The Annals of Probability
1990, Vol. 18, No. 1, 431–437

TWO BOOKS ON THE MALLIAVIN CALCULUS

DENIS R. BELL, *The Malliavin Calculus*, Pitman Monographs and Surveys in Pure and Applied Mathematics **34**, Longman Scientific and Technical, New York, 1987, x + 105 pages, $64.95.

S. WATANABE, *Lectures on Stochastic Differential Equations and Malliavin Calculus*, Tata Institute of Fundamental Research Lecture Notes, Springer, New York, 1984, 111 pages, $15.00.

REVIEW BY DANIEL L. OCONE

Rutgers University

The Malliavin calculus is a recent development in stochastic process theory that provides a method for studying the probability densities of random vectors defined on Wiener space. Malliavin initiated the field [Malliavin (1978a, b)] in an effort to achieve a probabilistic proof that solutions to stochastic differential equations whose vector fields satisfy Hörmander's hypoellipticity conditions admit C^∞ probability densities. To explain, let $\xi(x, t, W)$ be the solution to

$$(1) \qquad d\xi(t) = X_0(\xi(t))\,dt + \sum_1^d X_i(\xi(t)) \circ dW^i(t), \qquad \xi(x,0) = x.$$

In (1), $\xi(t) \in R^n$, $X_i(\xi) = (X_{i1}(\xi), \dots, X_{id}(\xi))$, $0 \le i \le d$, are C^∞ vector fields on R^d, $\{(W^1(t), \dots, W^d(t)),\ t \ge 0\}$ is a Brownian motion, and $\circ\, dW$ denotes the Stratonovich differential. As usual, we interpret the vector fields as operator-valued, i.e., $X_i(\xi) = \sum_1^d X_{ij}(\xi) v\partial/\partial\xi_j$. We want to find conditions on (1) that guarantee the existence and regularity of the density for $\xi(t)$. The traditional approach to this question relies on the theory of partial differential equations. Let $P_x(t, \cdot\,)$ denote the probability distribution of $\xi(x, t, W)$, that is, $P_x(t, A) = \text{Prob}(\xi(x, t, W) \in A)$. $P_x(t, \cdot\,)$ is a fundamental solution in the sense of distributions to the Fokker–Planck equation

$$(2) \qquad \frac{\partial}{\partial t} P_x(t,\cdot) = \left[\frac{1}{2}\sum_1^d X_i^2(\xi) + X_0(\xi)\right]^* P_x(t,\cdot),$$

where L^* denotes the formal adjoint of an operator L. Therefore we can use p.d.e. theory to study the existence of a density $p_x(t, \xi)$ satisfying $p_x(t, \xi)\,d\xi = P_x(t, d\xi)$. However there is also a differential geometric way to look at the problem. Let $[\cdot,\cdot\,]$ denote the Lie bracket between vector fields defined by $[X_i, X_j] = X_i \circ X_j - X_j \circ X_i$. Let Λ be the smallest linear space of vector fields closed under $[\cdot,\cdot\,]$ and containing $X_1, \dots, X_d$ and also $[X_0, X_1], \dots, [X_0, X_d], [X_0, [X_0, X_1]], \dots, [X_0, [X_0, X_d]]$, etc., that is, Λ is the ideal generated by $X_1, \dots, X_d$ in the Lie algebra generated by $X_0, \dots, X_d$. Let $\Lambda(\xi)$ denote

Received July 1989.

the vector fields of Λ at ξ. Roughly speaking, $\Lambda(x)$ represents the set of possible directions in which $\xi(x, dt, W)$ moves transverse to the flow of X_0 in an infinitesimal time, as W ranges over Wiener paths. Therefore, if the dimension of $\Lambda(x) = n$, $\xi(t)$ diffuses in all directions in R^n and this should have a smoothing effect leading to existence and regularity of densities. Such reasoning suggests the following result.

