CELEBRATIO MATHEMATICA

S. R. S. Varadhan

DANIEL W. STROOCK

LARGE DEVIATIONS

2012

msp

msp

LARGE DEVIATIONS

DANIEL W. STROOCK

Suppose that $\{\mu_n : n \in \mathbb{N}\}$ is a sequence of Borel probability measures on a Polish space Ω, and assume that, as $n \to \infty$, μ_n degenerates to the point mass δ_{ω_0} at $\omega_0 \in \Omega$. Then, it is reasonable to say that, at least for large n, neighborhoods of ω_0 represent "typical" behavior and that their complements represent "deviant" behavior; it is often important to know how fast their complements are becoming deviant. Finding a detailed solution to such a problem usually entails rather intricate analysis. However, if one's interest is in behavior which is "highly deviant", in the sense that it is dying out at an exponential rate, and if one is satisfied with finding the exponential rate at which it is disappearing, then one is studying *large deviations* and life is much easier. Indeed, instead of trying to calculate the asymptotic limit of quantities like $\mu_n\big(B(\omega_0, r)^{\complement}\big)$ (where $B(\omega, r)$ denotes the ball of radius r centered at ω), one is trying to calculate

$$\lim_{n \to \infty} \frac{1}{n} \log \mu_n\big(B(\omega_0, r)^{\complement}\big),$$

which is an inherently simpler task.

To develop the intuition for this type of analysis, remember that we are dealing here with probability measures. Thus, the only way that the μ_n can degenerate to δ_{ω_0} is that more and more of their mass is concentrated in a neighborhood of ω_0. In the nicest situation, this concentration is taking place because

$$\mu_n(d\omega) = \frac{1}{Z_n} e^{-nI(\omega)} \lambda(d\omega),$$

where $I(\omega) > I(\omega_0) \geq 0$ for $\omega \neq \omega_0$, and λ is some reference measure. Indeed, assuming that

$$\lim_{n \to \infty} \frac{1}{n} \log Z_n = 0,$$

then

$$\lim_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) = \lim_{n \to \infty} \log \|\mathbf{1}_\Gamma e^{-I}\|_{L^n(\lambda)} = \log \|\mathbf{1}_\Gamma e^{-I}\|_{L^\infty(\lambda)} = -\operatorname*{essinf}_{\omega \in \Gamma} I(\omega).$$

That is,

$$(1) \qquad \lim_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) = -\operatorname*{essinf}_{\omega \in \Gamma} I(\omega).$$

Of course, for many applications (for example, to number theory, geometry, or statistical mechanics) the non-appearance of Z_n in the answer would mean that one has thrown out the baby with the wash. On the other hand, because it is so crude, the type of thinking used in the previous remark predicts correct results even when it has no right to. To wit, suppose that $\Omega = \{\omega \in C([0, 1]; \mathbb{R}) : \omega(0) = 0\}$ and μ_n is the

distribution of $\omega \in \Omega \longmapsto n^{-1/2}\,\omega \in \Omega$ under standard Wiener measure. Clearly, the μ_n are degenerating to the point-mass at the path $\mathbf{0}$ which is identically 0. Moreover, Feynman's representation of μ_n is

$$\mu_n(d\omega) = \frac{1}{Z_n} e^{-n\,I(\omega)}\,\lambda(d\omega),$$

where

$$I(\omega) = \frac{1}{2}\int_0^1 |\dot\omega(t)|^2 dt$$

and λ is the Lebesgue measure on Ω. Ignoring the fact that none of this is mathematically kosher and proceeding formally, one is led to the guess that (1) may nonetheless be true, at least after one has taken into account some of its obvious flaws. In particular, there are two sources of concern. The first of these is the almost-sure non-differentiability of Wiener paths. However this objection is easily overcome by simply defining $I(\omega) = \infty$ unless ω has a square-integrable derivative. The second objection is that λ does not exist and therefore the "ess" before the "inf" has no meaning. This objection is more serious, and its solution requires greater subtlety. In fact, it was Varadhan's solution to this problem which was one of his seminal contributions to the whole field of large deviations. Namely, our derivation of (1) was purely measure-theoretic: we took no account of topology. On the other hand, not even the sense in which the μ_n are degenerating can be rigorously described in purely measure-theoretic terms. The best that one can say is that they are tending weakly to $\delta_{\mathbf{0}}$. Thus, one should suspect that (1) must be amended to reflect the topology of Ω, and that topology should appear in exactly the same way that it does in weak convergence. With this in mind, one can understand Varadhan's answer that (1) should be replaced by

$$(2) \qquad -\inf_{\omega \in \Gamma^\circ} I(\omega) \le \varliminf_{n\to\infty} \frac{1}{n}\log\mu_n(\Gamma) \le \varlimsup_{n\to\infty} \frac{1}{n}\log\mu_n(\Gamma) \le -\inf_{\omega\in\overline\Gamma} I(\omega).$$

