Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Some Problems of Large Deviations

S. R. S. Varadhan

1. Introduction. The problem of large deviations arises naturally in probability theory in different contexts. We shall first look at some typical examples.

If $X_1, X_2, \ldots, X_n, \ldots$ is a sequence of independent and identically distributed random variables with mean zero, then for any $a>0$

$$\lim_{n\to\infty} P\left[\frac{X_1+\ldots+X_n}{n} \geqslant a\right] = 0 \tag{1.1}$$

by the law of large numbers. Under the assumption that for every real θ

$$M(\theta) = E[e^{\theta X_i}] < \infty$$

Cramér [3] showed that the probability in (1.1) has the specific behavior:

$$\lim_{n\to\infty} \left\{P\left[\frac{X_1+\ldots+X_n}{n} \geqslant a\right]\right\}^{1/n} = \varrho_a \tag{1.2}$$

exists and is given by

$$\varrho_a = \inf_{\theta\geqq 0} M(\theta)e^{-\theta a}.$$

For the second example we consider the process $x_\varepsilon(t)=\varepsilon\beta(t)$ defined in the time interval $0\leqslant t\leqslant 1$, where $\beta(\cdot)$ is the standard one-dimensional Brownian motion. We denote by P_ε the measure corresponding to $x_\varepsilon(t)$, which lives on the space $C_0[0, 1]$ of continuous functions vanishing at $t=0$. As $\varepsilon\to 0$, $x_\varepsilon(t)$ tends to the zero function uniformly and therefore

$$\lim_{\varepsilon\to 0} P_\varepsilon(A) = 0 \tag{1.3}$$

for sets A that are disjoint from some neighborhood of the zero function. One can again show that for a large class of sets $A \subset C_0[0, 1]$

$$\lim_{\varepsilon \to 0} 2\varepsilon^2 \log P_\varepsilon(A) = -h(A) \tag{1.4}$$

exists where

$$h(A) = \inf_{f \in A} \int_0^1 [f'(t)]^2 \, dt \tag{1.5}$$

and the infimum is taken over absolutely continuous functions in f in A with a square integrable derivative. Such a result in a slightly different form can be found in **[14]**.

For the third example we will consider a Markov chain on a finite state space X and transition probabilities π_{ij}. For simplicity we shall assume that $\pi_{ij} > 0$ for all i and j. Let $\vec{q} = \{q_i\}$ be the unique invariant probability vector. If $X_0, X_1, \ldots, X_n, \ldots$ is a realization of the Markov chain, we denote by $f_i^{(n)}$ the frequency of occurrence of the state i during the first n steps of the Markov chain and by $p_i^{(n)}$ the proportion of visits to the state i during the same time. The vector $\vec{p}^{(n)}$ with components $\{p_i^{(n)}\}$ is a random probability distribution on X and is the empirical distribution based on the first n steps. By the ergodic theorem, for large values of n, the vector $\vec{p}^{(n)}$ is close to the invariant probability vector $\vec{q}$ with a very high probability. The theorem on large deviations states in this context that for suitable sets A in space of probability distributions on X

$$\lim_{n \to \infty} \frac{1}{n} \log P\{\vec{p}^{(n)} \in A\} = -h(A) \tag{1.6}$$

exists, where

$$h(A) = \inf_{p \in A} I(\vec{p}) \tag{1.7}$$

and the function $I(\vec{p})$ is defined for probability distributions $\vec{p}$ on X by

$$I(\vec{p}) = -\inf_{u > 0} \sum_{i \in X} p_i \log \left(\frac{\sum_{j \in X} \pi_{ij} u_j}{u_i} \right) \tag{1.8}$$

where the infimum is taken over vectors $u = \{u_i\}$ which have strictly positive components. The function $I(\cdot)$, while it is not explicit, is a nonnegative convex function of $\vec{p}$ which vanishes only when $\vec{p}$ equals the invariant probability $\vec{q}$.

The first example can be generalized by allowing the random variables to take values in more general linear spaces. The second can be generalized to cover processes $x_\varepsilon(t)$ governed by stochastic differential equations of the form

$$dx_\varepsilon(t) = \varepsilon \sigma(x_\varepsilon(t)) \, d\beta(t) + b(x_\varepsilon(t)) \, dt,$$
$$x_\varepsilon(0) = x.$$

Actually one can even combine the two classes in a single large class of problems of large deviations for Markov processes. See for instance **[15]**–**[20]**, **[1]**, **[2]**, and **[13]**.

As I am hoping that Professor Ventcel will cover this ground in some detail I shall proceed to a discussion of the ideas connected with the third example.

