An applicaiton of Laplace's method to the large deviation principles for posterior distributions

長崎大学　式見拓仙

パラメータ ϑ は事前分布 $\pi(\theta),\ \theta \in \Theta$ をもち，$X_1, X_2, \ldots$ は $\vartheta = \theta$ のもとで分布 $f(x \,|\, \theta)$ をもつ i.i.d. 確率変数列とする．さらに，$L_n(\theta)$ を尤度関数とし，$K_n(t)$ を $X_1, \ldots, X_n$ のもとでの事後のキュムラント母関数とすると，

$$K_n(nt) = \log \frac{\displaystyle\int_\Theta e^{nt\theta} L_n(\theta)\pi(\theta)d\theta}{\displaystyle\int_\Theta L_n(\theta)\pi(\theta)d\theta} = \log \frac{\displaystyle\int_\Theta \exp\left[n\left(\frac{1}{n}\log L_n(\theta) + t\theta\right)\right]\pi(\theta)d\theta}{\displaystyle\int_\Theta \exp\left[n \cdot \frac{1}{n}\log L_n(\theta)\right]\pi(\theta)d\theta}. \tag{1}$$

$\log L_n(\theta)/n \to h(\theta) =: E[\log f(X_1 \,|\, \theta) \,|\, \vartheta]$ であるから，(1) の分母分子の積分はそれぞれ $\int_\Theta e^{nh(\theta)}\pi(\theta)d\theta$, $\int_\Theta e^{n[h(\theta)+t\theta]}\pi(\theta)d\theta$ で近似でき，それらはさらに Laplace 近似を適用することにより，

$$\int_\Theta e^{nh(\theta)}\pi(\theta)d\theta \sim \pi(\hat\theta)e^{nh(\hat\theta)}\sqrt{\frac{2\pi}{-nh''(\hat\theta)}}$$
$$\int_\Theta e^{n[h(\theta)+t\theta]}\pi(\theta)d\theta \sim \pi(\hat\theta(t))e^{n[h(\hat\theta(t))+t\hat\theta(t)]}\sqrt{\frac{2\pi}{-nh''(\hat\theta(t))}}$$

のように近似できる．ただし，$\hat\theta, \hat\theta(t)$ はそれぞれ $h(\theta), h(\theta)+t\theta$ の最大値を与える点である．よって，

$$\frac{K_n(nt)}{n} \to K(t) =: h(\hat\theta(t)) + t\hat\theta(t) - h(\hat\theta) \in [-\infty, \infty].$$

$K^*(t^*)$ を $K(t)$ の Legendre 変換を $K^*(t^*)$ とし，また，$\mathcal{D}_K = \{t \in \mathbb{R} : K(t)\}$ とおく．$\mathcal{E}_{K^*}$ を，$K^*(t)$ の exposed point でその exposing hyperplane が $\mathcal{D}_K$ の内点となるようなもの全体とする．もし，0 が $\mathcal{D}_K$ の内点ならば，Gärtner-Ellis の定理より，$K^*(t^*)$ は $\mathbb{R}$ 上のレート関数であり，また

$$\limsup_{n\to\infty} \frac{1}{n}\log P(\vartheta \in F \,|\, X_1, \ldots, X_n) \leqslant -\inf_{t^* \in F} K^*(t^*) \quad (\forall F \subset \mathbb{R} \text{ closed}),$$
$$\liminf_{n\to\infty} \frac{1}{n}\log P(\vartheta \in G \,|\, X_1, \ldots, X_n) \geqslant -\inf_{t^* \in G\cap\mathcal{E}_{K^*}} K^*(t^*) \quad (\forall G \subset \mathbb{R} \text{ open})$$

が成り立つ．さらに，$\mathcal{D}_{K^*} \subset \Theta$ ならば，$K^*(\theta),\ \theta \in \Theta$ は Θ 上のレート関数であり，Θ 上の large deviation upper bound が成り立つ．もし，$\Theta \subset \mathcal{E}_{K^*}$ ならば，Θ 上の large deviation lower bound が成り立つ．上記のアプローチが有効な簡単な具体例を 1 つ挙げておく．

例． $X_1, X_2, \ldots$ は $\vartheta = \theta \in \Theta =: (0,1)$ のもとで i.i.d. で，$X_1 \sim Ber(\theta)$ であるとする．このとき，ϑ の事後分布は Θ 上で大偏差原理を満たし，レート関数は以下のように求められる．

$$K^*(\theta) = \vartheta\log\frac{\vartheta}{\theta} + (1-\vartheta)\log\frac{1-\vartheta}{1-\theta}, \quad \theta \in \Theta.$$

$Ber(p)$ の $Ber(q)$ に関する相対エントロピーを $R(p \,|\, q) = p\log p/q + (1-p)\log((1-p)/(1-q))$ とおくと，$K^*(\theta) = R(\vartheta \,|\, \theta)$ である．他方，Cramér の定理より，標本平均 $\bar X_n$ は大偏差原理を満たす．そのレート関数を $I(x)$ とおくと，$I(x) = R(x \,|\, \vartheta),\ x \in [0,1]$ である．$K^*(\theta)$ と $I(x)$ とでは $R(\cdot \,|\, \cdot)$ の変数の役割が入れ代わっている．