Monroe Donsker provided the original inspiration for this type of analysis of rescaled Wiener measure, and his student Schilder was the first to obtain rigorous results. However, it was Varadhan [1967] who first realized that Schilder's work could be viewed in the context of large deviations, and it was he who gave and proved the validity of the formulation in (2). Indeed, a strong case can be made for saying that the modern theory of large deviations was born in [Varadhan 1967]. In particular, (2) quickly became the archetype for future results; families $\{\mu_n : n \in \mathbb{N}\}$ for which (2) are now said to satisfy *the large deviation principle with rate function I*. In addition, it was in the same article that Varadhan proved how to pass from (2) to the sort of results which Schilder had proved. Namely, he proved that if (2) holds with a rate function I which has compact level sets (that is, $\{\omega : I(\omega) \le R\}$ is compact for each $R \in [0, \infty)$) and $F : \Omega \longrightarrow \mathbb{R}$ is a bounded, continuous function, then

$$(3) \qquad \lim_{n\to\infty} \mathbb{E}^{\mu_n}\big[e^{nF}\big] = \sup_{\omega\in\Omega}\big(F(\omega) - I(\omega)\big).$$

This result, which is commonly called *Varadhan's lemma*, is exactly what one would expect from the model case when $\mu_n(d\omega) = (1/Z_n)\,e^{-nI}\,\lambda(d\omega)$; its proof in general is quite easy, but one would be hard put to overstate its importance. Not only is it a practical computational tool, but it provides a link between the theory of large deviations and convex analysis. Specifically, when Ω is a closed, convex subset of a topological vector space E, then, under suitable integrability assumptions, Varadhan's lemma combined

with the inversion formula for Legendre transforms often can be used to identify the rate function I as the Legendre transform

$$\Lambda^*(\omega) = \sup_{\lambda \in E^*} \{\langle \omega, \lambda \rangle - \Lambda(\lambda) : \lambda \in E^*\},$$

where

$$\lambda \in E^* \longrightarrow \Lambda(\lambda) \equiv \lim_{n \to \infty} \frac{1}{n} \log \mathbb{E}^{\mu_n}[e^{n\lambda(\omega)}].$$

Had he only laid the foundation for the field, Varadhan's impact on the study of large deviations would have been already profound. However, he did much more. Perhaps his deepest contributions come from his recognition that large deviations underlie and explain phenomena in which nobody else even suspected their presence. The depth of his understanding is exemplified by his explanation of Marc Kac's famous formula for the principle eigenvalue of a Schrödinger operator. Donsker had been seeking such an explanation for years, but it was not until he joined forces with Varadhan that real progress was made on the problem. Prior to their article [Donsker and Varadhan 1975b], all applications (Schilder's theorem, including its extensions and improvements by Varadhan, as well as the many beautiful articles by Freidlin and Wentcel) of large deviations to diffusion theory had been based on the observation that, during a short time interval, "typical" behavior of a diffusion is given by the solution to an ordinary differential equation. Thus, the large deviations in these applications come from the perturbation of an ordinary differential equation by a Gaussian-like noise term. The large deviations in [Donsker and Varadhan 1975b] have an entirely different origin. Instead of short-time behavior of the diffusion paths themselves, the quantity under consideration is the long-time behavior of their empirical distribution. In this case, "typical" behavior is predicted by ergodic theory, and the large deviations are those of the empirical distribution from ergodic behavior. The situation in [Donsker and Varadhan 1975b] is made more challenging by the fact that there really is no proper ergodic behavior of Brownian motion on the whole of space, since, in so far as possible, the empirical distribution of a Brownian path is trying to become the normalized Lebesgue measure. What saves the day is the potential term in the Schrödinger operator, whose presence penalizes paths that attempt to spread out too much.

