1. Sanov's Theorem

Here we consider a sequence of i.i.d. random variables with values in some complete separable metric space $\mathcal{X}$ with a common distribution α. Then the sample distribution

$$\beta_n = \frac{1}{n}\sum_{j=1}^{n} \delta_{x_j}$$

maps $\mathcal{X}^n \to \mathcal{M}(\mathcal{X})$ and the product measure α^n will generate a measure P_n on the space $\mathcal{M}(\mathcal{X})$ which is the distribution of the empirical distribution. The law of large numbers implies with a little extra work that $P_n \to \delta_\alpha$, i.e. for large n, the empirical distribution β_n is very close to the true distribution α. Close here is in the sense of weak convergence. We want to prove a large deviation result for $\{P_n\}$.

Theorem 1.1. *The sequence $\{P_n\}$ satisfies a large deviation principle on $\mathcal{M}(\mathcal{X})$ with the rate function $I(\beta)$ given by*

$$I(\beta) = h_\alpha(\beta) = h(\alpha;\beta) = +\infty$$

unless $\beta << \alpha$ and $\frac{d\beta}{d\alpha}|\log \frac{d\beta}{d\alpha}|$ is in $L_1(\alpha)$. Then

$$I(\beta) = \int \frac{d\beta}{d\alpha}\log\frac{d\beta}{d\alpha}d\alpha = \int \log\frac{d\beta}{d\alpha}d\beta$$

Before we begin the proof of the theorem we prove a lemma that is useful.

Lemma 1.2. *Let α, β be two probability distributions on a measure space $(\mathcal{X}, \mathcal{B})$. Let $B(\mathcal{X})$ be the space of bounded measurable functions on $(\mathcal{X}, \mathcal{B})$. Then*

$$I(\beta) = \sup_{f \in B(\mathcal{X})} \left[\int f(x)d\beta(x) - \log\int e^{f(x)}d\alpha(x)\right]$$

Proof. The function $x\log x$ and $e^x - 1$ are dual to each other in the sense that

$$\begin{aligned} x\log x - x + 1 &= \sup_y [x\,y - (e^y - 1)] \\ e^y - 1 &= \sup_x [x\,y - (x\log x - x + 1)] \end{aligned}$$

If $b(x) = \frac{d\beta}{d\alpha}$, then

$$\begin{aligned} \int f(x)d\beta &= \int f(x)b(x)d\alpha(x) \\ &\le \int \big[[b(x)\log b(x) - b(x) + 1] + [e^{f(x)} - 1]\big]d\alpha(x) \\ &= I(\beta) + \int e^{f(x)}d\alpha(x) \end{aligned}$$

Writing $f(x) = f(x) - c + c$ and optimizing with respect to c,

$$\int f(x) d\beta \leq \inf_c [I(\beta) + c + \int [e^{f(x)-c} - 1]] \, d\alpha(x)]$$
$$= I(\beta) + \log \int e^{f(x)} d\alpha(x)$$

with the choice of $c = \log \int e^{f(x)} d\alpha(x)$. On the other hand if for every f,

$$\int f(x) d\beta(x) \leq C + \log e^{f(x)} d\alpha(x)$$

taking $f(x) = \lambda \mathbf{1}_A(x)$,

$$\beta(A) \leq \frac{1}{\lambda}[C + \log[e^{\lambda}\alpha(A) + (1 - \alpha(A))]$$

With $\lambda = -\log \alpha(A)$,

$$\beta(A) \leq \frac{C + 2}{\log \frac{1}{\alpha(A)}}$$

proving $\beta << \alpha$. Let $b(x) = \frac{d\beta}{d\alpha}$.

$$\int f(x) b(x) d\alpha(x) \leq C + \log \int e^{f(x)} d\alpha(x)$$

Pick $f(x) = \log b(x)$. Log $b(x)$ may not be a bounded function. But one can truncate and pass to the limit. Note that $(b \log b)^-$ is bounded. The trouble is only from the positive part $(b \log b)^+$ and this is controlled by the upper bound. If X is a nice metric space, we can replace $B(X)$ by $C(X)$ the space of bounded continuous functions. One can use Lusin's theorem to approximate $b(x)$ by bounded continuous functions with respect to both α and β and pass to the limit. We will the have

$$I(\beta) = \sup_{f \in C(\mathcal{X})} \left[\int f(x) d\beta(x) - \log \int e^{f(x)} d\alpha(x) \right]$$

$\square$

We now turn to the proof of the theorem.

