Notes for Probability Reading Seminar

Probability group of UW-Madison

October 25, 2015

1 Large deviations for Markov chains

Given by Elnur Emrah in September 2015 Madison

1.1 Large deviation principle

Let $(X_i)_{i\in\mathbb{N}}$ be i.i.d. random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ such that $\mathbf{E}\, X_i = 0$ and $\mathbf{Var} X_i = 1$. Let $S_n = n^{-1}\sum_{i=1}^n X_i$ and μ_n denote the distribution of S_n for $n \in \mathbb{N}$. For example, consider $X_i \sim \mathcal{N}(0,1)$. Then $S_n \sim \mathcal{N}(0, 1/n)$ and

$$\mathbf{P}(|S_n| \geqslant \ell) = \frac{2}{\sqrt{2\pi}}\int_{\ell\sqrt{n}}^{+\infty} e^{-y^2/2}dy = e^{-\frac{\ell^2}{2}n+o(n)}.$$

Similarly for $0 < \ell < \ell'$, we have

$$\mathbf{P}(|S_n| \in [\ell, \ell']) = e^{-\frac{\ell^2}{2}n+o(n)}.$$

It is natural to ask if X is a general random variable with measure μ, what should we put on the r.h.s. ? Motivation: for measure μ, what is the following I?

$$\mathbf{P}(|S_n| \in [\ell, \ell']) = e^{-I(\ell)n+o(n)}. \tag{1}$$

Definition 1.1. We say that μ_n satisfies an LDP with a *rate function* I *if $I : \mathbb{R} \to [0, +\infty]$ is lower semicontinuous and, for all Borel sets $B \subset \mathbb{R}$, we have*

$$\begin{aligned} -\inf_{x\in B^0} I(x) &\leqslant \liminf_{n\to+\infty} \frac{\log \mu_n(B)}{n} \quad \textit{(lower bound)} \\ -\inf_{x\in \overline{B}} I(x) &\geqslant \limsup_{n\to+\infty} \frac{\log \mu_n(B)}{n} \quad \textit{(upper bound).} \end{aligned}$$

Here, B^0 and $\overline{B}$ denote the interior and the closure of B. Recall that I is lower-semicontinuous if the sublevel set $\{I \leqslant \alpha\}$ is closed for any $\alpha < +\infty$. This condition is equivalent to $\liminf_{y\to x} I(y) \geqslant I(x)$ for any $x \in \mathbb{R}$.

Remark: it may take a while to understand this form. Here is the equivalent expressions, which is very useful for me

$$e^{-(\inf_{x\in\bar{B}} I(x))\cdot n+o(n)} \geqslant \mu_n(B) \geqslant e^{-\left(\inf_{x\in B^0} I(x)\right)\cdot n+o(n)}$$

Note: you have to use $\bar{B}$ in l.h.s, and use B^0 in l.h.s

The definition of LDP can be given for sequences of measures on arbitrary topological spaces. I will refer to LDP for measures on Euclidean spaces below.

There are many basic properties of LDP in Prof. Varadhan, which some one can introduce to us in the future.

The answer to (1):

$$I(x) = \sup_{\lambda\in\mathbb{R}} \{\lambda x - \log \mathbf{E}^{\mu}[e^{\lambda X}]\} = \sup_{\lambda\in\mathbb{R}} \left\{\lambda x - \frac{1}{n}\log \mathbf{E}^{\mu_n}[e^{\lambda X}]\right\}.$$

where μ is measure of X_i and μ_n is measure of $\sum_i X_i$.

A useful tool to establish LDP is Gärtner-Ellis theorem. We consider the following setup. Let $(Z_n)_{n\in\mathbb{N}}$ be a sequence of random vectors in $\mathbb{R}^d$. Let μ_n denote the distribution of Z_n. Consider the log-moment generating function $\Lambda_n(\lambda) = \log \mathbf{E}[e^{\lambda\cdot Z_n}]$ for $\lambda \in \mathbb{R}^d$. We assume that the following conditions hold:

1. The limit $\Lambda(\lambda) = \lim_{n\to+\infty} n^{-1}\Lambda_n(n\lambda) \in (-\infty, +\infty]$ exists.
2. $0 \in D^0_\Lambda$, where $D_\Lambda = \{\lambda \in \mathbb{R}^d : \Lambda(\lambda) < \infty\}$.
3. Λ is differentiable on D^0_Λ.
4. (Steepness condition) For any $x \in \partial D_\Lambda$, $\lim_{\substack{\lambda\to x\\ \lambda\in D^0_\Lambda}} |\nabla\Lambda(\lambda)| = +\infty$.

