BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 40, Number 1, Pages 89–97
S 0273-0979(02)00968-0
Article electronically published on October 16, 2002

STOCHASTIC ANALYSIS AND APPLICATIONS

S. R. S. VARADHAN

1. Introduction

The world we live in has never been very predictable, and randomness has always been part of our lives. There is ample evidence that our ancestors did enjoy playing games of chance, and the early development of probability theory had to do with these games. The study of probability has always been motivated by potential applications. Besides gambling, the stimulus has come from life insurance, pricing of annuities, statistical modeling of errors in observations, genetics, etc.

By 1900, there had developed a fairly well understood body of work, although it was not within an axiomatic framework. This prompted Hilbert to pose the axiomatization of probability theory as one of his famous problems. Although it was not until 1933 in [13] that Kolmogorov axiomatized probability theory by making it part of measure theory, the subject continued to blossom in the early part of the twentieth century at the hands of people like Lévy, Khintchin, and Wiener.

2. Infinite dimensional analysis

Wiener, in [27], was the first person to construct a measure, corresponding to what we now call Brownian motion, on the space of R^d valued continuous functions on $[0,1]$, thereby making it possible to integrate legitimately in function spaces.

The basic ingredient in the construction of the Wiener measure is the kernel

$$p(t,x,y) = \frac{1}{(2\pi t)^{\frac{d}{2}}} \exp\left[-\frac{\|y-x\|^2}{2t}\right]$$

and its relationship to the Wiener measure P through the formula

$$P[x(\cdot) : x(t_1) \in A_1, \ldots, x(t_n) \in A_n]$$
$$= \int_{A_1} \cdots \int_{A_n} p(t_1,0,x_1)p(t_2-t_1,x_1,x_2)\cdots p(t_n-t_{n-1},x_{n-1},x_n)dx_1\cdots dx_n.$$

The function p is of course the fundamental solution of the heat equation

$$\frac{\partial p}{\partial t} = \frac{1}{2}\Delta p,$$

and this creates a basic connection between the Wiener measure and the Laplace operator. Kolmogorov observed, in [12], that this relationship can be extended to a large class of probability measures (Markov or diffusion processes) that generalize

Received by the editors September 7, 2001, and, in revised form, February 21, 2002.
2000 *Mathematics Subject Classification.* Primary 60-02.

the Wiener measure on the one hand, and a large class of differential operators (second order parabolic) that generalize the heat equation on the other:

$$\frac{\partial p}{\partial t} = \frac{1}{2}\sum_{i,j} a_{i,j}(x)\frac{\partial^2 p}{\partial x_i \partial x_j} + \sum_j b_j(x)\frac{\partial p}{\partial x_j}$$

or

$$\frac{\partial p}{\partial t} = \frac{1}{2}\sum_{i,j} \frac{\partial^2}{\partial y_i \partial y_j}[a_{i,j}(y)p] - \sum_j \frac{\partial}{\partial y_j}[b_j(y)p].$$

These are referred to respectively as Kolmogorov's backward and forward equations. Itô, in 1944, described a mapping of the space of trajectories that will directly transform the Wiener measure into the more general measure that corresponds to a given operator

$$\mathcal{L} = \frac{1}{2}\sum_{i,j} a_{i,j}(x)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_j b_j(x)\frac{\partial}{\partial x_j}.$$

This transformation is defined in [9] by means of a system of stochastic differential equations

$$dy_i(t) = \sum_j \sigma_{i,j}(y(t))dx_j(t) + b_i(y(t))dt$$

with $y(0) = z$, where

$$\sigma(x)\sigma^*(x) \equiv a(x).$$

One big advantage of Itô's construction is that it bypasses the PDE, thereby opening up the possibility of saying something about the solution of the PDE by studying the properties of the SDE.