2. Large deviation for occupation times. Let the state space X of the Markov chain be a complete separable metric space instead of a finite set. Let the transition probabilities be given by $\pi(x, dy)$. We shall denote by π, the corresponding operator on the space $B(X)$ of bounded measurable functions on X, defined by

$$(\pi f)(x) = \int f(y)\pi(x, dy). \tag{2.1}$$

We denote by $\mathcal{M}$ the space of all probability distributions on X and we will view $\mathcal{M}$ as a complete separable metric space with weak convergence as the underlying convergence notion. We make the following assumptions on $\pi(x, dy)$.

(H1) *(Feller Property)*. If f is bounded and continuous then so is πf.

(H2) *(Strong transitivity)*. There is a reference measure $\alpha(dy)$ on X such that $\alpha(dy)$ and $\pi(x, dy)$ are mutually absolutely continuous for every $x \in X$.

(The above condition can be relaxed somewhat.)

For $\mu \in \mathcal{M}$ we define $I(\mu)$ by

$$I(\mu) = -\inf_{u \in U} \int_X \log\left(\frac{\pi u}{u}\right)(x)\mu(dx) \tag{2.2}$$

where U consists of functions in the space $C(X)$ of bounded continuous functions which have a positive lower bound. One can verify that $I(\cdot)$ is a convex, nonnegative, lower semicontinuous functional on $\mathcal{M}$ which vanishes only at invariant probability distributions. The chain of course may not possess an invariant probability distribution in which case $I(\cdot)$ is never zero.

We shall denote by ω, a realization $X_0, X_1, \dots, X_n, \dots$ of the chain and by $L_n(\omega, \cdot)$ the empirical distribution based on the first n steps

$$L_n(\omega, E) = \frac{1}{n} \sum_{j=1}^{n} \chi_E(X_j), \quad E \subset X. \tag{2.3}$$

Fixing an arbitrary starting point $x \in X$ for the chain we have the measure P_x corresponding to the process on the space Ω of all possible realizations. One can view $L_n(\cdot, \cdot)$ as a map of Ω into $\mathcal{M}$ and this induces a measure $Q_{n,x}$ on $\mathcal{M}$ which is the distribution of the empirical distribution. If A is a subset of $\mathcal{M}$ then we are interested in the asymptotic behavior of $Q_{n,x}(A)$ as $n \to \infty$. We have the following results:

If $C \subset \mathcal{M}$ is compact, then

$$\limsup_{n \to \infty} \frac{1}{n} \log Q_{n,x}[C] \leqslant -\inf_{\mu \in C} I(\mu). \tag{2.4}$$

If $G \subset \mathcal{M}$ is open, then

$$\limsup_{n \to \infty} \frac{1}{n} \log Q_{n,x}[G] \geqslant -\inf_{\mu \in G} I(\mu). \tag{2.5}$$

The reason that (2.4) can be established only for compact sets is that we have no assumption of positive recurrence. Perhaps the measures $Q_{n,x}$ are dissipative and (2.4) really measures the rate of dissipation. It is even possible that $\inf_\mu I(\mu)>0$ and in that case (2.4) cannot hold when C is taken to be the whole space $\mathcal{M}$. However when X is not compact, we can impose a strong positive recurrence condition that will enable us to prove (2.4) for closed sets.

There are continuous time analogs to these results for Markov processes $x(t)$ with transition probabilities $p(t, x, dy)$ on a state space X. We have the corresponding operators

$$(T_t f)(x) = \int f(y)\, p(t, x, dy) \tag{2.6}$$

we make the following assumptions regarding the semigroup $\{T_t\}$.

(H1*) T_t maps the space of bounded continuous functions into itself.

(H2*) There is a reference measure $\alpha(dy)$ such that $\alpha(dy)$ and $p(t, x, dy)$ are mutually absolutely continuous for every $t>0$ and $x\in X$.

(H3*) The strongly continuous center of the semigroup T_t is sufficiently large.

We denote by L the infinitesimal generator acting on the domain $\mathcal{D}\subset C(X)$. We denote by $\mathcal{D}^+$ those functions in $\mathcal{D}$ with a positive lower bound. We define the I-function as the analog of (2.2) by

$$I(\mu) = -\inf_{u\in\mathcal{D}^+} \int \left(\frac{Lu}{u}\right)(x)\, \mu(dx). \tag{2.7}$$

We look at the occupation distribution

$$L_t(\omega, E) = \frac{1}{t}\int_0^t \chi_E(x(s))\, ds \tag{2.8}$$

as the analog of (2.3) and the distribution $Q_{t,x}$ of $L_t(\cdot,\cdot)$ as the analog of $Q_{n,x}$. We then have the exact analogs of (2.4) and (2.5).

For C compact in $\mathcal{M}$,

$$\limsup_{t\to\infty} \frac{1}{t}\log Q_{t,x}(C) \leq -\inf_{\mu\in C} I(\mu). \tag{2.9}$$

For G open in $\mathcal{M}$,

$$\liminf_{t\to\infty} \frac{1}{t}\log Q_{t,x}(G) \geq -\inf_{\mu\in G} I(\mu). \tag{2.10}$$

Again if X is not compact one can impose a strong recurrence condition and obtain (2.9) for closed sets instead of just compact sets.