Theorem 1.1 (Gärtner-Ellis theorem). *Under assumptions (a)-(d), (μ_n) satisfy an LDP with convex, good (i.e. sublevel sets are compact) rate function Λ^*, the Legendre-Fenchel transform of Λ given by*

$$\Lambda^*(x) = \sup_{\lambda\in\mathbb{R}^d} \{\lambda\cdot x - \Lambda(\lambda)\}.$$

This theorem, in fact, a special case of the Gärtner-Ellis theorem; see, for example, [1], [2] for the full theorem and its proof.

Example: Sum of i.i.d. random variables.

1.2 Application to the Markov chains

We now present an application of this theorem to the Markov chains in discrete time with finite state space. We introduce some notation first. The state space is $[N] = \{1, \ldots, N\}$. Let $\Pi = [\pi(i,j)]_{i,j\in[N]}$ be a stochastic matrix, that is, $\pi(i,j) \geqslant 0$ and $\sum_j \pi(i,j) = 1$ for each $i \in [N]$. Let P^π_σ denote the Markov probability measure with transition matrix Π and initial state at $\sigma \in [N]$. Let Y_n denote the state the chain visits at time n. We have

$$P^\pi_\sigma(Y_1 = y_1, \ldots, Y_n = y_n) = \pi(\sigma, y_1)\pi(y_1, y_2)\ldots\pi(y_{n-1}, y_n)$$

for any path $(y_1, \ldots, y_n)$ in the state space. We assume that Π irreducible; this means that for each (i,j), there exists $m(i,j) \in \mathbb{N}$ such that $\Pi^{m(i,j)}(i,j) > 0$.

Our goal is to obtain an LDP for random variables $Z_n = n^{-1}\sum_{i=1}^n f(Y_i)$, where $f : [N] \to \mathbb{R}^d$ is a given function.

For the computation Λ, the limiting log-moment generating function, we will utilize the following result. For a vector u, we will write $u \gg 0$ if all components of u are positive.

Theorem 1.2 (Perron-Frobenius). *Let $B = [B(i,j)]_{i,j\in[N]}$ be an irreducible matrix with positive entries. Then B has a real eigenvalue ρ (called the Perron-Frobenius eigenvalue) with the following properties.*

(i) $|\lambda| \leqslant \rho$ *for any eigenvalue of* B.

(ii) There exist a left eigenvector u and a right eigenvector v corresponding to ρ such that $u \gg 0$ and $v \gg 0$.

(iii) ρ has multiplicity 1.

(iv) For all $i \in [N]$ and $\varphi \gg 0$, we have

$$\lim_{n\to+\infty} \frac{1}{n} \log \left[\sum_{j=1}^{N} B^n(i,j)\varphi_j \right] = \lim_{n\to+\infty} \frac{1}{n} \log \left[\sum_{i=1}^{N} B^n(i,j)\varphi_i \right] = \log \rho.$$

Proof of (iv). Let $c = \min_j \varphi_j / \min_j v_j$, where v is the right eigenvector corresponding to ρ. We have

$$\sum_{j=1}^{N} B^n(i,j)\varphi_j \geqslant \sum_{j=1}^{N} B^n(i,j)v_j c = c\rho^n v_i$$

Taking logarithms, dividing through by n and letting $n \to +\infty$ yields

$$\liminf_{n\to+\infty} \frac{1}{n} \log \left[\sum_{j=1}^{N} B^n(i,j)\varphi_j \right] \geqslant \log \rho.$$

We similarly obtain that the limsup is bounded by $\log \rho$. □

Theorem 1.3. *For Markov chain, random variables $Z_n = n^{-1} \sum_{i=1}^{n} f(Y_i)$ satisfy LDP with a rate function $I(x)$ with*

$$I(x) = \sup_{\lambda\in\mathbb{R}^d} \{\lambda \cdot x - \log \rho(\Pi_\lambda)\}$$

where $\Pi_\lambda = [\pi_\lambda(i,j)]_{i,j\in[N]}$ defined by $\pi_\lambda(i,j) = \pi(i,j)e^{\lambda\cdot f(j)}$

Proof: We now turn to LDP for the Markov chain (Y_n). We have

$$\begin{aligned}
\frac{\Lambda_n(n\lambda)}{n} &= \frac{1}{n} \log E_\sigma^\pi \left[\exp\left(\sum_{i=1}^{n} \lambda \cdot f(Y_i) \right) \right] \\
&= \frac{1}{n} \log \left[\sum_{(y_1,\ldots,y_n)\in[N]^n} \exp\left(\sum_i \lambda \cdot f(y_i) \right) \prod_i \pi(y_{i-1}, y_i) \right] \\
&= \frac{1}{n} \log \left[\sum_{(y_1,\ldots,y_n)\in[N]^n} \prod_i \pi(y_{i-1}, y_i) e^{\lambda\cdot f(y_i)} \right],
\end{aligned}$$

