Proc. Indian Acad. Sci. (Math. Sci.) Vol. 118, No. 2, May 2008, pp. 161–182.

Large deviations: An introduction to 2007 Abel prize

S RAMASUBRAMANIAN

Stat.-Math. Unit, Indian Statistical Institute, 8th Mile, Mysore Road,
Bangalore 560 059, India
E-mail: ram@isibang.ac.in; sunramin@gmail.com

MS received 24 July 2007

Abstract. 2007 Abel prize has been awarded to S R S Varadhan for creating a unified theory of large deviations. We attempt to give a flavour of this branch of probability theory, highlighting the role of Varadhan.

Keywords. Large deviation principle (LDP); rate function; Cramer's theorem; Sanov's theorem; Esscher transform/tilt; convex conjugates; Laplace's method; Varadhan's lemma; weak convergence of probability measures; empirical distribution; Himilton–Jacobi equation; Burger's equation; variational formula; sample path LDP; Brownian motion/diffusion; Markov processes; ergodicity; Wentzell–Freidlin theory; exit problem; Feynman–Kac formula; occupation time; principal eigenvalue; Donsker–Varadhan theory.

1. Introduction

The award of the prestigious Abel prize for 2007 to Professor S R S Varadhan has been widely acclaimed, especially among the mathematical community in India. The Abel prize, perhaps the highest award in mathematics, has been instituted by the Norwegian Academy of Science and Letters since 2003; this annual award is patterned roughly along the lines of Nobel prize in sciences. The citation says that Varadhan is being given the award "for his fundamental contributions to probability theory and in particular for creating a unified theory of large deviations".

Large deviations is a part of probability theory; it provides asymptotic estimates for probabilities of rare events. It may be pointed out that the strong law of large numbers and the central limit theorem, the versatile classical limit theorems of probability theory, concern typical events. As large deviation estimates deal with probabilities of rare events the methods needed are more subtle. Moreover, context specific techniques play a major role though there are quite a few general principles. In this write-up we attempt to give an elementary introduction to large deviations, of course, highlighting the role of Varadhan.

2. Actuarial origin and theorems of Cramer and Sanov

Suppose $X_1, X_2, \ldots$ are independent identically distributed real-valued random variables, real-valued i.i.d.'s for short. Let $F(x) = P(X_i \leq x)$, $x \in \mathbb{R}$ denote their common distribution function, and $m = \int_{\mathbb{R}} x \, \mathrm{d}F(x)$ their common mean (expectation or 'average value'). If m exists then the strong law of large numbers states that $\frac{1}{n}(X_1 + \cdots + X_n) \to m$

with probability 1 as $n \to \infty$. This forms a basis for the validity of many statistical and scientific procedures of taking averages. If $a > m$ then the above implies that

$$\lim_{n\to\infty} P\left(\frac{1}{n}(X_1 + \cdots + X_n) > a\right) = 0. \tag{2.1}$$

However one would like to know at what rate convergence in (2.1) takes place.

Denote $S_n = X_1 + \cdots + X_n,\ n \geq 1$. As the central limit theorem (CLT) explains the prevalence of Gaussian distribution in various aspects of nature, one wonders if CLT can shed more light. By the classical CLT for sums of i.i.d.'s, assuming that the common variance is 1, we have for $a > m$,

$$\begin{aligned} P\left(\frac{1}{n}S_n > a\right) &= P\left(\frac{1}{\sqrt{n}}(S_n - nm) > \sqrt{n}(a - m)\right) \\ &\approx 1 - \Phi(\sqrt{n}(a - m)) \to 0, \text{ as } n \to \infty \end{aligned}$$

where

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, \mathrm{e}^{-\frac{1}{2}y^2} \mathrm{d}y,\ x \in \mathbb{R}. \tag{2.2}$$

Here $\Phi(\cdot)$ is the standard normal (Gaussian) distribution function. So the CLT is not powerful enough to discern the rate.

The event $\{S_n > na\}$ is a typical 'rare event' of interest in insurance. For example, X_i can denote the claim amount of policy holder i in a given year, and hence S_n denotes the total claim amount of n policy holders. Assuming a big portfolio for the insurance company (that is, n is very large), any estimate for $P(S_n > na)$, where $a > m$, gives information about the 'right tail' of the total claim amount payable by the company in a year.

As another illustration from insurance, S_n can be regarded as the cumulative net payout (that is, claim payment minus income from premiums and interests) in n years. The initial capital u_0 of the company is generally quite large. If S_n exceeds u_0 then the company is ruined.

It is easy to see why actuaries would be interested in the tail behaviour of S_n. They would like to have an idea of how bad an extremely bad year can be, and perhaps fine-tune premium rates or reinsurance levels. It is no wonder that the problem attracted the attention of the great Swedish probabilist Harald Cramer, a pioneer in putting statistics as well as insurance modelling on firm mathematical foundations. However, F Esscher, a Scandinavian actuary, may have been the first to look at the problem and come up with some interesting ideas (in 1932) which were later sharpened/extended by Cramer.

To appreciate Cramer's result let us first look at two examples.

Example 2.1. Let X_i be i.i.d. random variables such that $P(X_i = 0) = P(X_i = 1) = \frac{1}{2}$; that is, $\{X_i\}$ is a sequence of i.i.d. Bernoulli random variables with parameter $\frac{1}{2}$. Note that $m = \frac{1}{2}$. Let $a \in \left(\frac{1}{2}, 1\right]$. It is easily seen that (as S_n has a binomial distribution)

$$2^{-n} Q_n(a) \leq P(S_n \geq na) \leq (n+1)2^{-n} Q_n(a),$$

where $Q_n(a) = \max_{k \geq an}\binom{n}{k}$. As $a > \frac{1}{2}$, the maximum is attained at $k_0 =$ the smallest integer $\geq an$. Using Stirling's formula $n! = n^n \mathrm{e}^{-n}\sqrt{2\pi n}\left(1 + O\left(\frac{1}{n}\right)\right)$ it is not difficult to see that

$$\lim_{n\to\infty} \frac{1}{n} \log Q_n(a) = -a\log a - (1-a)\log(1-a).$$

From this it follows that

$$\lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq na) = -[\log 2 + a\log a + (1-a)\log(1-a)].$$

Example 2.2. Let $\{X_i\}$ be i.i.d. $N(0,1)$ random variables so that $P(X_i \in A) = \int_A \mathfrak{n}(x)\,\mathrm{d}x,\ A \subset \mathbb{R}$ where

$$\mathfrak{n}(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\}, \quad x \in \mathbb{R}. \tag{2.3}$$

Note that $P(X_i \leq x) = \Phi(x)$ where Φ is given by (2.2). In this case $m = 0$. Note that the *empirical mean* $\frac{1}{n}\,S_n = \frac{1}{n}\,\sum_{i=1}^n X_i$ has the $N\big(0, \frac{1}{n}\big)$ distribution. Hence by the properties of normal distributions, for any $a > 0$,

$$P\left(\left|\frac{1}{n}S_n\right| \geq a\right) = 2[1 - \Phi(a\sqrt{n})]. \tag{2.4}$$

Now for any $y > 0$ clearly

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

Integrating the above over $[z, \infty)$, where $z > 0$,

$$\left(\frac{1}{z} - \frac{1}{z^3}\right)\mathfrak{n}(z) < [1 - \Phi(z)] < \frac{1}{z}\mathfrak{n}(z). \tag{2.5}$$

From (2.4) and (2.5) it easily follows that

$$\lim_{n\to\infty} \frac{1}{n} \log P\left(\left|\frac{1}{n}S_n\right| \geq a\right) = -\frac{1}{2}\,a^2. \tag{2.6}$$

Thus, the probability of the rare event $\left\{\left|\frac{1}{n}S_n\right| \geq a\right\}$ is of the order $\exp\big(-\frac{1}{2}na^2\big)$. This is a typical large deviations statement, and $\frac{1}{2}\,a^2$ is an example of a rate function. □

Cramer's theorem is about an analogue of the above for sums of i.i.d. random variables. Assume that the moment generating function (or the Laplace transform) of X_1 exists, that is,

$$M(t) \triangleq E[\mathrm{e}^{tX_1}] = \int_{\mathbb{R}} \mathrm{e}^{tx}\,\mathrm{d}F(x) < \infty,\ \forall\, t \in \mathbb{R}. \tag{2.7}$$

For $a > E(X_1)$, by Chebyshev's inequality

$$\begin{aligned} P\left(\frac{1}{n}S_n \geq a\right) &= P\left(\exp\left(\frac{\theta}{n}S_n\right) \geq \mathrm{e}^{\theta a}\right) \\ &\leq \mathrm{e}^{-\theta a} E\left[\exp\left(\frac{\theta}{n}S_n\right)\right] = \mathrm{e}^{-\theta a}\left(M\left(\frac{\theta}{n}\right)\right)^n \end{aligned}$$

for any $\theta > 0$. Putting $\theta = nt$ we now see that (as θ is arbitrary)

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{n}S_n \geq a\right) &\leq \inf_{t\geq 0}[-ta + \log M(t)] \\ &= -\left\{\sup_{t\geq 0}[ta - \log M(t)]\right\} \\ &= -\left\{\sup_{t\in\mathbb{R}}[ta - \log M(t)]\right\},\end{aligned}$$

where the last step follows as $a > E(X_1)$. In Example 2.2 we have $M(t) = e^{\frac{1}{2}t^2}$ and hence $\frac{1}{2}a^2 = \sup_{t\in\mathbb{R}}[ta - \log M(t)]$. A similar comment is true also of Example 2.1.

In fact we have the following theorem, which is the starting point in large deviations.

Theorem 2.3 (Cramer, 1938). *Let $\{X_i\}$ be real-valued i.i.d.'s having finite moment generating function $M(\cdot)$. Then for any $a > E(X_1)$,*

$$\lim_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{n}S_n \geq a\right) = -I(a), \tag{2.8}$$

where

$$I(a) \triangleq \sup\{at - \log M(t): t \in \mathbb{R}\}. \tag{2.9}$$

Similarly for $a < E(X_1)$,

$$\lim_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{n}S_n \leq a\right) = -I(a). \tag{2.10}$$

□

For a proof, see [H], [DZ] and [V2].