The upshot of Donsker and Varadhan's analysis [1975b] is a new variational formula for the principal eigenvalue. Although their formula reduces to the classical one in the case of self-adjoint operators, it has the advantage that it relies entirely on probabilistic reasoning (that is, the minimum principle) and, as they showed in [Donsker and Varadhan 1975c], is therefore equally valid for operators which are not self-adjoint. More important, it launched a program which produced a spectacular sequence of articles. The general theory was developed in [Donsker and Varadhan 1975c] and [Donsker and Varadhan 1976], each one raising the level[1] of abstraction and, at the same time, revealing more fundamental principles. However, they did not content themselves with general theory. On the contrary, they applied their theory to solve a remarkably varied set of problems, ranging from questions about the range of a random walk in [Donsker and Varadhan 1979] to questions coming from mathematical physics about function-space integrals in [Donsker and Varadhan 1976] and [Donsker and Varadhan 1981], with each abstraction designed to tackle a specific problem.

1: Among their disciples, these levels are designated as levels I, II, and III. Because it reminds me, depending on my mood, of the voice one hears on the elevator as one ascends from "boys clothing" to "women's lingerie" at Macy's or of the levels of hell through which Dante descended, I have never subscribed to this terminology.

As is nearly always the case when breaking new ground, the applications required ingenious modifications of the general theory. To give an indication of just how ingenious these modifications had to be, consider the "Wiener sausage" calculation in [Donsker and Varadhan 1975a]. The problem there, which grew out of a question posed by the physicist J. Luttinger, was to find the asymptotic volume of the tubular neighborhood of a Brownian path as the time goes to infinity and the diameter of the neighborhood goes to 0. If one thinks about it, one realizes that this volume can be computed by looking at a neighborhood in the space of measures of the empirical distribution. However, the neighborhood that one needs is the one determined by the variation norm, whereas their general theory deals with the weak topology. Thus, except in one dimension where local time comes to the rescue, they had to combine their general theory with an intricate approximation procedure in order to arrive at their goal. Their calculation in [Donsker and Varadhan 1976] is a true *tour de force*, only exceeded by their solution to the polaron problem in [Donsker and Varadhan 1981].

In conclusion, it should be emphasized that Varadhan's contributions to the theory of large deviations were to both its foundations and its applications. Because of his work, the subject is now seen as a basic tool of analysis, not simply an arcane branch of probability and statistics. With 20/20 hindsight, it has become clear that large deviations did not always provide the most efficient or best approach to some of the problems which he solved, but there can be no doubt that his insights have transformed the field forever.

References

[Donsker and Varadhan 1975a] M. D. Donsker and S. R. S. Varadhan, "Asymptotics for the Wiener sausage", *Comm. Pure Appl. Math.* **28**:4 (1975), 525–565. MR 0397901 (53 #1757a)

[Donsker and Varadhan 1975b] M. D. Donsker and S. R. S. Varadhan, "On a variational formula for the principal eigenvalue for operators with maximum principle", *Proc. Nat. Acad. Sci. U.S.A.* **72** (1975), 780–783. MR 0361998 (50 #14440)

[Donsker and Varadhan 1975c] M. D. Donsker and S. R. S. Varadhan, "Asymptotic evaluation of certain Markov process expectations for large time. I & II", *Comm. Pure Appl. Math.* **28** (1975), 1–47 & 279–301. MR 0386024 (52 #6883)

[Donsker and Varadhan 1976] M. D. Donsker and S. R. S. Varadhan, "Asymptotic evaluation of certain Markov process expectations for large time. III", *Comm. Pure Appl. Math.* **29**:4 (1976), 389–461. MR 0428471 (55 #1492)

[Donsker and Varadhan 1979] M. D. Donsker and S. R. S. Varadhan, "On the number of distinct sites visited by a random walk", *Comm. Pure Appl. Math.* **32**:6 (1979), 721–747. MR 539157 (81j:60080)

[Donsker and Varadhan 1981] M. D. Donsker and S. R. S. Varadhan, "The polaron problem and large deviations", *Phys. Rep.* **77**:3 (1981), 235–237. New stochasitic methods in physics. MR 639028

[Varadhan 1967] S. R. S. Varadhan, "Diffusion processes in a small time interval", *Comm. Pure Appl. Math.* **20** (1967), 659–685. MR 0217881 (36 #970)

Daniel W. Stroock

dws@math.mit.edu

Department of Mathematics
Massachusetts Institute of Technology
Cambridge, MA 02139
USA