It is not as simple as it sounds. The functions $x_j(t)$ are not of bounded variation (almost surely with respect to the Wiener measure), and therefore even in integrated form it is a task to make sense of the dx integrals

$$y_i(t) = z_i + \int_0^t \sum_j \sigma_{i,j}(y(s))dx_j(s) + \int_0^t b_i(y(s))ds.$$

Itô developed his theory of stochastic integration to take care of this sticky point, including a stochastic calculus with special rules

$$dx_i(t)dx_k(t) = \delta_{i,k}dt,$$
$$dx_i(t)dt = 0,$$

and

$$dy_i(t)dy_j(t) = a_{i,j}(y(t))dt.$$

In integrated form the identity

$$f(y(t)) - f(y(0)) = \int_0^t \langle \sigma^*(y(s))\nabla f(y(s)), dx(s)\rangle + \int_0^t (\mathcal{L}f)(y(s))ds$$

is known as Itô's formula. An advantage of Itô's approach is that the matrix of coefficients $\{a_{i,j}(x)\}$ can be semi-definite. Often in PDE, nondegeneracy is needed.

Another big advantage of Itô's approach is the possibility of dealing with infinite systems just as easily as finite ones. With suitable regularity and proper definition of norms, this can be realized. The PDE, on the other hand, is hard to study directly in infinite dimensions.

By a combination of methods from analysis (PDE) and probability (SDE), diffusion processes in finite dimensional spaces were thoroughly studied during 1950–1980.

The Wiener measure provides us with a reasonable integration theory in the infinite dimensional space $\Omega = C_0[[0,1]; R^d]$. However it is not invariant with respect to translations on Ω, and in fact is not even quasi-invariant with respect to all translations. It is however quasi-invariant with respect to translations from the following dense subspace $\mathcal{H} \subset \Omega$:

$$\mathcal{H} = \left\{ h : h(t) = \int_0^t f(s)ds, \int_0^1 \|f(s)\|^2 ds < \infty \right\}.$$

One can therefore talk about the smoothness of a function, defined only almost everywhere, in the direction $h \in \mathcal{H}$. The gradient operator ∇_h exists, and it makes sense to talk of $\|\nabla_h u\|_p$ existing in the L_p sense. Since $\mathcal{H}$ has a natural Hilbert structure, it also makes sense to talk about the squared gradient

$$\sum_i \|\nabla_{e_i} u\|_2^2,$$

where $\{e_i\}$ is an orthonormal basis of $\mathcal{H}$. There are natural Sobolev spaces and corresponding spaces of distributions. Since we are in infinite dimensions, no amount of L_p regularity will yield continuity, but a very nice theory of differential analysis can still be built up. This was initiated by Malliavin in [16] and developed further by others, including Kusuoka and Stroock in [14], [15], and [20].

A particularly interesting byproduct is Hörmander's theory of hypoellipticity, which can be explained in terms of differential calculus on Wiener space and an integration by parts formula. Using Itô's theory, one can write the fundamental solution $p(t,z,y)$ of

$$\frac{\partial}{\partial t} - \mathcal{L}$$

as

$$E^P[\delta(y_z(t) - y)].$$

The idea is to differentiate as many times as we need to with respect to z and y, integrate by parts to get rid of unwanted derivatives, and express the answers as Wiener integrals and estimate them.

The general method bears the name of Malliavin calculus and is an important tool in the study of regularity of maps defined as stochastic integrals from one infinite dimensional space to another. Clearly the study of infinite dimensional spaces locally modeled after Wiener spaces will rely heavily on such a calculus.

One area from where problems are likely to come is nonlinear analysis in an infinite dimensional setting.

Probability theory has always generated its problems by its contact with other areas. There are very few problems that are generated by its own internal structure. This is partly because, once stripped of everything else, a probability space is essentially the unit interval with Lebesgue measure. If you have seen one you have seen them all.

3. Large discrete random structures

Perhaps the most studied example of a discrete random structure is the Ising model [8]. In its simplest version we have a finite lattice $\Lambda_N \subset Z^d$, a state space

$\Omega_N = \{-1,1\}^{\Lambda_N}$ of maps $\xi(\cdot) : \Lambda_N \to \{-1,1\}$, and a probability measure depending on a small number of parameters. Given the two parameters J and H, the Ising measure on Ω_N is the probability measure defined by

$$p^g_{N,J,H}(\xi) = \frac{1}{Z^g_N(J,H)} \exp[-JE^g_N(H,\xi)],$$

where

$$E^g_N(H,\xi) = \sum_{\substack{x,y\in\Lambda_N \\ |x-y|=1}} |\xi(x)-\xi(y)|^2 + 2 \sum_{\substack{x\in\Lambda_N \\ y\notin\Lambda_N \\ |x-y|=1}} |\xi(x)-g(y)|^2 + H \sum_{x\in Z_N} \xi(x).$$

The interesting questions concern the behavior of the measures p_N on Ω_N as $N \to \infty$ and the absence or persistence of the influence of the boundary conditions g, after the infinite volume limit has been taken.

