Z}$ with rate function given by

$$J(z) = \inf\{I(x) : \ f(x) = z\},$$

and $J(z) = 0$ if, and only if, $z = f(\mu \times \lambda)$. In particular, the law of $f(X_n)$ converges to $f(\mu \times \lambda)$ exponentially fast. Thus, by 'smooth' in this context we mean the function of the ξ_k can be represented as a continuous function of the encoding X_n.

In principle, we can regard any random structure as a functional of some pure, homogeneous, underlying randomness. If we are interested in proving an LDP, why not just have one LDP, and deduce all others by contraction? Although this is not a realistic proposal, it is philosophically attractive.

In this short note, we consider a class of Markov chains for which the large deviation theory is well understood. However, it has always been studied by considering directly the evolution of the chain. Motivated by the above discussion, we would like to view the chain as a functional of a collection of i.i.d. uniform random variables and derive the LDP by contraction.

2 how

Let $\{U_i,\ i = 1, \dots\}$ be a sequence of i.i.d., Uniform ([0,1]) random variables. Define the following 'empirical measure' on $S \stackrel{def}{=} [0,1]^2$:

$$X_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{(U_i, i/n)} \tag{1}$$

where δ_x denotes a point mass at x. We equip $M_1(S)$ with the weak topology, and denote by λ the Lebesgue measure on $[0,1]$. Barbe and Broniatowski [1] prove that the sequence X_n satisfies the LDP in $M_1(S)$ with good rate function given by

$$I(\nu) = H(\nu|\lambda^2) \tag{2}$$

provided $\nu \ll \lambda^2$ and $\nu([0,1], \cdot) = \lambda$; otherwise $I(\nu) = \infty$. We can construct a Markov chain from the U_i as follows. Let $p : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ be a Markov transition density and for each $x \in \mathbb{R}$ denote by $g(\cdot, x)$ the inverse of the distribution function associated with the density $p(x, \cdot)$. We assume that $g : [0,1] \times \mathbb{R} \to \mathbb{R}$ is a bounded Lipschitz continuous function (see Lemma 3 below). Denote by $D([0,1])$ the space of right continuous functions on $[0,1]$ having left limits. Consider the function $F : D([0,1]) \times M_1(S) \to D([0,1])$

$$F(f, \nu) \stackrel{def}{=} \int_0^{\cdot} \int_0^1 g(u, f(s))\, \nu(du, ds). \tag{3}$$

Note that $F(\cdot, X_n)$ has a unique fixed point in $D([0,1])$, which we denote by $\Phi(X_n)$, and we have the following representation. Consider a Markov chain $Z(k)$ in $\mathbb{R}$ with $Z(0) = 0$ and with transition density given by $p_n(x,y) = p(x/n, y/n)$. Then $\{Z([sn])/n, 0 \le s \le 1\}$ has the same law as $\Phi(X_n)$.

Now, if Φ were defined and continuous on all of $M_1(S)$, we could immediately deduce the LDP for $\Phi(X_n)$ by contraction. However, this is not the case. What we *can* prove is that Φ is defined and continuous on the set where the rate function is finite. Note however that X_n does not belong to this set. We therefore introduce the following absolutely continuous modification of X_n:

$$\begin{aligned} Y_n = n \sum_{i=1}^{n} \mathbf{1}_{\{U_i \ge 1/n\}} \mathbf{1}_{\{[U_i - 1/n, U_i) \times [\frac{i-1}{n}, \frac{i}{n})\}} \cdot \lambda^2 \\ + \sum_{i=1}^{n} \mathbf{1}_{\{U_i < 1/n\}} \mathbf{1}_{\{[0, U_i) \times [\frac{i-1}{n}, \frac{i}{n})\}} \cdot \lambda^2 / U_i. \end{aligned}$$

The second term is included to ensure that $Y_n \in M_1(S)$.

