Extension of LD to $\mathbb{R}^d$ and dependent process. Gärtner-Ellis Theorem

Content.

1. Large Deviations in may dimensions
2. Gärtner-Ellis Theorem
3. Large Deviations for Markov chains

1 Large Deviations in $\mathbb{R}^d$

Most of the developments in this lecture follows Dembo and Zeitouni book [1]. Let $X_n \in \mathbb{R}^d$ be i.i.d. random variables and $A \subset \mathbb{R}^d$. Let $S_n = \sum_{1\le i\le n} X_n$ The large deviations question is now regarding the existence of the limit

$$\lim_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\frac{S_n}{n}\in A).$$

Given $\theta \in \mathbb{R}^d$, define $M(\theta) = \mathbb{E}[\exp(\langle \theta, X_1\rangle)]$ where $\langle\cdot,\cdot\rangle$ represents the inner product of two vectors: $\langle a,b\rangle = \sum_i a_i b_i$. Define $I(x) = \sup_{\theta\in\mathbb{R}^d}(\langle\theta,x\rangle - \log M(\theta))$, where again $I(x)=\infty$ is a possibility.

Theorem 1 (Cramér's theorem in multiple dimensions)**.** *Suppose $M(\theta) < \infty$ for all $\theta \in \mathbb{R}^d$. Then*

(a) for all closed set $F \subset \mathbb{R}^d$,

$$\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\frac{S_n}{n}\in F) \le -\inf_{x\in F} I(x)$$

(b) for all open set $U \subset \mathbb{R}^d$

$$\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(\frac{S_n}{n}\in U) \ge -\inf_{x\in U} I(x)$$

Unfortunately, the theorem does not hold in full generality, and the additional condition such as $M(\theta) < \infty$ for all θ is needed. Known counterexamples are somewhat involved and can be found in a paper by Dinwoodie [2] which builds on an earlier work of Slaby [5]. The difficulty arises that there is no longer the notion of monotonicity of $I(x)$ as a function of the vector x. This is not the tightest condition and more general conditions are possible, see [1]. The proof of the theorem is skipped and can be found in [1].

Let us consider an example of application of Theorem 2. Let $X_n \stackrel{d}{=} N(0,\Sigma)$ where $d = 2$ and

$$\Sigma = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}, \; F = \{(x_1, x_2) : 2x_1 + x_2 \geq 5\}.$$

Goal: prove that the limit $\lim_n \frac{1}{n} \log \mathbb{P}(\frac{S_n}{n} \in F)$ exists, and compute it.

By the upper bound part,

$$\limsup_n \frac{1}{n} \log \mathbb{P}(\frac{S_n}{n} \in F) \leq -\inf_{x \in F} I(x)$$

We have

$$M(\theta) = \mathbb{E}[\exp(\langle \theta, X \rangle)]$$

Letting $\stackrel{d}{=}$ denote equality in distribution, we have

$$\begin{aligned} \langle \theta, X \rangle &\stackrel{d}{=} N(0, \theta^T \Sigma \theta) \\ &= N(0, \theta_1^2 + \theta_1 \theta_2 + \theta_2^2), \end{aligned}$$

where $\theta = (\theta_1, \theta_2)$. Thus

$$\begin{aligned} M(\theta) &= \exp(\frac{1}{2}(\theta_1^2 + \theta_1 \theta_2 + \theta_2^2)) \\ I(x) &= \sup_{\theta_1, \theta_2} (\theta_1 x_1 + \theta_2 x_2 - \frac{1}{2}(\theta_1^2 + \theta_1 \theta_2 + \theta_2^2)) \end{aligned}$$

Let

$$g(\theta_1, \theta_2) = \theta_1 x_1 + \theta_2 x_2 - \frac{1}{2}(\theta_1^2 + \theta_1 \theta_2 + \theta_2^2).$$

From $\frac{d}{d\theta_j} g(\theta_1, \theta_2) = 0$, we have that

$$x_1 - \theta_1 - \frac{1}{2}\theta_2 = 0, \quad x_2 - \theta_2 - \frac{1}{2}\theta_1 = 0,$$

from which we have

$$\theta_1 = \frac{4}{3}x_1 - \frac{2}{3}x_2, \quad \theta_2 = \frac{4}{3}x_2 - \frac{2}{3}x_1$$