If the transition probabilities $p(t, x, dy)$ are such that they have a symmetric density with respect to the reference measure $\alpha(dy)$ then the semigroup T_t is a family of self adjoint contractions in the Hilbert space of functions that are square integrable with respect to the measure $\alpha(dy)$. Then the infinitesimal generator L can be thought of as a nonpositive self adjoint operator. Therefore the operator $\sqrt{-L}$ is well defined. In such a context one can show that $I(\mu)$ is finite if and only if μ is absolutely continuous with respect to α and the square integrable function $(d\mu/d\alpha)^{1/2}$ is in the domain of $\sqrt{-L}$. In such a case

$$I(\mu) = \left\| \sqrt{-L} \left(\frac{d\mu}{d\alpha}\right)^{1/2} \right\|^2 \tag{2.11}$$

where $\| \ \|$ is the mean square norm with respect to α. For the Brownian motion in R^d this becomes

$$I(f) = I(\mu) = \frac{1}{8}\int \frac{|\nabla f|^2}{f}\, dx = \frac{1}{2}\int |\nabla \sqrt{f}|^2\, dx \tag{2.12}$$

where $\mu(dx) = f(x)\, dx$.

The details of these results can be found in [4], [5], [21].

3. Connections with the principal eigenvalue. Let us consider the case when X is compact. If $x(t)$ is the Markov process corresponding to the generator L and $I(\cdot)$ is the corresponding I-functional defined by (2.7) then one can show [see [15]] that

$$\begin{aligned} \lambda(V) &= \lim_{t\to\infty} \frac{1}{t} \log E_x \left[\exp\left\{\int_0^t V(x(s))\, dx\right\}\right] \\ &= \sup_{\mu\in\mathcal{M}} \left[\int V(x)\,\mu(dx) - I(\mu)\right] \end{aligned} \tag{3.1}$$

for all $V\in C(X)$. The quantity $\lambda(V)$ can also be identified as the principal eigenvalue (i.e. the point in the spectrum with the largest real part) of the operator $L+V$. From the maximum principle for L one can conclude that $\lambda(V)$ is a convex functional of V and $I(\cdot)$ is its conjugate convex functional. This is the explanation of why $I(\cdot)$ is readily computable in the self-adjoint case. The variational formula (3.1) in that case reduces to the classical Rayleigh–Ritz variational formula for the principal eigenvalue of self adjoint operators. See in this context [6] and [12].

4. Applications. If $\beta(s)$ is the d-dimensional Brownian motion then the Wiener sausage $C_\varepsilon^t(\omega)$ for the trajectory $\omega=\beta(\cdot)$ up to time t is the set

$$C_\varepsilon^t(\omega) = \left\{y\colon \inf_{0\leqq s\leqq t} |y-\beta(s)| \leqslant \varepsilon\right\}.$$

A problem that comes from statistical mechanics [11] is the behavior of

$$E[\exp\{-\nu|C_\varepsilon^t(\omega)|\}]$$

where $|A|$ stands for the Lebesgue measure of the set A. Using the techniques developed in the preceding section (see [7], [8]) one can show that

$$\lim_{t\to\infty} \frac{1}{t^{d/(d+2)}} \log E[\exp\{-\nu|C_\varepsilon^t(\omega)|\}] = -k(\nu, d) \tag{4.1}$$

exists and compute $k(\nu, d)$ as

(4.2)
$$k(\nu, d) = \inf_{\substack{f \in L_1(R^d) \\ f \geqq 0}} [\nu|x: f(x) > 0| + I(f)]$$

where $I(f)$ is given by (2.12).

One can consider a random walk on the lattice Z^d in R^d of points with integral coordinates. If the distribution of the single step has mean zero and covariance identity then under mild irreducibility assumptions the number $D(n)$ of distinct sites visited by the random walk in the first n steps can be shown to satisfy

(4.3)
$$\lim_{n \to \infty} \frac{1}{n^{d/(d+2)}} \log E[\exp\{-\nu D(n)\}] = -k(\nu, d)$$

where $k(\nu, d)$ is again given by (4.2). There are analogs when the Brownian motion is replaced by a symmetric stable process (see [9] for details).