Proof. First we note that $I(\beta)$ is convex and lower semi-continuous in the topology of weak convergence of probability distributions.

To prove that $D_\ell = \{\beta : I(\beta) \leq \ell\}$ are compact sets, we need to produce a compact set K_ϵ such that $\beta(K_\epsilon^c) \leq \epsilon$ for all $\beta \in D_\ell$. Let us pick K_ϵ so that $\alpha(K_\epsilon^c) \leq e^{-\frac{\ell+2}{\epsilon}}$. Then for any $\beta \in D_\ell$,

$$\alpha(K_\epsilon^c) \leq \frac{\ell + 2}{\log \frac{1}{\alpha(K_\epsilon)}} \leq \epsilon$$

We now show that given any β and any $\epsilon > 0$, there is an open set U_β, a small neighborhood around β so that

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(U_\beta) \le -I(\beta) + 2\epsilon$$

First we pick f so that $\int f(x)\, d\beta(x) - \log \int e^{f(x)} d\alpha(x) \ge I(\beta) - \epsilon$ Then

$$E[e^{n\int f(x) d\beta_n(x)}] = E[\exp[f(x_1) + f(x_2) + \cdots + f(x_n)]] = \left[\int e^{f(x)} d\alpha(x)\right]^n$$

If $U_\beta = \{\gamma : |\int f(x) d\beta - \int f(x) d\gamma(x)| < \epsilon\}$, then by Tchebechev's inequality

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(U_\beta) \le -\int f(x)\, d\beta + \epsilon + \log \int e^{f(x)} d\alpha(x) \le -I(\beta) + 2\epsilon$$

If D is any compact subset of $\mathcal{M}(\mathcal{X})$, then since D can be covered by a finite number of U_β we can conclude that for any compact $D \subset \mathcal{M}(\mathcal{X})$

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(D) \le -\inf_{\beta\in D} I(\beta) + 2\epsilon$$

Since ϵ is arbitrary we actually have

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(D) \le -\inf_{\beta\in D} I(\beta)$$

If we can show that for any $\ell < \infty$ there is some compact set $D_\ell \subset \mathcal{M}(\mathcal{X})$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(D_\ell^c) \le -\ell$$

then it would follow that for any closed set C

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(C) \le -\inf_{\beta\in C} I(\beta)$$

by writing $P(C) \le P(C \cap D_\ell) + P(D_\ell^c)$. Then

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(C) \le \max\{-\inf_{\beta\in C} I(\beta), -\ell\}$$

and by letting $\ell \to \infty$ we would get the upper bound. Let us pick K_j so that $\alpha(K_j^c) \le \epsilon_j$. Let

$$D = \{\gamma : \gamma(K_j^c) \le \delta_j \text{ forall } j \ge 1\}$$

$$\begin{aligned} P_n(D^c) &\le \sum_{j=1}^{\infty} P[\beta_n(K_j^c) \ge \delta_j] \\ &= \sum_j [B(n, \epsilon_j) \ge n\delta_j] \\ &\le \sum_j [\epsilon_j e^{\theta_j} + (1-\epsilon_j)]^n e^{-n\theta_j\, \delta_j} \end{aligned}$$

Since we can choose $\delta_j \downarrow 0, \epsilon_j \downarrow 0$ and $\theta_j \uparrow \infty$ arbitrarily, we can do it in such a way that

$$\sum_j [\epsilon_j e^{\theta_j} + (1-\epsilon_j)]^n e^{-n\theta_j\,\delta_j} \le e^{-\ell\, n}$$

For instance $\delta_j = \frac{1}{j}$, $\theta_j = j(\ell + \log 2 + j)$, $\epsilon_j = e^{-\theta_j}$ will do it.