Then

$$I(x_1, x_2) = \frac{2}{3}(x_1^2 + x_2^2 - x_1x_2).$$

So we need to find

$$\begin{aligned} &\inf_{x_1,x_2} \frac{2}{3}(x_1^2 + x^2 - x_1x_2) \\ &s.t. \quad 2x_1 + x_2 \geq 5 \quad (x \in F) \end{aligned}$$

This becomes a non-linear optimization problem. Applying the Karush-Kuhn-Tucker condition, we obtain

$$\begin{aligned} &\min f \quad s.t. \quad g \leq 0 \\ &\bigtriangledown f + \mu \bigtriangledown g = 0, \qquad (1) \\ &\mu g = 0 \\ &\mu < 0. \qquad (2) \end{aligned}$$

which gives

$$(\frac{4}{3}x_1 - \frac{2}{3}x_2, \frac{4}{3}x_2 - \frac{2}{3}x_1) + \mu(2,1) = 0, \quad \mu(2x_1 + x_2 - 5) = 0.$$

If $2x_1 + x_2 - 5 \neq 0$, then $\mu = 0$ and further $x_1 = x_2 = 0$. But this violates $2x_1 + x_2 \geq 5$. So we have $2x_1 + x_2 - 5 = 0$ which implies $x_2 = 5 - 2x_1$. Thus, we have a one dimensional unconstrained minimization problem:

$$\min \quad \frac{2}{3}x_1^2 + \frac{2}{3}(5 - 2x_1)^2 - x_1(5 - 2x_1)$$

which gives $x_1 = \frac{10}{11}$, $x_2 = \frac{35}{11}$ and $I(x_1, x_2) = 5.37$. Thus

$$\limsup_n \frac{1}{n} \log \mathbb{P}(\frac{S_n}{n} \in F) \leq -5.37$$

Applying the lower bound part of the Cramér's Theorem we obtain

$$\begin{aligned} &\liminf_n \frac{1}{n} \log \mathbb{P}(\frac{S_n}{n} \in F) \qquad (3) \\ &\quad \geq \liminf_n \frac{1}{n} \log \mathbb{P}(\frac{S_n}{n} \in F^o) \qquad (4) \\ &\quad =\geq \qquad - \inf_{2x_1+x_2>5} I(x_1, x_2) \\ &= - 5.37 \text{ (by continuity of } I). \end{aligned}$$

Combining, we obtain

$$\begin{aligned}\lim_n \frac{1}{n}\log \mathbb{P}(\frac{S_n}{n}\in F) &= \liminf_n \frac{1}{n}\log \mathbb{P}(\frac{S_n}{n}\in F)\\ &= \limsup_n \frac{1}{n}\log \mathbb{P}(\frac{S_n}{n}\in F)\\ &= -5.37.\end{aligned}$$

2 Gärtner-Ellis Theorem

The Gärtner-Ellis Theorem deals with large deviations event when the sequence X_n is not necessarily independent. One immediate application of this theorem is large deviations for Markov chains, which we will discuss in the following section.

Let X_n be a sequence of not necessarily independent random variables in $\mathbb{R}^d$. Then in general for $S_n = \sum_{1\le k\le n} X_k$ the identity $\mathbb{E}[\exp(\langle\theta, S_n\rangle)] = (\mathbb{E}[\exp(\langle\theta, X_1\rangle)])^n$ does not hold. Nevertheless there exists a broad set of conditions under which the large deviations bounds hold. Thus consider a general sequence of random variable $Y_n \in \mathbb{R}^d$ which stands for $(1/n)S_n$ in the i.i.d. case. Let $\phi_n(\theta) = \frac{1}{n}\log \mathbb{E}[\exp(n\langle\theta, Y_n\rangle)]$. Note that for the i.i.d. case

$$\begin{aligned}\phi_n(\theta) &= \frac{1}{n}\log \mathbb{E}[\exp(n\langle\theta, n^{-1}S_n\rangle)]\\ &= \frac{1}{n}\log M^n(\theta)\\ &= \log M(\theta)\\ &= \log \mathbb{E}[\exp(\langle\theta, X_1\rangle)].\end{aligned}$$

