BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 24, Number 2, April 1991

0273-0979/91 $1.00 + $.25 per page

Large deviations, by Jean-Dominique Deuschel and Daniel W. Stroock. Academic Press, New York, 1989, 300 pp., $34.95. ISBN 0-12-213150-9

The primary concern of the theory of large deviations is the precise estimation of the probabilities of certain classes of rare events. There is usually a natural parameter in the problem which can be assumed to be large. For example, this parameter could denote the size of the system, or if one is dealing with random perturbation of deterministic systems, the noise level could be related to the inverse of this parameter. Either the model or the event or sometimes both depend on this parameter, and the probability usually goes to zero exponentially fast in the parameter. The theory is concerned with the determination of the exact exponential decay rate. Very often the constant can be calculated explicitly in terms of quantities of physical significance. It is the existence of explicit formulae that makes the subject attractive to mathematicians and physicists. Many of the problems of equilibrium and nonequilibrium statistical mechanics have interpretations in terms of the theory of large deviations.

Here are two typical examples of the theory. Let $x_1, x_2, \dots x_n$ be n independent identically distributed (i.i.d.) random variables having $F(x)$ for their common distribution function. Cramer [1] proved in 1937 that

$$\lim_{n\to\infty} \frac{1}{n} \log P\left[a \le \frac{x_1 + \cdots + x_n}{n} \le b\right] = -c$$

exists for $-\infty \le a < b \le \infty$ and the constant c can be explicitly calculated in terms of F according to the following recipe:

$$c = \inf_{a\le x\le b} I(x)$$

where

$$I(x) = \sup_\theta[\theta x - \log M(\theta)]$$

and

$$M(\theta) = \int e^{\theta x} dF(x).$$

The Scandinavian school in the 1930s was interested in the calculation of risks in the insurance business and Cramer's theorem was an outgrowth of that. The second example which was proved by Schilder [4] in 1964 is concerned with the distribution of $\sqrt{\varepsilon}\beta(\cdot)$ on $C_0[0, 1]$ where $\beta(\cdot)$ is standard Brownian motion and $\varepsilon > 0$ is a small parameter. The distribution of $\sqrt{\varepsilon}\beta(\cdot)$ is a scaled Wiener measure on the space $C_0[0, 1]$ of continuous functions on $[0, 1]$ which vanish at the origin. We denote this measure by P_ε. If A is a set of trajectories not containing the function identically zero, then we expect $P_\varepsilon(A)$ to tend to zero, and Schilder's theorem roughly speaking states that

$$\lim_{\varepsilon \to 0} \varepsilon \log P_\varepsilon(A) = -c(A)$$

exists for nice sets A and

$$c(A) = \inf_{f \in A} I(f)$$

where $I(f) = \frac{1}{2}\int_0^1 [f^1(t)]^2 dt$ if $f(0) = 0$ and $f(t)$ is absolutely continuous with a square integrable derivative. $I(f) = \infty$ otherwise. The function $I(f)$ is of course the classical action function and its appearance is significant.

The developments in the theory of large deviations over the last twenty-five years have been regarding generalizations and variations of these two types of results. An abstract formulation of the principle of large deviation is the following.

We have a family P_λ of probability measures on a topological space X and a rate function $I(x)$ on X satisfying

(1) $0 \leq I(x) \leq \infty$.

(2) $I(x)$ is lower semicontinuous.

(3) For every $\ell < \infty$, $\{x : I(x) \leq \ell\}$ is a compact subset of X.

One says that P_λ satisfies a large deviation principle with rate function $I(\cdot)$ if the following holds:

(i) For every closed set $C \subset X$

$$\limsup_{\lambda \to \infty} \frac{1}{\lambda} \log P_\lambda(C) \leq -\inf_{x \in C} I(x).$$

(ii) For every open set $G \subset X$

$$\liminf_{\lambda \to \infty} \frac{1}{\lambda} \log P_\lambda(G) \geq -\inf_{x \in G} I(x).$$

The book by Deuschel and Stroock which grew out of earlier lecture notes by Stroock starts out with the two basic examples. Then it goes into the Ventcel-Freidlin theory of small random perturbations, where one considers a stochastic differential equation of the form

$$dx_\varepsilon(t) = b(x_\varepsilon(t))dt + \sqrt{\varepsilon}\sigma(x_\varepsilon(t))d\beta(t).$$

and a large deviation principle for the solution $x_\varepsilon(\cdot)$ is established.