Note that $\log M(t)$ is a convex function. One can show that $I(\cdot)$ is also convex and that

$$\log M(t) = \sup\{ta - I(a): a \in \mathbb{R}\}. \tag{2.11}$$

Thus $\log M(t)$ and $I(a)$ are *convex conjugates*. The rate function $I(\cdot)$ is also known as the *Fenchel–Legendre transform* of the logarithmic moment generating function $\log M(\cdot)$.

As seen above, the upper bound in (2.8) is an easy consequence of Chebyshev's inequality. The key idea in the proof of the lower bound is an 'exponential tilting' or Esscher transform of the distribution, a device having its origins again in insurance problems. With $F(\cdot)$ and $M(\cdot)$ as above, for each fixed $t \in \mathbb{R}$ the Esscher transform is defined by

$$d\tilde{F}_t(x) = \frac{1}{M(t)} e^{tx} \, dF(x).$$

Under the tilted distribution the rare event $\left\{\frac{1}{n}S_n \geq a\right\}$ becomes a typical event, thereby facilitating analysis (see [H]).

See [F] for an account of Cramer's theorem in the context of the central limit problem; also illustrations from risk theory are sprinkled often in [F]. For detailed account of insurance models, and for the role played by Esscher transform in estimating ruin probabilities, see [RSST]. According to Varadhan, variations of Esscher transform is a recurring theme in large deviations.

Under the hypothesis of Cramer's theorem the rate function I has the following properties:

(i) I has compact level sets, that is, $I^{-1}([0, c])$ is compact for all $c < \infty$; in particular I is lower semicontinuous;
(ii) $I(z) \geq 0$ with equality if and only if $z = E(X_1)$;
(iii) I is convex on $\mathbb{R}$.

If X_i has the Bernoulli distribution with parameter $0 < p < 1$, then $I(a) = a \log\left(\frac{a}{p}\right) + (1-a) \log\left(\frac{1-a}{1-p}\right)$, for $a \in [0, 1]$, and $I(a) = \infty$, otherwise. Similarly, if X_i has the Poisson distribution with parameter $\lambda > 0$, then $I(a) = \lambda - a + a \log\left(\frac{a}{\lambda}\right)$, for $a \geq 0$, and $I(a) = \infty$ otherwise.

With the assumptions as in Theorem 2.3, if $a > m$ note that $I(z) \geq I(a)$ for all $z \geq a$. So the result (2.8) can be rephrased as, denoting $A = [a, \infty)$,

$$\lim_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{n} S_n \in A\right) = -\inf_{z\in A} I(z). \tag{2.12}$$

Such a statement holds also for nice subsets A. □

Even a reasonable formulation of Cramer's theorem in $\mathbb{R}^d$ was perhaps achieved only when the general framework for large deviations was given by Varadhan. Finer aspects of convexity, like a minimax theorem, are needed in the proof (see [V2]).

In 1957, Russian probabilist Sanov established an important extension of Cramer's theorem to empirical distributions of real-valued i.i.d.'s. We shall describe it briefly. Let μ be a probability measure on $\mathbb{R}$. For $y \in \mathbb{R}$ we shall denote by δ_y the Dirac measure concentrated at y. Let $\{Y_i : i \geq 1\}$ be a real-valued i.i.d. sequence defined on a probability space $(S, \mathcal{F}, P)$ with common distribution μ. Set

$$N_n^Y(\omega, \cdot) \triangleq \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i(\omega)}(\cdot), \;\omega \in S, \;n \geq 1. \tag{2.13}$$

For each n, ω note that $N_n^Y(\omega, \cdot)$ is a probability measure on $\mathbb{R}$; $\{N_n^Y\}$ is called the family of *empirical distributions* of $\{Y_i\}$.

Let $\mathcal{M}(\mathbb{R})$ denote the set of all probability measures on the real line $\mathbb{R}$. This is a closed convex subset of the topological vector space of all finite signed measures on $\mathbb{R}$ with the topology of weak convergence of measures; that is ν_n converges to ν, denoted $\nu_n \Rightarrow \nu$, if and only if $\int f \, d\nu_n \to \int f \, d\nu$ for all $f \in C_b(\mathbb{R})$.

Denote $X_i(\omega) = \delta_{Y_i(\omega)} \in \mathcal{M}(\mathbb{R})$. Hence $N_n^Y = \frac{1}{n} \sum_{i=1}^n X_i, \; n \geq 1$, is a family of random variables taking values in $\mathcal{M}(\mathbb{R})$. For any n, ω note that $N_n^Y(\omega, (-\infty, y]) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, y]}(Y_i(\omega)) \triangleq F_n(y, \omega)$ for all $y \in \mathbb{R}$; so for fixed n, ω note that $F_n(\cdot, \omega)$ is the distribution function of the probability measure $N_n^Y(\omega, \cdot)$. By law of large numbers $F_n(y, \omega) \to F(y) \triangleq \mu((-\infty, y])$, for any $y \in \mathbb{R}$ as $n \to \infty$ for a.e. ω. That is, $N_n^Y(\omega, \cdot)$ converges, in the topology of $\mathcal{M}(\mathbb{R})$, to μ as $n \to \infty$, for a.e. ω. So questions concerning probabilities of rare events, like $P(N_n^Y \notin U)$ where U is neighbourhood of μ, become meaningful.

By analogy with Cramer's theorem the rate would involve the logarithmic moment generating function of X_i, and its convex conjugate. As X_i is $\mathcal{M}(\mathbb{R})$-valued random variable its logarithmic moment generating function is a function on $C_b(\mathbb{R})$ (= space of bounded continuous functions on $\mathbb{R}$) given by

$$\log M(g) = \log E[\exp\langle g, X_i\rangle] = \log E[\exp\langle g, \delta_{Y_i}\rangle]$$

$$= \log \int_{\mathbb{R}} \mathrm{e}^{g(y)}\, \mathrm{d}\mu(y), \; g \in C_b(\mathbb{R}). \tag{2.14}$$

We expect the rate function to be given by

$$I(\nu) = \sup\{\langle g, \nu\rangle - \log M(g) \colon g \in C_b(\mathbb{R})\}, \; \nu \in \mathcal{M}(\mathbb{R}). \tag{2.15}$$

(Here $\langle g, \nu\rangle = \int_{\mathbb{R}} g(y)\, \mathrm{d}\nu(y)$.) In fact a more tractable expression for the rate function can be given. Define for $\nu \in \mathcal{M}(\mathbb{R})$,

$$J(\nu|\mu) \triangleq \begin{cases} \int_{\mathbb{R}} f(y)[\log f(y)]\, \mathrm{d}\mu(y), & \text{if } f := \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \text{ exists} \\ +\infty, & \text{otherwise} \end{cases} \tag{2.16}$$

where $\mathrm{d}\nu/\mathrm{d}\mu$ is the Radon Nikodym derivative of ν with respect to μ. In the case of Bernoulli distribution with parameter $p \in (0, 1)$ it is easily seen that $I(a) = J(\nu|\mu), \; 0 < a < 1$ with $\mu =$ Bernoulli $(p), \; \nu =$ Bernoulli (a). $J(\nu|\mu)$ is referred to as the *relative entropy* of ν with respect to μ; it is also called *Kullback–Leibler information* in statistics.

The preceding heuristics suggest the following:

Theorem 2.4 (Sanov, 1957). *Let the empirical distributions* $\{N_n^Y\}$ *be given by* (2.13). *Then for any convex open set* $A \subset \mathcal{M}(\mathbb{R})$,

$$\lim_{n\to\infty} \frac{1}{n} \log P(N_n^Y \in A) = -\inf\{I(\nu) \colon \nu \in A\}, \tag{2.17}$$

where the rate function I *is given by* (2.15). *Moreover* $I(\nu) = J(\nu|\mu), \; \nu \in \mathcal{M}(\mathbb{R})$ *with* $J(\cdot|\mu)$ *given by* (2.16). □

For proof and extensions, see [DZ] and [DS].

Though the two results are equivalent, a difference between the set up of Cramer's theorem and that of Sanov's theorem is worth pointing out. The former deals with deviations away from a number (i.e. the mean) and the rate is a function on $\mathbb{R}$, whereas the latter concerns deviations away from a measure and hence the rate function is defined on the space of probability measures. As Sanov's theorem represents a greater conceptual sophistication, it is sometimes referred to as an example of *level 2 large deviations* (see [V3], [SW] and [DZ]). However, the importance of Sanov's work was perhaps fully realised only when the infinite dimensional version was proved by Donsker and Varadhan in the mid 70's; see [DS].

Another major development was initiated by Chernoff in 1952. He initiated a program in which questions about asymptotic efficiency of statistical tests and performance of estimators were analyzed using large deviations. Further developments/refinements of the results of Cramer, Sanov and Chernoff were made in the late 50's/early 60's. These include in particular the works of R R Bahadur, R Ranga Rao and J Sethuraman at the Indian Statistical Institute.

These and related works ensured the statistical pedigree of large deviations. For accounts of the above, see [B], [H], [DZ] and [DS]. Moreover these must have made an impression on Varadhan, who was a graduate student at the Indian Statistical Institute, Kolkata during 1959–62.

3. "Asymptotic probabilities and differential equations"

The title of this section is borrowed from that of the landmark 1966 paper of Varadhan [V1].

Perhaps the best known example of a convex conjugate pair is the Lagrangian and the Hamiltonian from mechanics via calculus of variations. In mechanics, the Lagrangian denotes the difference between the kinetic energy and the potential energy, while the Hamiltonian is the sum of the two.

