Some aspects of Large Deviations

S. Ramasubramanian
Stat.-Math. Unit
Indian Statistical Institute
8th Mile, Mysore Road
Bangalore - 560 059

1 Introduction

Large Deviations is a part of probability theory; it deals with asymptotic computations of small probabilities (that is, probabilities of rare events) on an exponential scale. Large deviations estimates have proved to be very useful in diverse contexts. The first basic result was proved by Cramer in 1938; it was generalized later by Sanov in 1961; these were in the context of sums of independent identically distributed random variables (i.i.d.'s).

However, rapid development of the subject became possible only after a general theory was formulated in the right degree of abstraction by Varadhan since the late 60's / early 70's.

Cramer's theorem is proved in Section 2 to illustrate the statistical origins of the subject. In Section 3 basic definitions are given, followed by Varadhan's lemma; a 'toy example' is presented to indicate that probabilities of rare events can decisively influence asymptotic expectations. Large deviation principle for Brownian motion process is considered in Section 4. In Section 5, a brief introduction is given to the Freidlin-Wentzell theory; proof is given only in a special case.

For detailed information/exposition an interested reader may consult the references given at the end.

2 Cramer's theorem

We begin with an example

Example 2.1 *:* Let $X_1, X_2, \ldots$ be i.i.d.'s each having a standard normal distribution:

$$P(X_i \in A) \equiv \text{Prob. } (X_i \in A) = \int_A \psi(x)dx, A \in \mathcal{B}(I\!R)$$

where

$$\psi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2), x \in I\!R. \qquad (2.1)$$

Let $S_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ be the empirical mean. Note that S_n has the $N(0, \frac{1}{n})$ distribution (= normal distribution with mean 0 and variance $\frac{1}{n}$) and hence by the Chebyshev inequality, for any $\delta > 0$

$$P(|S_n| \geq \delta) \to 0, \text{as } n \to \infty. \qquad (2.2)$$

Since $\sqrt{n}S_n \sim N(0,1)$ for any n note that

$$\lim_{n\to\infty} P(\sqrt{n}S_n \in A) = \int_A \psi(x)dx. \qquad (2.3)$$

It is also clear that for any $\delta > 0$

$$P(|S_n| \geq \delta) = 2[1 - \Phi(\delta\sqrt{n})] \qquad (2.4)$$

where Φ is the distribution function of the $N(0,1)$ distribution, viz.

$$\Phi(x) = \int_{-\infty}^{x} \psi(t)dt. \qquad (2.5)$$

Fact: For any $z > 0$

$$\left(\frac{1}{z} - \frac{1}{z^3}\right)\psi(z) < [1 - \Phi(z)] < \frac{1}{2}\psi(z). \qquad (2.6)$$

Proof of fact: For any $y > 0$ clearly

$$\left(1 - \frac{3}{y^4}\right) \psi(y) < \psi(y) < \left(1 + \frac{1}{y^2}\right) \psi(y).$$

Integrating the above expression over $[z, \infty]$ we get (2.6) $\square$

From (2.4), (2.6) it is easily seen that for any $\delta > 0$

$$\lim_{n\to\infty} \frac{1}{n} \log P(|S_n| \geq \delta) = -\frac{1}{2}\delta^2. \qquad (2.7)$$

Now (2.2) says that for any $\delta > 0, |S_n| \geq \delta$ is a rare event (for large n); (2.3) implis that the "typical" value of S_n is of the order $\frac{1}{\sqrt{n}}$. But (2.7) means that with small probability (of the order $\exp(-n\delta^2/2)), S_n$ takes relatively large values; thus (2.7) is an example of a large deviations statement, and $\frac{1}{2}\delta^2$ is an exmaple of "rate function". $\square$

By the strong law of large numbers and the central limit theorem respectively both (2.2) and (2.3) remain valid for any i.i.d. sequence $\{X_i\}$ with second moments (after suitable centering and scaling). One may wonder if an analogue of (2.7) holds for non-normal i.i.d.'s $X_1, X_2, \ldots$

Theorem 2.2 *(Cramér): Let $X_1, X_2, \ldots$ be real valued i.i.d.'s. Assume that*

$$\varphi(t) := E(e^{tX_1}) < \infty, \ \forall\ t \in I\!R. \qquad (2.8)$$

Let $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then for any $a > E(X_1)$,

$$\lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq a) = -I(a), \qquad (2.9)$$

where

$$I(z) := \sup\{zt - \log\varphi(t) : t \in I\!R\}. \qquad (2.10)$$

Similarly for $a < E(X_1)$

$$\lim_{n\to\infty} \frac{1}{n} \log P(S_n \leq a) = -I(a). \qquad (2.11)$$

Proof: We may assume that X_i's are nondegenerate; (otherwise the theorem is trivial). Also note that it is enough to consider $a = 0, E(X_i) < 0$. Put $\rho = \inf\{\varphi(t) : t \in I\!R\}$. Observe that $I(0) = -\log \rho$ (with $I(0) = \infty$ if $\rho = 0$). So we need to prove that

$$\lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq 0) = \log \rho. \qquad (\ 2.12\ )$$

Let $F(x) = P(X_i \leq x), x \in I\!R$ be the distribution function of X_i. Then $\varphi(t) = E(e^{tX_i}) = \int\limits_{I\!R} e^{tx} dF(x)$. It is easily seen that $\varphi \in C^\infty(I\!R), \varphi'(t) = \int\limits_{I\!R} x e^{tx} dF(x), \varphi''(t) = \int\limits_{I\!R} x^2 e^{tx} dF(x)$. Hence φ is strictly convex and $\varphi'(0) = E(X_i) < 0$.

We now distinguish three cases.

Case (i): $P(X_i < 0) = 1$. In this case $\varphi(\cdot)$ is strictly decreasing, $\lim\limits_{t\to\infty} \varphi(t) = 0$ and hence $\log \rho = -\infty$; as $P(S_n \geq 0) = 0$, clearly (2.12) holds proving the theorem.

Case (ii): $P(X_1 \leq 0) = 1, P(X_1 = 0) > 0$. Since X_i is nondegenerate, $\varphi(\cdot)$ is again strictly decreasing. Observe that $\lim\limits_{t\to\infty} \varphi(t) = \inf \varphi(t) = \rho = P(X_1 = 0) > 0$, and that $P(S_n \geq 0) = P(X_1 = \ldots = X_n = 0) = \rho^n$. Hence (2.12) follows.

