Large Deviations – a mini-course –

Frank den Hollander Leiden University, The Netherlands



P.C. Mahalanobis Lectures 2015–2016, Indian Statistical Institute, Delhi & Bangalore & Kolkata.

PART I

A quick overview of the basic theory

PART II

An application to drawing random words from random letter sequences

PART I

Basic theory

\S BACKGROUND

Large deviation theory is the study of unlikely events. Since such events are always realised

in the least unlikely of all the unlikely ways

they can be captured with the help of variational principles. This gives the theory a particularly elegant structure.

Large deviation theory originated in statistical physics and via insurance and finance mathematics reached probability theory and statistics. It can now be found in almost any scientific discipline. A classical setting for large deviations is the following. Let $(X_i)_{i \in \mathbb{N}}$ be i.i.d. \mathbb{R} -valued random variables with

$$\mathbb{E}(X_1) = \mu \in \mathbb{R}, \qquad \forall \operatorname{ar}(X_1) = \sigma^2 \in (0, \infty).$$

Let

$$\bar{X}_n = n^{-1}(X_1 + \dots + X_n)$$

be the empirical average of the *n*-sample. Then, as $n \to \infty$,

LLN:
$$X_n \to \mu$$
 P-a.s.
CLT: $\sqrt{n}(\bar{X}_n - \mu) \to \sigma N(0, 1)$ in P-probability,

where N(0, 1) is the standard normal random variable.

Under the assumption that $\mathbb{E}(e^{\lambda X_1}) < \infty$ for all $\lambda \in \mathbb{R}$, large deviation theory says that, as $n \to \infty$,

$$\mathbb{P}(\bar{X}_n \approx \nu) = \mathrm{e}^{-nI(\nu) + o(n)},$$

where \approx means close in a proper sense, and

 $\nu \mapsto I(\nu)$

is a rate function that achieves a unique zero at $\nu = \mu$.

Large deviations away from the mean are exponentially costly at a rate that depends on the size of the deviation.

The rate function captures the cost of the large deviations.

\S LARGE DEVIATION PRINCIPLE

DEFINITION (Large Deviation Principle)

A family of probability measures $(\mu_{\epsilon})_{\epsilon>0}$ on a Polish space \mathcal{X} is said to satisfy the large deviation principle (LDP) with rate function $I: \mathcal{X} \to [0, \infty]$ if

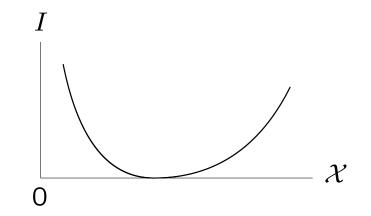
(i) I has compact level sets and is not identically infinite, (ii) $\liminf_{\epsilon \downarrow 0} \epsilon \log \mu_{\epsilon}(O) \ge -I(O)$ for all $O \subseteq \mathcal{X}$ open, (iii) $\limsup_{\epsilon \downarrow 0} \epsilon \log \mu_{\epsilon}(C) \le -I(C)$ for all $C \subseteq \mathcal{X}$ closed,

where $I(S) = \inf_{x \in S} I(x), S \subseteq \mathcal{X}$.

Informally, the LDP says that if $B_{\delta}(x)$ is the open ball of radius $\delta > 0$ centred at $x \in \mathcal{X}$, then

$$\mu_{\epsilon}(B_{\delta}(x)) = \mathrm{e}^{-[1+o(1)]I(x)/\epsilon}$$

when $\epsilon \downarrow 0$ followed by $\delta \downarrow 0$.



Paradigmatic picture of a rate function with a unique zero.