The next step is Sanov's theorem:

Let $x_1, \dots, x_n, \dots$ be a sequence of i.i.d random variables with values in some metric space X with a common distribution α. Let us look at the empirical distribution $\frac{1}{n}[\delta_{x_1} + \cdots + \delta_{x_n}]$ as a random measure i.e. a random variable with values in the space $\mathcal{M}$ of probability measures on X. Sanov's theorem is a large deviation theorem for the empirical distribution with a rate function $I_\alpha(\mu)$ on $\mathcal{M}$ which is just relative entropy

$$\begin{aligned} I_\alpha(\mu) &= \int \left(\frac{d\mu}{d\alpha}\log\frac{d\mu}{d\alpha}\right) d\alpha \quad \text{if } \mu << \alpha, \\ &= \infty \qquad\qquad\qquad\qquad \text{otherwise.} \end{aligned}$$

Sanov's theorem is, in some sense, an abstract version of Cramer's theorem, and the fact that entropy comes up as the rate function is the crux of the connection with statistical mechanics and thermodynamics.

The rest of the book is devoted to generalizations and variations of Sanov's theorem. One can drop independence and assume that $x_1, \dots, x_n, \dots$ forms a Markov chain on the space X with transition probabilities $\pi(x, dy)$. Under suitable assumptions on the ergodic behavior of the chain, it is possible to establish a large deviation principle for the empirical distribution. The rate function is given by

$$I(\mu) = \sup_{u>0} \int \log\left(\frac{u}{\pi u}\right)(x)d\mu(x).$$

Here the sup is taken over all bounded uniformly positive functions and $\pi(u)(x) = \int u(y)\pi(x, dy)$. In particular if $\pi(x, dy) = \alpha(dy)$, i.e. in the i.i.d case, $I(\mu)$ reduces to relative entropy $I_\alpha(\mu)$ of Sanov's theorem. If one replaces the Markov chain by a continuous time Markov Process with generator L and the empirical distribution by its analog

$$\ell_t(A) = \frac{1}{t}\int_0^t \chi_A(x(s))\,ds$$

again one has a large deviation theorem with a rate function

$$I(\mu) = -\inf_{u>0}\int \left(\frac{Lu}{u}\right)(x)\mu(dx).$$

These results, developed by Donsker and Varadhan and to a lesser extent by Gartner, are treated very well in the book. A certain amount of hard analysis is required to handle the ergodicity requirements. These problems of suitable ergodicity conditions for the Markovian case as well as mixing conditions for the non-Markovian case take up the last chapters of the book.

The book contains an extensive list of references as well as detailed historical comments.

Those interested in connections with statistical mechanics should read references [2] or [3].

REFERENCES

[1] H. Cramer, *Sur un nouveau théorème-limite de la théorie des probabilités*, Actualités Sci. Indust. **736** (1938), 5–23. Colloque consacre a la theorie des probabilities, vol. 3, Hermann, Paris.

[2] R. S. Ellis, *Entropy, large deviations and statistical mechanics*, Springer-Verlag, Berlin, 1985.

[3] O. E. Lanford, *Entropy and equilibrium states in classical statistical mechanics*, Statistical Mechanics and Mathematical Problems (A. Lenard, ed.), Lecture Notes in Phys. vol. 20, Springer-Verlag, Berlin, 1973, pp. 1–113.

[4] M. Schilder, *Some asymptotics formulae for Wiener integrals*, Trans. Amer. Math. Soc. **125** (1966), 63–85.

S.R.S. VARADHAN
COURANT INSTITUTE OF MATHEMATICAL SCIENCES

BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 24, Number 2, April 1991

0273-0979/91 $1.00 + $.25 per page

Stochastic calculus in manifolds, by Michel Emery. Springer-Verlag, Berlin, New York, 1989, 151 pp., $29.00. ISBN 3-540-51664-6

I am glad, but also a little embarrassed to present this book because Emery's work is very closely connected with Paul André Meyer's and mine, these two last ones being also much intertwined. A large part of the book is an exposition of previous work, but also much of the material is new. Anyway, the presentation is always original and interesting. I always prefer intrinsic formulations for manifolds "à la Bourbaki," giving the expression in coordinates