Case (iii): $P(X_1 < 0) > 0, P(X_1 > 0) > 0$. In this case note that $\lim\limits_{t\to\pm\infty} \varphi(t) = \infty$. Since φ is strictly convex there exists a unique minimum $\tau \in I\!R$ so that $\varphi(\tau) = \rho, \varphi'(\tau) = 0$.

Now by Chebyshev inequality, independence of X_i's it follows

$$\begin{aligned} P(S_n \geq 0) &= P(\tau n S_n \geq 0) = P(e^{\tau n S_n} \geq 1) \\ &\leq E(e^{\tau n S_n}) = [E(e^{\tau X_1})]^n = [\varphi\tau)]^n = \rho^n. \end{aligned}$$

Therefore

$$\limsup_{n\to\infty} \frac{1}{n} \log P(S_n \geq 0) \leq \log \rho. \qquad (\ 2.13\ )$$

This proves the upper bound.

It remains to prove the lower bound in case (iii). Put

$$\hat{F}(x) = \frac{1}{\rho} \int\limits_{(-\infty,x]} e^{\tau y} dF(y). \tag{2.14}$$

Verify that $\hat{F}$ is a distribution function; the function $\hat{F}$ is called the *Cramér transform* of F. Let $\hat{X}_1, \hat{X}_2, \ldots$ be an i.i.d. sequence with distribution $\hat{F}$; (note that one can find such a sequence by Kolmogorov consistency theorem). We need a few lemmas.

Lemma 2.3 *:* $E(\hat{X}_i) = \hat{\mu} = 0, \mathrm{Var}(\hat{X}_i) = \hat{\sigma}^2 \in (0, \infty)$.

Proof: Put $\hat{\varphi}(t) = E(e^{t\hat{X}_1})$. Then $\hat{\varphi}(t) = \int\limits_{I\!R} e^{tx} d\hat{F}(x) = \frac{1}{\rho} \int\limits_{I\!R} e^{tx} e^{\tau x} dF(x) = \frac{1}{\rho}\varphi(t+\tau) < \infty$ for all $t \in I\!R$.

So $\hat{\varphi} \in C^{\infty}(I\!R); E(\hat{X}_i) = \hat{\varphi}'(0) = \frac{1}{\rho}\varphi'(\tau) = 0$, and $\mathrm{Var}(\hat{X}_i) = E(\hat{X}_i^2) = \hat{\varphi}''(0) = \frac{1}{\rho}\varphi''(\tau) \in (0, \infty)$.

Lemma 2.4 *Put* $\hat{S}_n = \frac{1}{n} \sum\limits_{i=1}^{n} \hat{X}_i$*. Then*

$$\begin{aligned} P(S_n \geq 0) &= P\left(\sum_{i=1}^{n} X_i \geq 0\right) \\ &= \rho^n E[1_{[0,\infty)}(n\hat{S}_n) \cdot \exp(-\tau n \hat{S}_n)]. \end{aligned}$$

Proof:

$$\begin{aligned} P(S_n \geq 0) &= P(\sum_{i=1}^{n} X_i \geq 0) \\ &= \int\limits_{H^n} dF(x_1) dF(x_2) \ldots dF(x_n) \\ &= \int\limits_{H^n} (\rho e^{-\tau x_1} d\hat{F}(x_1)) \ldots (\rho e^{-\tau x_n} d\hat{F}(x_n)) \\ &= \rho^n \int\limits_{H^n} \exp(-\tau(x_1 + \ldots + x_n)) d\hat{F}(x_1) \ldots d\hat{F}(x_n) \\ &= \rho^n E[1_{[0,\infty)}(n\hat{S}_n) \cdot \exp(-\tau n \hat{S}_n)] \end{aligned}$$

where $H_n = \{(x_1, \ldots, x_n) : x_1 + \ldots + x_n \geq 0\}$. $\square$

Lemma 2.5 *:*

$$\liminf_{n\to\infty} \frac{1}{n} \log E[1_{[0,\infty)}(n\hat{S}_n) \cdot \exp(-\tau n \hat{S}_n)] \geq 0.$$

Proof: In view of Lemma 2.3, note that the central limit theorem is applicable to $\hat{X}_1, \hat{X}_2, \ldots$ Now choose $C > 0$ such that $\Phi(C) - \Phi(0) > \frac{1}{4}$ where Φ is given by (2.5). Therefore

$$\begin{aligned}
& E[1_{[0,\infty)}(n\hat{S}_n) \cdot \exp(-\tau n\hat{S}_n)] \\
= \; & E\left[1_{[0,\infty)}\left(\frac{1}{\sqrt{n}\hat{\sigma}}\sum_{i=1}^{n}\hat{X}_i\right) \cdot \exp\left(-\tau\sqrt{n}\hat{\sigma}\frac{1}{\sqrt{n}\hat{\sigma}}\sum_{i=1}^{n}\hat{X}_i\right)\right] \\
\geq \; & E\left[1_{[0,C]}\left(\frac{1}{\sqrt{n}\hat{\sigma}}\sum_{i=1}^{n}\hat{X}_i\right) \cdot \exp\left(-\tau\sqrt{n}\hat{\sigma}\frac{1}{\sqrt{n}\hat{\sigma}}\sum_{i=1}^{n}\hat{X}_i\right)\right] \\
\geq \; & \exp(-\tau\sqrt{n}\hat{\sigma}C) \cdot P\left(0 \leq \frac{1}{\sqrt{n}\hat{\sigma}}\sum_{i=1}^{n}\hat{X}_i \leq C\right) \\
> \; & [\Phi(C) - \Phi(0)]\exp(-\tau\sqrt{n}\hat{\sigma}C) \\
> \; & \frac{1}{4}\exp(-\tau\sqrt{n}\hat{\sigma}C)
\end{aligned}$$

for all large n. The lemma now follows. $\square$

Proof of theorem (contd): By the two preceding lemmas we now get

$$\liminf_{n\to\infty} \frac{1}{n} \log P(S_n \geq 0) \geq \log \rho. \qquad (2.15)$$

Clearly (2.13) and (2.15) imply (2.12) in case (iii). This completes the proof of the theorem. $\square$

If $a > E(X_1)$ from (2.10) it is clear that $I(z) \geq I(a)$ for all $z \geq a$; therefore the statement (2.9) of Cramer's theorem can be written as

$$\lim_{n\to\infty} \frac{1}{n} \log P(S_n \in A) = -\inf\{I(z) : z \in A\} \qquad (2.16)$$

with $A = [a, \infty)$.