This is an area that has been very well studied as a statistical model for phase transitions. While a lot is known, there are still some basic issues that are unresolved, even in the simplest model. There are issues of universality, etc., that are still open. The infinite volume limits of these types of models provide natural examples of random fields in Z^d that are correlated. The recent results concerning surface tension and the Wulff construction are excellent examples of interesting detailed analysis of such systems in the strongly correlated region.

The subject of random graphs had its originals in a paper of Erdös and Rényi [3]. The basic model is of a graph with n vertices in which some randomly chosen edges are turned on while the others are shut down. The subtle interplay between the probability with which the individual edges are turned on and the qualitative behavior of the resulting graph for large values of n has been a fascinating subject for study ever since the paper of Erdös and Rényi.

Percolation. If we take the standard square lattice in Z^d with $d \geq 2$ and select the nearest neighbor edges independently with probability p, there is a qualitative change in the nature of the resultant infinite graph as p changes. For small p the graph is just a collection of finite clusters, whereas for p large enough, with probability 1, there is an infinite cluster. There are detailed studies regarding the nature of this transition at a critical value $0 < p_c < 1$ of p. See [5].

The random cluster models of Fortuin and Kasteleyn [4] provide a unified point of view, connecting percolation and Ising type models.

Recently there have been exciting developments [24], [1], and [11] in situations that appear at first glance to be different, but lead to the same phenomenon. These have to do with (non-Gaussian) asymmetric fluctuations of certain extremal problems.

1. Random matrix theory. We look at the largest eigenvalue λ_N of a random symmetric matrix with an orthogonally invariant Gaussian distribution.

2. We look at the length l_N of the longest increasing subsequence of a random permutation of $1, 2, \ldots, N$.

3. Random growth models. We have a collection of independent random variables $X_{i,j}$ at each site in Z^2 and a path π connecting $(0,0)$ and (M,N) that moves only either up or to the right, and we define

$$G(M,N) = \max_{\pi} \sum_{(i,j)\in\pi} X_{i,j}.$$

All these random variables exhibit non-Gaussian asymmetric fluctuations and after rescaling have the same limiting distribution, known as the Tracy-Widom distribution. Its distribution function is given by

$$F(s) = \exp\left[-\int_s^\infty (x-s)[u(x)]\right]^2 dx,$$

where u is the solution of the Painlevé II equation

$$u'' = 2u^3 + xu$$

with $u \simeq \mathrm{Ai}(x)$ as $x \to \infty$.

The theory, as developed by P. Deift and others, depends on exact formulas and a precise analysis of the asymptotic behavior of solutions of associated Riemann-Hilbert problems.

To a probabilist there is a mystery here, and it is very likely that these results will one day be part of a larger class of limit theorems.

Disordered systems. Random environments, etc. One of the earliest examples of a problem involving disorder is a random walk in a random environment. Let us take a family of independent random variables π_j, $-\infty < j < \infty$, with $0 < \pi_j < 1$, and consider a random walk on Z with transition probabilities $p_{x,x+1} = \pi_x$ and $p_{x,x-1} = 1 - \pi_x$. The random choices are made once and fixed for all time. The asymptotic behavior of such a walk can be studied in detail for the one dimensional nearest neighbor case. Other problems of a similar nature range from the difficult to the impossibly difficult.

Roughly speaking, these problems deal with probability measures $\mu_{\mathbf{a}}$, where $\mathbf{a}$ is a large or an infinite set of parameters $\{a_x\}$ indexed by $x \in \Lambda$. If $a_x \equiv a$ (a constant), the situation is well understood. In problems of disorder, $\{a_x\}$ are assumed to be random, chosen independently for each x and with a common distribution. One is interested in properties that are shared by $\mu_{\mathbf{a}}$ for almost all $\mathbf{a}$. A random walk in a random environment is one example. See [21].