Here we consider a simplified Hamiltonian. We assume $H\colon \mathbb{R} \to \mathbb{R}$ is a uniformly convex function; that is $H''(\cdot) \geq c > 0$. This function plays the role of the *Hamiltonian*. Let $T > 0$ be fixed. We consider the terminal value problem for the *Hamilton–Jacobi equation*

$$\left.\begin{aligned} U_t - H(U_x) &= 0, \text{ in } (0,T)\times\mathbb{R} \\ U(T,x) &= G(x),\ x \in \mathbb{R} \end{aligned}\right\}. \tag{3.1}$$

Here $G(\cdot)$ is a known continuous function, and U_t, U_x denote respectively the derivatives with respect to t, x. Let L denote the corresponding *Lagrangian*, that is

$$L(z) = \sup\{zx - H(x)\colon x \in \mathbb{R}\},\ z \in \mathbb{R} \tag{3.2}$$

is the convex conjugate of H. It is known from calculus of variations that the weak solution U of (3.1) is given by the 'variational principle'

$$U(t,x) = \sup\left\{G(w(T)) - \int_t^T L(\dot{w}(s))\,\mathrm{d}s\colon w(t) = x,\ w \text{ is } C^1\right\}, \tag{3.3}$$

where $\dot{w}(s) = \mathrm{d}w/\mathrm{d}s(s)$.

Remarks.

(i) In calculus of variations one considers the initial value problem for $U_t + H(U_x) = 0$. The quantity $\int_0^t L(\dot{w}(s))\,\mathrm{d}s$ is called an 'action functional'. The analogue of (3.3) is then an infimum, and hence is called the principle of least action. The reason for our considering the 'backward problem' (3.1) is that the expression (3.3) can be readily tied up with Varadhan's lemma later.

(ii) In optimal control theory, the modern avatar of calculus of variations, a cost functional is minimised/maximised as in (3.3), and a nonlinear PDE like (3.1) is derived via a dynamic programming principle.

(iii) PDE in (3.1) can also arise as a tool for solving initial/terminal value problems for certain scalar conservation laws of the form $u_t - (H(u))_x = 0$. In fact this served as the motivation for Varadhan [V1]. For example, if $H(x) = \frac{1}{2}x^2$ then the inviscid Burger's equation $u_t - uu_x = 0$ is transformed into an equation like (3.1) by taking certain indefinite integrals. See Varadhan [V1] for a brief discussion on this, and Evans [E] for a detailed account.

Let $D_u[0,T] = \{w\colon w$ right continuous on $[0,T]$ into $\mathbb{R}$, and $w(t-)$ exists for each $t\}$, with the topology of uniform convergence. Let $D_{ac} = \{w \in D_u[0,T]\colon w(t) = w(0) +$

$\int_0^t \xi(s)\,\mathrm{d}s,\ 0 \le t \le T$ and $\int_0^T |L(\xi(s))|\mathrm{d}s < \infty\}$. Define the function $I\colon D_u[0,T] \to [0,\infty]$ by

$$I(w) = \begin{cases} \int_0^T L(\dot{w}(s))\,\mathrm{d}s, & \text{if } w \in D_{ac} \\ +\infty, & \text{otherwise.} \end{cases} \tag{3.4}$$

Then it is not difficult to show that $I(\cdot)$ given by (3.4) has properties similar to the rate function of Cramer's theorem, albeit on a more complicated space.

An expression similar to r.h.s. of (3.3) crops up naturally in *Laplace's method* in classical asymptotic analysis. Assuming appropriate integrability conditions, and denoting $\|\cdot\|_k$ for the norm in $L^k(\mathbb{R})$, note that

$$\lim_{n\to\infty} \frac{1}{n} \log\left(\int_{\mathbb{R}} \mathrm{e}^{n\gamma(x)}\mathrm{d}x\right) = \lim_{n\to\infty} \log\|\mathrm{e}^{\gamma(\cdot)}\|_n = \log\|\mathrm{e}^{\gamma(\cdot)}\|_\infty$$
$$= \sup\{\gamma(x)\colon x \in \mathbb{R}\} \tag{3.5}$$

for any nice function $\gamma(\cdot)$ on $\mathbb{R}$. (Use of $\|\cdot\|_n \to \|\cdot\|_\infty$ in (3.5) was suggested by R Bhatia in place of an earlier argument.) In particular, if $\gamma(x) = g(x) - I(x)$ where g is a bounded continuous function and $I(\cdot) \ge 0$ is like a rate function then

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{\mathbb{R}} \mathrm{e}^{ng(x)}\, \mathrm{e}^{-nI(x)}\,\mathrm{d}x = \sup\ \{g(x) - I(x)\colon x \in \mathbb{R}\}. \tag{3.6}$$

Note the similarity between the right-hand sides of (3.3) and (3.6). In addition, for each n suppose $\mathrm{d}P_n(x) = \mathrm{e}^{-nI(x)}\mathrm{d}x$ is a probability measure. Then for large a, by similar analysis on $[a,\infty)$,

$$\lim_{n\to\infty} \frac{1}{n} \log P_n([a,\infty)) = \lim_{n\to\infty} \frac{1}{n} \log \int_a^\infty \mathrm{e}^{-nI(x)}\,\mathrm{d}x$$
$$= \sup\{-I(x)\colon x \ge a\} = -\inf\{I(x)\colon x \ge a\}. \tag{3.7}$$

Note the resemblance between (3.7) and (2.12). Clearly (3.5)–(3.7) suggest that there could be a close connection between suitable families of probability measures (like those encountered in Cramer's theorem), and approximation schemes for solving differential equations (like (3.1)).

See [E] for an application of Laplace's method in the asymptotics of viscous Burger's equation.

At this stage it is convenient to introduce Varadhan's unifying framework for large deviations. The idea is to characterize the limiting behaviour of a family $\{P_\epsilon\}$ of probability measures as $\epsilon \downarrow 0$, in terms of a rate function. Let (S,d) be a complete separable metric space, and $\mathcal{F}$ denote its Borel σ-algebra. The required abstraction is contained in the following two key definitions.

DEFINITION 3.1

A function $I\colon S \to [0,\infty]$ is called a *rate function* if $I \not\equiv \infty$ and if the level set $\{x \in S\colon I(x) \le c\}$ is compact in S for each $c < \infty$.

In particular, a rate function is lower semicontinuous; that is, $I^{-1}([0,c])$ is closed in S for all $c < \infty$, which is equivalent to $\liminf_{n\to\infty} I(x_n) \ge I(x)$ whenever $x_n \to x$ in S.

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 rate function* I if

(a) I is a rate function,
(b) for every closed set $C \subseteq S$,

$$\limsup_{\epsilon \to 0} \epsilon \log P_\epsilon(C) \leq - \inf_{y \in C} I(y), \tag{3.8}$$

(c) for every open set $A \subseteq S$,

$$\liminf_{\epsilon \to 0} \epsilon \ \log \ P_\epsilon(A) \geq - \inf_{y \in A} I(y). \tag{3.9}$$

Remarks.

(i) Let $\{X_i\}$ be as in Cramer's theorem and P_n denote the distribution of $\frac{1}{n}$ $(X_1 + \cdots + X_n)$, for $n \geq 1$. Then, with $\epsilon = n^{-1}$, Cramer's theorem says that $\{P_n\}$ satisfies LDP with rate function given by (2.9) (see (2.12)).

(ii) In Sanov's theorem $S = \mathcal{M}(\mathbb{R})$; it is a complete separable metric space in the topology of weak convergence; see [P]. Also $\epsilon = n^{-1}$, P_n = distribution of N_n^Y, $n \geq 1$. So Sanov's theorem says that $\{P_n\}$ satisfies LDP with relative entropy given by (2.16) as the rate function.

(iii) In the place of (3.8), (3.9) the more intuitive stipulation that

$$\lim_{\epsilon \to 0} \epsilon \log \ P_\epsilon(M) = - \inf_{y \in M} I(y)$$

turns out too strong to be useful. For example, if P_ϵ is nonatomic for all ϵ, then taking $M = \{x\}$, $x \in S$, the above can hold only if $I(\cdot) \equiv \infty$. This would rule out most of the interesting cases. It turns out that (3.8), (3.9) are enough to yield a rich theory.

(iv) The framework of complete separable metric space is known to be optimal for a rich theory of weak convergence; see [P]. In the case of large deviations too this seems to be so.

(v) We can now formulate Cramer's theorem in $\mathbb{R}^d$. Let $\{X_i\}$ be $\mathbb{R}^d$-valued i.i.d.'s with finite moment generating function M. Let $I(z) = \sup\{\langle\theta, z\rangle - \log M(\theta) : \theta \in \mathbb{R}^d\}$ for $z \in \mathbb{R}^d$. Let P_n denote the distribution of $\frac{1}{n}(X_1 + \cdots + X_n)$, $n = 1, 2, \ldots$. Then $\{P_n\}$ satisfies the LDP with rate function I (see [V2] or [DZ] for a proof).

The following elementary result gives a way of getting new families satisfying LDP's through continuous maps. This is also a main reason for not insisting that the rate function be convex.

Theorem 3.3 (Contraction principle). *Let $\{P_\epsilon\}$ satisfy the LDP with a rate function $I(\cdot)$. Let $(\hat{S}, \hat{d})$ be a complete separable metric space, and $\pi : S \to \hat{S}$ a continuous function. Put $\hat{P}_\epsilon = P_\epsilon \pi^{-1}$, $\epsilon > 0$. Then $\{\hat{P}_\epsilon\}$ also satisfies the LDP with the rate function*

$$\hat{I}(y) = \begin{cases} \inf\{I(x) : x \in \pi^{-1}(y)\}, & \textit{if } \pi^{-1}(y) \neq \phi \\ \infty, & \textit{otherwise}. \end{cases}$$

□

Note. In the above π can also depend on ϵ, with some additional assumptions (see [V2]).

