## C8.6 Limit Theorems and Large Deviations in Probability

## Sheet 3 HT 2021 (Sections 8, 9)

**1**. Recall that if $\mu$ is a probability distribution on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ which is exponential integrable, i.e. $\int_{\mathbb{R}} e^{\lambda x}\mu(dx) < \infty$ for every real number $\lambda$, then

$$I_\mu(x) = \sup_\lambda \left(\lambda x - \log \int_{\mathbb{R}} e^{\lambda x}\mu(dx)\right)$$

for $x \in \mathbb{R}$. There is a similar definition for a distribution on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$.

(i) Let $a < b$ be two numbers, $p \in (0, 1)$, and $\mu = p\delta_a + (1-p)\delta_b$. Show that

$$I_\mu(x) = \begin{cases} \frac{x-a}{b-a}\log\frac{x-a}{1-p} + \frac{b-x}{b-a}\log\frac{b-x}{p} - \log(b-a) & \text{if } x \in (a,b); \\ -\log p & \text{if } x = a; \\ -\log(1-p) & \text{if } x = b; \\ \infty & \text{if } x \notin [a,b]. \end{cases}$$

(ii) Let $\mu$ be the exponential distribution, so with a pdf $e^{-x}$ for $x \geq 0$. Show that

$$I_\mu(x) = \begin{cases} x - \log x - 1 & \text{for } x > 0; \\ \infty & \text{for } x \leq 0. \end{cases}$$

(iii) Suppose $\mu$ is the normal distribution $N(a, \sigma^2)$, show that

$$I_\mu(x) = \frac{1}{2\sigma^2}|x-a|^2$$

for every $x \in \mathbb{R}$.

State the corresponding Cramér's large deviation principle for each case.

**2**. (*Large deviation principle for Gaussian measures*) Let $\boldsymbol{\sigma} = (\sigma_{ij})$ be a positive-definite and symmetric $d \times d$ matrix, whose inverse matrix is denoted by $\boldsymbol{\sigma}^{-1} = (\sigma^{ij})$. Let $\mu$ be the normal distribution with mean $\boldsymbol{a} = (a_i)$ and co-variance matrix $\boldsymbol{\sigma}$, denoted by $N(\boldsymbol{a}, \boldsymbol{\sigma})$, which is the probability measure on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ whose pdf is given by

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d \det \boldsymbol{\sigma}}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})\cdot\boldsymbol{\sigma}^{-1}(\boldsymbol{x}-\boldsymbol{a})}$$

for $\boldsymbol{x} \in \mathbb{R}^d$, where $\boldsymbol{x} \cdot \boldsymbol{\sigma}^{-1}\boldsymbol{y} = \sum_{i,j} x_i \sigma^{ij} y_j$. The characteristic function

$$\int_{\mathbb{R}^d} e^{\sqrt{-1}\boldsymbol{t}\cdot\boldsymbol{x}}\mu(d\boldsymbol{x}) = \exp\left[\sqrt{-1}\boldsymbol{t}\cdot\boldsymbol{a} - \frac{1}{2}\boldsymbol{t}\cdot\boldsymbol{\sigma}\boldsymbol{t}\right]$$

for all $\boldsymbol{t}$.

(i) Let $X_1, \cdots, X_n$ be independent with the same distribution $\mu$, and $\mu_n$ be the distribution of $S_n = \frac{1}{n}(X_1 + \cdots + X_n)$. Show that the distribution $\mu_n \sim N(\boldsymbol{a}, n^{-1}\boldsymbol{\sigma})$.

(ii) For $\varepsilon > 0$ consider the scaling mapping $\mathit{\Gamma}_\varepsilon : \mathbb{R}^d \to \mathbb{R}^d$ by sending $\boldsymbol{x}$ to $\sqrt{\varepsilon}\boldsymbol{x}$. Then both $\mathit{\Gamma}_\varepsilon$ and its inverse $\mathit{\Gamma}_\varepsilon^{-1}$ are Borel measurable. Let $\nu_\varepsilon = \mu \circ \mathit{\Gamma}_\varepsilon^{-1}$. Show that $\nu_\varepsilon$ has a normal $N(\sqrt{\varepsilon}\boldsymbol{a}, \varepsilon\boldsymbol{\sigma})$. Therefore, for the case where $\boldsymbol{a} = 0$, $\mu_n = \nu_{1/n}$ for any $n = 1, 2, \cdots$.