To prove the lower bound, we tilt the measure from P_n to Q_n based on i.i.d with β for each component. Let U_β be a neighborhood around β.

$$P_n(U_\beta) = \int_{\beta_n \in U_\beta} [b(x_1)b(x_2)\cdots b(x_n)]^{-1} d\beta(x_1)\cdots d\beta(x_n)$$

If $b(x) = 0$ on a set of positive α measure we would still have a lower bound

$$P_n(U_\beta) \ge \int_{\beta_n \in U_\beta} [b(x_1)b(x_2)\cdots b(x_n)]^{-1} d\beta(x_1)\cdots d\beta(x_n)$$

In any case

$$\begin{aligned} P_n(U_\beta) &\ge \int_{\substack{\beta_n \in U_\beta \\ |\int \log b(x)\,\beta_n(dx) - I(\beta)| \le \epsilon}} [b(x_1)b(x_2)\cdots b(x_n)]^{-1} d\beta(x_1)\cdots d\beta(x_n) \\ &\ge e^{-n[I(\beta)+\epsilon]} \int_{\substack{\beta_n \in U_\beta \\ |\int \log b(x)\,\beta_n(dx) - I(\beta)| \le \epsilon}} d\beta(x_1)\cdots d\beta(x_n) \\ &= e^{-n[I(\beta)+\epsilon]}(1 + o(1)) \end{aligned}$$

by the law of large numbers. This completes the proofs of both upper and lower bounds. $\square$

Sanov's theorem has the following corollary.

Corollary 1.3. *Let $\{X_i\}$ be i.i.d.r.v with values in a separable Banach space $\mathcal{X}$ with a common distribution α. Assume*

$$E[e^{\theta\|X\|}] < \infty$$

for all $\theta > 0$. Then the mean $\frac{1}{n}\sum_{i=1}^n X_i$ satisfies a large deviation principle with rate function

$$H(x) = \sup_{y \in \mathcal{X}^*} [< y, x > - \log \int e^{<y,x>}\, d\alpha(x)]$$

Before starting on the proof let us establish a certain formula.

Lemma 1.4. *Let α be a probability distribution on $\mathcal{X}$ with $\int e^{\theta|z|} d\alpha(z) < \infty$ for all $\theta > 0$. Let*

$$M(y) = \int e^{\langle y,z\rangle} d\alpha(z)$$

and

$$H(x) = \sup_{y \in \mathcal{X}^*} [\langle y, x\rangle - \log M(y)]$$

Then

$$H(x) = \inf_{b:\int zb(z)\,d\alpha(z)=x} \int b(z)\log b(z)\,d\alpha(z)$$

Remark 1.5. The proof will depend on the following minmax theorem from convex analysis. Let $F(x,y)$ be a function on $\mathcal{X}\times\mathcal{Y}$, which is convex and lower semi continuous in y for each x and concave and upper semi continuous in x for each y. Let $C_1\subset\mathcal{X}$ and $C_2\subset\mathcal{Y}$ be closed convex subsets. Let either C_1 or C_2 be compact or let either all the level sets $D_x^\ell=\{y: F(x,y)\le\ell\}$ or all the level sets $D_y^\ell=\{x: F(x,y)\ge\ell\}$ be compact. Then

$$\inf_{y\in C_2}\sup_{x\in C_1} F(x,y) = \sup_{x\in C_1}\inf_{y\in C_2} F(x,y)$$

Proof. We consider the function

$$F(y,b(\cdot)) = \langle y,x\rangle - \int \langle y,z\rangle b(z)d\alpha(z) + \int b(z)\log b(z)d\alpha(z)$$

on $\mathcal{X}^*\times\mathcal{N}$, where $\mathcal{N}$ is the set of $b(z)$ that are nonnegative, $\int b(z)\,d\alpha(z)=1$ and $\int b(z)\log b(z)d\alpha(z)<\infty$.

$$\sup_y\inf_b F(y,b(\cdot)) = \sup_y[\langle y,x\rangle - \log M(y)] = H(x)$$

while

$$\sup_y\big[\langle y,x\rangle - \int\langle y,z\rangle d\alpha(z)\big] = \infty$$

unless $x=\int z\,d\alpha(z)$, in which case it is 0. Therefore

$$\inf_b\sup_y F(y,b(\cdot)) = \inf_{b:\int zb(z)\,d\alpha(z)=x}\int b(z)\log b(z)\,d\alpha(z)$$

It is not hard to verify the conditions for the minmax theorem to be applicable. ☐

Proof. (of corollary) **Step 1.** First we assume that $\|X_i\|$ is bounded by ℓ. Suppose β_n is a sequence of probability measures supported in the ball of radius ℓ and β_n converges weakly to β. If $x_n=\int zd\beta_n(z)$ then $< y, x_n > \to < y, x >= \int zd\beta(z)$ for each $y\in\mathcal{X}^*$. By Prohorov's condition there is a compact set K_ϵ that has probability at least $(1-\epsilon)$ under all the β_n as well as β. If $x_n^\epsilon=\int_{K^\epsilon} z\,d\beta_n(z)$, x_n is in the convex hull of K^ϵ and 0, which is compact. More over $\|x_n^\epsilon - x_n\|\le\epsilon\ell$. This is enough to conclude that $\|x_n-x\|\to 0$.

