Exam Large Deviation Theory

January 30, 2017

Exercise 1 Let $I : \mathbb{R} \to (-\infty, \infty]$ and $J : \mathbb{R} \to (-\infty, \infty]$ be good rate functions such that

$$\sup_{x\in\mathbb{R}} \big[\lambda x - I(x)\big] = \sup_{x\in\mathbb{R}} \big[\lambda x - J(x)\big] \qquad (\lambda \in \mathbb{R}).$$

(a) Assume that I is convex. Prove that $I \leq J$.

(b) Assume that I is strictly convex in the sense that for any $x < y$ with $I(x), I(y) < \infty$ and $0 < p < 1$, one has $I(px + (1-p)y) < pI(x) + (1-p)I(y)$. Prove that $I = J$.

Hint: You can use the following facts. For $a, \lambda \in \mathbb{R}$, let $l_{a,\lambda}$ denote the affine function $l_{a,\lambda}(x) := a + \lambda x$ $(x \in \mathbb{R})$. Then for any function $I : \mathbb{R} \to (-\infty, \infty]$

$$\overline{I}(x) := \sup\{l_{a,\lambda}(x) : a, \lambda \in \mathbb{R},\ l_{a,\lambda} \leq I\} \qquad (x \in \mathbb{R})$$

is the convex hull of I. In particular, if I is convex and lower semi-continuous, then $I = \overline{I}$. Also, if I is a lower semi-continuous function that is not convex, then $\overline{I}$ is not strictly convex.

In the following exercise you can use the result of Exercise 1, as well as the following proposition, the proof of which is entirely analogous to the proof of Proposition 1.29 of the lecture notes. Below, if I is a normalized good rate function on a Polish space E, and $\mathcal{D} \subset \mathcal{C}_{\rm b}(E)$, then we say that $\mathcal{D}$ *determines* I if for any other normalized good rate function J,

$$\|f\|_{\infty,I} = \|f\|_{\infty,J} \quad \forall f \in \mathcal{D} \quad \text{implies} \quad I = J.$$

Proposition Let E be a Polish space, let μ_n be probability measures on E, let s_n be positive constants converging to infinity, and let I be a normalized good rate function on E. Assume that $\mathcal{D} \subset \mathcal{C}_b(E)$ determines I and that

(i) The sequence $(\mu_n)_{n\geq 1}$ is exponentially tight with speed s_n.

(ii) $\lim_{n\to\infty} \|f\|_{s_n,\mu_n} = \|f\|_{\infty,I}$ for all $f \in \mathcal{D}$.

Then the measures μ_n satisfy the large deviation principle with speed s_n and rate function I.

Exercise 2 Let $K \subset \mathbb{R}$ be a compact interval and let $(X_i)_{i\geq 1}$ be a sequence of random variables taking values in K. Assume that the limit

$$\Gamma(\lambda) := \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\big[e^{\lambda \sum_{i=1}^n X_i}\big]$$

exists for each $\lambda \in \mathbb{R}$, and that $\Gamma : \mathbb{R} \to (-\infty, \infty]$ is an element of the class Conv_∞ defined in Section 2.1 of the Lecture Notes. Prove that the measures

$$\mathbb{P}\Big[\frac{1}{n}\sum_{i=1}^{n} X_i \in \cdot\,\Big]$$

satisfy the Large Deviations Principle with speed n and good rate function I given by

$$I(x) := \sup_{\lambda \in \mathbb{R}} \big[\lambda x - \Gamma(\lambda)\big] \qquad (x \in \mathbb{R}).$$

Warning: We do *not* assume that the random variables $(X_i)_{i\geq 1}$ are independent.