Recall that $\{P_\epsilon\}$ *converges weakly* to P (denoted $P_\epsilon \Rightarrow P$) if

$$\lim_{\epsilon\to 0} \int_S f(x)\, \mathrm{d}P_\epsilon(x) = \int_S f(x)\, \mathrm{d}P(x)$$

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

$$\limsup_{\epsilon\to 0}\ P_\epsilon(C) \le P(C),\ C \subseteq S,\ C \text{ closed},$$

$$\liminf_{\epsilon\to 0}\ P_\epsilon(A) \ge P(A),\ A \subseteq S,\ A \text{ open}.$$

The formal similarity between (3.8), (3.9) and the above suggests that LDP may be suitable for handling convergence of integrals of exponential functionals. Indeed we have the following fundamental result, which is the key to diverse applications.

Theorem 3.4 (Varadhan's lemma (1966)). *Let $\{P_\epsilon\}$ satisfy the LDP with a rate function $I(\cdot)$. Then for any bounded continuous function g on S,*

$$\lim_{\epsilon\to 0}\ \epsilon \log\left[\int_S \exp\left(\frac{1}{\epsilon} g(x)\right)\, \mathrm{d}P_\epsilon(x)\right] = \sup\{g(x) - I(x) : x \in S\}. \tag{3.10}$$

$\square$

Thus Varadhan's lemma is an extension of Laplace's method to an abstract setting (see [V1], [V2] and [DZ] for proof and discussion). Moreover the factor $\exp\left(\frac{1}{\epsilon} g(\cdot)\right)$ in (3.10) is reminiscent of the Esscher tilt. This strategy highlights the contribution of 'rare events', that is, sets with very small P_ϵ-measure where $g(\cdot)$ may take large values.

Example 3.5. We now present a 'toy example', taken from den Hollander [H], to indicate that probabilities of rare events can decisively influence asymptotic expectations. Let $\{X_i\}$ be i.i.d.'s such that $P\left(X_i = \frac{1}{2}\right) = P\left(X_i = \frac{3}{2}\right) = \frac{1}{2}$. Let P_n denote the distribution of $\frac{1}{n}(X_1 + \cdots + X_n),\ n \ge 1$. By Cramer's theorem $\{P_n\}$ satisfies the LDP with rate function

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

Now

$$E\left(\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^n\right) = \int_{\left[\frac{1}{2},\frac{3}{2}\right]} \exp(n \log x)\, \mathrm{d}P_n(x) = \int_{\mathbb{R}} \exp(n\ g(x))\, \mathrm{d}P_n(x),$$

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

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \log E\left[\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^n\right] &= \sup\left\{\log x - I(x) : x \in \left[\frac{1}{2}, \frac{3}{2}\right]\right\} \\ &\triangleq b,\ \text{say}. \end{aligned} \tag{3.11}$$

It can be shown easily that $b > 0$. By the law large numbers $\frac{1}{n}(X_1 + \cdots + X_n) \to 1$ with probability 1. So one might naively expect l.h.s. of (3.11) to be zero. However, as shown above it is not so. Thus the asymptotic expectation is determined not by the typical (or almost sure) behaviour but by the rare event when $\frac{1}{n}\sum_{i=1}^{n} X_i$ takes values near x^*, the value where supremum is attained in (3.11). □

Exit problem, discussed in the next section, gives a more concrete example where rare event determines the quantity/characteristic of interest.

Now, as $I(\cdot)$ given by (3.4) is a rate function, the similarity between the r.h.s. of (3.3) and (3.10) is quite striking. In fact, if we can have a family $\{Q_n\}$ of probability measures on $D_u[0, T]$ satisfying LDP with rate function given by (3.4), then (3.10) gives an approximation scheme for the solution to (3.1).

Suppose $H(\theta) = c \log M_F(\theta)$, $\theta \in \mathbb{R}$ where M_F is the moment generating function (Laplace transform) of a probability distribution F on $\mathbb{R}$. Let $\{X_i\}$ be an i.i.d. sequence with distribution F, and $S_n = X_1 + \cdots + X_n$, $n \geq 1$. For $n = 1, 2, \ldots$ define the stochastic process $Z_n(t) = \frac{1}{n} S_{[nt]}$, $0 \leq t \leq T$; (here $[z]$ denotes the integer part of z). Then the sample paths (or trajectories) of $Z_n(\cdot)$ are in $D_u[0, T]$. Let Q_n denote the probability measure induced by the process Z_n on $D_u[0, T]$; (Q_n may be called the *distribution of the process* $Z_n(\cdot)$). It can be proved that $\{Q_n\}$ satisfies LDP with rate function given by (3.4); this is basically a functional version of Cramer's theorem, proved in Varadhan [V1].

If the Hamiltonian is not a logarithmic moment generating function then the approximation scheme, though similar in spirit, is more involved. But once again, it uses processes with independent increments. Hamilton–Jacobi equations (of the type (3.1)) with non-zero right side can also be handled (see [V1]).

Even at the pain of repetition, it may be worth mentioning the following. Thanks to the work of Hopf, Lax and Oleinik, it was shown only in the late 50's/early 60's that U given by (3.3) is the weak solution, in a suitable sense, to (3.1). In more modern jargon (3.3) gives the viscosity solution to (3.1) (see [E] for a detailed discussion on this circle of ideas). Varadhan [V1] has given an approximation scheme for (3.3) in terms of probabilistic objects. On the way, a unifying framework for large deviations has been synthesized, with Varadhan's lemma set to play a key role.

4. Sample path LDP, Wentzell–Freidlin theory and exit problem

We begin with LDP in connection with the best known stochastic process, viz. the Brownian motion.

Let $C_0[0, T] = \{w\colon w \text{ continuous on } [0, T] \text{ into } \mathbb{R}, \text{ and } w(0) = 0\}$. Let $X(t, w) = w(t)$, $0 \leq t \leq T$, $w \in C_0[0, T]$ denote the coordinate projections. Let P denote the standard *one-dimensional Wiener measure* on $C_0[0, T]$. So $P(C_0[0, T]) = 1$, and under P, for $0 < t_1 < t_2 < \cdots < t_k < T$ the random variables $X(t_1), X(t_2) - X(t_1), \ldots, X(t_k) - X(t_{k-1})$ are independent having respectively $N(0, t_1)$, $N(0, t_2 - t_1), \ldots, N(0, t_k - t_{k-1})$ distributions. That is, under P the stochastic process $\{X(t)\colon 0 \leq t \leq T\}$ is a *standard one-dimensional Brownian motion*.

For $\epsilon > 0$, let P_ϵ denote the distribution of the process $\{\sqrt{\epsilon}X(t)\colon t \geq 0\}$; so for any Borel set $A \subseteq C_0[0, T]$ note that $P_\epsilon(A) = P\left(\frac{1}{\sqrt{\epsilon}}A\right)$. Clearly $P_\epsilon \Rightarrow \delta_0$ as $\epsilon \downarrow 0$ where δ_0 is the probability measure concentrated on the function which is identically 0.

We look at an example, some aspects of which had been alluded to earlier, to justify why an LDP for $\{P_\epsilon\}$ may be useful.

Example 4.1. Let $T > 0$ and g be a continuous function on $\mathbb{R}$. Consider the terminal value problem for the *viscous Burger's equation*:

$$\left.\begin{aligned} &u_t^\epsilon - u^\epsilon u_x^\epsilon + \tfrac{1}{2}\epsilon\, u_{xx}^\epsilon = 0, \text{ in } (0,T)\times\mathbb{R} \\ &u^\epsilon(T,x) \qquad\qquad = g(x),\ x\in\mathbb{R} \end{aligned}\right\}. \tag{4.1}$$

Here $\epsilon > 0$ is a parameter. As $\epsilon \downarrow 0$ we expect u^ϵ to converge to the solution to the equation

$$u_t - \left(\frac{1}{2}u^2\right)_x = 0, \text{ in } (0,T)\times\mathbb{R} \tag{4.2}$$

with terminal value $u(T,x) = g(x),\ x\in\mathbb{R}$.

The conservation law (4.2) is also called *inviscid Burger's equation*. If we denote by $U^\epsilon,\ G$ the indefinite integrals (with respect to the space variable) of $u^\epsilon,\ g$ respectively, then

$$U_t^\epsilon - \frac{1}{2}(U_x^\epsilon)^2 + \frac{1}{2}\epsilon\, U_{xx}^\epsilon = 0, \text{ in } (0,T)\times\mathbb{R} \tag{4.3}$$

with terminal value $U^\epsilon(T,x) = G(x),\ x\in\mathbb{R}$. If U^ϵ solving (4.3) can be obtained then, $u^\epsilon = U_x^\epsilon$ solves (4.1). The nonlinear equation (4.3) can be transformed into the heat equation by the *Cole–Hopf transformation* $V^\epsilon(t,x) = \exp(-\frac{1}{\epsilon}U^\epsilon(t,x))$. Then

$$\left.\begin{aligned} &V_t^\epsilon + \tfrac{\epsilon}{2}V_{xx}^\epsilon = 0, \text{ in } (0,T)\times\mathbb{R} \\ &V^\epsilon(T,x) \quad = \exp\left(-\tfrac{1}{\epsilon}G(x)\right) \end{aligned}\right\}. \tag{4.4}$$

Once heat equation is encountered, can probability be far behind?