It may be two optimistic to expect (2.9) or even (2.16) to carry over to more general situations. The appropriate approach, essentially due to Varadhan, is outlined in the next section.

3 The Large Deviation Principle

The large deviation principle (LDP) characterizes the limiting behaviour of a family $\{P_\epsilon\}$ of probability measures as $\epsilon \downarrow 0$, in terms of a rate function. As $\epsilon \to 0$, typically P_ϵ will converge weakly to a degenerate probability measure; all measures will be on some complete separable metric space.

Let (S, d) be a complete separable metric space; let $\mathcal{F}$ denote the Borel σ-algebra of S. Recall that a function $f : S \to [-\infty, \infty]$ is *lower semicontinuous* if f has closed level sets; that is, $\{x \in S : f(x) \leq c\}$ is closed for all $c \in \mathbb{R}$; (or equivalently $\liminf\limits_{n\to\infty} f(x_n) \geq f(x)$ whenever $x_n \to x$ in S). Note that a lower semicontinuous function attains a minimum on every nonempty compact set.

The required abstraction is contained in the following two key definitions.

Definition 3.1 *:* A function $I : S \to [0, \infty]$ is called a *rate function* if $I \not\equiv \infty$ and if for each $c \in \mathbb{R}$ the level set $\{x \in S : I(x) \leq c\}$ is compact in S. (In particular a rate function is lower semicontinuous.)

Definition 3.2 *:* Let $\{P_\epsilon : \epsilon > 0\}$ be a family of probability measures on $(S, \mathcal{F})$. The family $\{P_\epsilon\}$ is said to *satisfy the large deviation principle (LDP) with a rate function* I if
(a) I is a rate function;
(b) for every closed set $C \subseteq S$

$$\limsup_{\epsilon\to 0} \epsilon \log P_\epsilon(C) \leq -I(C), \tag{3.1}$$

(c) for every open set $G \subseteq S$

$$\liminf_{\epsilon\to 0} \epsilon \log P_\epsilon(G) \geq -I(G), \tag{3.2}$$

where

$$I(H) = \inf\{I(y) : y \in H\}, H \subseteq S. \tag{3.3}$$

The following comments are fairly easy to verify / explain certain aspects of the definition.

1) The rate function is unique for any family $\{P_\epsilon\}$ satifying the LDP.
2) Let $X_1, X_2, \ldots$ be i.i.d.'s as in Cramer's theorem; let μ be their common distribution; (that is, $\mu = PX_i^{-1}$ for all i). Let $S_n = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$, and P_n be the distribution of S_n; that is, $P_n(A) = PS_n^{-1}(A) = \mu^{*(n)}(nA)$ for any Borel set $A \subseteq I\!R$, where $\mu^{*(n)} = \mu * \ldots * \mu$ is the n-fold convolution of μ. Then Cramer's theorem basically says that $\{P_n\}$ satisfies LDP with rate function given by (2.10).
3) In the place of (3.1), (3.2), the more intuitive/naive stipulation that

$$\lim_{\epsilon\to 0} \epsilon \log P_\epsilon(H) = -I(H), H \in \mathcal{F} \tag{3.4}$$

turns out too strong to be satisfactory. For example, if P_ϵ is nonatomic for all $\epsilon > 0$ then taking $H = \{x\}$ for any $x \in S$, (3.4) can hold only if $I(\cdot) \equiv \infty$. This would leave out many interesting cases. It turns out that (3.1), (3.2) are enough to yield a rich theory. Of course, for $H \in \mathcal{F}$, if $I(H^\circ) = I(\bar{H})$ then (3.4) holds for H; such a set is said to be I-*continuous*. (For instance, if I is continuous, than all H satisfying $H \subseteq \bar{H}^\circ$ are I-continuous; this includes all open sets.)
4) Though it is possible to make sense of Definition 3.2 on an arbitrary topological space, it does not lead to a rich theory. Without the structure of a complete separable metric space many results do not carry over. Conversely as more structure is added stronger results are possible, as indicated by Cramer's theorem.

Recall that $P_\epsilon \Rightarrow P$, that is, $\{P_\epsilon\}$ converges weakly to P if

$$\lim_{\epsilon\to 0} \int\limits_S f(x)\, dP_\epsilon(x) = \int\limits_S f(x)\, dP(x)$$

for any bounded continuous function f. Also $P_\epsilon \Rightarrow P$ is equivalent to

$$\begin{aligned}\limsup_{\epsilon\to 0} P_\epsilon(C) &\leq P(C), C \subseteq S, C \text{ closed},\\ \liminf_{\epsilon\to 0} P_\epsilon(G) &\geq P(G), G \subseteq S, G \text{ open}.\end{aligned}$$

The formal similarity between (3.1), (3.2) and the above suggests that LDP may be suitable for handling convergence of integrals of exponential functionals. This is indeed so as given below.

Theorem 3.3 *(Varadhan's lemma): Let $\{P_\epsilon\}$ satisfy LDP with a rate function $I(\cdot)$. Then for any bounded continuous function g on S*

$$\begin{aligned}
&\lim_{\epsilon\to 0} \epsilon \log \left[\int_S \exp(\frac{1}{\epsilon} g(x)) dP_\epsilon(x) \right] \\
&\quad = \sup\{[g(x) - I(x)] : x \in S\}. \qquad (3.5)
\end{aligned}$$

Proof: Upper bound: Let $\delta > 0$. There exist an integer $N = N(\delta)$, sets $C_1, \ldots, C_N$ so that C_i are closed, $S = \bigcup_{i=1}^{N} C_i, g(x) \le g_j + \delta$ for all $x \in C_j$, where $g_j = \inf\{g(y) : y \in C_j\}, 1 \le j \le N$. Then

$$\begin{aligned}
\int_S \exp(\frac{1}{\epsilon} g(x)) dP_\epsilon(x) &\le \sum_{j=1}^{N} \int_{C_j} \exp(\frac{1}{\epsilon} g(x)) dP_\epsilon(x) \\
&\le \sum_{j=1}^{N} P_\epsilon(C_j) \exp(\frac{1}{\epsilon}(g_j + \delta)).
\end{aligned}$$

Hence it follows, as $\{P_\epsilon\}$ satisfies LDP,

$$\begin{aligned}
& \limsup_{\epsilon\to 0} \epsilon \log \int_S \exp(\frac{1}{\epsilon} g(x)) dP_\epsilon(x) \\
\le\ & \sup_{1\le j\le N} [(g_j + \delta) - I(C_j)] \\
=\ & \sup_{1\le j\le N} [\sup_{x\in C_j} (g(x) - I(x))] + \delta = \sup_x (g(x) - I(x)) + \delta
\end{aligned}$$

As $\delta > 0$ is arbitrary, upper bound follows.