The Sherrington-Kirkpatrick or S-K model consists of a family of probability measures $\mu_N = \mu_N^{\beta,h,\{g_{i,j}\}}$ on $\{0,1\}^N$ depending on parameters $\beta, h, \{g_{i,j}\}$:

$$\mu_N\{\epsilon\} = \frac{1}{Z_N}\exp[-\beta H_N(\epsilon)],$$

where

$$H_N(\epsilon) = -\frac{1}{\sqrt{N}}\sum_{1\le i<j\le N} g_{i,j}\epsilon_i\epsilon_j + h\sum_{1\le i\le N}\epsilon_i;$$

$\{g_{i,j}\}$ is a realization of i.i.d. standard Gaussian random variables.

If $h = 0$ and $\beta < 1$, the direction of the vector ϵ is rather random, so that

$$E^{\mu_N\times\mu_N}\left[\frac{(\epsilon.\epsilon')^2}{N^2}\right]$$

is small for most $\{g_{i,j}\}$. This is not expected to be true for $\beta > 1$, although what is known is a slightly weaker form of the result for $\beta \gg 1$.

This is just one of several similar problems. See the recent work of Talagrand [22] and [23] for detailed comments.

The similar problem in the case of short range interactions, where

$$H_N(\epsilon) = -\sum_{(x,y):|x-y|=1} g_{x,y}\epsilon_x\epsilon_y + h\sum_x \epsilon_x$$

and the summation extends over nearest neighbors in Z^d, is even less understood. See for instance the article [17] by Newman and Stein.

4. Scaling limits

Another area that has been active is the dynamical behavior of large or infinite interacting systems. The evolution may have conserved quantities, in which case the evolution of slow modes can be studied under hydrodynamical scaling, averaging out the fast modes with respect to appropriate invariant distributions.

Fairly general methods involving entropy, rate of entropy production and relative entropy have been developed.

In the case of systems without conserved quantities, one of the issues is whether the rate of approach to equilibrium slows down as the system size increases. This can be measured in terms of either the spectral gap or the log-Sobolev constant. Conditions for these estimates to hold uniformly are known. They can be used to estimate the accuracy of simulations. See Stroock and Zegarliński [19].

In the presence of conserved quantities, the spectral gaps and the log-Sobolev constants become degenerate with increasing volume, and the rate can be guessed by looking at the slow modes. That this is in fact the case is a useful tool in the analysis of these systems and is known for several models. See Yau [29], or Cancrini and Martinelli [2].

A typical class of models that illustrates the scaling behavior is known as simple exclusion processes. See [25], [26] and [28].

One has, as physical space, either Z^d or Z_N^d, the periodic lattice in Z^d of N^d elements. Each site can have either one particle or no particle. The state is described by $\{\eta(x) : x \in Z_N^d\}$.

The dynamics is specified by the generator

$$(\mathcal{A}f)(\eta) = \sum_{x,y} \eta(x)(1-\eta(y))p(y-x)[f(\eta^{x,y}) - f(\eta)].$$

The scaling $x \to \frac{x}{N}$ of space imbeds Z_N^d into the d-torus $\mathcal{T}^d$. The micro-state η can be replaced by a macro-state

$$r_N(t,\cdot) = \frac{1}{N^d}\sum_x \delta_{\frac{x}{N}},$$

which as $n \to \infty$ turns into the particle density $\rho(\theta)$.

Assuming that the initial configuration leads to the macro-state $\rho_0(\theta)$, the dynamical behavior of the large system should be describable in terms of $\rho_0(\cdot)$.

There are two cases. If $m = \sum zp(z)$, either $m = 0$ or $m \neq 0$. Let us first look at $m \neq 0$. The special case $d = 1$, $p(1) = 1$ is typical. In this case one needs to speed up the time scale by a factor of N as well, and the micro-state at time t is, with probability nearly 1, close to the macro-state $\rho(t,\theta)$, which is the unique entropic solution of the Burgers equation

$$\rho_t + [\rho(1-\rho)]_\theta = 0.$$

The secondary issues of probabilities of large deviations can also be studied for this model. See Jensen [10]. Given $\rho(\cdot,\cdot)$, the question of estimating the probability

$$P[r_N(\cdot,\cdot) \simeq \rho(\cdot,\cdot)] = \exp[-N\, I(\rho(\cdot,\cdot)) + o(N)]$$