It is known that the solution to (4.4) can be written in terms of the heat kernel or equivalently the Brownian motion; see [KS]. Indeed

$$\begin{aligned} V^\epsilon(t,x) &= \int_{\mathbb{R}} \left[\exp\left(-\frac{1}{\epsilon}G(y)\right)\right]\cdot \frac{1}{\sqrt{2\pi\epsilon(T-t)}}\exp\left\{-\frac{(y-x)^2}{2\epsilon(T-t)}\right\}\mathrm{d}y \\ &= \int_{\mathbb{R}} \left[\exp\left(-\frac{1}{\epsilon}G(x+z)\right)\right]\cdot \frac{1}{\sqrt{2\pi\epsilon(T-t)}}\exp\left\{-\frac{z^2}{2\epsilon(T-t)}\right\}\mathrm{d}z \\ &= \int_{\mathbb{R}} \left[\exp\left(-\frac{1}{\epsilon}G(x+z)\right)\right]\cdot \mathrm{d}P_\epsilon X_{T-t}^{-1}(z) \\ &= \int_{C_0[0,T]} \left[\exp\left(-\frac{1}{\epsilon}G(x+X(T-t,w))\right)\right]\mathrm{d}P_\epsilon(w). \end{aligned} \tag{4.5}$$

Inverting the Cole–Hopf transformation we see that

$$\begin{aligned} -U^\epsilon(t,x) &= \epsilon\log V^\epsilon(t,x) \\ &= \epsilon\log\left[\int_{C_0[0,T]}\exp\left\{\frac{1}{\epsilon}[-G(x+X(T-t,w))]\right\}\mathrm{d}P_\epsilon(w)\right]. \end{aligned} \tag{4.6}$$

Clearly r.h.s. of (4.6) suggests that limit of U^ϵ as $\epsilon \downarrow 0$ can be handled using Varadhan's lemma, once it is shown that $\{P_\epsilon : \epsilon > 0\}$ satisfies the LDP and the rate function is identified.

Also as $\epsilon \downarrow 0$ we expect U^ϵ to converge to the solution of the Hamilton–Jacobi equation

$$U_t - \frac{1}{2}(U_x)^2 = 0, \text{ in } (0, T) \times \mathbb{R} \tag{4.7}$$

with the terminal value $U(T, x) = G(x),\ x \in \mathbb{R}$. From this, solution to (4.2) can be obtained by differentiating with respect to x. Here the Hamiltonian is $H(y) = \frac{1}{2}y^2$ and hence the Lagrangian is $L(z) = \frac{1}{2}z^2$. Note that the approximation scheme suggested here is somewhat different from the one discussed in the preceding section. This problem was considered by Donsker and his student Schilder at the Courant Institute around 1965, serving as another motivation for [V1]. □

Now define I_B: $C_0[0, T] \to [0, \infty]$ by

$$I_B(w) = \begin{cases} \frac{1}{2}\int_0^T |\dot{w}(s)|^2 \mathrm{d}s, & \text{if } w \in D_{ac} \cap C_0[0, T] \\ \infty, & \text{otherwise} \end{cases} \tag{4.8}$$

where D_{ac} is as in (3.4) with $L(x) = \frac{1}{2}x^2$; so I_B is the restriction of I given by (3.4) to $C_0[0, T]$ with $L(x) = \frac{1}{2}x^2$.

Theorem 4.2 (Schilder 1966). $\{P_\epsilon : \epsilon > 0\}$ *satisfies the LDP with rate function* I_B *given by* (4.8). *An analogous result also holds for the* d*-dimensional Brownian motion.* □

An important ingredient of the proof is the Cameron–Martin formula which gives the Radon–Nikodym derivative of translation by an absolutely continuous function with respect to the Wiener measure (see Varadhan [V2] for a proof). In view of Example 2.2 and Cramer's theorem the rate function I_B may not be surprising. Theorem 4.2 is an example of a *sample path large deviations principle*. This is a level 1 LDP like Cramer's theorem.

A far reaching generalization of the above is the LDP for diffusion processes, again a sample path LDP, due to Wentzell and Friedlin (1970); some special cases had been considered earlier by Varadhan. A *diffusion process* can be represented as a solution to a stochastic differential equation. Let $\{X(t): t \geq 0\}$ denote a standard d-dimensional Brownian motion, where $d \geq 1$ is an integer. 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

$$\mathrm{d}Z(t) = \sigma(Z(t))\, \mathrm{d}X(t) + b(Z(t))\, \mathrm{d}t \tag{4.9}$$

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

$$Z(t) = z_0 + \int_0^t \sigma(Z(s))\, \mathrm{d}X(s) + \int_0^t b(Z(s))\, \mathrm{d}s. \tag{4.10}$$

Here expressions of the form $\int_0^t \xi(s)\, \mathrm{d}X(s)$ denote Ito integrals. When $\sigma,\ b$ are Lipschitz continuous, by a Picard iteration unique solution to (4.10) can be obtained. The diffusion process given by (4.10) is a *Markov process*; that is, if the 'present' is known, then the 'past' and the 'future' of the process are independent; this is also called the memoryless property (see [KS]). The sample paths of the diffusion are continuous; so the process $\{Z(t): 0 \leq t \leq T\}$ induces a probability measure on $C([0, T]: \mathbb{R}^d)$.

Theorem 4.3 (Wentzell–Freidlin, 1970). *Let σ, b be Lipschtz continuous. Assume that $a(\cdot) \triangleq \sigma(\cdot)\,\sigma(\cdot)^{\dagger}$ is uniformly positive definite; that is, $\exists\ \lambda_0 > 0$ such that $\langle a(x)\xi, \xi\rangle \geq \lambda_0|\xi|^2$, for all $x, \xi \in \mathbb{R}^d$. Let $x \in \mathbb{R}^d$ be fixed. For $\epsilon > 0$ consider the diffusion process*

$$\mathrm{d}Z^{\epsilon,x}(t) = \sqrt{\epsilon}\,\sigma(Z^{\epsilon,x}(t))\,\mathrm{d}X(t) + b(Z^{\epsilon,x}(t))\,\mathrm{d}t \tag{4.11}$$

with initial value $Z^{\epsilon,x}(0) = x$. Let $Q_{\epsilon,x}$ denote the probability measure induced on $C([0,T]\colon \mathbb{R}^d)$ by the process $\{Z^{\epsilon,x}(t)\colon 0 \leq t \leq T\}$. Then $\{Q_{\epsilon,x}\colon \epsilon > 0\}$ satisfies LDP with the rate function

$$I_x(w) = \begin{cases} \frac{1}{2}\int_0^T \langle \dot{w}(t) - b(w(t)),\ a^{-1}(w(t))\,(\dot{w}(t) - b(w(t)))\rangle \mathrm{d}t, & \text{if } w \in D^x \\ \infty, & \text{otherwise} \end{cases} \tag{4.12}$$

where

$$D^x = \left\{ w \in C([0,T]\colon \mathbb{R}^d)\colon w(t) = x + \int_0^t \xi(s)\,\mathrm{d}s,\ \xi \in L^2[0,T] \right\}.$$

□

Remark. If $\sigma(\cdot) \equiv$ identity matrix, $b(\cdot) \equiv 0$ then the above reduces to Schilder's theorem. In fact, if $\sigma(\cdot) \equiv$ constant, then the above result is a simple consequence of Schilder's theorem and the contraction principle. So the expression (4.12) may not be surprising; however the proof in the general case involves a delicate approximation (see [FW] and [V2]).

We next indicate a connection between diffusions and second order elliptic/parabolic PDE's. With σ, a, b as in Theorem 4.3 define the elliptic differential operator L by

$$Lg(x) = \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x)\,\frac{\partial^2 g(x)}{\partial x_i \partial x_j} + \sum_{i=1}^{d} b_i(x)\,\frac{\partial g(x)}{\partial x_i}, \tag{4.13}$$

where $a(\cdot) = ((a_{ij}(\cdot)))$. The operator L is called the *infinitesimal generator* of the diffusion process $Z(\cdot)$ given by (4.9) and (4.10). The probabilistic behaviour of the diffusion is characterized by L . In particular, the transition probability density function of $Z(\cdot)$ is the fundamental solution to the parabolic operator $\frac{\partial}{\partial t} + L$. (For example, the generator corresponding to Brownian motion is the d-dimensional Laplacian $\frac{1}{2}\Delta := \frac{1}{2}\sum_{i=1}^{d}\frac{\partial^2}{\partial x_i^2}$, and the heat kernel is the corresponding transition probability density function.) (see [KS]).

Let $G \subset \mathbb{R}^d$ be a bounded smooth domain. Consider the (Dirichlet) boundary value problem

$$Lu(x) = -g(x),\ x \in G \text{ with } u(x) = f(x),\ x \in \partial G \tag{4.14}$$

where g, f are known functions. Then the unique solution to (4.14) can be written as

$$u(x) = E\left[f(Z(\tau)) + \int_0^{\tau} g(Z(s))\,\mathrm{d}s \,\middle|\, Z(0) = x \right],\ x \in \bar{G} \tag{4.15}$$

where Z is the diffusion given by (4.9), (4.10), and $\tau = \inf\{t > 0\colon Z(t) \notin G\} =$ first exit time from G. Note that r.h.s. of (4.15) denotes taking expectation given that $Z(0) = x$.

The random variable τ is an example of a *stopping time*. As $a(\cdot)$ is uniformly positive definite, we have $\tau < \infty$ with probability 1. Thus (4.15) gives a stochastic representation to the solution to the Dirichlet problem (4.14); this can be proved using stochastic calculus. (see [KS]).

For $\epsilon > 0,\ x \in \mathbb{R}^d$ let $Z^{\epsilon,x}(\cdot)$ be given by (4.11). The infinitesimal generator L_ϵ of the diffusion $Z^{\epsilon,x}$ is given by

$$L_\epsilon v(x) = \frac{\epsilon}{2} \sum_{i,j=1}^{d} a_{ij}(x) \frac{\partial^2 v}{\partial x_i \partial x_j}(x) + \sum_{i=1}^{d} b_i(x) \frac{\partial v}{\partial x_i}(x). \tag{4.16}$$

For each $\epsilon > 0$, with the same $g,\ f$, the problem

$$L_\epsilon u_\epsilon(x) = -g(x),\ x \in G \ \text{ with } u_\epsilon(x) = f(x),\ x \in \partial G \tag{4.17}$$

has the solution

$$u_\epsilon(x) = E\left[f(Z^{\epsilon,x}(\tau_\epsilon)) + \int_0^{\tau_\epsilon} g(Z^{\epsilon,x}(s))\, \mathrm{d}s \right],\ x \in \bar{G} \tag{4.18}$$

where $\tau_\epsilon = \inf\{t > 0\colon Z^{\epsilon,x}(t) \notin G\}$. In particular, $x \mapsto E_x(\tau_\epsilon) := E(\tau_\epsilon \mid Z^{\epsilon,x}(0) = x)$ is the solution to (4.17) with $f \equiv 0,\ g \equiv 1$.