Lower bound: Let $\delta > 0$. Choose $y \in S$ such that $g(y) - I(y) \ge \sup_x [g(x) - I(x)] - \frac{\delta}{2}$. One can find an open neighbourhood U of y such that $g(x) \ge g(y) - \frac{\delta}{2}$ for all $x \in U$. Then

$$\begin{aligned}
& \liminf_{\epsilon\to 0} \epsilon \log \int_S \exp[\frac{1}{\epsilon} g(x)] dP_\epsilon(x) \\
\ge\ & \liminf_{\epsilon\to 0} \epsilon \log \int_U \exp[\frac{1}{\epsilon} g(x)] dP_\epsilon(x)
\end{aligned}$$

$$
\begin{aligned}
&\geq \liminf_{\epsilon \to 0}[(g(y) - \frac{\delta}{2}) + \epsilon \log P_\epsilon(U)] \\
&\geq g(y) - \frac{1}{2}\delta - I(U) \geq g(y) - I(y) - \frac{\delta}{2} \\
&\geq \sup_x [g(x) - I(x)] - \delta.
\end{aligned}
$$

Since $\delta > 0$ is arbitrary we are done. □

Example 3.4 *:* Let $X_1, X_2, \dots$ be i.i.d.'s such that $P(X_i = \frac{1}{2}) = P(X_i = \frac{3}{2}) = \frac{1}{2}$. Let $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, and P_n the distribution of $S_n, n = 1, 2, \dots$ By Cramer's theorem $\{P_n\}$ satisfies LDP (taking $\epsilon = \frac{1}{n}$) with rate function

$$
I(z) = \begin{cases} \log 2 + (z - \frac{1}{2})\log(z - \frac{1}{2}) + (\frac{3}{2} - z)\log(\frac{3}{2} - z), \text{ if } \frac{1}{2} \leq z \leq \frac{3}{2} \\ \infty, \text{ otherwise.} \end{cases}
$$

Now

$$
\begin{aligned}
E(S_n^n) &= \int_{I\!R} x^n dP_n(x) = \int_{[\frac{1}{2},\frac{3}{2}]} \exp(n \log x) dP_n(x) \\
&= \int_{I\!R} \exp(ng(x)) dP_n(x)
\end{aligned}
$$

where $g(\cdot)$ is a bounded continuous function on $I\!R$ such that $g(x) = \log x, \frac{1}{2} \leq x \leq \frac{3}{2}$. By Varadhan's lemma

$$
\begin{aligned}
\lim_{n \to \infty} \frac{1}{n} \log E(S_n^n) &= \sup_{x \in I\!R}[g(x) - I(x)] \\
&= \sup\{\log x - I(x) : \frac{1}{2} \leq x \leq \frac{3}{2}\} = b. \qquad (3.6)
\end{aligned}
$$

Putting $h(x) = \log x - I(x)$, note that $h(1) = 0, h'(1) > 0$; hence $b > 0$. Though this example is somewhat artifical, it is instructive. By the law of large numbers $S_n \to 1$ a.s. Hence one might naively expect l.h.s. of (3.6) to be zero. However, as shown above it is not so. Thus $E(S_n^n)$ is basically determined (for large n) not by the almost sure behaviour but by the rare event when S_n takes values near b. □

We conclude this section with following

Theorem 3.5 *: (Contraction Principle): Let $\{P_\epsilon\}$ satisfy the LDP with a rate function $I(\cdot)$. Let $(\hat{S}, \hat{d})$ be another complete separable metric space; let $\pi : S \to \hat{S}$ be a continuous function. Put $\hat{P}_\epsilon = P_\epsilon \pi^{-1} \epsilon > 0$. Then $\{\hat{P}_\epsilon\}$ also satisfies the LDP with a rate function*

$$\hat{I}(y) = \inf\{I(x) : \pi(x) = y\}. \tag{3.7}$$

(If $y \notin$ Range (π) then $\hat{I}(y) = \infty$.)

Proof: For any $c \geq 0$ it is not difficult to see that $\pi(\{x : I(x) \leq c\}) = \{y : \hat{I}(y) \leq c\}$ and hence that $\hat{I}(\cdot)$ is a rate function. For any $\hat{A} \subset \hat{S}$ note that $\inf\{\hat{I}(y) : y \in \hat{A}\} = \inf\{I(x) : x \in \pi^{-1}(\hat{A})\}$; consequently by continuity of π, analogues of (3.1), (3.2) are easily verified for $\{\hat{P}_\epsilon\}$. $\square$

4 LDP for Brownian motion

Let $C_0[0,1] = \{w : [0,1] \to I\!\!R : w$ continuous and $w(0) = 0\}$. For $f, g \in C_0[0,1]$ put $d(f,g) = \|f - g\| = \sup\limits_{0 \leq t \leq 1} |f(t) - g(t)|$; $(C_0[0,1], d)$ is a complete separable metric space. $X(t, w) = w(t), 0 \leq t \leq 1, w \in C_0[0,1]$ denote the coordinate projections; $\mathcal{B}$ is the Borel σ-algebra of $C_0[0,1]$.