Step 2. We truncate and write $X=Y^\ell+Z^\ell$ where X^ℓ is supported in the ball of radius ℓ. Since $E[e^{\theta\|Z^\ell\|}]\to 1$ as $\ell\to\infty$ for all $\theta>0$, we see that

$$\limsup_{\ell\to\infty}\limsup_{n\to\infty}\frac{1}{n}\log P[\frac{1}{n}\sum_{j=1}^n\|Z_j^\ell\|\ge\epsilon] \le \limsup_{\ell\to\infty}\inf_{\theta>0}[-\theta\epsilon+\log E[e^{\theta\|Z^\ell\|}] = -\infty$$

Step 3. Finally

$$P[\frac{1}{n}\sum_{j=1}^n X_j\in C] \le P[\frac{1}{n}\sum_{j=1}^n Y_j^\ell\in C^\epsilon] + P[\frac{1}{n}\sum_{j=1}^n\|Z_j^\ell\|\ge\epsilon]$$

If we denote by

$$H^\ell(x) = \sup_{y\in\mathcal{X}^*}[< y, x > -\log E[e^{<y,Y^\ell>}]]$$

then

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log P[\frac{1}{n}\sum_{j=1}^n X_j \in C] &\le \max\{-\inf_{x\in\overline{C^\epsilon}} H^\ell(x), \limsup_{n\to\infty}\log P[\frac{1}{n}\sum_{j=1}^n \|Z_j^\ell\| \ge \epsilon]\}\\ &\le -\liminf_{\epsilon\to 0}\liminf_{\ell\to\infty}\inf_{x\in\overline{C^\epsilon}} H^\ell(x)\\ &= -\inf_{x\in C} H(x)\end{aligned}$$

The last step needs the fact that if for some $x_\ell \in \overline{C^\epsilon}$, $H^\ell(x_\ell)$ is bounded, then x_ℓ^ϵ is a compact sequence and any limit point x^ϵ will be in $\overline{C^\epsilon}$ and as $\epsilon \to 0$ this will produce a limit point in $x \in C$ with $H(x) \le \liminf_{\epsilon\to 0}\liminf_{\ell\to\infty} H^\ell(x_\ell^\epsilon)$. The following lemma justifies the step.

Lemma 1.6. *Let $\{\mu_\sigma\}$, $\sigma \in \mathcal{S}$, be a collection of probability measures on $\mathcal{X}$ satisfying, the tightness condition that for any given $\epsilon > 0$, there exists a compact set $K^\epsilon \subset \mathcal{X}$, such that $\mu_\sigma(K^\epsilon) \ge 1 - \epsilon$ for all $\sigma \in \mathcal{S}$,and*

$$\sup_{\sigma\in\mathcal{S}}\int e^{\theta\|z\|} d\mu_\sigma(z) = m(\theta) < \infty$$

for every $\theta > 0$. Let

$$H_\sigma(x) = \sup_{y\in\mathcal{X}^*}[< y, x > -\log\int e^{<y,z>}\, d\mu_\sigma(z)]$$

Then the set $D_\ell = \cup_{\sigma\in\mathcal{S}}\{x : H_\sigma(x) \le \ell\}$ has a compact closure in $\mathcal{X}$.

Proof. If $H_\sigma(x_\sigma) \le \ell$ then there exists b_σ such that $\int b_\sigma(z)\log b_\sigma(z)\, d\mu_\sigma(z) \le \ell$ and $x_\sigma = \int z d\beta_\sigma(z) = \int z\, b_\sigma(z) d\mu_\sigma(z)$. From the entropy bound $I(\beta_\sigma) \le \ell$ and the tightness of $\{\mu_\sigma\}$ it follows that $\{\beta_\sigma\}$ is tight and $\int \|z\| d\beta_\sigma(z)$ uniformly integrable. Therefore $\{x_\sigma\}$ is compact. □