Loosely speaking Gärtner-Ellis Theorem says that when convergence

$$\phi_n(\theta) \to \phi(\theta) \tag{5}$$

takes place for some limiting function ϕ, then under certain additional technical assumptions, the large deviations principle holds for rate function

$$I(x) \triangleq \sup_{\theta\in\mathbb{R}}(\langle\theta, x\rangle - \phi(x)). \tag{6}$$

Formally,

Theorem 2. *Given a sequence of random variables Y_n, suppose the limit $\phi(\theta)$ (5) exists for all $\theta \in \mathbb{R}^d$. Furthermore, suppose $\phi(\theta)$ is finite and differentiable* *everywhere on $\mathbb{R}^d$. Then the following large deviations bounds hold for I defined by (6)*

$$\limsup_n \frac{1}{n} \log \mathbb{P}(Y_n \in F) \leq -\inf_{x \in F} I(x), \quad \textit{for any closed set } F \subset \mathbb{R}^d.$$

$$\liminf_n \frac{1}{n} \log \mathbb{P}(Y_n \in U) \geq -\inf_{x \in U} I(x), \quad \textit{for any open set } U \subset \mathbb{R}^d.$$

As for Theorem 1, this is not the most general version of the theorem. The version above is established as exercise 2.3.20 in [1]. More general versions can be found there as well.

Can we use Chernoff type argument to get an upper bound? For $\theta > 0$, we have

$$\begin{aligned} \mathbb{P}(\frac{1}{n} Y_n \geq a) &= \mathbb{P}(\exp(\theta Y_n) \geq \exp(\theta n a)) \\ &\leq \exp(-n(\theta a - \phi_n(\theta))) \end{aligned}$$

So we can get an upper bound

$$\sup_{\theta \geq 0} (\theta a - \phi_n(a))$$

In the i.i.d. case we used the fact that $\sup_{\theta \geq 0}(\theta a - M(\theta)) = \sup_\theta(\theta a - M(\theta))$ when $a > \mu = \mathbb{E}[X]$. But now we are dealing with the multidimensional case where such an identity does not make sense.

3 Large Deviations for finite state Markov chains

Let X_n be a finite state Markov chain with states $\Sigma = \{1, 2, \ldots, N\}$. The transition matrix of this Markov chain is $P = (P_{i,j}, 1 \leq i, j \leq N)$. We assume that the chain is irreducible. Namely there exists $m > 0$ such that $P_{i,j}^{(m)} > 0$ for all pairs of states i, j, where $P^{(m)}$ denotes the m-the power of P representing the transition probabilities after m steps. Our goal is to derive the large deviations bounds for the empirical means of the Markov chain. Namely, let $f : \Sigma \to \mathbb{R}^d$ be any function and let $Y_n = f(X_n)$. Our goal is to derive the large deviations bound for $n^{-1} S_n$ where $S_n = \sum_{1 \leq i \leq n} Y_k$. For this purpose we need to recall the Perron-Frobenious Theorem.

Theorem 3. *Let $B = (B_{i,j}, 1 \leq i, j \leq N)$ denote a non-negative irreducible matrix. Namely, $B_{i,j} \geq 0$ for all i, j and there exists m such that all the elements of B^m are strictly positive. Then B possesses an eigenvalue ρ called Perron-Frobenious eigenvalue, which satisfies the following properties.*

1. *$\rho > 0$ is real.*

2. *For every e-value λ of B, $|\lambda| \leq \rho$, where $|\lambda|$ is the norm of (possibly complex) λ.*

3. *The left and right e-vectors of B denoted by μ and ν corresponding to ρ, are unique up to a constant multiple and have strictly positive components.*

This theorem can be found in many books on linear algebra, for example [4].