§ UNIQUENESS OF THE RATE FUNCTION

Suppose that I, J are both rate functions for the same $(\mu_{\epsilon})_{\epsilon>0}$. Pick $x \in \mathcal{X}$ and consider the open balls $B_N = B_{1/N}(x)$, $N \in \mathbb{N}$. Then

$$-I(x) \leq -I(B_{N+1}) \leq \liminf_{\epsilon \downarrow 0} \epsilon \log \mu_{\epsilon}(B_{N+1})$$

$$\leq \limsup_{\epsilon \downarrow 0} \epsilon \log \mu_{\epsilon}(\operatorname{cl}(B_{N+1})) \leq -J(\operatorname{cl}(B_{N+1}) \leq -J(B_{N}).$$

Since J has compact level sets it is lower semi-continuous, so that $\lim_{N\to\infty} J(B_N) = J(x)$. Hence $I(x) \ge J(x)$.

By interchanging I and J we get the opposite bound and hence I(x) = J(x). Since $x \in \mathcal{X}$ is arbitrary, this proves that I = J.

§ WEAK LDP

If the level sets of *I* are assumed to be closed rather than compact, and the inequality for the limpsup is assumed to hold for compact sets rather than closed sets, then it is said that the weak LDP holds.

Strengthening a weak LDP to an LDP requires establishing exponential tightness, i.e., proving that for every $N < \infty$ there exists a compact set $K_N \subseteq \mathcal{X}$ such that

 $\limsup_{\epsilon \downarrow 0} \epsilon \log \mu_{\epsilon}([K_N]^c) \leq -N.$

\S DUALITY

The LDP is the workhorse for the computation of averages of exponential functionals.

THEOREM (Varadhan's Lemma)

If $(\mu_{\epsilon})_{\epsilon>0}$ satisfies the LDP on \mathcal{X} with rate function I, then $\lim_{\epsilon\downarrow 0} \epsilon \log \int_{\mathcal{X}} e^{F(x)/\epsilon} \mu_{\epsilon}(\mathrm{d}x) = \Lambda_{F}, \qquad \forall F \in C_{b}(\mathcal{X}),$

where $C_b(\mathcal{X})$ is the space of bounded continuous functions on \mathcal{X} , and

$$\Lambda_F = \sup_{x \in \mathcal{X}} [F(x) - I(x)].$$

Varadhan's Lemma can be easily extended to functions F that are unbounded and discontinuous, provided certain tail estimates on μ_{ϵ} are available.

Varadhan's Lemma has the following inverse.

THEOREM (Bryc's Lemma)

Suppose that $(\mu_{\epsilon})_{\epsilon>0}$ is exponentially tight and the limit in Varadhan's Lemma exists for all $F \in C_b(\mathcal{X})$. Then $(\mu_{\epsilon})_{\epsilon\geq0}$ satisfies the LDP with rate function I given by

$$I(x) = \sup_{F \in C_b(\mathcal{X})} [F(x) - \Lambda_F], \qquad x \in \mathcal{X}.$$

§ FORWARD PRINCIPLES

There are several ways to generate one LDP from another. We give three examples.

THEOREM (Contraction Principle)

Let $(\mu_{\epsilon})_{\epsilon>0}$ satisfy the LDP on \mathcal{X} with rate function I. Let \mathcal{Y} be a second Polish space, and $T: \mathcal{X} \to \mathcal{Y}$ a continuous map from \mathcal{X} to \mathcal{Y} . Then the family of probability measures $(\nu_{\epsilon})_{\epsilon>0}$ on \mathcal{Y} defined by

$$\nu = \mu \circ T^{-1}$$

satisfies the LDP on $\mathcal Y$ with rate function J given by

$$J(y) = \inf_{\substack{x \in \mathcal{X} \\ T(x) = y}} I(x), \qquad y \in \mathcal{Y}.$$

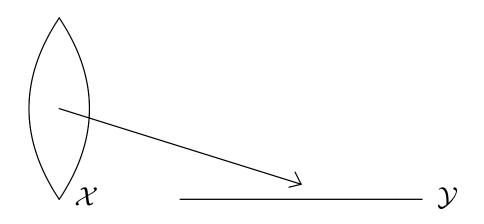


Illustration of the contraction principle.

