3.3 Sample path large deviations

In many problems, the interest is actually in rare events that depend on a collection of random variables, or, more generally, on a random process.

Interest often lies in the probability that a path of a random process hits a particular set.

We will briefly addressed such questions. We start with the simplest example, namely sample path large deviations for random walks.

$X_1, X_2, \dots$ be a sequence of i.i.d random vectors taking values in $\mathbb{R}^d$, $\Lambda(\lambda) = \mathbb{E}[e^{\langle\lambda, X_1\rangle}] < \infty$ for all $\lambda \in \mathbb{R}^d$.

We know that Cramér's theorem gives large deviation principle for the mean $\frac{1}{n}\sum_{i=1}^{n} X_i$ (mean vector).

Similarly, the LDP for $1/n \sum_{i=1}^{[n/2]} X_i$ can be obtained.

Here, we are seeking the large deviations joint behaviour of a family of random variables indexed by $t \geq 0$.

Define

$$Z_n(t) = \frac{1}{n}\sum_{i=1}^{[nt]} X_i \quad , 0 \leq t \leq 1,$$

and let μ_n be the law of Z_n in $L_\infty([0,1]) = L_\infty([0,1]; \mathbb{R}^d)$ (essentially bounded functions

$[0,1] \to \mathbb{R}^d)$.

Theorem 16 (Mogulskii 1976):

The sequence $(\mu_n)_{n\geq 1}$ satisfies in $L_\infty([0,1])$ the LDP with the good rate function

$$I(\phi) = \begin{cases} \int_0^1 \Lambda^*(\dot{\phi}(t))\, dt & \text{, if } \phi \in \mathcal{AC},\ \phi(0)=0 \\ \infty & \text{, otherwise} \end{cases},$$

where $\mathcal{AC}$ denotes the space of absolutely continuous functions.

Remarks: (1) $\phi : [0,1] \to \mathbb{R}^d$ absolutely continuous implies that ϕ is differentiable almost everywhere (it is an integral of an $L_1([0,1])$ function).

(2) μ_n are supported on the space of functions continuous from the right and having left limits and $\mathcal{D}_I$ is a subset. Henceforth the LDP holds also when the space is equipped with the $\|\cdot\|_\infty$ - topology.

(3) Extensions to stochastic processes with jumps at random times exists.

Proof: We give a sketch of the proof and refer to chapter 5 in [DZ98] for further details.

Step 1: $\tilde{\mu}_n$ law of $\tilde{Z}_n$ in $L_\infty([0,1])$

$$\tilde{Z}_n(t) = Z_n(t) + \left(t - \frac{[nt]}{n}\right) X_{[nt]+1}$$

Then $(\mu_n)_{n\geq 1}$ and $(\tilde{\mu}_n)_{n\geq 1}$ are exponentially equivalent, i.e., LDP of one ~~infers~~ implies the LDP of the other.
Recall that $(\mu_n), (\nu_n)$ are exponentially equivalent if there is a coupling, that is a probability measure space $(\Omega, \mathcal{B}_n, P_n)$ such that $P_n^{(1)} = \mu_n$ and $P_n^{(2)} = \nu_n$ and $\overline{\lim_{n\to\infty}} \, 1/n \log P_n(\Gamma_\delta) = -\infty$, where,

for $\delta > 0$, $\Gamma_\delta = \{(x,y) \in E \times E : d(x,y) > \delta\}$.

Step 2: $X := \{\omega : [0,1] \to \mathbb{R}^d : \omega(0) = 0\}$, and equip X with the topology of pointwise convergence on $[0,1]$.
$(\tilde{\mu}_n)_{n\geq 1}$ LDP on X with rate good function I.

Step 3: $(\tilde{\mu}_n)_{n\geq 1}$ exponentially tight in $C_0([0,1])$ (space of all continuous $\omega : [0,1] \to \mathbb{R}^d$, such that $\omega(0) = 0$, equipped with supremum norm topology).

We outline step 2 and refer to [DZ98] for the remaining ones and the proof of the theorem.

Let $\mathcal{J}$ denote the collection of all ordered finite subsets of $(0,1]$. For any $j = \{0 < t_1 < t_2 \cdots < t_{|j|} \leq 1\} \in \mathcal{J}$ and $f: [0,1] \to \mathbb{R}^d$, let $p_j(f)$ denote the vector $(f(t_1), f(t_2), \ldots, f(t_{|j|})) \in (\mathbb{R}^d)^{|j|}$. Then $(\mu_n \circ p_j^{-1})_{n \geq 1}$ satisfies the LDP in $(\mathbb{R}^d)^{|j|}$ with the good rate function

$$I_j(z) = \sum_{\ell=1}^{|j|} (t_\ell - t_{\ell-1}) \Lambda^* \left(\frac{z_\ell - z_{\ell-1}}{t_\ell - t_{\ell-1}} \right),$$

where $z = (z_1, \ldots, z_{|j|})$ and $t_0 = 0$, $z_0 = 0$.