THEOREM 1. *If dimension* $\Lambda(x) = n$, $p_x(\xi, t)$ *exists and is a* C^∞ *function.*

The condition that dimension $\Lambda(\xi) = n$ for all ξ is Hörmander's celebrated condition for hypoellipticity of the operator $\partial/\partial t - (\Sigma_1^d X_i^2 + X_0)$, and one can prove a weaker version of Theorem 1 using Hörmander's theorem. Malliavin's achievement was to show how to prove Theorem 1 by a stochastic argument making use of the differential geometric intuition. Moreover, in so doing he invented a strategy for exploring existence and regularity of densities that applies in principle to any random vector defined on Wiener space, not just to solutions of stochastic d.e.'s.

Malliavin's work generated considerable interest and inspired a number of researchers, most notably Stroock, Kusuoka, Watanabe and Bismut, who took up Malliavin's ideas and developed them into their present form. The achievements of the field to date include a complete probabilistic proof of Hörmander's theorem for second order operators and numerous new applications to problems which did not submit to techniques from pure analysis, for example, existence of densities for infinite systems of stochastic d.e.'s, existence of densities for multiple Wiener integrals, regularity of conditional densities in nonlinear filtering, regularity of Poisson kernels for degenerate elliptic operators, extensions of the theory to Poisson driven stochastic d.e.'s, and connections to large deviation analysis on Wiener space. This activity has also inspired more widespread interest in functional analysis on infinite-dimensional Gaussian spaces, which lies behind the Malliavin calculus. For example, Nualart and Pardoux (1988) have recently developed the theory of the Skorohod stochastic integral, which generalizes Itô's integral to nonadapted integrands, using the functional analytic setup of the Malliavin calculus.

The fundamental criterion for existence and regularity of densities can be easily stated and motivated. Consider first a finite-dimensional, nonstochastic version of the problem. Let F: $R^m \to R^n$, where $m > n$, and let λ^k denote Lebesgue measure on R^k. We want a condition implying that $\lambda^m \circ F^{-1}$ be absolutely continuous with respect to λ^n. To this end we introduce the co-area formula, a result well known in geometric measure theory. Let $DF(x) = [\partial F_i/\partial x_j]_{1 \le i \le n, 1 \le j \le m}$ denote the differential of F. Then, with suitable hypothesis, if $A \subset R^n$,

$$(3) \quad \int 1_{\{F(x) \in A\}} \sqrt{\det DF(x)\, DF^T(x)}\, \lambda^m(dx) = \int_A \mathscr{H}^{m-n}(F^{-1}(y))\lambda^n(dy),$$

where $\mathscr{H}^{m-n}(F^{-1}(y))$ is the $(m-n)$-dimensional Hausdorff measure of $F^{-1}(y)$. It follows that

$$(4) \qquad \det DF(x)\, DF^T(x) > 0 \quad \text{for a.e. } x$$

implies that $\lambda^m \circ F^{-1} \ll \lambda^n$. Indeed, if (4) holds, then from (3), $\lambda^n(A) = 0$ must imply that $\int 1_{\{F(x) \in A\}} \lambda^m(dx) = \lambda^m \circ F^{-1}(A) = 0$ also.

The principal criterion of the Malliavin calculus is a generalization and refinement of condition (4). We replace R^m by $C_0[0, T]$, the space of continuous paths on $[0, T]$ starting at 0, and λ^m by Wiener measure, which we denote by μ. If F: $C_0[0, T] \to R^n$, the probability distribution of F is $\mu \circ F^{-1}$; thus, asking that F admit a probability density is equivalent to asking that $\mu \circ F^{-1} \ll \lambda^n$. To generalize (4) requires a notion of differentiability of Wiener functionals, but here we must be subtle. We might at first try working with functionals F that are Fréchet differentiable with respect to the sup norm topology on $C_0[0, T]$. However, such a theory would not cover many interesting cases. For example, solutions $\xi(x, t, W)$ to (1) will not in general even be continuous functions in W if the vector fields $X_1, \dots, X_d$ do not commute, that is, if $[X_i, X_j] \neq 0$ for some $1 \leq i, j \leq d$. Moreover, having differentiability of Wiener functionals in arbitrary directions in $C_0[0, T]$, as in Fréchet differentiable functionals, makes no sense probabilistically because F is defined only up to sets of Wiener measure 0; hence, for a fixed continuous function h, $F(W + \varepsilon h)$ is well defined only if $W(\cdot) + \varepsilon h(\cdot)$ is absolutely continuous with respect to Wiener measure, that is, only if $\mu(A) = 0$ implies $\mu(W + \varepsilon h \in A) = 0$ for any set A. But this can be the case for all ε if and only if h is in the space H of continuous function satisfying

$$(5) \qquad h(t) = \int_0^t h'(s)\, ds \quad \text{where} \quad \int_0^T |h'(s)|^2\, ds < \infty.$$