For $\mu, \nu \in M_1(S)$, define

$$\rho(\mu,\nu) = \inf \left\{ \int |x-y| \pi(dx, dy) : \right. \\ \left. \pi \in M_1(S^2), \ \int_S \pi(\cdot, dy) = \mu, \ \int_S \pi(dx, \cdot) = \nu \right\};$$

this is a metric on $M_1(S)$ which is consistent the weak topology (see Rachev [3] for a fascinating survey.) Clearly, $\rho(X_n, Y_n) < 1/n$ almost surely, so that the sequences X_n and Y_n are ρ-exponentially equivalent; it follows that the sequence Y_n also satisfies the LDP in $M_1(S)$ (see, for example, [2]) with the same rate function. Let

$$A_1(S) = \{\nu \in M_1(S) : \nu \ll \lambda^2, \ \nu([0,1], \cdot) = \lambda\}.$$

Note that $Y_n \in A_1(S)$ for each n, and the rate function I is infinite outside $A_1(S)$; we can therefore restrict the LDP for Y_n to the space $A_1(S)$ (again, see [2]).

Denote by $C([0,1])$ the space of continuous functions on $[0,1]$, equipped with the uniform topology, and note that $F : C([0,1]) \times A_1(S) \to C([0,1])$.

Lemma 1 *The function $F(\cdot, \nu)$ has a unique fixed point $\Phi(\nu) \in C([0,1])$ for each $\nu \in A_1(S)$. Moreover, the mapping $\Phi : A_1(S) \to C([0,1])$ is continuous.*

We can now apply the contraction principle to get that the sequence $\Phi(Y_n)$ satisfies the LDP in $C([0,1])$ with good rate function given by

$$J(f) = \inf\{I(\nu) : \ \Phi(\nu) = f\}. \tag{4}$$

The LDP for $\Phi(X_n)$ (the scaled Markov chain) follows from the next lemma.

Lemma 2 *The sequences $\Phi(X_n)$ and $\Phi(Y_n)$ are exponentially equivalent in $C([0,1])$.*

The proofs are given below. Using these, we obtain

Theorem 3 *The scaled Markov chain $\Phi(X_n)$ satisfies the LDP in $C([0,1])$ with good rate function $J(f)$ given in (4).*

Proof. Since Y_n satisfies the LDP, by Lemma 1 the contraction principle applies [2] and $\Phi(Y_n)$ satisfies the LDP with rate function (4). By Lemma 2, so does $\Phi(X_n)$ [2]. ■

We conclude our discussion with a technical point and proofs.

Lemma 4 *If the transition distribution function*

$$F_x(\alpha) = \int_{-\infty}^{\alpha} p(x,y)dy$$

is Lipschitz in x, uniformly in α, and the transition densities $p(x,\cdot)$ are uniformly bounded away from zero and have compact support, the g is bounded and Lipschitz continuous.

Idea of proof. Draw a picture of $F_x(\alpha)$ and note that $g(x,\cdot)$ is obtained by reflecting the graph about the diagonal. ■

Proof of Lemma 1. Since $\nu([0,1],\cdot) = \lambda$ and g is Lipschitz, we have:

$$\begin{aligned} |F(f_1,\nu) - F(f_2,\nu)| &\le \int_0^t \int_0^1 |g(u,f_1(s)) - g(u,f_2(s))|\nu(du,ds) \\ &\le \int_0^t C|f_1(s) - f_2(s)|ds \\ &\le tC\|f_1 - f_2\|_\infty. \end{aligned}$$