To see this define the following random objects:

$$Z_n^j = (Z_n(t_1), \ldots, Z_n(t_{|j|})) \quad \text{and}$$

$$Y_n^j = (Z_n(t_1), Z_n(t_2) - Z_n(t_1), \ldots, Z_n(t_{|j|}) - Z_n(t_{|j|-1}))$$

$Y_n^j \mapsto Z_n^j$ is ~~map~~ onto itself, continuous and one to one, hence contraction principle applies: LDP for $(Y_n^j)_{n \geq 1}$ with rate function

$$\Lambda_j^*(y) := \sum_{\ell=1}^{|j|} (t_\ell - t_{\ell-1}) \Lambda^* \left(\frac{y_\ell}{t_\ell - t_{\ell-1}} \right), \quad y \in (\mathbb{R}^d)^{|j|}$$

It is easy to see that

$$\Lambda_j^*(y) = \sup_{\lambda \in (\mathbb{R}^d)^{|j|}} \{ \underbrace{\langle \lambda, y \rangle}_{\text{in } (\mathbb{R}^d)^{|j|}} - \Lambda_j(\lambda) \}, \text{ where}$$

$\Lambda_j(\lambda) = \sum_{\ell=1}^{|j|} (t_\ell - t_{\ell-1}) \Lambda(\lambda_\ell)$. Thus Λ_j^* is the Fenchel-Legendre transform of the finite and differentiable function Λ_j. The LDP for $(Y_n^j)_{n\geq 1}$ follows from the Gärtner-Ellis theorem (Theorem 10), since by independence of X_i, $\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}[e^{n\langle\lambda, Y_n^j\rangle}] = \Lambda_j(\lambda)$.

$(\mu_n \circ p_j^{-1})_{n\geq 1}$ and $(\tilde{\mu}_n \circ p_j^{-1})_{n\geq 1}$ are exponentially equivalent in $(\mathbb{R}^d)^{|j|}$, thus we also get the following LDP: For any $j \in J$, $(\tilde{\mu}_n \circ p_j^{-1})_{n\geq 1}$ satisfies the LDP in $(\mathbb{R}^d)^{|j|}$ with the good rate function I_j above.

We now turn to the proof of step 2. For $i, j \in J$, $i = \{s_1, \dots, s_{|i|}\} \leq j = \{t_1, \dots, t_{|j|}\}$ iff for any ℓ, $s_\ell = t_{q(\ell)}$ for some $q(\ell)$.

Then, for $i \leq j \in J$, the projection $p_{ij} : (\mathbb{R}^d)^{|j|} \to (\mathbb{R}^d)^{|i|}$ is defined in the natural way.

Let X denote the projective limit of $\{Y_j = (\mathbb{R}^d)^{|j|}\}_{j\in J}$ with respect to the projections p_{ij}.

Applying the Dawson-Gärtner theorem (we had no time to discuss that and henceforth refer to [DZ98]) we obtain the rate function

$$I_X = \sup_{0=t_0<t_1<\dots<t_k\leq 1} \sum_{\ell=1}^{k} (t_\ell - t_{\ell-1}) \Lambda^*\left(\frac{f(t_\ell)-f(t_{\ell-1})}{t_\ell - t_{\ell-1}}\right)$$

where $f:[0,1]\to\mathbb{R}^d$, $f(0)=0$, $f(t)=x_{\{t\}}$.
Note that each point in the projective limit space $\mathcal{X}$, i.e., each $x=(x_j)_{j\in J}$, may be identified with the map f.

It remains to show that $I_{\mathcal{X}}=I$.
The convexity of Λ^* implies by Jensen's inequality that $I(\phi)\geq I_{\mathcal{X}}(\phi)$. As for the opposite inequality, first consider $\phi\in\mathcal{AC}$.
Let $g(t):=\frac{d\phi(t)}{dt}\in L_1([0,1])$, and define

$$g^k(t)=k\int_{[kt]/k}^{([kt]+1)/k} g(s)\,ds \quad, t\in[0,1), k\geq 1, \quad g^k(1)=\int_{1-1/k}^{1} g(s)ds$$

$$I_{\mathcal{X}}(\phi)\geq\liminf_{k\to\infty}\sum\frac{1}{k}\Lambda^*\left(k\left(\phi\left(\frac{l}{k}\right)-\phi\left(\frac{l-1}{k}\right)\right)\right)$$

$$=\liminf_{k\to\infty}\int_0^1\Lambda^*(g^k(t))\,dt$$

By Lebesgue's theorem, $\lim_{k\to\infty}g^k(t)=g(t)$ almost everywhere in $[0,1]$. Hence, by Fatou's lemma and the lower semicontinuity of Λ^*,

$$\liminf_{k\to\infty}\int_0^1\Lambda^*(g^k(t))\,dt\geq\int_0^1\liminf_{k\to\infty}\Lambda^*(g^k(t))\,dt$$
$$\geq\int_0^1\Lambda^*(g(t))\,dt=I(\phi)$$