For $\epsilon = 0$ note that the equation (4.11) becomes the ODE

$$\mathrm{d}z(t) = b(z(t))\, \mathrm{d}t,\ \text{ with } z(0) = x. \tag{4.19}$$

We make the following assumption:

(A) There exists $x_0 \in G$ (an interior point) such that for any $x \in \bar{G}$ the solution $z(\cdot)$ to (4.19) with initial value $z(0) = x$ satisfies $z(t) \in G$ for all $t > 0$ and $\lim_{t\to\infty} z(t) = x_0$; that is, x_0 is the *unique stable equilibrium point* in $\bar{G}$ of the ODE (4.19).

Some questions of interest are: What happens to u_ϵ as $\epsilon \downarrow 0$? In particular, what about $E_x(\tau_\epsilon)$ as $\epsilon \downarrow 0$? What can one say about the hitting distribution on ∂G in the limit?

For small ϵ, the trajectories of the diffusion $Z^{\epsilon,x}(\cdot)$ are close to the deterministic trajectory $z(\cdot)$ with very high probability. And, as the deterministic trajectory $z(\cdot)$ does not exit G at all, a reasonable guess would be that the system $Z^{\epsilon,x}$ tends to stay inside G for small ϵ. In such an eventuality note that limiting exit time and exit place are not defined.

To get a handle on the problem, we proceed differently. By continuity of sample paths, $Z^{\epsilon,x}(\tau_\epsilon)$ is ∂G-valued. So for any $\epsilon > 0$, the hitting distribution, i.e. the distribution of $Z^{\epsilon,x}(\tau_\epsilon)$ is a probability measure on ∂G. Since ∂G is compact this family of probability measures has limit points.

To appreciate the importance of the problem let us look at two situations. The first example is from chemistry, which is the origin of the 'exit problem'. It is known that molecules need to overcome a potential barrier to be able to participate in a chemical reaction. As the molecules are in motion, their energy is modelled by a diffusion of the type $Z^{\epsilon,x}(\cdot)$, oscillating about a stable state; here $\epsilon > 0$ is the so-called Arrhenius factor. The potential barrier θ is represented by the diameter of the domain G. In general, $\epsilon \ll \theta$. So exit from the 'right end' of G for small ϵ means reaction will proceed. Hence the asymptotic rate of exit at the right end of the potential well, as $\epsilon \downarrow 0$, gives a very good estimate of reaction rate (see [Kp], [Sc] for more background information and ad hoc ϵ-expansion method due to Kramers).

The second example is from engineering, concerning track loss in radar systems. In such a system the observed tracing error, due to evasive maneuvres of the target as well as to observation noise, is modelled by a diffusion of the type $Z^{\epsilon,x}(\cdot)$. Here ϵ gives the variance parameter in the observation noise. As radar systems are quite sophisticated this parameter is very small compared to the actual tracing error. Since the observation device has a limited field of view, $Z^{\epsilon,x}(\cdot)$ ceases to model the observation process as soon as the tracking error exits from the field of view. So exiting the domain in this case is an undesirable event. Hence information on probability of exit, mean time of exit, exiting place on ∂G, etc. may be useful in designing optimal devices (see [DZ] for a detailed discussion).

Motivated by the rate function in Theorem 4.3, for $0 < t < \infty$ define

$$I_t(y(\cdot)) = \frac{1}{2}\int_0^t \langle(\dot{y}(s) - b(y(s))),\ a^{-1}(y(s))(\dot{y}(s) - b(y(s)))\rangle\, \mathrm{d}s$$

if y is absolutely continuous with square integrable derivative $\dot{y}$. Set

$$\varphi_t(x, y) = \inf\{I_t(y(\cdot)) \colon y(0) = x,\ y(t) = y,\ y \text{ absolutely continuous},\ \dot{y} \text{ square integrable}\},$$

$$\varphi(x, y) = \inf_{t>0} \varphi_t(x, y),\ x, y \in \mathbb{R}^d.$$

Heuristically $\varphi_t(x, y)$ can be interpreted as the cost of forcing the diffusion $Z^{\epsilon,x}(\cdot)$ to be at the point y at time t. Define the function $\bar{\varphi}$ by

$$\bar{\varphi}(y) = \varphi(x_0, y),\ y \in \bar{G} \tag{4.20}$$

where x_0 is the unique stable equilibrium point of the ODE (4.19) as in (A).

Theorem 4.4 (Wentzell–Freidlin, 1970). *Let σ, b, a be as in Theorem* 4.3. *Assume* (*A*). *Further assume that there exists $\bar{y} \in \partial G$ such that $\bar{\varphi}(\bar{y}) < \bar{\varphi}(y)$ for $y \in \partial G$, $y \neq \bar{y}$. Then the following hold.*

(i) *For any $x \in G$, $Z^{\epsilon,x}(\tau_\epsilon) \to \bar{y}$ with probability* 1. *So the hitting distribution converges* (*weakly*) *to $\delta_{\bar{y}}$ as $\epsilon \downarrow 0$, whatever be the starting point.*

(ii) *Let u_ϵ be the solution to* (4.17) *for continuous boundary data f, and $g \equiv 0$. Then* $\lim_{\epsilon\downarrow 0} u_\epsilon(x) = f(\bar{y})$ *for any $x \in G$.* (*Part* (*ii*) *is an immediate consequence of part* (*i*).)

(iii) $\lim_{\epsilon\downarrow 0} \epsilon \log E_x(\tau_\epsilon) = \bar{\varphi}(\bar{y})$, *for any $x \in G$.* □

The intuitive explanation is along the following lines. "Any wandering away from x_0 has an overwhelmingly high probability of being pushed back to x_0, and it is not the time spent near any part of ∂G that matters but the *a priori* chance for a direct, fast exit due to a rare segment in the Brownian motion's path" – see p. 198 of [DZ]. This is possible since $p_0(t, x, y) > 0$ for any $t > 0$ (however small), any $x, y \in \mathbb{R}^d$ (however large $|x - y|$ may be), where p_0 is the heat kernel. For proof and discussion, see [V2], [FW] and [DZ].

An interesting application of sample path LDP arises in queueing networks. Identification of the rate function in a general setting is not very easy. But once the rate function is found and is regular, it can be used to characterize decoupling of data sources and to define 'effective bandwidth' of each source; this is of importance to traffic engineering in communication networks. Moreover the most likely way in which buffers overflow can be determined by considering 'minimizing large deviation paths' for certain reflected Brownian motion processes that arise as heavy traffic models for queueing networks (see [SW], [AD], [RD], [DR] and references therein).

5. LDP for occupation times: Prelude to Donsker–Varadhan theory

We next look at large deviation methods in connection with principal eigenvalues.

Let $V(\cdot)$ be a continuous periodic function on $\mathbb{R}$ with period 2π, and consider the problem

$$\frac{\partial u}{\partial t}(t,x) = \frac{1}{2}\,\frac{\partial^2 u}{\partial x^2}(t,x) + V(x)\,u(t,x),\ t > 0,\ x \in \mathbb{R}$$

with the initial value $u(0,x) = 1$. By Feynman–Kac formula the solution is given by

$$\begin{aligned} u(t,x) &= E_x\left[\exp\left\{\int_0^t V(X(s))\,\mathrm{d}s\right\}\right] \\ &\triangleq E\left[\exp\left\{\int_0^t V(X(s))\,\mathrm{d}s\right\}\middle| X(0) = x\right], \end{aligned}$$

where $X(\cdot)$ denotes one-dimensional Brownian motion; this can be proved using stochastic calculus; see [KS].

Since V and the initial value are periodic, $x \mapsto u(t,x)$ is also periodic. Note that $Y(t) \triangleq X(t) \bmod 2\pi,\ t \geq 0$ is the Brownian motion on the 1-dimensional torus (circle) $\mathbb{T}$. So the problem as well as the solution can be considered on $\mathbb{T}$ rather than on $\mathbb{R}$. In other words, the problem is basically

$$\begin{aligned} &\frac{\partial u}{\partial t}(t,\theta) = Au(t,\theta) \triangleq \left(\frac{1}{2}\frac{\partial^2}{\partial \theta^2} + V(\theta)\right) u(t,\theta),\ t > 0,\ \theta \in \mathbb{T} \\ &u(0,\theta) = 1,\ \theta \in \mathbb{T} \end{aligned} \tag{5.1}$$

and the solution, by Feyman–Kac formula, is

$$u(t,\theta) = E\left[\exp\left\{\int_0^t V(Y(s))\mathrm{d}s\right\}\middle| Y(0) = \theta\right],\ t \geq 0,\ \theta \in \mathbb{T}. \tag{5.2}$$

The one-dimensional Schrödinger operator $A \triangleq \frac{1}{2}\frac{\partial^2}{\partial\theta^2} + V(\theta)$ is an unbounded operator with domain $(A) \subset L^2(\mathbb{T})$. It is known from the theory of second-order elliptic differential equations that A^{-1} is a bounded self-adjoint compact operator. So by spectral theory A has a sequence $\{\lambda_i\}$ of eigenvalues, and a corresponding sequence $\{\psi_i(\cdot)\}$ of eigenfunctions such that $\lim_{m\to\infty} \lambda_m = -\infty,\ \lambda_1 > \lambda_2 \geq \lambda_3 \geq \ldots$, the principal eigenvalue λ_1 is of multiplicity one and the corresponding eigenfunction $\psi_1(\cdot) > 0$ (see [E] or [K]). The semigroup $\{T_t\}$ corresponding to (5.1) can be formally written as $\{\mathrm{e}^{tA}\}$ and hence by spectral theory again

$$u(t,\theta) = (\mathrm{e}^{tA}\ 1)(\theta) = \sum_{k=1}^{\infty} \mathrm{e}^{\lambda_k t}\langle\psi_k, 1\rangle\psi_k(\theta), \tag{5.3}$$

where 1 denotes the function which is identically 1 on $\mathbb{T}$, and $\langle\cdot,\cdot\rangle$ denotes the inner product in $L^2(\mathbb{T})$. As $\lambda_1 > \lambda_i,\ i \geq 2$ and $\psi_1 > 0$, from (5.3) we have

$$u(t,\theta) = \mathrm{e}^{\lambda_1 t}\langle\psi_1, 1\rangle\psi_1(\theta)\ [1 + O(\mathrm{e}^{-(\lambda_1-\lambda_2)t})]$$

and consequently

$$\lim_{t\to\infty} \frac{1}{t} \log u(t,\theta) = \lambda_1. \tag{5.4}$$

This is a result due to Kac.