§ PROOF OF CONTRACTION PRINCIPLE

Since T is continuous, T^{-1} maps open sets into open sets and closed sets into closed sets. Pick $C \subset \mathcal{Y}$ closed and write

$$\limsup_{\epsilon \downarrow 0} \epsilon \log \nu_{\epsilon}(C) = \limsup_{\epsilon \downarrow 0} \epsilon \log \mu_{\epsilon}(T^{-1}(C))$$
$$\leq -I(T^{-1}(C)) = -\inf_{x \in T^{-1}(C)} I(x) = -\inf_{y \in C} J(y) = -J(C).$$

A similar argument applies for $O \subset Y$ open. Since the rate function is unique, this identifies J as the rate function.

Moreover, J inherits compact levels set from I because T is continuous.

THEOREM (Exponential Tilting)

Let $(\mu_{\epsilon})_{\epsilon>0}$ satisfy the LDP on \mathcal{X} with rate function I, and let $F \in C_b(\mathcal{X})$. Then the family of probability measures $(\nu_{\epsilon})_{\epsilon>0}$ on \mathcal{X} defined by

$$\nu_{\epsilon}(dx) = \frac{1}{N_{\epsilon}} e^{F(x)/\epsilon} \mu_{\epsilon}(dx), \qquad N_{\epsilon} = \int_{X} e^{F(x)/\epsilon} \mu_{\epsilon}(dx),$$

satisfies the LDP on \mathcal{X} with rate function J given by

$$J(x) = \Lambda_F - [F(x) - I(x)], \qquad x \in \mathcal{X}.$$

THEOREM (Dawson-Gärtner projective limit LDP)

Let $(\mu_{\epsilon})_{e>0}$ be a family of probability measures on \mathcal{X} . Let $(\pi^N)_{N\in\mathbb{N}}$ be a nested family of projections acting on \mathcal{X} such that $\bigcup_{N\in\mathbb{N}}\pi^N$ is the identity. Let

$$\mathcal{X}^N = \pi^N \mathcal{X}, \qquad \mu_{\epsilon}^N = \mu_{\epsilon} \circ (\pi^N)^{-1}, \qquad N \in \mathbb{N}.$$

If, for each $N \in \mathbb{N}$, the family $(\mu_{\epsilon}^{N})_{\epsilon>0}$ satisfies the LDP on \mathcal{X}^{N} with rate function I^{N} , then $(\mu_{\epsilon})_{\epsilon>0}$ satisfies the LDP on \mathcal{X} with rate function I given by

$$I(x) = \sup_{N \in \mathbb{N}} I^N(\pi^N x), \qquad x \in \mathcal{X}.$$

Since

$$I^{N}(y) = \inf_{\{x \in \mathcal{X}: \pi^{N}(x) = y\}} I(x), \qquad y \in \mathcal{X}^{N},$$

the supremum defining I is non-decreasing in N because the projections are nested.

The projective limit LDP can be used to extend a suitably nested sequence of LDP's on finite-dimensional spaces to an LDP on an infinite-dimensional space.

\S Special structures

LDPs can be formulated on general topological spaces \mathcal{X} , although this comes at the cost of more technicalities. Conversely, more can be said when \mathcal{X} has more structure.

For instance, if \mathcal{X} is a vector space, then the rate function can be identified as the Legendre transform of a generalised cumulant generating function. When $\mathcal{X} = \mathbb{R}^d$, we have the following.

THEOREM (Gärtner-Ellis Theorem)

Let $(\mu_{\epsilon})_{\epsilon>0}$ be a family of probability measures on \mathbb{R}^d , $d \ge 1$, with the following properties:

(i) $\phi(u) = \lim_{\epsilon \downarrow 0} \epsilon \log \int_{\mathbb{R}^d} e^{\langle u, x \rangle / \epsilon} \mu_{\epsilon}(dx)$ exists in \mathbb{R} for all $u \in \mathbb{R}^d$, where $\langle \cdot, \cdot \rangle$ denotes the standard inner product on \mathbb{R}^d . (ii) $u \mapsto \phi(u)$ is differentiable on \mathbb{R}^d .