Let P denote the standard (one dimensional) Wiener measure, viz. $P(C_0[0,1]) = 1$, and under P, for $0 < t_1 < t_2 < \ldots < t_k < 1$ the k-dimensional random variable $(X(t_1), X(t_2), \ldots, X(t_k))$ has the k-dimensional normal (Gaussian) distribution with density function

$$g(z) = \left(\frac{1}{2\pi}\right)^{k/2} \frac{1}{\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}\langle z, \Sigma^{-1} z\rangle\right\}, z \in I\!\!R^k$$

where $\Sigma = ((t_i \wedge t_j))_{1 \leq i,j \leq k}$. Under P note that $X(0) = 0$ a.s. and $X(t_1), X(t_2) - X(t_1), \ldots, X(t_k) - X(t_{k-1})$ are independent one dimensional normal (Gaussian) random variables with mean zero and respective variances $t_1, t_2 - t_1, \ldots, t_k - t_{k-1}$; that is, under P the stochastic process $\{X(t) : 0 \leq t \leq 1\}$ is a standard Brownian motion. See [KS] for details concerning existence/construction and various properties.

For $\epsilon > 0$ let P_ϵ denote the distribution of the process $\{\sqrt{\epsilon} X(t) : t \geq 0\}$; so for $A \in \mathcal{B}$ we have $P_\epsilon(A) = P(\frac{1}{\sqrt{\epsilon}} A)$. Clearly $P_\epsilon \Rightarrow \delta_0$ as $\epsilon \downarrow 0$ where δ_0

is the probability measure concentrated on the function which is identically zero.

Define $I : C_0[0,1] \to [0,\infty]$ by

$$I(f) = \begin{cases} \frac{1}{2} \int\limits_0^1 |\dot{f}(s)|^2 ds, & f \in H \\ \infty, & f \notin H \end{cases} \qquad (4.1)$$

where $H = \{f \in C_0[0,1] : f$ is absolutely continuous and $\dot{f}$ is square intergrable$\}$.

To prove that I is a rate function we need

Lemma 4.1 : *For $f \in C_0[0,1]$*

$$2I(f) = \sup_n n \sum_{k=1}^{n} \left| f\left(\frac{k}{n}\right) - f\left(\frac{k-1}{n}\right) \right|^2 . \qquad (4.2)$$

Proof: Set $\zeta_n(f) = n \sum\limits_{k=1}^{n} |f(\frac{k}{n}) - f(\frac{k-1}{n})|^2$ and $\zeta(f) = \sup\limits_n \zeta_n(f)$. Note that $\zeta_n(f) = \int\limits_0^1 |\dot{f}^{(n)}(r)|^2 dr$ where $f^{(n)}$ is the piecewise linear function such that $f^{(n)}(\frac{k}{n}) = f(\frac{k}{n}), k = 0,1,\ldots,n$ and $n = 1,2,\ldots$ Therefore $\zeta(f) < \infty$ implies that $\{\dot{f}^{(n)} : n = 1,2,\ldots\}$ is a bounded set in $L^2[0,1]$. So Banach-Alaoglu theorem assures that there exists $\dot{f} \in L^2[0,1]$ such that (by taking a subsequence if necessary) $\langle \dot{f}^{(n)}, \varphi \rangle \to \langle \dot{f}, \varphi \rangle$ for all $\varphi \in L^2[0,1]$. Hence

$$\begin{aligned} f(t) &= \lim_n f^{(n)}(t) = \lim_n \langle \dot{f}^{(n)}, 1_{[0,t]} \rangle \\ &= \int_0^t \dot{f}(s)\ ds. \end{aligned}$$

Thus f is absolutely continuous. Moreover $\|\dot{f}\|^2 = \lim\limits_n \langle \dot{f}^{(n)}, \dot{f} \rangle \leq \sqrt{\zeta(f)} \|\dot{f}\|$, and hence $I(f) = \frac{1}{2}\|\dot{f}\|^2 \leq \frac{1}{2}\zeta(f)$. Thus we have proved that if $f \in C_0[0,1]$ and $\zeta(f) < \infty$, then $f \in H$ and $2I(f) \leq \zeta(f)$.

To complete the proof we need to show that $\zeta(f) \leq 2I(f)$ for $f \in H$. For

$n = 1, 2, \ldots$ observe that by Schwarz inequality

$$\zeta_n(f) \le n \cdot \frac{1}{n} \sum_{k=1}^{n} \int_{(k-1)/n}^{k/n} |\dot{f}(r)|^2 dr = 2I(f)$$

whence it follows that $\zeta(f) \le 2I(f)$. □

Lemma 4.2 : *I is a rate function.*

Proof: For $\alpha < \infty$ let $\psi_I(\alpha) = \{f \in C_0[0,1] : I(f) \le \alpha\}$. For $f \in \psi_I(\alpha)$ using Schwarz inequality it is easily seen that $|f(t) - f(s)| \le \sqrt{2\alpha}\sqrt{t-s}, 0 \le s < t \le 1$. Therefore in view of Ascoli's theorem it is now enough to prove that $\psi_I(\alpha)$ is closed. By preceding lemma $\psi_I(\alpha) = \bigcap_{n=1}^{\infty} \{f : \zeta_n(f) \le 2\alpha\}$. Continuity of f implies that of ζ_n and hence the required conclusion. □

Lemma 4.3 : *For any closed set $F \subseteq C_0[0,1]$ we have $I(F^\delta) \uparrow I(F)$ as $\delta \downarrow 0$, where F^δ denotes δ-neighbourhood of F.*

Proof: Note that $F^\delta = \{w : d(w, F) < \delta\}$ where d is the metric on $C_0[0,1]$, and $I(F^\delta), I(F)$ are defined by (3.3). Clearly $I(F^\delta)$ increases as $\delta \downarrow 0$. Suppose $I(F^\delta) \le a < I(F)$ for all $\delta > 0$. Choose $f_n \in F^{1/n}$ so that $I(f_n) \le I(F^{1/n}) + \frac{1}{n} \le a + \frac{1}{n}$, for $n = 1, 2, \ldots$ Since $\psi_I(a+1)$ is compact by the preceding lemma, $\{f_n\}$ has a convergent subsequence $\{f_{n_k}\}$; let $f = \lim_k f_{n_k}$. Then $f \in F$, and by lower semicontinuity of I we have $I(f) \le \liminf_{k \to \infty} I(f_{n_k}) \le a < I(F)$. This contradicts the definition of $I(F)$ unless $a = I(F)$. Thus $I(F^\delta) \uparrow I(F)$. □

We state below two results concerning the Wiener measure without proof; see [KS] for proofs.