$\blacksquare$

Brownian motion sample path large deviations:

Let B_t, $t \in [0,1]$ (time horizon), denote a standard Brownian motion in $\mathbb{R}^d$. Consider the process

$B_\varepsilon(t) = \sqrt{\varepsilon}\, B_t$, and let ν_ε be the probability

measure induced by B_ε on $C_0([0,1])$, the space of all continuous functions $\phi : [0,1] \to \mathbb{R}^d$ such that $\phi(0) = 0$, equipped with the supremum norm topology.

B_ε is a candidate for an LDP similar to the following one (compare with Z_n above)

$$Y_\varepsilon(t) = \varepsilon \sum_{i=1}^{[t/\varepsilon]} X_i, \quad 0 \le t \le 1, \qquad \text{and } \varepsilon \downarrow 0.$$

First observe the following elementary fact.

Lemma 17: For any $d \in \mathbb{N}$ and any $\tau, \varepsilon, \delta > 0$

$$\mathbb{P}\Big(\sup_{0 \le t \le \tau} |B_\varepsilon(t)| \geq \delta \Big) \le 4d\, e^{-\delta^2/2d\tau\varepsilon}.$$

Proof:

$$\mathbb{P}\Big(\sup_{0 \le t \le \tau} |B_\varepsilon(t)| \geq \delta \Big) = \mathbb{P}\Big(\sup_{0 \le t \le \tau} |B_t|^2 \geq \varepsilon^{-1}\delta^2 \Big)$$

$$\le d\, \mathbb{P}\Big(\sup_{0 \le t \le \tau} (B_t)_1^2 \geq \frac{\delta^2}{d\varepsilon} \Big) \quad \text{because}$$

$\{x \in \mathbb{R}^d : |x|^2 \geq \alpha\} \subset \bigcup_{i=1}^{d} \{x \in \mathbb{R}^d : |x_i|^2 \geq \frac{\alpha}{d}\}$.

The laws of B_t

and $\sqrt{\tau}\, B_{t/\tau}$ are identical, we get by time rescaling

$$\mathbb{P}\Big(\sup_{0 \leq t \leq \tau} |B_\varepsilon(t)| \geq \delta\Big) \leq d\,\mathbb{P}\Big(\|(B_t)_1\| \geq \frac{\delta}{\sqrt{d\tau\varepsilon}}\Big).$$

Now define $B_t^* = (B_t)_1$, where B_t^* is a one-dimensional Brownian motion.

B_t^* and $-B_t^*$ have the same law in $C_0([0,1])$,

$$\mathbb{P}(\|B_t^*\| \geq \eta) \leq 2\mathbb{P}\Big(\sup_{0 \leq t \leq 1} B_t^* \geq \eta\Big)$$

$$= 4\mathbb{P}(B_1^* \geq \eta) \leq 4e^{-\eta^2/2}$$, where the equality is Désiré André's reflection principle. ∎

~~Thus~~

Thus, $\|B_\varepsilon\| \xrightarrow[\varepsilon \downarrow 0]{} 0$ in probability (actually, almost surely) and exponentially fast in $1/\varepsilon$.

<u>Theorem 18: (Schilder)</u>

$(\nu_\varepsilon)_{\varepsilon > 0}$ satisfies, in $C_0([0,1])$, the LDP with good rate function

$$I_B(\phi) = \begin{cases} \frac{1}{2}\int_0^1 |\dot{\phi}(t)|^2\,dt, & \phi \in H_1, \\ \infty, & \text{otherwise.} \end{cases}$$

Remarks: (1) $H_1 = \{ \phi : \phi(t) = \int_0^t f(s)\,ds : f \in L_2([0,1]) \}$ is the space of all absolutely continuous functions with square integrable derivative equipped with the norm $\|g\|_{H_1} = [\int_0^1 |\dot{g}(t)|^2 dt]^{1/2}$.

(2) The LDP is for $\varepsilon \to 0$.