This concludes the proof of the corollary □

Gaussian distributions on a Banach space are defined by a mean and covariance. Suppose X is a Gaussian random variable with mean 0 and some covariance $B(y_1, y_2) = E[\langle y_1, X\rangle\langle y_2, X\rangle]$. Then the distribution of $X_1 + X_2 + \cdots + X_n$ is the same as that of $\sqrt{n}X$. If there is a large deviation principle for $\frac{1}{n}\sum_{j=1}^n X_j$, this gives a Gaussian tail behavior for X.

$$\limsup_{\lambda\to\infty}\frac{1}{\lambda^2}\log P[X \in \lambda C] \le -\inf_{x\in C} H(x)$$

where

$$H(x) = \sup_{y\in\mathcal{X}^*}[\langle y, x\rangle - \frac{1}{2}B(y, y)]$$

According to our result we only need to check that $E[e^{\theta\|X\|}] < \infty$ for all $\theta > 0$. There is a simple argument due to Fernique which shows that any Gaussian distribution on a Banach space must necessarily be such that $E[e^{\theta\|X\|^2}] < \infty$ for some $\theta > 0$. The proof uses the independence of $X + Y$ and $X - Y$. If $\|X\| \geq \ell$ and $\|Y\| \leq a$ we must have $\|X + Y\|$ and $\|X - Y\| \geq \ell - a$. Therefore

$$P[\|X + Y\| \geq \ell - a \cap \|X - Y\| \geq \ell - a] = P[\|X\| \geq \frac{\ell - a}{\sqrt{2}}]^2 \geq P[\|X\| \geq \ell]P[\|Y\| \leq a]$$

We can pick a so that $P[\|Y\| \leq a] \geq \frac{1}{2}$. We then have

$$P[\|X\| \geq \ell] \leq 2P[\|X\| \geq \frac{\ell - a}{\sqrt{2}}]^2$$

or

$$P[\|X\| \geq \sqrt{2}\ell + a] \leq 2P[\|X\| \geq \ell]^2.$$

Iterating this inequality will give the necessary bound. Define $\ell_{n+1} = \sqrt{2}\ell_n + a$. Then for $T(\ell) = P[\|X\| \geq \ell]$ we have

$$T(\ell_{n+1}) \leq 2[T(\ell_n)]^2 \leq [2T(\ell_1)]^{2^n}$$

With proper choice of a and ℓ_1, $\ell_n \simeq C\,2^{\frac{n}{2}}$ and $T(\ell_n) \leq \delta^{2^n}$ for some $\delta < 1$. That does it.

2. Schilder's Theorem.

One of the advantages in formulating the large deviation principle in fairly general terms is the ability to apply it in several infinite dimensional contexts. The first such example is result concerning the behavior of Brownian motion with a small parameter. Let us consider the family of stochastic process $\{x_\epsilon(t)\}$ defined by

$$x_\epsilon(t) = \sqrt{\epsilon}\,\beta(t)$$

or equivalently

$$x_\epsilon(t) = \beta(\epsilon t)$$

for t in some fixed time interval, say $[0, 1]$ where $\beta(\cdot)$ is the standard Brownian motion. The distributions of $x_\epsilon(\cdot)$ induce a family of scaled Wiener processes on $C[0, 1]$ that we denote by Q_ϵ. We are interested in establishing an LDP for Q_ϵ as $\epsilon \to 0$. The rate function will turn out to be

$$I(f) = \frac{1}{2}\int_0^1 [f'(t)]^2 dt$$

if $f(0) = 0$ and $f(\cdot)$ is absolutely continuous in t with a square integrable derivative $f'(\cdot)$. Otherwise

$$I(f) = +\infty$$

The main theorem of this section due to M.Schilder is the following

Theorem 2.1. *The family Q_ϵ satisfies an LDP with rate function $I(\cdot)$.*

Proof. First let us note that as soon as we have a bound on the rate function say $I(f) \leq \ell$, then f satisfies a Holder inequality

$$|f(t) - f(s)| = |\int_s^t f'(\sigma)\, d\sigma| \leq |t-s|^{\frac{1}{2}} \left(\int_s^t [f'(\sigma)]^2 d\sigma\right)^{\frac{1}{2}} \leq \sqrt{2\ell}\, |t-s|^{\frac{1}{2}}.$$

The lower semi-continuity of $I(\cdot)$ is obvious and since $f(0) = 0$, by the Ascoli-Arzela theorem, the level sets are totally bounded and hence compact.