(5) is the key to the proper concept of differentiability. Endow H with the inner product $\langle h, g \rangle = \int_0^T h'(s) \cdot g'(s)\, ds$. Very roughly, if G: $C_0[0, T] \to R$, G is said to be differentiable if there is an H-valued $DG(W)$ such that $\lim_{\varepsilon \to 0} \varepsilon^{-1}[G(W + \varepsilon h) - G(W)] = \langle DG(W), h \rangle$ for all $h \in H$. Similarly, one can define higher order derivatives $D^2G = D(DG)$, D^3G, etc. Of course, this definition of D is not rigorous, but it conveys the idea. Clearly, D defines a notion of derivative much weaker than that of Fréchet. For example, the solution $\xi(t)$ to (1), even though it may not be continuous in W, will be differentiable with respect to D.

Now suppose that each component of F: $C_0[0, T] \to R^n$ is differentiable. The analogue of the matrix appearing in (4) is the nonnegative definite matrix $[\langle DF_i, DF_j \rangle]_{\{1 \leq i, j \leq n\}}$, which is called the Malliavin covariance matrix. In analogy to (4), almost sure positivity of the Malliavin covariance matrix will imply the existence of a density for F; see Bouleau and Hirsch (1986). This is the reason for the term "covariance matrix"; indeed, if F is the Gaussian solution of a linear stochastic d.e., then the Malliavin covariance matrix equals the covariance matrix of F. The most important achievement of the Malliavin calculus, however, is a condition for regularity of the density. It is shown that F admits a C^∞ density if

(H.1) F is infinitely differentable w.r.t. D, and F and all its derivatives are in $L^p(\mu)$ for all $p > 1$,

(H.2) $(\det[\langle DF_i, DF_j\rangle])^{-1} \in L^p(\mu)$ for all $p > 1$.

Thus, given an F on Wiener space, such as the solution $\xi(x, t, W)$ of (1), one calculates the Malliavin covariance matrix and tries to verify (H.1) and (H.2) in order to prove existence of a smooth density. This is the strategy behind the probabilistic proof of Theorem 1. The task of proving (H.1) and (H.2) for a particular example is usually very difficult and analytically involved. It is this stage of the subject that gives it its reputation for complexity and difficulty.

The Malliavin calculus is not one unified theory: Broadly speaking there are two approaches, which we shall call, respectively, the Stroock–Shigekawa approach and the Bismut approach, after their main developers; see Stroock (1981), Shigekawa (1980) and Bismut (1981). Although we motivated the Malliavin covariance matrix with the co-area formula, the common point of departure for both approaches is a different theorem from analysis. [However, for a direct connection to the co-area formula, see Bouleau and Hirsch (1986).] If ν is a finite measure on R^n, then $d\nu/d\lambda^n$ exists and $d\nu/d\lambda^n \in C^{N-n-1}(R^n)$ if there exists a constant C such that

$$\left| \int \frac{\partial^{|\alpha|}\phi(x)}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}} \nu(dx) \right| \le C \sup_x |\phi(x)| \tag{6}$$

for all infinitely differentiable functions ϕ with compact support and for all multiindices $\alpha = (\alpha_1, \ldots, \alpha_n)$ such that $|\alpha| \le N$. Suppose that $\nu = \mu \circ F^{-1}$ is the probability distribution of a Wiener functional. Then the integral in the l.h.s. of (6) may be written

$$\int_{C_0[0,T]} \frac{\partial^{|\alpha|}\phi(F(W))}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}} \mu(dW) = E\left[\frac{\partial^{|\alpha|}\phi(F)}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha^n}} \right].$$