Now the bilinear form associated with A is

$$\begin{aligned} B[f,g] = \langle Af, g\rangle &= \int_{\mathbb{T}} \frac{1}{2} f''(\theta)\ g(\theta)\mathrm{d}\theta + \int_{\mathbb{T}} V(\theta)\ f(\theta)\ g(\theta)\mathrm{d}\theta \\ &= -\int_{\mathbb{T}} \frac{1}{2} f'(\theta)\ g'(\theta)\mathrm{d}\theta + \int_{\mathbb{T}} V(\theta)\ f(\theta)\ g(\theta)\ \mathrm{d}\theta, \end{aligned} \tag{5.5}$$

where in the last step we have used integration by parts and periodicity. It is known by the classical Rayleigh–Ritz variational formula (see [E] or [K]) that the principal eigenvalue λ_1 can be given, in view of (5.5) by

$$\begin{aligned} \lambda_1 &= \sup\{B[g,g]\colon g \text{ differentiable},\ \|g\|_{L^2} = 1\} \\ &= \sup\left\{\int_{\mathbb{T}} V(\theta)\ g^2(\theta)\ \mathrm{d}\theta - \frac{1}{2}\int_{\mathbb{T}} |g'(\theta)|^2 \mathrm{d}\theta\colon g \text{ differentiable},\ \|g\|_{L^2} = 1\right\}. \end{aligned} \tag{5.6}$$

Similar analysis is possible also on $\mathbb{R}$ if $\lim_{x\to\pm\infty} V(x) = -\infty$. The above discussion basically means that the Perron–Frobenius theorem for nonnegative irreducible matrices goes over to self-adjoint second-order elliptic operators.

A natural question, whose implications turn out to be far reaching is: Is there a direct way of getting (5.6) from (5.2) without passing through differential equation (5.1) or the interpretation of the limit in (5.4) as an eigenvalue? If it is possible to do so, then one can replace $\int_0^t V(Y(s))\mathrm{d}s$ by more general functionals of the form $F(Y(t))$ depending on Brownian paths and hope to calculate $\lim_{t\to\infty} \frac{1}{t} \log E[\exp(F(Y(t)))]$. In such a case there may be no connection with differential equations. Moreover one can also consider processes other than Brownian motion. Donsker's firm conviction that something deep was going on here propelled the investigation along these lines.

Put $f(\theta) = g^2(\theta)$. Then what we seek can be written as

$$\begin{aligned} &\lim_{t\to\infty} \frac{1}{t} \log E\left[\exp\left\{t\left\{\frac{1}{t}\int_0^t V(Y(s))\ \mathrm{d}s\right\}\right\}\middle| Y(0) = y\right] \\ &= \sup\left\{\int_{\mathbb{T}} V(\theta)\ f(\theta)\ \mathrm{d}\theta - \frac{1}{8}\int_{\mathbb{T}} \frac{1}{f(\theta)}|f'(\theta)|^2 \mathrm{d}\theta\colon \|f\|_{L^1} = 1,\ f \geq 0\right\} \end{aligned} \tag{5.7}$$

for any $y \in \mathbb{T}$. If (5.7) can be considered as a special case of (3.10) then our purpose would be served by Varadhan's lemma. Also (5.7) implies that the factor $\exp(\int_0^t V(Y(s))\mathrm{d}s)$ in the Feyman–Kac formula (5.2) can be viewed upon as an Esscher tilt.

Towards this, let $\Omega = \{w\colon [0,\infty) \to \mathbb{T}\colon w \text{ continuous}\}$; this can be taken as the basic probability space. Define $Y(t,w) = w(t),\ t \geq 0,\ w \in \Omega$. For $y \in \mathbb{T}$, let P_y denote the probability measure on Ω making $\{Y(t)\colon t \geq 0\}$ a Brownian motion on $\mathbb{T}$ starting at y; that is, P_y is the distribution induced on Ω by the Brownian motion on the torus starting at y. For $t \geq 0,\ w \in \Omega,\ A \subseteq \mathbb{T}$ define

$$M(t, w, A) = \frac{1}{t} \int_0^t I_A(Y(s, w))\, \mathrm{d}s, \tag{5.8}$$

denoting the proportion of time that the trajectory $Y(\cdot, w)$ spends in the set A during the period $[0, t]$. This is called the *occupation time*. Note that $A \mapsto M(t, w, A)$ is a probability measure on the torus. Let $\mathcal{M}(\mathbb{T})$ denote the space of probability measures on $\mathbb{T}$, endowed with the topology of weak convergence of probability measures. For $t \geq 0$ fixed, let M_t denote the map $w \mapsto M(t, w, \cdot) \in \mathcal{M}(\mathbb{T})$. Let $Q_t^{(y)} \triangleq P_y M_t^{-1}$ denote the distribution of M_t. So $Q_t^{(y)}$ is a probability measure on $\mathcal{M}(\mathbb{T})$; in other words, $Q_t^{(y)} \in \mathcal{M}(\mathcal{M}(\mathbb{T}))$, for any $t \geq 0,\ y \in \mathbb{T}$.

Now observe that

$$\frac{1}{t} \int_0^t V(Y(s, w))\, \mathrm{d}s = \int_{\mathbb{T}} V(\theta)\, M(t, w, \mathrm{d}\theta) \tag{5.9}$$

and consequently

$$\begin{aligned} &E\left[\exp\left\{t\left\{\frac{1}{t}\int_0^t V(Y(s))\,\mathrm{d}s\right\}\right\}\middle| Y(0) = y\right] \\ &\quad = \int_\Omega \left[\exp\left\{t \int_{\mathbb{T}} V(\theta)\, M(t, w, \mathrm{d}\theta)\right\}\right] \mathrm{d}P_y(w) \\ &\quad = \int_{\mathcal{M}(\mathbb{T})} \left[\exp\left\{t\int_{\mathbb{T}} V(\theta)\, \mu(\mathrm{d}\theta)\right\}\right] \mathrm{d}Q_t^{(y)}(\mu) \\ &\quad = \int_{\mathcal{M}(\mathbb{T})} \exp(t\Psi(\mu))\, \mathrm{d}Q_t^{(y)}(\mu), \end{aligned} \tag{5.10}$$

where

$$\Psi(\mu) = \int_{\mathbb{T}} V(\theta)\, \mu(\mathrm{d}\theta), \ \ \mu \in \mathcal{M}(\mathbb{T}). \tag{5.11}$$

Note that (5.9), (5.10) imply that l.h.s. of (5.7) is of the same form as l.h.s. of (3.10) with $S = \mathcal{M}(\mathbb{T}),\ \epsilon = \frac{1}{t},\ P_\epsilon = Q_t^{(y)},\ g(x) = \Psi(\mu)$. It can be shown that $I_0(\cdot)$ defined by

$$I_0(\mu) = \begin{cases} \frac{1}{8}\int_{\mathbb{T}} \frac{1}{f(\theta)} |f'(\theta)|^2 \mathrm{d}\theta, \ \text{if } \mathrm{d}\mu(\theta) = f(\theta)\,\mathrm{d}\theta, & \text{and } f \text{ differentiable} \\ \infty, & \text{otherwise} \end{cases} \tag{5.12}$$

is the rate function on $\mathcal{M}(\mathbb{T})$; note that $\mathcal{M}(\mathbb{T})$ is a compact metric space by Prohorov's theorem. In fact we have the following:

Theorem 5.1 (Donsker–Varadhan, 1974). *For any $y \in \mathbb{T}$, the family $\{Q_t^{(y)}\colon t \geq 0\}$ of probability measures on $\mathcal{M}(\mathbb{T})$, induced by the occupation time functionals of Brownian motion on $\mathbb{T}$, satisfies LDP with rate function I_0 given by* (5.12). *Consequently, by Varadhan's lemma* (5.7) *holds.* □

For proof, see [DV1]. Moreover asymptotics of functionals of the form $\Psi(M(t, w, \mathrm{d}\theta))$ can be described. Like Sanov's theorem, the above result of Donsker and Varadhan is a level 2 LDP.

The basic space in the above set up is the torus which has a canonical measure, viz. the rotation invariant (Haar) measure $\mathrm{d}\theta$. The basic process is the Brownian motion on the torus. Its generator is the Laplacian which is uniformly elliptic and self-adjoint. Hence the normalized Haar measure on the torus turns out to be the unique ergodic probability measure for the basic process. This important fact has played a major role in the background.

The above result is the proverbial tip of the iceberg. It led to an extensive study, by Donsker and Varadhan, of LDP for occupation times for Markov chains and processes. Some of the results were also independently obtained by Gartner [G]. This in turn formed the basis for providing a variational formula for the principal eigenvalue of a not necessarily self-adjoint second-order elliptic differential operator, a solution to the problem of Wiener sausage, etc. However it is not powerful enough to deal with the polaron problem from statistical physics.