The following corollary for the Perron-Frobenious Theorem shows that the essentially the rate of growth of the sequence of matrices B^n is ρ^n. Specifically,

Corollary 1. *For every vector $\phi = (\phi_j, 1 \leq j \leq N)$ with strictly positive elements, the following holds*

$$\lim_n \frac{1}{n} \log \left[\sum_{1 \leq j \leq N} B^n_{i,j} \phi_j \right] = \lim_n \frac{1}{n} \log \left[\sum_{1 \leq j \leq N} B^n_{j,i} \phi_j \right] = \log \rho.$$

Proof. Let $\alpha = \max_j \nu_j, \beta = \min_j \nu_j, \gamma = \max_j \phi_j, \delta = \min_j \phi_j$. We have

$$\frac{\gamma}{\beta} B^n_{i,j} \nu_j \geq B^n_{i,j} \phi_j \geq \frac{\delta}{\alpha} B^n_{i,j} \nu_j.$$

Therefore,

$$\begin{aligned} \lim_n \frac{1}{n} \log \left[\sum_{1 \leq j \leq N} B^n_{i,j} \phi_j \right] &= \lim_n \frac{1}{n} \log \left[\sum_{1 \leq j \leq N} B^n_{i,j} \nu_j \right] \\ &= \lim_n \frac{1}{n} \log(\rho^n \nu_i) \\ &= \log \rho. \end{aligned}$$

The second identity is established similarly. □

Now, given a Markov chain X_n, a function $f : \Sigma \to \mathbb{R}^d$ and vector $\theta \in \mathbb{R}^d$, consider a modified matrix $P_\theta = (e^{\langle \theta, f(j) \rangle} P_{i,j}, 1 \leq i, j \leq N)$. Then P_θ is an irreducible non-negative matrix, since P is such a matrix. Let $\rho(P_\theta)$ denote its Perron-Frobenious eigenvalue.

Theorem 4. *The sequence $\frac{1}{n} S_n = \frac{1}{n} \sum_{1 \leq i \leq k} f(X_n)$ satisfies the large deviations bounds with rate function $I(x) = \sum_{\theta \in \mathbb{R}^d} (\langle \theta, x \rangle - \log \rho(P_\theta))$. Specifically,* *for every state $i_0 \in \Sigma$, closed set $F \subset \mathbb{R}^d$ and every open set $U \subset \mathbb{R}^d$, the following holds:*

$$\limsup_n \frac{1}{n} \log \mathbb{P}(n^{-1}S_n \in F | X_0 = i_0) \leq - \inf_{x \in F} I(x),$$
$$\liminf_n \frac{1}{n} \log \mathbb{P}(n^{-1}S_n \in U | X_0 = i_0) \leq - \inf_{x \in U} I(x).$$

Proof. We will show that the sequence of functions $\phi_n(\theta) = \frac{1}{n} \log \mathbb{E}[e^{\langle \theta, S_n \rangle}]$ has a limit ϕ which is finite and differentiable everywhere. Given the starting state i_0 we have

$$\begin{aligned} \log \mathbb{E}[e^{\langle \theta, S_n \rangle}] &= \log \sum_{i_1, \ldots, i_n \in \Sigma} P_{i,i_1} P_{i_1,i_2} \cdots P_{i_{n-1},i_n} \prod_{1 \leq k \leq n} e^{\langle \theta, f(i_k) \rangle} \\ &= \log \left[\sum_{1 \leq j \leq N} P_\theta^n(i_0, j) \right], \end{aligned}$$

where $P_\theta^n(i,j)$ denotes the i,j-th entry of the matrix P_θ^n. Letting $\phi_j = 1$ and applying Corollary 1, we obtain

$$\lim_n \phi_n(\theta) = \log \rho(P_\theta).$$

Thus the Gärtner-Ellis can be applied provided the differentiability of $\log \rho(P_\theta)$ with respect to θ can be established. The Perron-Frobenious theory in fact can be used to show that such a differentiability indeed takes place. Details can be found in the book by Lancaster [3], Theorem 7.7.1. □

References

[1] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Springer, 1998.

[2] IH Dinwoodie, *A note on the upper bound for iid large deviations*, The Annals of Probability **19** (1991), no. 4, 1732–1736.

[3] Peter Lancaster and Miron Tismenetsky, *Theory of matrices*, vol. 2, Academic press New York, 1969.

[4] E Seneta, *Non-negative matrices andmarkov chains*, Springer2Verlag, New York (1981).

[5] M Slaby, *On the upper bound for large deviations of sums of iid random vectors*, The Annals of Probability (1988), 978–990.

15.070J / 6.265J Advanced Stochastic Processes
Fall 2013