Exercise 3 Let $(X_i)_{i\geq 1}$ denote an i.i.d. sequence of random variables taking values in the set $\{A, T, C, G\}$ with common law μ defined by

$$\mu(A) = 0.3, \quad \mu(T) = 0.3, \quad \mu(C) = 0.2, \quad \text{and} \quad \mu(G) = 0.2.$$

Let

$$R_n := \frac{1}{n}\sum_{i=1}^{n} 1_{\{X_i = X_{i+1}\}}$$

denote the frequency, among the first $n+1$ letters in the sequence, with which a letter is followed by the same letter. Indicate a method to calculate the limit

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}[R_n \geq 0.5].$$

Solutions

Ex 1 The assumption implies that

$$l_{a,\lambda} \le I \quad \Leftrightarrow \quad \sup_{x\in\mathbb{R}} \big[a + \lambda x - I(x)\big] \le 0 \quad \Leftrightarrow \quad \sup_{x\in\mathbb{R}} \big[\lambda x - I(x)\big] \le -a \quad \Leftrightarrow \quad l_{a,\lambda} \le J,$$

which shows that $\overline{I} = \overline{J}$. If I is convex, then this implies that $I = \overline{J}$, so part (a) follows from the fact that $\overline{J} \le J$. If I is strictly convex, then $I = \overline{J}$ implies that J is convex and hence $I = \overline{J} = J$, solving part (b).

Ex 2 For each $\lambda \in \mathbb{R}$, let l_λ denote the linear function $l_\lambda(x) := \lambda x$ $(x \in K)$. Then, for any rate function J, one has

$$\log \|e^{l_\lambda}\|_{\infty,J} = \log \sup_{x\in K} e^{-J(x)} |e^{\lambda x}| = \sup_{x\in K} \big[\lambda x - J(x)\big].$$

In view of this, Exercise 1 shows that if $I : K \to \mathbb{R}$ is a strictly convex good rate function, then the functions $\mathcal{D} := \{e^{l_\lambda} : \lambda \in \mathbb{R}\}$ determine I. (Note that $e^{l_\lambda} \in \mathcal{C}_{\rm b}(K)$ by the assumption that K is compact.) Let

$$I(x) := \sup_{\lambda\in\mathbb{R}} \big[\lambda x - \Gamma(\lambda)\big] \qquad (x \in \mathbb{R})$$

denote the Legendre transform of Γ. Since $\Gamma \in \mathrm{Conv}_\infty$, Proposition 2.3 in the Lecture Notes tells us that $I \in \mathrm{Conv}_\infty$ and that Γ is the Legendre transform of I, i.e.,

$$\Gamma(\lambda) = \sup_{x\in K} \big[\lambda x - I(x)\big] \qquad (\lambda \in \mathbb{R}).$$

Let μ_n denote the law of $\frac{1}{n}\sum_{i=1}^n X_i$. Then our assumptions say that

$$\log \|e^{l_\lambda}\|_{n,\mu_n} = \log \frac{1}{n} \log \int |e^{l_\lambda}|^n \mathrm{d}\mu_n = \mathbb{E}\big[e^{\lambda \sum_{i=1}^n X_i}\big] \underset{n\to\infty}{\longrightarrow} \Gamma(\lambda) = \sup_{x\in K} \big[\lambda x - I(x)\big],$$

and hence $\|e^{l_\lambda}\|_{n,\mu_n} \to \|e^{l_\lambda}\|_{\infty,I}$ for all $\lambda \in \mathbb{R}$. In particular, applying this to $\lambda = 0$, we see that I is normalized. Since K is compact, the measures μ_n are automatically exponentially tight, so applying the proposition, we obtain that the μ_n satisfy the Large Deviations Principle with speed n and good rate function I.

Ex 3 Let $S := \{A, T, C, G\}$ and let $M_n^{(2)}$ be the pair empirical measures of the sequence $(X_i)_{i\ge1}$, i.e.,

$$M_n^{(2)}(x, y) := \frac{1}{n} \sum_{i=1}^n 1_{\{(X_i,X_{i+1})=(x,y)\}} \qquad (x, y \in S).$$

Let $\mathcal{M}_1(S^2)$ denote the space of probability distributions on S^2, define $O \subset \mathcal{M}_1(S^2)$ by

$$O := \big\{\nu \in \mathcal{M}_1(S^2) : \sum_{x\in S} \nu(x, x) > 0.5\big\},$$

and let $\overline{O}$ denote the closure of O. (Since $\mathcal{M}_1(S^2)$ is a finite dimensional space, there is only one reasonable topology on it.) Let ν^1, ν^2 denote the first and second marginal of a probability distribution $\nu \in \mathcal{M}_1(S^2)$ and set $\mathcal{V} := \{\nu \in \mathcal{M}_1(S^2) : \nu^1 = \nu^2\}$. Then Theorem 3.2 or Theorem 3.16 (a) of the Lecture Notes tells us that