The strategy of the Malliavin calculus is to prove inequalities like (6) by somehow integrating by parts over Wiener space in order to remove the derivatives from ϕ, and the Stroock–Shigekawa and Bismut approaches distinguish themselves from each other by the way they do this. The Stroock–Shigekawa approach is functional analytic and builds on the theory of analysis on infinite-dimensional Gaussian spaces developed by Gross, Kree, Kuo and others. The Itô–Wiener decomposition of $L^2(\mu)$ and a generalization of the Ornstein–Uhlenbeck operator L to $L^2(\mu)$ play a prominent role in the theory and they are used to define Sobolev spaces of differentiable Wiener functionals. Integration by parts is then accomplished by using the Ornstein–Uhlenbeck operator, which is self-adjoint, and the relation $2\langle DG, DG\rangle = L(FG) - GL(F) - FL(G)$. Alternatively, following the book of Watanabe under review, one can define an adjoint δ to D satisfying $E[(\langle DF, DG\rangle] = E[F\delta(DF)]$; the two methods are equivalent because $L = -\delta D$. To see how integration by parts works, let F take values in R. Then $D(\phi(F)) = \phi'(F)DF$, and so $\phi'(F) = \langle DF, DF\rangle^{-1}\langle D(\phi(F)), DF\rangle$. By using δ, $E[\phi'(F)] = E[\phi(F)\delta(\langle DF, DF\rangle^{-1}DF)]$. We see immediately how nondegeneracy and differentiability of the Malliavin covariance matrix enter into the derivation of (6). The Bismut approach, on the other hand, relies on Girsanov's transformation. The idea is to represent the effects of a variation of the sample paths W in a random direction $\tilde{h}$ by a transformation of measures. In this way, derivatives on ϕ are transferred to functional derivatives on the Girsanov density. The Bismut approach is not used to prove the general criterion of existence of C^∞ densities sketched above. Rather it is used on a case by case basis. The trick is the proper choice of $\tilde{h}$, which is always the one that leads to the appearance of Malliavin's covariance matrix. The Bismut approach is particularly suited to extension to random functionals of Poisson processes.

The two books under review here are both introductions to the Malliavin calculus. Their contents overlap only partly and, taken together, they provide a fairly complete picture of the field as it stands today. The book by Watanabe contains the lecture notes of a course he gave at the Tata Institute in India in 1984. In it, he first develops the functional analysis on infinite-dimensional Gaussian spaces necessary for defining the Malliavin covariance matrix and proving conditions (H.1) and (H.2) for existence of C^∞ densities. His approach to (H.1) and (H.2) differs from those sketched above in that it does not use (6). But don't despair! It is not a completely different version of the Malliavin calculus. It is rather a variant of the functional analytic approach. Watanabe introduces D and defines Sobolev spaces of differentable functionals using the Ornstein–Uhlenbeck operator in the spirit of Stroock and Shigekawa. However, he then observes that where there are Sobolev spaces there are dual spaces of distributions, which in this context operate on Wiener functionals. Now let δ_y denote the Dirac delta function at $y \in R^n$. Watanabe shows that (H.1) and (H.2) imply that $\delta_y(F)$ can be made sense of as a distribution in one of these dual spaces, and that $y \to \delta_y(F)$ is C^∞ in an appropriate weak sense. We may thus define $p_F(y) = E[\delta_y(F)] = \delta_y(F)(1)$, where $\delta_y(F)(1)$ is the action of $\delta_y(F)$ on the function 1. It is intuitively reasonable that $p_F(y)$ is the density of F and so existence of a smooth density follows. The second part of the book is entirely taken up with the lengthy proof of Theorem 1, which Watanabe presents very clearly. The longest and most difficult step in this proof has recently been simplified by Norris (1986); Norris's proof appears in the book by Bell.