Then $(\mu_{\epsilon})_{\epsilon>0}$ satisfies the LDP on \mathbb{R}^d with a convex rate function ϕ^* given by

$$\phi^*(x) = \sup_{u \in \mathbb{R}^d} [\langle u, x \rangle - \phi(u)], \qquad x \in \mathbb{R}^d.$$

There is a version of the Gärtner-Ellis Theorem where the domain of ϕ is not all of \mathbb{R}^d , in which case some additional assumptions must be made.

Two special cases deserve to be mentioned:

- Cramér's Theorem
- Sanov's Theorem

Let $(X_i)_{i \in \mathbb{N}}$ be i.i.d. \mathbb{R} -valued random variables with law ρ . Let $\mathcal{M}_1(\mathbb{R})$ denote the space of probability measures on \mathbb{R} (which is a subset of the vector space of signed measures on \mathbb{R}). Cramér's Theorem:

Let μ_n denote the law of the empirical average

$$\bar{X}_n = n^{-1} \sum_{i=1}^n X_i \in \mathbb{R}.$$

If

$$M(\lambda) = \int_{\mathbb{R}} e^{\lambda x} \rho(dx) < \infty \qquad \forall \ \lambda \in \mathbb{R},$$

then $(\mu_n)_{n\in\mathbb{N}}$ satisfies the LDP on \mathbb{R} with rate $\epsilon = n^{-1}$ and rate function

$$I(x) = \sup_{\lambda \in \mathbb{R}} [\lambda x - \log M(\lambda)], \qquad x \in \mathbb{R}.$$

Examples:

1. $\rho = N(0, 1)$: $M(\lambda) = e^{\lambda^2/2}, \lambda \in \mathbb{R},$ $I(x) = \frac{1}{2}x^2, x \in \mathbb{R}.$

2.
$$\rho = \frac{1}{2}(\delta_{-1} + \delta_{+1})$$
:
 $M(\lambda) = \cosh(\lambda), \ \lambda \in \mathbb{R},$
 $I(x) = \frac{1}{2}(1+x)\log(1+x) + \frac{1}{2}(1-x)\log(1-x), \ x \in [-1,+1].$

3. $\rho = \text{POISSON}(m)$:

$$M(\lambda) = \exp[m(e^{\lambda} - 1)], \ \lambda \in \mathbb{R},$$

$$I(x) = x \log(x/m) - (x - m), \ x \in [0, \infty).$$

Sanov's Theorem:

Let μ_n denote the law of the empirical distribution

$$L_n = n^{-1} \sum_{i=1}^n \delta_{X_i} \in \mathcal{M}_1(\mathbb{R}).$$

Then $(\mu_n)_{n\in\mathbb{N}}$ satisfies the LDP on $\mathcal{M}_1(\mathbb{R})$ with rate $\epsilon = n^{-1}$ and rate function

$$I(\nu) = H(\nu | \rho) = \int_{\mathbb{R}} \nu(\mathrm{d}x) \log\left[\frac{\mathrm{d}\nu}{\mathrm{d}\rho}\right](x), \qquad \nu \in \mathcal{M}_1(\mathbb{R}),$$

with the right-hand side infinite when ν is not absolutely continuous with respect to ρ .

 $H(\nu | \rho)$ is called the relative entropy of ν with respect to ρ .

With the help of the Dawson-Gärtner projective limit LDP an infinite-dimensional version of Sanov's theorem can be derived. The key object of interest is the empirical process

$$R_n = n^{-1} \sum_{i=1}^n \delta_{\theta^i(X_1,\dots,X_n)} \text{per} \in \mathcal{M}_1^*(\mathbb{R}^{\mathbb{N}}),$$

where

$$(X_1,\ldots,X_n)^{\mathsf{per}} = \underbrace{X_1,\ldots,X_n},\underbrace{X_1,\ldots,X_n},\ldots$$

is the periodic extension of the first n elements of $(X_i)_{i \in \mathbb{N}}$, θ is the left-shift acting on $\mathbb{R}^{\mathbb{N}}$, and $\mathcal{M}_1^*(\mathbb{R}^{\mathbb{N}})$ is the space of θ -invariant probability measures on $\mathbb{R}^{\mathbb{N}}$. Sanov's Theorem at the process level:

Let μ_n denote the law of the empirical process R_n . Then $(\mu_n)_{n \in \mathbb{N}}$ satisfies the LDP on $\mathcal{M}_1^*(\mathbb{R}^{\mathbb{N}})$ with rate $\epsilon = n^{-1}$ and rate function

$$I(Q) = H(Q \mid \rho^{\otimes \mathbb{N}}) = \lim_{N \to \infty} N^{-1} H(\pi^N Q \mid \rho^N),$$

where $\pi^N Q$ is the projection of Q onto the first N coordinates.

 $H(Q | \rho^{\otimes \mathbb{N}})$ is called the specific relative entropy of Q with respect to $\rho^{\otimes \mathbb{N}}$.

The rate functions

$$x \mapsto I(x), \quad \nu \mapsto I(\nu), \quad Q \mapsto I(Q)$$

are all convex. The first two are strictly convex, the third is affine. There is a deep connection with information theory.

It is possible to generalize the three LDPs to sequences $(X_i)_{i \in \mathbb{N}}$ that are not i.i.d., for instance, Markov sequences. However, some mixing properties are needed to ensure proper rate functions. In general these rate functions are no longer convex.

§ LITERATURE

S.R.S. Varadhan, *Large Deviations and Applications*, Society for Industrial and Applied Mathematics, 1984.

R.S. Ellis, *Entropy, Large Deviations, and Statistical Physics*, Springer, 1985.

J.-D. Deuschel and D.W. Stroock, *Large Deviations*, Academic Press, 1989.

A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Springer, 1998.

F. den Hollander, *Large Deviations*, American Mathematical Society, 2000.

+ ...

PART II

An application

§ MOTIVATION

Let

$$S = (S_k)_{k \in \mathbb{N}_0}, \qquad S' = (S'_k)_{k \in \mathbb{N}_0},$$

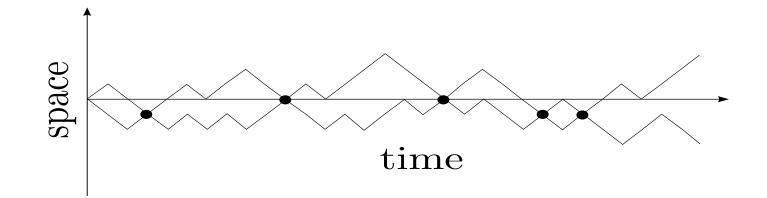
be two independent random walks on \mathbb{Z}^d , $d \ge 1$, starting at 0, with a symmetric transition kernel $p(\cdot, \cdot)$. Let

$$V = \sum_{k \in \mathbb{N}_0} \mathbf{1}_{\{S_k = S'_k\}}$$

denote their collision local time. Then

$$\mathbb{P}(V < \infty) = 1$$

if and only if $p(\cdot, \cdot)$ is transient.



Counting collisions of two random walks.

Define

$$z_1 = \sup \left\{ z \ge 0 \colon \mathbb{E} \left[z^V \mid S \right] < \infty \ S\text{-a.s.} \right\},$$
$$z_2 = \sup \left\{ z \ge 0 \colon \mathbb{E} \left[z^V \right] < \infty \right\}.$$

It is obvious that $z_2 \leq z_1$. For recurrent random walk $z_1 = z_2 = 1$.

QUESTION Under what conditions on $p(\cdot, \cdot)$ is it the case that $z_2 < z_1$?

Suppose that $n \mapsto p^{2n}(0,0)$ is regularly varying at infinity.

THEOREM $z_2 < z_1$ when $\sum_{n \in \mathbb{N}} n p^{2n}(0,0) < \infty$.

CONJECTURE $z_2 < z_1$ if and only if $\sum_{n \in \mathbb{N}} p^{2n}(0,0) < \infty$.