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}[R_n \in \overline{O}] &= - \inf_{\nu \in \mathcal{V}\cap\overline{O}} H(\nu|\nu^1 \otimes \mu), \\
\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}[R_n \in O] &= - \inf_{\nu \in \mathcal{V}\cap O} H(\nu|\nu^1 \otimes \mu),
\end{aligned}$$

where H denotes the relative entropy function and $\nu^1 \otimes \mu$ denotes the product measure of ν^1 and μ. By the continuity of the function $\mathcal{V} \ni \nu \mapsto H(\nu|\nu^1 \otimes \mu)$ (which follows from the fact that $\mu > 0$) and the fact that $\overline{O}$ is the closure of its interior, it follows that the liminf and limsup are equal (see Remark 1 below Proposition 1.7 in the Lecture Notes), so

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}[R_n \geq 0.5] = - \inf_{\nu \in \mathcal{V}\cap\overline{O}} H(\nu|\nu^1 \otimes \mu).$$

To evaluate the minumum, we need to minimize $\nu \mapsto H(\nu|\nu^1 \otimes \mu)$ over $\mathcal{V}$ subject to the constraint

$$\sum_{x\in S} \nu(x,x) \geq 0.5.$$

We use the method of Lagrange multipliers and try to maximize

$$\lambda \sum_{x\in S} \nu(x,x) - H(\nu|\nu^1 \otimes \mu)$$

for all possible values of $\lambda \in \mathbb{R}$, and then choose λ so that $\sum_{x\in S} \nu(x,x) = 0.5$. Define $\phi : S^2 \to \mathbb{R}$ by $\phi(x,y) := 1_{\{x=y\}}$. Then we want to find

$$r_\lambda := \sup_{\nu\in\mathcal{V}} \big[\lambda\nu\phi - H(\nu|\nu^1 \otimes \mu)\big]$$

By Lemma 3.25 of the Lecture Notes, r_λ is the Perron-Frobenius eigenvalue of the matrix $\big(A(x,y)\big)_{x,y\in S}$ defined by

$$A(x,y) = \mu(y) e^{\lambda\phi(x,y)} \qquad (x,y \in S).$$

Moreover, if h denotes the associated (right) Perron-Frobenius eigenvector, then the supremum is attained in $\nu = \pi \otimes A_h$, where A_h is the probability kernel defined in (3.17) and π is its invariant law. Based on the symmetry of the problem, we must have $h(A) = h(T)$ and $h(C) = h(G)$. Since

$$\begin{aligned}
Ah(A) &= 0.3e^\lambda h(A) + 0.3h(T) + 0.2h(C) + 0.2h(G), \\
Ah(C) &= 0.3h(A) + 0.3h(T) + 0.2e^\lambda h(C) + 0.2h(G),
\end{aligned}$$

our eigenvector equation simplifies to

$$\begin{pmatrix} 0.3(1+e^{\lambda}) & 0.4 \\ 0.6 & 0.2(1+e^{\lambda}) \end{pmatrix} \begin{pmatrix} h(A) \\ h(C) \end{pmatrix} = r \begin{pmatrix} h(A) \\ h(C) \end{pmatrix}.$$

To find the eigenvalues, we must solve

$$\det \begin{pmatrix} 0.3(1+e^{\lambda}) - r & 0.4 \\ 0.6 & 0.2(1+e^{\lambda}) - r \end{pmatrix} = 0,$$

or equivalently

$$\big(3(1+e^{\lambda}) - r\big)\big(2(1+e^{\lambda}) - r\big) - 4 * 6 = 0.$$

Now the formulas become somewhat unpleasant, but in principle we have a formula for the largest root to this quadratic equation. Using this, we can solve h and A_h. Then we still have to find the invariant law π of A_h and choose λ such that $\sum_x \nu(x,x) = \sum_x (\pi \otimes A_h)(x,x)$ equals 0.5. In principle, this can be done, but not within the timeframe of an exam.