Another application is to derive laws of the iterated logarithm for local times. Let $\omega = \beta(\cdot)$ be one dimensional Brownian motion. Let

$$l(t, y) = \int_0^t \delta(\beta(s) - y)\, ds$$

be the local time of the Brownian path as a function of t and y. We define

$$L(t, y) = \frac{1}{\sqrt{t \log\log t}}\, l\left(t, y\sqrt{\frac{t}{\log\log t}}\right).$$

We can view $L(t, y)$ as a random probability density on the line. One can show that the set of limit points of $L(t, \cdot)$ as $t \to \infty$ coincides as functions of y (in the topology of uniform convergence on compact sets) almost surely with the set of subprobability densities $f(\cdot)$ satisfying

$$\int f(x)\, dx \leq 1 \quad \text{and} \quad \frac{1}{8}\int \frac{[f'(x)]^2}{f(x)}\, dx \leq 1.$$

These results and similar ones for certain stable processes can be found in [10].

5. General remarks. In the case of continuous time processes the I-functional plays the role of the Dirichlet integral to which it reduces in the self adjoint case. For instance $G \subset X$ is an open set with compact closure and one is interested in the exponential decay rate of $P_x[x(s) \in G, 0 \leq s \leq t]$ as $t \to \infty$. One can under mild conditions compute it as

$$\lambda(G) = \inf_{\substack{\mu \in \mathcal{M} \\ \mu(G) = 1}} I(\mu).$$

There is a close connection, at least in spirit, to the theory developed by Lanford in [13] in the context of statistical mechanics.

If we start with a Markov process and reverse it in time using an invariant probability distribution then the new reversed Markov process has the same I-functional as the old one. It is an interesting question to examine how much information about the process can be recovered by a knowledge of its I-functional.

In case the process is transient one can consider the total occupation time i.e.

$$L(\omega, E) = \int_0^\infty \chi_E(x(s))\, ds$$

which will be a σ-finite measure on the state space. The tail behavior of its distribution on the space of all σ-finite measures on the state space should again be related to the I-functional. The details of the connection are being currently worked out.

References

1. R. Azencott and G. Ruget, *Mélanges d'équation différentielles et grand écart à la loi des grands nombres*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **38** (1977), 1—54.

2. R. R. Bahadur and S. L. Zabell, *Large deviations of the sample mean in general vector spaces* (preprint).

3. H. Carmér, *On a new limit theorem in the theory of probability:* Colloquium on the Theory of Probability, Hermann, Paris, 1937.

4. M. D. Donsker and S. R. S. Varadhan, *Asymptotic evaluation of certain Markov process expectations for large time*, I, Comm. Pure Appl. Math. **28** (1975), 1—47.

5. ——— *Asymptotic evaluation of certain Markov process expectations for large time*, III, Comm. Pure Appl. Math. **29** (1977), 389—461.

6. ——— *On a variational formula for the principal eigenvalue for operators with maximum principle*, Proc. Nat. Acad. Sci. U.S.A. **72** (1975), 780—783.

7. ——— *Asymptotic evaluation of certain Markov process expectations for large time*, II, Comm. Pure Appl. Math. **28** (1975), 279—301.

8. ——— *Asymptotics for the Wiener sausage*, Comm. Pure Appl. Math. **28** (1975), 525—565.

9. ——— *On the number of distinct sites visited by a random walk* (preprint).

10. ——— *On laws of the iterated logarithms for local times*, Comm. Pure Appl. Math. **30** (1977), 707—753.

11. M. Kac, *Probabilistic methods in some problems of scattering theory*, Rocky Mountain J. Math. **4** (1974), 511—538.

12. T. G. Kurtz, *A variational formula for the growth rate of a positive operator semigroup* (preprint).

13. O. E. Lanford, *Entropy and equilibrium states in classical statistical mechanics*, Statistical Mechanics and Mathematical Problems, Lecture notes in Physics, vol. 20, Springer-Verlag, Berlin and New York, 1971, pp. 1—113.

14. M. Schilder, *Some asymptotic formulae for Wiener integrals*, Trans. Amer. Math. Soc. **125** (1966), 63—85.

15. S. R. S. Varadhan, *Asymptotic probabilities and differential equations*, Comm. Pure Appl. Math. **19** (1966), 261—286.

16. ——— *On the behavior of the fundamental solution of the heat equation with variable coefficients*, Comm. Pure Appl. Math. **20** (1967), 431—455.

17. ——, *Diffusion processes in a small time interval,* Comm. Pure Appl. Math. **20** (1967), 659—685.

18. A. D. Ventcel, *Rough limit theorems on large deviations for Markov stochastic processes,* I, Theor. Probability Appl. **21** (1976), 227—242.

19. ——, *Rough limit theorems on large deviations for Markov stochastic processes,* II, Theor. Probability Appl. **21** (1976), 499—512.

20. A. D. Ventcel and M. I. Freidlin, *On small random perturbations of dynamical systems,* Russian Math. Surveys **25** (1970), 1—55.

21. J. Gärtner, *On large deviations from the Invariant Measure,* Theor. Probability Appl. **22** (1977), 24—39.

Courant Institute of Mathematical Sciences
New York University
New York, New York 10012, U.S.A.