where $y_0 = \sigma$. We observe that the matrix $\Pi_\lambda = [\pi_\lambda(i,j)]_{i,j\in[N]}$ defined by $\pi_\lambda(i,j) = \pi(i,j)e^{\lambda\cdot f(j)}$ has positive entries and is irreducible because it is obtained from such a matrix Π by multiplying each entry with a positive number. Hence,

$$\frac{\Lambda_n(n\lambda)}{n} = \frac{1}{n} \log \left[\sum_{y_n=1}^{N} \Pi_\lambda^n(\sigma, y_n) \right] \to \log \rho(\Pi_\lambda)$$

as $n \to +\infty$, by the Perron-Frobenius theorem (applied with $\varphi = (1,\ldots,1)$). Since the Perron-Frobenius eigenvalue is positive, we have $\Lambda(\lambda) = \log \rho(\Pi_\lambda) \in (-\infty, +\infty)$ for all $\lambda \in \mathbb{R}^d$. Hence, (a), (b) hold and (d) is vacuously true. To check differentiability of Λ, we consider the characteristic equation

$$0 = \det[xI - \Pi_\lambda] = x^N + a_{N-1}(\lambda)x^{N-1} + \ldots + a_1(\lambda)X + a_0(\lambda),$$

where coefficients a_i are smooth functions of λ. Let $F(x,\lambda)$ denote the function of $(x,\lambda) \in \mathbb{R}^{d+1}$ on the far right-hand side. We have $F(\Lambda(\lambda),\lambda) = 0$ and, because the Perron-Frobenius eigenvalue has multiplicity 1, $\partial_x F(\Lambda(\lambda),\lambda) \neq 0$. Hence, it follows from the the implicit function theorem that Λ is a smooth function of λ.

Then, the conclusion from the Gärtner-Ellis theorem is that μ_n (the distribution of Z_n) satisfy an LDP with rate function $I(z) = \sup_{\lambda\in\mathbb{R}^d}\{\lambda\cdot z - \log\rho(\Pi_\lambda)\}$.

□

1.3 Key words:

1. Large deviation principle.
2. rate function
3. Gärtner-Ellis theorem
4. Perron-Frobenius theorem
5. Legendre-Fenchel transform
6. LDP of Markov chain

1.4 Exercise:

Exercise 1. Let μ_n be probability measures on $\mathbb{R}$ and $I : \mathbb{R} \to [0,+\infty]$ be a function (not necessarily lower semicontinuous). Define $\tilde{I}(x) = \min\{I(x), \liminf_{y\to x} I(y)\}$ for $x \in \mathbb{R}$.

(a) Show that $\tilde{I}$ is lower semicontinuous. (Hence, the assumption of lower semicontinuity is not restrictive. $\tilde{I}$ is called the lower semicontinuous regularization of I).

(b) Suppose that the lower and the upper bounds above hold for all Borel sets $B \subset \mathbb{R}$. Show that these bounds still hold if I is replaced with $\tilde{I}$, that is,

$$-\inf_{x\in B^0}\tilde{I}(x) \leqslant \liminf_{n\to+\infty}\frac{\log\mu_n(B)}{n}$$
$$-\inf_{x\in\overline{B}}\tilde{I}(x) \geqslant \limsup_{n\to+\infty}\frac{\log\mu_n(B)}{n}.$$

for all Borel sets $B \subset \mathbb{R}$. Moreover, $\tilde{I}$ is the unique lower semicontinuous function with range $[0,+\infty]$ that satisfy these bounds. (Hence, the rate function, if exists, is unique.)

Exercise 2. Let $\mathcal{M}_1([N])$ denote the set of probability measures on the set $[N] = \{1,\ldots,N\}$. We can identify each $\mu \in \mathcal{M}_1([N])$ with the vector $(\mu_j,\ldots,\mu_N)$, where $\mu_j = \mu(\{j\})$ for $j \in [N]$. The relative entropy of $q \in \mathcal{M}_1([N])$ with respect to $\mu \in \mathcal{M}_1([N])$ is defined as

$$H(q|\mu) = \sum_j q_j \log\left(\frac{q_j}{\mu_j}\right),$$

where we interpret $0\log 0$ and $0\log(0/0)$ as 0. Suppose that $q_j > 0$ for all $j \in [N]$. Show that

$$H(q|\mu) = \sup_{\substack{u\in\mathcal{M}_1([N])\\ u_i>0}} \sum_j q_j \log\left(\frac{u_j}{\mu_j}\right).$$

References

[1] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, second edition, 1998.

[2] T. Seppäläinen. Large deviations for increasing sequences on the plane. *Probab. Theory Relat. Fields*, 112(2):221–244, 1998.