Proof: $\widehat{B}_\varepsilon(t) := B_\varepsilon(\varepsilon[t/\varepsilon])$

is Y_ε defined above $\left(\text{i.e., } Y_\varepsilon(t) = \varepsilon \sum_{i=1}^{[t/\varepsilon]} X_i \right)$, for the particular choice of X_i, which are standard Normal random variables in $\mathbb{R}^d$ (zero mean and covariance matrix the identity).

Theorem 16 can be easily applied to show that the probability laws of $\widehat{B}_\varepsilon$ satisfy the LDP in $L_\infty([0,1])$ with the good rate function $I(\phi) = \begin{cases} \int_0^1 \Lambda^*(\dot\phi(t))\,dt, & \text{if } \phi \in \mathcal{AC}, \phi(0)=0 \\ \infty & \text{, otherwise.} \end{cases}$

For standard Normal variables considered here, we compute

$$\Lambda(\lambda) = \log \mathbb{E}[e^{\langle \lambda, X_1 \rangle}] = \frac{1}{2}|\lambda|^2, \text{ implying}$$

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d} \{ \langle \lambda, x \rangle - 1/2 |\lambda|^2 \} = \frac{1}{2}|x|^2.$$

Hence, $\mathcal{D}_I = H_1$, and the rate function I specialises to I_B.

For any $\delta > 0$,

$$\mathbb{P}(\|B_\varepsilon - \hat{B}_\varepsilon\| \geq \delta) \leq \left(\left[\tfrac{1}{\varepsilon}\right] + 1\right) \mathbb{P}\left(\sup_{0 \leq t \leq \varepsilon} |B_\varepsilon(t)| \geq \delta\right)$$

$$\leq 4d\, \varepsilon^{-1}(1+\varepsilon)\, e^{-\delta^2/(2d\varepsilon^2)}$$

(follows by time-homogeneity of increments of the Brownian motion and by Lemma 17).

Consequently, $\limsup_{\varepsilon \to 0} \varepsilon \log \mathbb{P}(\|B_\varepsilon - \hat{B}_\varepsilon\| \geq \delta) = -\infty$,

and by exponential equivalence, it follows that $(\nu_\varepsilon)_{\varepsilon > 0}$ satisfies the LDP in $L_\infty([0,1])$ with the good rate function I_B. The restriction to $C_0([0,1])$ follows since $B_\varepsilon \in C_0([0,1])$ with probability one ■

The preceding results can be applied and extended also to multivariate random walks (multi-index random process) and Brownian sheet.

Real-world applications are quite numerous, among which we just mention asymptotic performance analysis of communication systems (systems of queues in digital networks).

The book by Shwartz and Weiss "Large Deviations For Performance Analysis: Queues, Communication and Computing, 1995, Chapman / Hall.

Another application concerns perturbed dynamical systems (Freidlin-Wentzell) which we will study later in chapter 6.

Another area of direct real-world applications are large exceedances in $\mathbb{R}^d$ (applications related to DNA sequence matching, queueing networks, and abrupt change detection in dynamical systems):

X_i, i.i.d. $\mathbb{R}^d$-valued random variables with zero mean and finite logarithmic moment generating function $\Lambda(\lambda)$.

Let $b \neq 0$ be a given drift vector in $\mathbb{R}^d$ and define

$$Z_n = \sum_{i=1}^{n} (X_i - b) .$$

The increments $Z_n - Z_m$ of the random walk are $\approx -(n-m)b + \xi_{n-m}$, where ξ_{n-m} has zero mean and covariance of the order $(n-m)$. Thus, with $n-m$ large, it becomes unlikely that $Z_n - Z_m$ is far away from $-(n-m)b$. Interest lies in the probability of the rare events $\{Z_n - Z_m \in A/\varepsilon\}$, for small ε, where A is a closed set that does not intersect the typical ray $\{-\alpha b\}_{\alpha \geq 0}$.

$$Y_t^\varepsilon = \varepsilon \sum_{i=1}^{[t/\varepsilon]} (X_i - b) \quad , t \in [0, \infty) ,$$

rare events $Y_t^\varepsilon - Y_\tau^\varepsilon \in A$.

$T_\varepsilon = \inf\{t : \exists s \in [0,t)$ such that $Y_t^\varepsilon - Y_s^\varepsilon \in A\}$

$\tau_\varepsilon = \sup\{s \in [0, T_\varepsilon) : Y_{T_\varepsilon}^\varepsilon - Y_s^\varepsilon \in A\} - \varepsilon$; $L_\varepsilon = T_\varepsilon - \tau_\varepsilon$.

LDP analysis provides estimates and shows $\varepsilon \log T_\varepsilon$ and L_ε converge.