with an explicit I has a simple answer: $I(\rho(\cdot,\cdot)) = \infty$ unless $\rho(\cdot,\cdot)$ is also a weak solution, and then $I(\rho(\cdot,\cdot))$ is the precise amount by which the entropy condition is violated for the convex functional

$$h(\rho) = \rho \log \rho + (1-\rho)\log(1-\rho);$$

that is,

$$I(\rho(\cdot,\cdot)) = \int_0^T \int [[h(\rho)]_t + [g(\rho)]_x]^+ dxdt.$$

In particular, one can show that

$$\lim_{N\to\infty} \frac{1}{N} \log E[\exp\{hN(t)\}] = \begin{cases} \frac{h}{4} & \text{if } h \geq 0, \\ \tanh \frac{h}{4} & \text{if } h \leq 0, \end{cases}$$

and one can compare the fluctuation theory with the large deviation estimate.

The case $m = 0$ exhibits more complex behavior. There exists $A(\rho)$ such that the equality

$$\rho_t = \frac{1}{2} \nabla A(\rho) \nabla \rho$$

describes the evolution of the macroscopic state in the new time scale, which now needs to be speeded up by a factor of N^2. The calculation of $A(\rho)$ is in general rather complicated. However, if p is symmetric, $A(\rho)$ simplifies to $A = \sum z \otimes z p(z)$.

Again the problems of large deviation as well as the motion of a tracer particle have satisfactory solutions. $S(\rho)$ is the self-diffusion coefficient in equilibrium, i.e., $\frac{x(t)}{\sqrt{t}} \simeq \beta$, and β has dispersion $S(\rho)$. Instead of just looking at the evolution of the macro-state, we may wish to consider the random process

$$R_N = \frac{1}{N^d} \sum_j \delta_{x_j(\cdot)},$$

which describes the collective history of all the particles. The limit

$$\lim_{N\to\infty} R_N = Q$$

exists. Both Q and the rate function $I(R)$ in the large deviation estimate

$$P[R_N \simeq R] = \exp[-N^d I(R) + o(N)]$$

have explicit exact descriptions. See Quastel, Rezakhanlou, and Varadhan [18].

The models that we have described have a fairly regular structure. The physical space is the lattice Z^d with connections between nearby sites. The rates are translation invariant.

One can imagine disordered versions of these problems. There have been some studies where the disorder is only in the rates. See [6] and [7]. But a more challenging problem is when the underlying structure itself is some sort of a random graph.

Whereas problems coming from physics present a more regular structure, problems from the social sciences present far more disorder.

Today more sophisticated models are being used in finance, insurance and other areas involving an analysis of uncertainty and risk. If we wish to fully understand

how large systems behave and why, one possible approach is to specify the model at the microscopic level and develop methods that allow us to make predictions of macroscopic behavior.