The presence of the gap $z_2 < z_1$ implies the existence of an intermediate phase for the long-time behavior of

- branching and migrating populations
- linearly interacting diffusions
- copolymers near selective interfaces.

The reason is that certain key quantities for these systems, such as mean growth rate, mean value in equilibrium or free energy per monomer, have a

random walk representation!

\S Letters, words and sentences

Joint work with:

Matthias Birkner (Mainz), Andreas Greven (Erlangen)

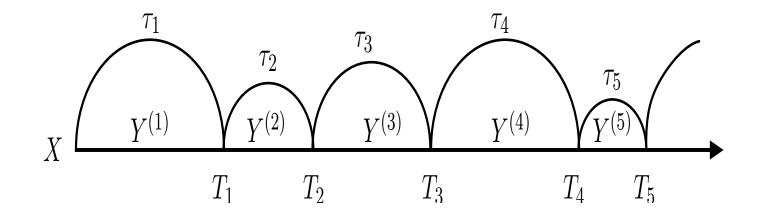
Let

- $X = (X_k)_{k \in \mathbb{N}}$ be an i.i.d. sequence of letters drawn from a countable alphabet E according to a law ν .
- $\tau = (\tau_i)_{i \in \mathbb{N}}$ be an i.i.d. sequence of renewal times drawn from \mathbb{N} according to a law ρ .

Put $T_0 = 0$, and for $i \in \mathbb{N}$ define

$$T_i = \tau_1 + \dots + \tau_i,$$

 $Y^{(i)} = (X_{T_{i-1}+1}, \dots, X_{T_i}).$



Cutting words from a letter sequence according to a renewal process.

 $Y = (Y^{(i)})_{i \in \mathbb{N}}$ is an i.i.d. sequence of words taking values in the sentence space $\tilde{E}^{\mathbb{N}}$, where $\tilde{E} = \bigcup_{n \in \mathbb{N}} E^n$ is the word space. Our key object of interest will be the empirical process of N-tuples of words

$$R_N = \frac{1}{N} \sum_{i=1}^N \delta_{\widetilde{\theta}^i(Y^{(1)}, \dots, Y^{(N)})^{\mathsf{per}}} \in \mathcal{M}_1^*(\widetilde{E}^{\mathbb{N}}),$$

where per denotes periodic extension and $\tilde{\theta}$ is the left-shift acting on $\tilde{E}^{\mathbb{N}}$.

By the ergodic theorem, we have

weak
$$-\lim_{N \to \infty} R_N = q^{* \otimes \mathbb{N}} \qquad \mathbb{P} - a.s.$$

with

$$q^*((x_1,\ldots,x_n)) = \rho(n)\nu(x_1)\cdots\nu(x_n),$$

$$n \in \mathbb{N}, x_1,\ldots,x_n \in E.$$

THEOREM (annealed LDP)

The family of probability distributions $\mathbb{P}(R_N \in \cdot)$, $N \in \mathbb{N}$, satisfies the LDP on $\mathcal{M}_1^*(\tilde{E}^{\mathbb{N}})$ with rate N and with rate function

$$Q \mapsto H(Q \,|\, q^{* \otimes \mathbb{N}}),$$

the specific relative entropy of Q w.r.t. $q^* \otimes \mathbb{N}$. This rate function is affine, has compact level sets and has a unique zero at $Q = q^* \otimes \mathbb{N}$.

\S CONDITIONING ON THE LETTER SEQUENCE

Let $\kappa \colon \tilde{E}^{\mathbb{N}} \to E^{\mathbb{N}}$ denote the concatenation map that glues a sequence of words into a sequence of letters. For $Q \in \mathcal{M}_1^*(\tilde{E}^{\mathbb{N}})$ such that

$$m_Q = \mathbb{E}_Q[\tau_1] < \infty,$$

define $\Psi_Q \in \mathcal{M}_1^*(E^{\mathbb{N}})$ as

$$\Psi_Q(\cdot) = \frac{1}{m_Q} \mathbb{E}_Q \left[\sum_{k=0}^{\tau_1 - 1} \delta_{\theta^k \kappa(Y)}(\cdot) \right],$$

where θ is the left-shift acting on $E^{\mathbb{N}}$.