Now let us turn to the proof of the upper bound. Suppose $C \subset \Omega$ is a closed subset of the space $\Omega = C[0,1]$. For any f in Ω and for any positive integer N, let us divide the interval $[0,1]$ in to N equal subintervals and construct the piecewise linear approximation $f_N = \pi_N f$ by matching $f_N = f$ at the points $\{\frac{j}{N} : 0 \leq j \leq N\}$ and interpolating linearly in between. An elementary calculation yields $I(\pi_N f) = \frac{1}{2}\sum_{j=0}^{N-1}[f(\frac{j+1}{N}) - f(\frac{j}{N})]^2$. We can estimate $Q_\epsilon(C)$ by

$$Q_\epsilon(C) = Q_\epsilon[f \in C] \leq Q_\epsilon[f_N \in C^\delta] + Q_\epsilon[\, \|f_N - f\| \geq \delta\,]$$

where $\|\ \|$ is the supremum norm on Ω, and C^δ is the δ neighborhood of the closed set C in the uniform metric. If we denote by

$$\ell_\delta = \inf_{f \in C^\delta} I(f) \quad \text{and} \quad \ell = \inf_{f \in C} I(f)$$

by the compactness of the level sets and the lower semicontinuity of $I(\cdot)$ it follows that

$$\ell_\delta \uparrow \ell \quad \text{as} \quad \delta \downarrow 0$$

Clearly

$$Q_\epsilon[f_N \in C^\delta] \leq Q_\epsilon[I_N(f) \geq \ell_\delta].$$

We know that, under Q_ϵ, $2I_N$ is essentially a sum of squares of N independent identically distributed Gaussians with mean 0, and has a scaled χ^2 distribution with N degrees of freedom.

$$Q_\epsilon[I_N(f) \geq \ell_\delta] = (\frac{1}{\epsilon})^N \frac{1}{\Gamma(\frac{N}{2})} \int_{\ell_\delta}^\infty u^{\frac{N}{2}-1} \exp[-\frac{u}{\epsilon}\ ]\, du$$

It is elementary to conclude that for fixed N and δ

$$\limsup_{\epsilon \to 0} \epsilon \log Q_\epsilon[I_N(f) \geq \ell_\delta] \leq -\ell_\delta.$$

As for the second term $Q_\epsilon[\ \|f_N - f\| \geq \delta\ ]$,

$$\|f_N - f\| \leq 2 \sup_{0 \leq j \leq N} \sup_{\frac{j}{N} \leq t \leq \frac{j+1}{N}} |f(t) - f(\frac{j}{N})|$$

and the events $\{ \sup_{\frac{j}{N}\le t\le \frac{j+1}{N}} |f(t)-f(\frac{j}{N})| \ge \frac{\delta}{2} \}$ all have the same probability;

$$\begin{aligned} Q_\epsilon\big[\sup_{\frac{j}{N}\le t\le \frac{j+1}{N}} |f(t)-f(\frac{j}{N})| \ge \frac{\delta}{2}\big] &= Q_\epsilon\big[\sup_{0\le t\le \frac{1}{N}} |f(t)| \ge \frac{\delta}{2}\big] \\ &\le 2Q_\epsilon\big[\sup_{0\le t\le \frac{1}{N}} f(t) \ge \frac{\delta}{2}\big] \quad \text{(by symmetry)} \\ &= 4Q_\epsilon\big[f(\frac{1}{N}) \ge \frac{\delta}{2}\big] \quad \text{(by reflection principle)} \end{aligned}$$

We can now easily estimate $Q_\epsilon[\ \|f_N - f\| \ge \delta\ ]$ by

$$Q_\epsilon[\ \|f_N - f\| \ge \delta\ ] \le 4NQ_\epsilon\big[f(\frac{1}{N}) \ge \frac{\delta}{2}\big]$$

and for fixed N and δ,

$$\limsup_{\epsilon\to 0} \epsilon \log Q_\epsilon\big[\|f_N - f\| \ge \delta\big] \le -\frac{N\delta^2}{8}$$

Now combining both terms

$$\limsup_{\epsilon\to 0} \epsilon \log Q_\epsilon[C] \le -\inf\{\ell_\delta, \frac{N\delta^2}{8}\}$$

Since N and δ are arbitrary we can let $N\to\infty$ first and then let $\delta\to 0$, to get the upper bound

$$\limsup_{\epsilon\to 0} \epsilon \log Q_\epsilon[C] \le -\ell$$

and we are done.