Proposition 4.4 : *For $a > 0, t > 0$*

$$P\left(\sup_{0 \le s \le t} |X(s)| \ge a\right) \le 2\exp(-a^2/2t). \qquad (4.3)$$

□

Remark: Inequality (4.3) can be proved using the exponential martingale $\exp\{\theta X(t) - \frac{\theta^2}{2}t\}$ and the submartingale inequality for appropriate θ.

Theorem 4.5 *: (Cameron-Martin): Let $g \in C_0[0,1]$ be absolutely continuous, say, $g(t) = \int_0^t \dot{g}(s)\ ds$ with square integrable derivative $\dot{g}$. Define the transformation $T : C_0[0,1] \to C_0[0,1]$ by $(Tw)(t) = w(t) + g(t) = w(t) + \int_0^t \dot{g}(s)\ ds$. Then PT^{-1} is equivalent to P and the Radon-Nikodym derivative is given by*

$$\frac{dPT^{-1}}{dP}(w) = \exp\left[\int_0^1 \dot{g}(s)\ dX(s,w) - \frac{1}{2}\int_0^1 |\dot{g}(s)|^2 ds\right] \qquad (4.4)$$

where $\int_0^1 \dot{g}(s) dX(s,w)$ is the Wiener integral of $\dot{g}$. □

Remark: Finite dimensional projections of PT^{-1} and P are clearly equivalent and the Radon-Nikodym derivatives can be computed easily. Above theorem involves taking limit of the finite dimensional derivatives in an appropriate L^2-sense. Wiener integral is a special case of stochastic / Ito integrals. If $g \in C^2$ then $\int_0^1 \dot{g}(s)\ dX(s,w)$ can be equivalently defined pathwise using Stieltjes integration, viz.

$$\begin{aligned}\int_0^1 \dot{g}(s) dX(s,w) &= \int_0^1 \dot{g}(s)\ dw(s) \\ &= w(1)\ \dot{g}(1) - \int_0^1 w(s)\ddot{g}(s)\ ds \qquad (4.5)\end{aligned}$$

as the r.h.s. of (4.5) is meaningful. □

Theorem 4.6 *(Schilder): $\{P_\epsilon\}$ satisfies the LDP with rate function $I(\cdot)$ given by (4.1).*

Proof: *Upper bound*: Let $F \subseteq C_0[0,1]$ be a nonempty closed set; for $\delta > 0, F^\delta$ denotes δ-neighbourhood of F. For any map $\pi : C_0[0,1] \to C_0[0,1]$ clearly

$$P_\epsilon(F) \le P_\epsilon(\{w : \pi(w) \in F^\delta\}) + P_\epsilon(\{w : \|\pi(w) - w\| \ge \delta\}). \qquad (\ 4.6\ )$$

For $n = 1, 2, \ldots$ we will take $\pi = \pi_n$ where $(\pi_n w)(t) = w(\frac{k}{n})$ if $t = \frac{k}{n}, k = 0, 1, \ldots, n$ and $\pi_n w$ is linearly interpolated on each interval $[\frac{k-1}{n}, \frac{k}{n}], k = 1, 2, \ldots, n$.

We first estimate $P_\epsilon(\{w : \pi_n(w) \in F^\delta\})$. Since $P_\epsilon(A) = P(\frac{1}{\sqrt{\epsilon}}A)$, using the notations and proof of Lemma 4.1, we get

$$\begin{aligned} P_\epsilon(\{w : \pi_n(w) \in F^\delta\}) &\le P_\epsilon(\{w : I(\pi_n w) \ge I(F^\delta)\}) \\ &= P_\epsilon\left(\{w : n \sum_{k=1}^{n} |w(\frac{k}{n}) - w(\frac{k-1}{n})|^2 \ge 2I(F^\delta)\}\right) \\ &= P\left(n \sum_{k=1}^{n} |X(\frac{k}{n}) - X(\frac{k-1}{n})|^2 \ge \frac{2I(F^\delta)}{\epsilon}\right). \end{aligned}$$

Under P note that $n \sum\limits_{k=1}^{n} |X(\frac{k}{n}) - X(\frac{k-1}{n})|^2$ has $\chi^2(n)$ distribution (that is, chi-square distribution with n degrees of freedom). Therefore from the above (after suitable change of variables)

$$\begin{aligned} P_\epsilon(\{w : \pi_n(w) \in F^\delta\}) \le&\ \text{Prob.}\left(\chi^2(n) \ge \frac{2I(F^\delta)}{\epsilon}\right) \\ =&\int_{I(F^\delta)/\epsilon}^{\infty} (\Gamma(n/2))^{-1} s^{\frac{n}{2}-1} e^{-s} ds \\ \le&\ \frac{(n/2)!}{\Gamma(n/2)} \left(\frac{I(F^\delta)}{\epsilon} \vee 1\right)^{\frac{n}{2}} \exp\left(-\frac{I(F^\delta)}{\epsilon}\right) \end{aligned} \qquad (\ 4.7\ )$$

for any even integer n, where the last step is obtained by repeated integration by parts.

Under $P, X(t+h) - X(h)$ has the same distribution as $X(t) - X(0) = X(t)$. Therefore by Proposition 4.4 we get for any n,

$$
\begin{aligned}
& P_\epsilon(\{w : \|\pi_n(w) - w\| \geq \delta\}) \\
\leq \; & \sum_{k=1}^{n} P_\epsilon \left(\{w : \sup_{\frac{k-1}{n} \leq t \leq \frac{k}{n}} |w(t) - w(\frac{k-1}{n})| \geq \frac{1}{2}\delta\} \right) \\
= \; & \sum_{k=1}^{n} P \left(\sup_{\frac{k-1}{n} \leq t \leq \frac{k}{n}} |X(t) - X(\frac{k-1}{n})| \geq \frac{\delta}{2\sqrt{\epsilon}} \right) \\
= \; & n \; P \left(\sup_{0 \leq t \leq \frac{1}{n}} |X(t)| \geq \frac{\delta}{2\sqrt{\epsilon}} \right) \\
\leq \; & 2n \; \exp \left(- \frac{n\delta^2}{8\epsilon} \right). \qquad (4.8)
\end{aligned}
$$