(iii) Suppose $\boldsymbol{a} = 0$, show that

$$I_\mu(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x} \cdot \boldsymbol{\sigma}^{-1}\boldsymbol{x}$$

for all $\boldsymbol{x} \in \mathbb{R}^n$. State Cramér's large deviation principle for $\mu_n$.

(iv) Now consider $\nu_\varepsilon = \mu \circ \Gamma_\varepsilon^{-1}$ for $\varepsilon \in (0,1)$, where $\mu \sim N(\boldsymbol{0}, \boldsymbol{\sigma})$. Let $n(\varepsilon) = \left[\frac{1}{\varepsilon}\right]$ the integer part of $1/\varepsilon$ for $\varepsilon \in (0,1)$, and $\gamma(\varepsilon) = \varepsilon n(\varepsilon)$. Then $1 - \varepsilon \leq \gamma(\varepsilon) \leq 1$ and $n(\varepsilon) \to \infty$ as $\varepsilon \downarrow 0$. Moreover $\nu_\varepsilon = \mu_{n(\varepsilon)} \circ \Gamma_{\gamma(\varepsilon)}^{-1}$. Hence show that $(\nu_\varepsilon)$ satisfies the large deviation

$$\limsup_{\varepsilon \to 0} \varepsilon \log\left(\nu_\varepsilon(F)\right) \leq -\inf_F I_\mu$$

for any closed subset $F \subset \mathbb{R}^d$, and

$$\limsup_{\varepsilon \to 0} \varepsilon \log\left(\nu_\varepsilon(G)\right) \geq -\inf_G I_\mu$$

for every open subset $G \subset \mathbb{R}^d$.

**3**. (i) Let $f : E \to E'$ be continuous mapping between two metric spaces $(E, \rho)$ and $(E', \rho')$. Suppose a family of random variables $Z^\varepsilon$ (where $\varepsilon \in (0,1)$) valued in $E$ satisfies a large deviation principle with a good rate function $I$:

$$\limsup_{\varepsilon \to 0} \varepsilon \log P\left[Z^\varepsilon \in F\right] \leq -\inf_F I$$

for every closed $F \subset E$, and

$$\limsup_{\varepsilon \to 0} \varepsilon \log P\left[Z^\varepsilon \in G\right] \geq -\inf_G I$$

for every open subset $G \subset E$.

Show that $X^\varepsilon = f(Z^\varepsilon)$ satisfies the large deviation with rate function

$$I'(s') = \inf\{I(s) : s \in E \text{ such that } f(s) = s'\}.$$

(ii) Let $\mu$ be the normal distribution $N(0, \boldsymbol{\sigma})$ and $T_{\boldsymbol{a}} : \boldsymbol{x} \to \boldsymbol{x} + \boldsymbol{a}$ where $\boldsymbol{a}$ is a fixed vector of $\mathbb{R}^d$. Then $\mu^{\boldsymbol{a}} = \mu \circ T_{\boldsymbol{a}}^{-1}$ has a normal distribution $N(\boldsymbol{a}, \boldsymbol{\sigma})$. Show that $\mu_\varepsilon^{\boldsymbol{a}} = \mu^{\boldsymbol{a}} \circ \Gamma_\varepsilon^{-1}$ satisfies a large deviation principle as $\varepsilon \downarrow 0$, where $\Gamma_\varepsilon \boldsymbol{x} = \sqrt{\varepsilon}\boldsymbol{x}$ for $\varepsilon > 0$.

**4**. Let $\{P_\varepsilon : \varepsilon > 0\}$ be a family of probability measures on a Polish space $(E, \rho)$ which is *exponentially tight,* that is, if for every $L > 0$ there is a compact set $K_L$ in $E$ such that

$$\limsup_{\varepsilon \downarrow 0} \varepsilon \log P_\varepsilon(E \setminus K_L) \leq -L. \tag{1}$$

If $\{P_\varepsilon : \varepsilon > 0\}$ satisfies the weak large deviation principle with a rate function $I$, that is,

$$\limsup_{\varepsilon \to 0} \varepsilon \log P_\varepsilon(F) \leq -\inf_F I$$

for every *compact* subset $F \subset E$, and

$$\limsup_{\varepsilon \to 0} \varepsilon \log P_\varepsilon(G) \geq -\inf_G I$$

for every open subset $G \subset E$.

(i) Show that $I$ is a good rate function.

(ii) $I$ governs the large deviations of $\{P_\varepsilon : \varepsilon > 0\}$.