Now we let $t = 1/2C$ and deduce that the mapping which takes $f \in C([1/2C])$ to the restriction of $F(f,\nu)$ to $[0,1/2C]$ is a contraction and thus has a unique fixed point in $C([0,1/2C])$; proceeding over successive intervals of length $1/2C$ we can recursively construct the unique fixed point $\Phi(\nu)$. It remains to show that Φ is continuous. We begin by proving that F is continuous. Suppose $(f_n,\nu_n) \to (f,\nu)$. Then, for each t,

$$\begin{aligned} F(f_n,\nu_n)(t) &= \int_0^t \int_0^1 g(u,f(s))\,\nu_n(du,ds) \\ &\quad + \int_0^t \int_0^1 g(u,f_n(s)) - g(u,f(s))\,\nu_n(du,ds) \\ &\to \int_0^t \int_0^1 g(u,f(s))\,\nu(du,ds) \equiv F(f,\nu) \end{aligned}$$

For the first term we are using the fact that g is bounded and continuous, f is continuous, and $\nu_n \to \nu$ weakly; for the second term we use the facts that g is Lipschitz, $f_n \to f$ uniformly, and ν_n is a probability measure. It is clear that the second term converges to zero uniformly in t. To see that the same is true of the first term, note that the mapping $t \mapsto F(f, \nu_n)$ is Lipschitz, so that convergence is uniform on compacts. We have thus established the continuity of F.

Now suppose $\nu_n \to \nu$ and set $f_n = \Phi(\nu_n)$. Observe that all fixed points f are Lipschitz (with the same Lipschitz constant) and have $f(0) = 0$: it follows that the set of fixed points is precompact, and so the sequence f_n has a convergent subsequence $f_{n(k)} \to f$, say. Thus, by the continuity of F,

$$\begin{aligned} F(f, \nu) &= \lim_{n(k)} F(f_{n(k)}, \nu_{n(k)}) \\ &= \lim_{n(k)} f_{n(k)} = f. \end{aligned}$$

Hence $\Phi(\nu) = f$. To complete the proof we note that, by uniqueness, any other convergent subsequence must converge to the same f. Thus, if f_n does not converge to f, it has a subsequence which converges to some $g \neq f$, a contradiction. ∎

Proof of Lemma 2. We will show that $\|\Phi(X_n) - \Phi(Y_n)\|_\infty < 3/n$ almost surely. Without loss of generality we assume that the Lipschitz constant of the function g is 1, and that it is bounded by 1. Fix $k \leq n$, and set

$$\epsilon = |\Phi(X_n)_{(k-1)/n} - \Phi(Y_n)_{(k-1)/n}|.$$

Then, since $\Phi(X_n)_0 = \Phi(Y_n)_0$ and $(1+1/n)^n \approx e$,

$$
\begin{aligned}
&|\Phi(X_n)_{k/n} - \Phi(Y_n)_{k/n}| \\
&= \left| \epsilon - \int_{(k-1)/n}^{k/n} \int_0^1 [g(u, \Phi(Y_n)_s) - g(u, \Phi(X_n)_{(k-1)/n})] Y_n(du, ds) \right| \\
&\leq \epsilon + \int_{(k-1)/n}^{k/n} \int_0^1 \left| \Phi(Y_n)_s - \Phi(X_n)_{(k-1)/n} \right| Y_n(du, ds) \\
&\leq \epsilon + \frac{1}{n} \left(\epsilon + \frac{1}{n} \right) \\
&= \epsilon \left(1 + \frac{1}{n} \right) + \frac{1}{n^2} \\
&\leq \frac{1}{n^2} \sum_{i=0}^{n} \left(1 + \frac{1}{n} \right)^i \\
&\leq \frac{1}{n^2} n \left(1 + \frac{1}{n} \right)^n \leq 3/n.
\end{aligned}
$$

We have thus proved our claim. ■

References

[1] Philippe Barbe and Michel Broniatowski. Deviation principle for set indexed processes with independent increments. Preprint.

[2] A. Dembo and O. Zeitouni. *Large Deviation Theory: Techniques and Applications.* Second Edition, Springer, 1998.

[3] S.T. Rachev. The Monge-Kantorovich mass transference problem and its stochastic applications. *Theor. Probab. Appl.* 29 (4) pp. 647–676, 1984.