For the lower bound, in order to show that for open sets $G\subset\Omega$

$$\liminf_{\epsilon\to 0} \epsilon \log Q_\epsilon(G) \ge -\inf_{g\in G} I(g),$$

because there is always a neighborhood N_g of g with $g\in N_g\subset G$, it is sufficient to show that, for any $g\in\Omega$ and $N\ni g$,

$$\liminf_{\epsilon\to 0} \epsilon \log Q_\epsilon(N) \ge -I(g).$$

Actually for any g with $I(g)<\infty$, and for any neighborhood $N\ni g$ and any $\delta>0$ there exist a smooth $h\in N$ with a neighborhood $\tilde{N}$ satisfying $h\in\tilde{N}\subset N$ and $I(h)\le I(g)+\delta$. So for the purpose of establishing the lowere bound there is no loss of generality in assuming that g is smooth. If we denote by $Q_{\epsilon,g}$ the measure that is obtained from Q_ϵ by the translation $f\to f-g$ mapping $\Omega\to\Omega$, then

$$Q_\epsilon[f:\|f-g\|<\delta] = Q_{\epsilon,g}[f:\|f\|<\delta] = \int_{\|f\|<\delta} \frac{dQ_{\epsilon,g}}{dQ_\epsilon}(f)\, dQ_\epsilon.$$

Cameron-Martin formula evaluates the Radon-Nikodym derivative

$$\frac{dQ_{\epsilon,g}}{dQ_\epsilon}(f) = \exp\big[-\frac{1}{\epsilon}\int_0^1 g'(t)df(t) - \frac{1}{2\epsilon}\int_0^1 [g'(t)]^2 dt\big]$$

The second term in the exponentiel is a constant and comes out of the integral as $\exp\left[-\frac{I(g)}{\epsilon}\right]$. As for the integral

$$\int_{\|f\|<\delta} \exp\big[-\frac{1}{\epsilon}\int_0^1 g'(t)df(t)\big]dQ_\epsilon$$

we first note that, because g is smooth we can integrate by parts to rewrite the "df" integral as

$$\int_0^1 g'(t)df(t) = g'(1)f(1) - \int_0^1 f(t)g''(t)dt.$$

We restrict the integration to the set $\{\|f\| < \delta\} \cap \{f : \int_0^1 g'(t)df(t) \le 0\}$ and use the symmetry of Q_ϵ under $f \to -f$ to conclude that

$$\int_{\|f\|<\delta} \exp\big[-\frac{1}{\epsilon}\int_0^1 g'(t)df(t)\big]dQ_\epsilon \ge \frac{1}{2}Q_\epsilon[\ \|f\| < \delta\ ].$$

Clearly, as $\epsilon \to 0$, $Q_\epsilon[\ \|f\| < \delta\ ] \to 1$ and it follows that

$$\liminf_{\epsilon\to 0} \epsilon \log Q_\epsilon[f : \|f - g\| < \delta] \ge -I(g)$$

and the proof of the lower bound is complete. □

Remark 2.2. A calculation involving the covariance $\min(s,t)$ of the Brownian Motion and its formal inverse quadratic form $\int_0^1 [f'(t)]^2 dt$ will allow us to write a formal expression

$$dQ_\epsilon = \exp\big[-\frac{1}{2\epsilon}\int_0^1 [f'(t)]^2 dt\big]\,\Pi_t\, df(t)$$

from which our rate function can be guessed. But the above density is with respect to an infinite dimensional Lebesgue measure that does not exist, and the density itself is an expression of doubtful meaning, since almost surely the Brownian paths are nowhere differentiable. Nevertheless the LDP is valid as we have shown. In fact the LDP we established for Gaussian distributions on a Banach space is applicable here and is not difficult to calculate for the covariance $\min\{s,t\}$ of Brownian motion

$$\sup\Big[\int_0^T f(t)g(t)dt - \frac{1}{2}\int_0^T\int_0^T \min(s,t)g(s)g(t)dsdt\Big] = \frac{1}{2}\int_0^T [f'(t)]^2 dt$$

provided $f(0) = 0$ and f' exists in $L_2[0,T]$.

Remark 2.3. This method of establishing the upper bound for some approximations that are directly carried out and controlling the error by super-exponential estimates will be a recurring theme. Similarly the lower bound is often established for a suitably dense set of smooth points. The trick of Cramer involving change of measure so that the large deviation becomes normal and getting a lower bound in terms of the Radon-Nikodym derivative will also be a recurring theme.