THEOREM (quenched LDP)

Suppose that

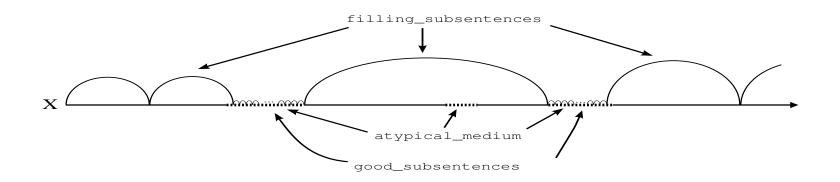
$$\lim_{n \to \infty} \frac{\log \rho(n)}{\log n} = -\alpha, \qquad \alpha \in (1, \infty).$$

Then, for $\nu^{\otimes \mathbb{N}}$ -a.s. all X, the family of conditional probability distributions $\mathbb{P}(R_N \in \cdot | X)$, $N \in \mathbb{N}$, satisfies the LDP on $\mathcal{M}_1^*(\tilde{E}^{\mathbb{N}})$ with rate N and with deterministic rate function $Q \mapsto I(Q) = H(Q | q^* \otimes \mathbb{N}) + (\alpha - 1) m_Q H(\Psi_Q | \nu^{\otimes \mathbb{N}})$

when $m_Q < \infty$, and I(Q) given by a truncation approximation when $m_Q = \infty$. This rate function is affine, has compact level sets and has a unique zero at $Q = q^* \otimes \mathbb{N}$.

HEURISTICS

Once X is fixed, the renewal process has to look for rare stretches in X on which it can realize $R_N \approx Q$ via a large number of small renewals. These stretches are reached via a small number of large renewals.



Looking for good subsentences and filling subsentences.

The quenched rate function is the sum of the annealed rate function and an entropy term involving randomized concatenation and the tail exponent of the renewal times.

The proof of the quenched LDP is complicated and requires

ergodic theory, combinatorics, large deviation theory, entropy estimates, compactification, mollification, projective limits, ...

\S back to collision local time

Write

$$z^{V} = ((z-1)+1)^{V}$$

= $1 + \sum_{N=1}^{V} (z-1)^{N} \frac{V(V-1)\cdots(V-N+1)}{N!}$
= $1 + \sum_{N=1}^{V} (z-1)^{N} \sum_{0 < j_{1} < \cdots < j_{N} < \infty} \mathbf{1}_{\{S_{j_{1}} = S'_{j_{1}}, \dots, S_{j_{N}} = S'_{j_{N}}\}}.$

Hence

$$\mathbb{E}\left[z^{V} \mid S\right] = 1 + \sum_{N=1}^{\infty} (z-1)^{N} F_{N}^{(1)}(X),$$
$$\mathbb{E}\left[z^{V}\right] = 1 + \sum_{N=1}^{\infty} (z-1)^{N} F_{N}^{(2)},$$

with

$$F_N^{(1)}(X) = \sum_{0 < j_1 < \dots < j_N < \infty} \mathbb{P} \left(S_{j_1} = S'_{j_1}, \dots, S_{j_N} = S'_{j_N} \mid X \right),$$
$$F_N^{(2)} = \mathbb{E} \left[F_N^{(1)}(X) \right],$$

where $X = (X_k)_{k \in \mathbb{N}}$ denotes the i.i.d. sequence of increments of S.