Fix $\delta > 0$; choose an even integer n so that $\eta = (\frac{n\delta^2}{8} - I(F^\delta)) > 0$; we may assume $\frac{I(F^\delta)}{\epsilon} \geq 1$ as we are interested in the case when $\epsilon \downarrow 0$. Then using (4.6) - (4.8) we get

$$
P_\epsilon(F) \leq exp \left(- \frac{I(F^\delta)}{\epsilon} \right) \cdot \{C(n,\delta)\epsilon^{-n/2} + 2n \; e^{-\eta/\epsilon}\}
$$

where $C(n,\delta) = \frac{(n/2)!}{\Gamma(n/2)} (I(F^\delta))^{n/2}$. Therefore

$$
\begin{aligned}
\epsilon \log P_\epsilon(F) \;\; \leq \;\; & -I(F^\delta) + \epsilon \log(C(n,\delta)\epsilon^{-n/2}) \\
& + \epsilon \log \left\{ 1 + \frac{2n}{C(n,\delta)} e^{-\eta/\epsilon} \epsilon^{n/2} \right\}.
\end{aligned}
$$

From the above it is clear that

$$
\limsup_{\epsilon \to 0} \epsilon \log P_\epsilon(F) \leq -I(F^\delta).
$$

Now letting $\delta \downarrow 0$ and using Lemma 4.3

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

proving the upper bound.

Lower bound: Let $G \subseteq C_0[0,1]$ be a nonempty open set. We may assume $I(G) < \infty$ for otherwise there is nothing to prove. As C^2-functions are dense, given $\alpha > 0$ we can find $f \in G \cap C^2$ so that $I(f) \leq I(G) + \alpha$. Choose $\delta_0 > 0$ such that $B(f, \delta_0) := \{g \in C_0[0,1] : \|f - g\| < \delta_0\} \subseteq G$. Then for $0 < \delta \leq \delta_0$

$$\begin{aligned} P_\epsilon(G) &\geq P_\epsilon(B(f,\delta)) = P\left(B(\frac{f}{\sqrt{\epsilon}}, \frac{\delta}{\sqrt{\epsilon}})\right) \\ &= PT_\epsilon^{-1}\left(B(0, \frac{\delta}{\sqrt{\epsilon}})\right) \end{aligned} \tag{4.9}$$

where $T_\epsilon : C_0[0,1] \to C_0[0,1]$ is defined by

$$(T_\epsilon w)(t) = w(t) - \frac{1}{\sqrt{\epsilon}} f(t) = w(t) - \frac{1}{\sqrt{\epsilon}} \int_0^t \dot{f}(s)\, ds.$$

By Theorem 4.5 and equations (4.4), (4.5)

$$\begin{aligned} \frac{dPT_\epsilon^{-1}}{dP}(w) &= \exp\left[-\frac{1}{\sqrt{\epsilon}}\{w(1)\dot{f}(1) - \int_0^1 w(s)\ddot{f}(s)\, ds\} - \frac{1}{2\epsilon}\int_0^1 |\dot{f}(s)|^2 ds\right] \\ &\geq \exp\left[-\frac{2\theta_f}{\sqrt{\epsilon}}(\sup_{0 \leq t \leq 1} |w(t)|) - \frac{I(f)}{\epsilon}\right\} \end{aligned} \tag{4.10}$$

where $\theta_f = \sup_{0 \leq t \leq 1} (|\dot{f}(t)| + |\ddot{f}(t)|)$. Now (4.9), (4.10) give

$$\begin{aligned} P_\epsilon(G) &= \int_{B(0,\delta/\sqrt{\epsilon})} \frac{dPT_\epsilon^{-1}}{dP}(w)\, dP(w) \\ &\geq \exp\left[-\frac{2\theta_f \delta}{\epsilon} - \frac{I(f)}{\epsilon}\right] \cdot P(B(0, \delta/\sqrt{\epsilon})) \end{aligned}$$

resulting in

$$\begin{aligned} \epsilon \log P_\epsilon(G) &\geq -I(f) - 2\theta_f \delta + \epsilon \log P(B(0, \delta/\sqrt{\epsilon})) \\ &\geq -I(G) - \alpha - 2\theta_f \delta + \epsilon \log P(B(0, \delta/\sqrt{\epsilon})). \end{aligned}$$

For any $\delta > 0$ note that $P(B(0, \delta/\sqrt{\epsilon})) \to 1$ as $\epsilon \downarrow 0$. Therefore for any $0 < \delta \leq \delta_0$, any $\alpha > 0$

$$\liminf_{\epsilon \to 0} \epsilon \log P_\epsilon(G) \geq -I(G) - \alpha - 2\theta_f \delta.$$

Since $\delta \in (0, \delta_0), \alpha > 0$ are arbitrary we get the required lower bound

$$\liminf_{\epsilon \to 0} \epsilon \log P_\epsilon(G) \geq -I(G).$$

$\square$

Remark: Schilder's theorem holds also for multidimensional Brownian motion (equivalently multidimensional Wiener measure) with the analogous rate function. The same proof works with the obvious modifications.

5 LDP for diffusions

In this section we consider LDP for diffusion processes, extending the result of the preceding section. Only a special case is proved.

Theorem 5.1 *(Freidlin-Wentzell): Let $P, \{X(t) : 0 \leq t \leq 1\}$ be as in Section 4. Let $b : \mathbb{R} \to \mathbb{R}$ be a bounded Lipschitz continuous function. For $\epsilon > 0$ let $\{X^\epsilon(t) : 0 \leq t \leq 1\}$ be the stochastic process satisfying the integral equation*

$$X^\epsilon(t) = \sqrt{\epsilon} X(t) + \int_0^t b(X^\epsilon(s))\, ds. \qquad (5.1)$$

Let $Q_\epsilon = P(X^\epsilon)^{-1}$ be the distribution of the process $X^\epsilon(\cdot)$. Then $\{Q_\epsilon\}$ satisfies the LDP with rate function

$$I(f) = \begin{cases} \frac{1}{2} \int_0^1 |\dot{f}(t) - b(f(t))|^2 dt, & \text{if } f \in H \\ \infty, & \text{if } f \notin H \end{cases} \qquad (5.2)$$

where H is as in Section 4.