For this a LDP at the process level is needed. This is called *level 3 large deviations*. A crowning achievement is the LDP for empirical distributions of Markov processes, due to Donsker and Varadhan. We briefly describe this far reaching extension of Theorem 5.1.

Note that (5.8) can also be written as

$$M(t, w, A) = \lim_{n\to\infty} \frac{1}{2^n} \sum_{k=1}^{2^n} \delta_{Y(tk2^{-n}, w)}(A).$$

On the r.h.s. of the above we have a sequence of empirical distributions. To handle large deviation problems, the proper way to extend the notion of empirical distribution to stochastic processes turns out to be as follows.

Let Σ be a complete separable metric space. Let $\Gamma = \{w\colon w$ right continuous on $(-\infty, \infty)$ into Σ, and $w(t-)$ exists for all $t\}$. Under the Skorokhod topology on bounded intervals, Γ is a complete separable metric space. Let Γ_+ denote the corresponding space of Σ-valued functions on $[0, \infty)$. For $r \in (-\infty, \infty)$ let θ_r denote the translation map on Γ given by $\theta_r w(s) = w(r+s)$.

For $w \in \Gamma,\ t > 0$ let w_t be such that $w_t(s+t) = w_t(s)$ for all $s \in (-\infty, \infty)$, and $w_t(s) = w(s)$ for $0 \le s < t$; that is, the segment of w on $[0, t)$ is extended periodically to get w_t. For $w \in \Gamma,\ t > 0,\ B \subset \Gamma$ define

$$R_{t,w}(B) = \frac{1}{t} \int_0^t I_B\ (\theta_r w_t)\mathrm{d}r. \tag{5.13}$$

It can be shown that $R_{t,w}(\theta_\sigma B) = R_{t,w}(B)$ for any $B \subseteq \Gamma,\ \sigma > 0$. So $B \mapsto R_{t,w}(B)$ is a translation invariant probability measure on Γ. Let $\mathcal{M}_S(\Gamma)$ denote the space of all translation invariant probability measures on Γ, with the topology of weak convergence. This is a complete separable metric space. For fixed $t \ge 0$, note that $w \mapsto R_{t,w}$ is a mapping from Γ into $\mathcal{M}_S(\Gamma)$. It is called the *empirical distribution functional.*

We write $M(t, w, A) = \frac{1}{t}\int_0^t I_A(w(s))\mathrm{d}s,\ A \subseteq \Sigma$ to denote the analogue of (5.8) in the present context. It can be seen that $M(t, w, \cdot) = R_{t,w}\pi_0^{-1}$, where π_0 is the projection from Γ onto Σ given by $w \mapsto w(0)$. Thus the occupation time functional is the marginal distribution of the empirical distribution functional.

Let $P_{0,x}$ denote the distribution of a Σ-valued ergodic Markov process starting from $x \in \Sigma$ at time 0; it is a probability measure on Γ_+. For $t \geq 0,\ x \in \Sigma$, let $\zeta_t^{(x)}$ be defined by

$$\zeta_t^{(x)}(E) = P_{0,x}\{w \in \Gamma\colon R_{t,w} \in E\},\ \ E \subseteq \mathcal{M}_S(\Gamma). \tag{5.14}$$

So $\zeta_t^{(x)}$ is a probability measure on $\mathcal{M}_S(\Gamma)$. As the Markov process is ergodic there is a unique invariant probability measure ν on Σ. Let $\mu \in \mathcal{M}_S(\Gamma)$ be the translation invariant measure on Γ with ν as its marginal distribution; that is, μ is the stationary Markov process with ν as its marginal distribution. Using ergodic theorem it can be proved that $\zeta_t^{(x)} \Rightarrow \delta_\mu$ as $t \to \infty$ for any $x \in \Sigma$, where δ_μ denotes the Dirac measure concentrated at μ.

A stochastic process, in particular a Markov process, can be identified with an appropriate element of $\mathcal{M}(\Gamma_+)$, while a stationary stochastic process can be identified with an element of $\mathcal{M}_S(\Gamma)$. For each ergodic Markov process we associate a stationary stochastic process. Since $\{P_{0,x}\}$ as well as $\{R_{t,w}\}$ represent stochastic processes, an LDP for $\{\zeta_t^{(x)}\} \subset \mathcal{M}(\mathcal{M}_S(\Gamma))$ is considered an example of the highest level large deviations.

It is a deep result due to Donsker and Varadhan (1983) that, under suitable conditions, for any $x \in \Sigma$ the family $\{\zeta_t^{(x)}\colon t \geq 0\}$ satisfies LDP with a rate function $H(\cdot)$ defined on $\mathcal{M}_S(\Gamma)$ in terms of a relative entropy function. For details, see [DV2], as well as [V2] and [DS]. The rate function $H(\cdot)$ is called *entropy with respect to the Markov process* $\{P_{0,x}\}$.

The process level LDP turned out to be essential for Donsker and Varadhan (1983) to solve the polaron problem. This problem involves showing that the limit

$$\eta(\alpha) \triangleq \lim_{t\to\infty} \frac{1}{t} \log E \left\{ \exp\left[\alpha \int_0^t \int_0^t \frac{\mathrm{e}^{-|s-r|}}{|B(s) - B(r)|}\, \mathrm{d}r\, \mathrm{d}s \right] \right\}$$

exists, where $B(\cdot)$ is the three-dimensional Brownian motion, and establishing a conjecture made in 1949 by Pekar concerning the asymptotics of $\eta(\alpha)$ as $\alpha \to \infty$. For a description of the polaron problem, see [R] (see [V2] and the references therein for details).

In this write-up we have attempted to give just a flavour of large deviations. While [V2] and [V3] give succinct overview, [DS], [DZ] and [H] are excellent textbooks on the subject; the latter two also discuss applications to statistics, physics, chemistry, engineering, etc. An interested reader may also look up [DE], [El], [FW], [O], [Sm], [SW] and [FK] for diverse aspects, applications and further references.

Acknowledgment

This is an expanded version of M N Gopalan Endowment Lecture given at the 22nd Annual Conference of Ramanujan Mathematical Society held at National Institute of Technology – Surathkal, Karnataka in June 2007. It is a pleasure to thank R Bhatia for his encouragement and useful discussions. The author thanks V S Borkar for constructive suggestions on an earlier draft. Thanks are also due to an anonymous referee for spotting an error and for useful suggestions.

References

[AD] Atar R and Dupuis P, Large deviations and queueing networks: methods for rate function identification, *Stochastic Process. Appl.* **84** (1999) 255–296

[B] Bahadur R R, Some Limit Theorems in Statistics (Philadelphia: SIAM) (1971)
[DZ] Dembo A and Zeitouni O, Large Deviations Techniques and Applications (Boston: Jones and Bartlett) (1993)
[DS] Deuschel J and Stroock D W, Large Deviations (Boston: Academic Press) (1989)
[DV1] Donsker M D and Varadhan S R S, Asymptotic evaluation of certain Wiener integrals for large time, in: Functional Integration and its Applications (ed.) A M Arthurs, (Oxford: Clarendon) (1975) pp. 15–33
[DV2] Donsker M D and Varadhan S R S, Asymptotic evaluation of certain Markov process expectations for large time, IV. *Comm. Pure Appl. Math.* **36** (1983) 183–212
[DE] Dupuis P and Ellis R S, A weak convergence approach to the theory of large deviations (New York: Wiley) (1997)
[DR] Dupuis P and Ramanan K, A time-reversed representation for the tail probabilities of stationary reflected Brownian motion, *Stochastic Process. Appl.* **98** (2002) 253–287
[El] Ellis R S, Entropy, large deviations and statistical mechanics (New York: Springer) (1985)
[E] Evans L C, Partial Differential Equations (Providence: Amer. Math. Society) (1998)
[F] Feller W, An Introduction to Probability Theory and its Applications, Vol. II (New Delhi: Wiley-Eastern) (1969)
[FK] Feng J and Kurtz T G, Large Deviations for Stochastic Processes (Providence: Amer. Math. Soc.) (2006)
[FW] Freidlin M I and Wentzell A D, Random perturbations of dynamical systems (New York: Springer) (1998)
[G] Gartner J, On large deviations from the invariant measure, *Theory Probab. Appl.* **22** (1977) 24–39
[H] den Hollander F, Large Deviations (Providence: Amer. Math. Soc.) (2000)
[Kp] van Kampen N G, Stochastic processes in Physics and Chemistry (Amsterdam: North Holland) (1981)
[KS] Karatzas I and Shreve S E, Brownian Motion and Stochastic Calculus (New York: Springer) (1988)
[K] Kesavan S, Topics in Functional Analysis and Applications (New Delhi: Wiley-Eastern) (1989)
[O] Orey S, Large deviations in ergodic theory, in Seminar on Stochastic Processes 12 (Basel: Birkhauser) (1985) pp. 195–249
[P] Parthasarathy K R, Probability measures on metric spaces (New York: Academic Press) (1967)
[RD] Ramanan K and Dupuis P, Large deviation properties of data streams that share a buffer, *Ann. Appl. Probab.* **8** (1998) 1070–1129
[R] Roepstorff G, Path integral approach to quantum physics: An introduction (Berlin: Springer) (1996)
[RSST] Rolski T, Schmidli H, Schmidt V and Teugels J, Stochastic processes for insurance and finance (New York: Wiley) (2001)
[Sc] Schuss Z, Theory and applications of stochastic differential equations (New York: Wiley) (1980)
[SW] Schwartz A and Weiss A, Large Deviations for Performance Analysis, Queues, Communications and Computing (London: Chapman and Hall) (1995)
[Sm] Simon B, Functional integration and quantum physics (New York: Academic Press) (1979)
[V1] Varadhan S R S, Asymptotic Probabilities and Differential Equations, *Comm. Pure Appl. Math.* **19** (1966) 261–286
[V2] Varadhan S R S, Large Deviations and Applications (Philadelphia: SIAM) (1984)
[V3] Varadhan S R S, Large deviations and applications, *Expo. Math.* **3** (1985) 251–272