References

1. Jinho Baik, Percy Deift, and Kurt Johansson, *On the distribution of the length of the longest increasing subsequence of random permutations*, J. Amer. Math. Soc. **12** (1999), 1119–1178. MR **2000e:**05006
2. N. Cancrini and F. Martinelli, *On the spectral gap of Kawasaki dynamics under a mixing condition revisited. Probabilistic techniques in equilibrium and nonequilibrium statistical physics*, J. Math. Phys. **41** (2000), 1391–1423.
3. P. Erdös and A. Renyi, *On the evolution of random graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **5** (1960), 17–61. MR **23:**A2338
4. C. M. Fortuin and P. W. Kasteleyn, *On the random cluster model. I: Introduction and relation to other models*, Physica **59** (1972), 535–564. MR **50:**12107
5. Geoffrey Grimmett, *Percolation*, 2nd ed., Springer-Verlag, Berlin, 1999. MR **2001a:**60114
6. Alice Guionnet and Boguslaw Zegarliński, *Decay to equilibrium in random spin systems on a lattice*, Comm. Math. Phys. **181** (1996), 703–732. MR **97j:**82105
7. ———, *Decay to equilibrium in random spin systems on a lattice. II*, J. Statist. Phys. **86** (1997), 899–904. MR **98h:**82046
8. E. Ising, *Beitrag zur Theorie des Ferromagnitismus*, Z. Phys. **31** (1926), 253–258.
9. Kiyosi Itô, *Stochastic integral*, Proc. Imp. Acad. Tokyo **20** (1944), 519–524. MR **7:**313c
10. Leif Jensen, *Large deviations of the asymmetric simple exclusion process*, Ph.D. thesis, Courant Inst., New York Univ., New York, 2000.
11. Kurt Johansson, *Shape fluctuations and random matrices*, Comm. Math. Phys. **209** (2000), 437–476. MR **2001h:**60177
12. A. N. Kolmogorov, *Über die analyttischen Methoden in der Wahrscheinlichkeitsrechnung*, Math. Ann. **104** (1931), 415–458.
13. ———, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer-Verlag, Berlin, 1933.
14. S. Kusuoka and D. Stroock, *Applications of the Malliavin calculus*. II, J. Fac. Sci. Univ. Tokyo Sect. IA Math. **32** (1985), 1–76. MR **86k:**60100b
15. ———, *Applications of the Malliavin calculus*. III, J. Fac. Sci. Univ. Tokyo Sect. IA Math. **34** (1987), 391–442. MR **89c:**60093
16. Paul Malliavin, *Stochastic calculus of variation and hypoelliptic operators*, Proc. Internat. Sympos. Stochastic Differential Equations (Res. Inst. Math. Sci., Kyoto Univ., Kyoto, 1976), Wiley, New York, 1978, pp. 195–263. MR **81f:**60083
17. C. M. Newman and D. L. Stein, *Equilibrium pure states and nonequilibrium chaos*, J. Statist. Phys. **94** (1999), 709–722. MR **2000d:**82040
18. Jeremy Quastel, Fraydoun Rezakhnalou, and S. R. S. Varadhan, *Large deviations for the symmetric simple exclusion process in dimensions* $d \geq 3$, Probab. Theory Related Fields **113** (1999), 1–84. MR **2000c:**60162
19. D. Stroock and B. Zegarliński, *On the ergodic properties of Glauber dynamics*, J. Statist. Phys. **81** (1995), 1007–1019. MR **96i:**60109
20. Daniel Stroock, *Applications of the Malliavin calculus*. I, Stochastic Analysis (Katata/Kyoto, 1982), North-Holland Math. Library, vol. 32, North-Holland, Amsterdam, 1984, pp. 271–306. MR **86k:**60100a
21. Alain-Sol Sznitman, *Brownian motion, obstacles and random media*, Springer-Verlag, Berlin, 1998. MR **2001h:**60147
22. Michel Talagrand, *Huge random structures and mean field models for spin glasses*, Proc. Internat. Congr. Math. (Berlin, 1998), Vol. I, Doc. Math. Extra Vol. I (1998), 507–536 (electronic). MR **2000c:**60164
23. ———, *The Sherrington-Kirkpatrick model: a challenge for mathematicians*, Probab. Theory Related Fields **110** (1998), 109–176. MR **99a:**60117
24. Craig A. Tracy and Harold Widom, *Level spacing distributions and the Bessel kernel*, Comm. Math. Phys. **161** (1994), 289–309. MR **95e:**82004
25. S. R. S. Varadhan, *The complex story of simple exclusion*, Itô's Stochastic Calculus and Probability Theory, Springer, Tokyo, 1996, pp. 385–400. MR **98g:**60058

26. ———, *Large deviations for interacting particle systems*, Perplexing Problems in Probability, 373–383, Progr. Probab., vol. 44, Birkhauser Boston, Boston, MA, 1999, pp. 373–383. MR **2001h:**60044
27. Norbert Wiener, *Differential space*, J. Math. and Phys. **2** (1923), 131–174.
28. Horng-Tzer Yau, *Scaling limit of particle systems, incompressible Navier-Stokes equation and Boltzmann equation*, Proc. Internat. Congr. Math. (Berlin, 1998), Vol. III Doc. Math. Extra Vol. III, 193–202 (electronic). MR **2000h:**82061
29. ———, *Logarithmic Sobolev inequality for lattice gases with mixing conditions*, Comm. Math. Phys. **181** (1996), 367–408. MR **98e:**82021

Courant Institute, New York University, New York, New York 10012-1110
E-mail address: Varadhan@cims.nyu.edu