We may write this out further as follows:

$$F_N^{(1)}(X) = \sum_{0 < j_1 < \dots < j_N < \infty} \prod_{i=1}^N p^{j_i - j_{i-1}} \left(0, \sum_{k=j_{i-1}+1}^{j_i} X_k \right)$$

$$= \sum_{0 < j_1 < \dots < j_N < \infty} \prod_{i=1}^N \rho(j_i - j_{i-1})$$

$$\times \exp\left[\sum_{i=1}^N \log\left(\frac{p^{j_i - j_{i-1}}(0, \sum_{k=j_{i-1}+1}^{j_i} X_k)}{\rho(j_i - j_{i-1})} \right) \right]$$

$$= \mathbb{E}\left[\exp\left(\sum_{i=1}^N \log f(Y_i) \right) \middle| X \right]$$

$$= \mathbb{E}\left[\exp\left(N \int_{\widetilde{E}} (\pi_1 R_N) (\mathrm{d}y) \log f(y) \right) \middle| X \right],$$

where

$$f((x_1, \dots, x_n)) = \frac{p^n(0, x_1 + \dots + x_n)}{\rho(n)},$$
$$n \in \mathbb{N}, x_1, \dots, x_n \in E = \mathbb{Z}^d,$$

and $\pi_1 R_N$ is the projection of R_N onto the first word. Similarly, we have

$$F_N^{(2)} = \mathbb{E}\left[\exp\left(N\int_{\widetilde{E}}(\pi_1 R_N)(\mathrm{d}y)\,\log f(y)\right)\right].$$

By picking

$$\rho(n) = \frac{p^n(0,0)}{G(0,0)-1}, \quad n \in \mathbb{N},$$

with G(0,0) the Green function at the origin, we obtain that f is bounded and continuous.

Consequently, the quenched LDP and the annealed LDP can be combined with Varadhan's Lemma to obtain the variational formulas ($z_i = 1 + e^{-r_i}$, i = 1, 2)

$$r_{1} = \lim_{N \to \infty} \frac{1}{N} \log F_{N}^{(1)}(X)$$

= $\sup_{Q} \left\{ \int_{\widetilde{E}} (\pi_{1}Q)(dy) \log f(y) - H(Q \mid q^{* \otimes \mathbb{N}}) - (\alpha - 1) m_{Q} H(\Psi_{Q} \mid \nu^{\otimes \mathbb{N}}) \right\}$ a.s.,

$$r_{2} = \lim_{N \to \infty} \frac{1}{N} \log F_{N}^{(2)}$$
$$= \sup_{Q} \left\{ \int_{\widetilde{E}} (\pi_{1}Q)(dy) \log f(y) - H(Q \mid q^{* \otimes \mathbb{N}}) \right\}$$

•

To prove that $z_2 < z_1$ it suffices to show that $r_1 < r_2$. The key to this gap are the following two facts:

• The variational formula for r_2 has a unique maximizer $\bar{Q}=\bar{q}^{\otimes\mathbb{N}}$ given by

$$\bar{q}((x_1, \dots, x_n)) = \frac{p^n(0, x_1 + \dots + x_n)}{G(0, 0) - 1} \prod_{k=1}^n p(0, x_k),$$
$$n \in \mathbb{N}, x_1, \dots, x_n \in E = \mathbb{Z}^d.$$

• If
$$m_{\bar{Q}} < \infty$$
, then $\Psi_{\bar{Q}} \neq \nu^{\otimes \mathbb{N}}$ (here $\nu(\cdot) = p(0, \cdot)$).

It turns out that $m_{ar{Q}} < \infty$ when $\sum_{n \in \mathbb{N}} n \, p^{2n}(0,0) < \infty.$

§ Literature

M. Birkner, A. Greven and F. den Hollander, Probab. Theory Relat. Fields 148 (2010) 403–456.

F. den Hollander, Proceedings of the ICM, Hyderabad, India, 2010.

M. Birkner and R. Sun, Ann. I. Henri Poincaré 46 (2010) 414–441.

N. Berger and F. Toninelli, Electronic J. Prob. 15 (2010) 654–683.

M. Birkner, A. Greven and F. den Hollander, Electronic J. Prob. 16 (2011) 552–586.

M. Birkner and R. Sun, Ann. I. Henri Poincaré 47 (2011) 259–293.

+ . . .