Proof: Define $F : C_0[0,1] \to C_0[0,1]$ as follows: for $w \in C_0[0,1], f = Fw \in C_0[0,1]$ is the unique solution to

$$f(t) = w(t) + \int_0^t b(f(s))\, ds, 0 \leq t \leq 1. \qquad (5.3)$$

As b is Lipschitz continuous it is easily verified that F is well defined and $Q_\epsilon = P(X^\epsilon)^{-1} = P_\epsilon F^{-1}$ where P_ϵ is as in Section 4. Let $f_1 = F(w_1), f_2 = F(w_2)$. Then

$$\begin{aligned} |f_1(t) - f_2(t)| &\leq |w_1(t) - w_2(t)| + \int_0^t |b(f_1(s)) - b(f_2(s))| ds \\ &\leq \|w_1 - w_2\| + K \int_0^t |f_1(s) - f_2(s)| ds \end{aligned}$$

and hence (by iteration) $\|f_1 - f_2\| \leq e^K \|w_1 - w_2\|$, where K is the Lipschitz constant. Thus F is continuous. So by Theorems 3.5 and 4.6 it follows that $\{Q_\epsilon\}$ satisfies the LDP with rate function

$$I(f) = \inf \left\{ \frac{1}{2} \int_0^1 |\dot{w}(s)|^2 ds : w \in H, f = F(w) \right\}. \tag{5.4}$$

It remains to identify (5.4) with (5.2). From the definition (5.3) of F it is clear that $Fw_1 = Fw_2$ implies $w_1 = w_2$. In (5.3) clearly w is absolutely continuous if and only if f is, and in such a case $\dot{f}(t) = \dot{w}(t) + b(f(t))$; consequently it is also clear that $\dot{f} \in L^2[0,1]$ if and only if $\dot{w} \in L^2[0,1]$. The proof is now complete. □

A large class of diffusion processes can be represented as solutions to stochastic differential equations. Let d be a positive integer; let $\{X(t) : t \geq 0\}$ denote a d-dimensional Brownian motion. Let $\sigma(\cdot), b(\cdot)$ respectively be $(d \times d)$ matrix valued, $\mathbb{R}^d$-valued functions on $\mathbb{R}^d$. The stochastic differential equation

$$dZ(t) = \sigma(Z(t))\ dX(t) + b(Z(t))\ dt \tag{5.5}$$

with initial value $Z(0) = z_0$ is interpreted as the stochastic integral equation

$$Z(t) = z_0 + \int_0^t \sigma(Z(s))\ dX(s) + \int_0^t b(Z(s))\ ds. \tag{5.6}$$

Here expressions of the form $\int_0^t f(s)\ dX(s)$ denote Ito integrals of appropriate scalar/ vector/ matrix valued processes f. When σ, b are bounded and Lipschitz continuous, by a Picard iteration, unique solution to (5.6) (in an appropriate sense) can be obtained. Clearly (5.1) is a special case of (5.6). The diffusion process given by (5.6) is a strong Markov process with "infinitesimal generator" L where L is gien by

$$Lg(x) = \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x)\frac{\partial^2 f(x)}{\partial x_i \partial x_j} + \sum_{i=1}^{d} b_i(x)\frac{\partial f(x)}{\partial x_i} \tag{5.7}$$

where $((a_{ij}(x)))_{1\le i,j\le d} = \sigma(x)\sigma^T(x)$. In particular the generator corresponding to (5.1) is $\frac{\epsilon}{2}\frac{d^2}{dx^2} + b(x)\frac{d}{dx}$; and the generator of the d-dimensional Brownian motion is the d-dimensional Laplacian $\frac{1}{2}\Delta := \frac{1}{2}\sum_{i=1}^{d}\frac{\partial^2}{\partial x_i^2}$. A good source for definitions, results, proofs, etc. is [KS].

We now state a generalization of the preceding theorem, without proof.

Theorem 5.2 *(Freidlin-Wentzell): Let σ, b be bounded and Lipschitz continuous. Let $x \in I\!R^d$. For $\epsilon > 0$ consider the diffusion*

$$dZ^{\epsilon,x}(t) = \sqrt{\epsilon}\sigma(Z^{\epsilon,x}(t))\ dX(t) + b(Z^{\epsilon,x}(t))\ dt,\ 0 \le t \le 1 \tag{5.8}$$

with initial value $Z^{\epsilon,x}(0) = x$. Assume further that $a(\cdot) := \sigma(\cdot)\sigma^T(\cdot)$ is uniformly positive definite. Let $Q_{\epsilon,x}$ denote the distribution of the process $Z^{\epsilon,x}(\cdot)$ in $C([0,1] : I\!R^d)$. Then $\{Q_{\epsilon,x} : \epsilon > 0\}$ satisfies the LDP with rate function

$$I_x(f) = \begin{cases} \frac{1}{2}\int_0^1 \langle(\dot{f}(t) - b(f(t))), a^{-1}(f(t))(\dot{f}(t) - b(f(t)))\rangle dt, & \\ & \text{if } f \in H^x \\ \infty & \text{otherwise} \end{cases} \tag{5.9}$$

where

$$H^x = \{f \in C([0,1] : I\!R^d) : f(t) = x + \int_0^t \varphi(s)ds, \varphi \in L^2[0,1]\}$$

□

The map on $C[0,1]$ given by $Z^{\epsilon,x}$ is only measurable and not continuous. So earlier proof cannot be applied directly. Proof involves an approximation where the coefficients are frozen for small durations of time, and estimating the correction terms. See [DZ], [V] for details.

In the context of the so called 'exit problem' for diffusions the rate function $I_x(\cdot)$ given by (5.9) plays an important role. See [DZ], [V], [FW].

References

[DZ] A. Dembo and O. Zeitouni: *Large Deviations Techniques and applications* (2nd edn.) Springer, New York, 1998.

[FW] M.I. Freidlin and A.D. Wentzell: *Random Perturbations of Dynamical Systems*, Springer, New York, 1984.

[H] F. den Hollander: *Large Deviations*, Amer. Math. Soc., Providence, 2000.

[KS] I. Karatzas and S.E. Shreve: *Brownian Motion and Stochastic Calculus*, Springer, New York, 1988.

[S] D.W. Stroock: *An introduction to the Theory of Large Deviations*, Springer, Berlin, 1984.

[V] S.R.S. Varadhan: *Large Deviations and applications*, SIAM, Philadelphia, 1984.