The style of Watanabe's book is very much "definition–theorem–proof" without a lot of casual motivation or example, and it is best read with some initial orientation to the field. However, it is clear, very complete and definitive. In particular, it contains a lengthy discussion of a difficult point in analysis on Wiener space having to do with the continuity of D between different order Sobolev spaces. Here, the treatment uses Meyer's inequalities, which are a very beautiful extension of Littlewood–Paley multiplier theory to Weiner space and which Watanabe treats fully. This material is not directly used in applications and the reader interested in the applications does not need to study it in detail.

The book by Bell is more introductory and broader in scope, and I think it is a good place to go for those seeking a first time introduction. The blurb on the jacket advertises "enough technical background to make the book accessible to readers without specialized knowledge in stochastic analysis." This is perhaps a slight exaggeration. Stochastic integration is reviewed in three brief pages! However, an elementary knowledge of stochastic integrals and stochastic differential equations and their connection to parabolic partial differential equations should suffice. Certainly anyone able to make some intelligent sense of this review will be well prepared to read Bell's book.

I think this book will be a successful introduction. Care is taken to motivate ideas and to get to the main results as directly and concisely as possible. Since its inception, the subject has undergone much simplification as it has become better understood, and the results of this process are evident in the book. Bell focuses on the main steps in the arguments without getting too involved in minor analytic details and thus he conveys the structure of the theory and brings out its important ideas very clearly.

Bell's book is also the most comprehensive survey of the field to date. He discusses and compares both the Stroock and the Bismut approaches and also reviews Malliavin's original paper. Bell does not treat Watanabe's method and, in fact, he does not define the operator D within the context of Wiener space functional analysis as in Watanabe. Instead, he follows a treatment of Stroock (1981) defining the Malliavin covariance matrix using the Ornstein–Uhlenbeck operator; recall that $2\langle DF, DG\rangle = L(FG) - GL(F) - FL(G)$. Bell also treats a variety of applications. Besides proving Theorem 1 using Norris's simplification, he also sketches an application of D. Michel to existence and regularity of conditional densities in the nonlinear filtering problem, a study due to Stroock of densities for infinite systems of interacting diffusions and some applications of his own to infinite-dimensional stochastic differential equations on Banach spaces.

The only major feature not covered by either the Bell or Watanabe book is the formulation of the theory for functionals of jump processes. However, this has been recently treated in book form by Bichteler, Gravereaux and Jacod (1987).

REFERENCES

BICHTELER, K., GRAVEREAUX, J.-B. and JACOD, J. (1987). *Malliavin Calculus for Processes with Jumps*. *Stochastic Monographs* **2**. Gordon and Breach, London.

BISMUT, J. M. (1981). Martingales, the Malliavin calculus and hypoellipticity under general Hörmander's conditions. *Z. Wahrsch. verw. Gebiete* **56** 469–505.

BOULEAU, N. and HIRSCH, F. (1986). Propriétés d'absolue continuité dans les espaces de Dirichlet et application aux équations différentielles stochastiques. *Séminaire de Probabilités XX*. *Lecture Notes in Math.* **1204** 131–161. Springer, New York.

MALLIAVIN, P. (1978a). Stochastic calculus of variation and hypoelliptic operators. *Proc. Intern. Symp. Stochastic Differential Equations, Kyoto, 1976* (K. Itô, ed.) 195–263. Wiley, New York.

MALLIAVIN, P. (1978b). C^k-hypoellipticity with degeneracy. I, II. In *Stochastic Analysis* (A. Friedman and M. Pinsky, eds.) 199–214, 327–340. Academic, New York.

NORRIS, J. (1986). Simplified Malliavin calculus. *Séminaire de Probabilités XX. Lecture Notes in Math.* **1204** 101–130. Springer, New York.

NUALART, D. and PARDOUX, E. (1988). Stochastic calculus with anticipating integrands. *Probab. Theory Related Fields* **78** 535–581.

SHIGEKAWA, I. (1980). Derivatives of Wiener functionals and absolute continuity of induced measures. *J. Math. Kyoto Univ.* **20** 263–289.

STROOCK, D. W. (1981). The Malliavin calculus, a functional analytic approach. *J. Funct. Anal.* **44** 212–257.

DEPARTMENT OF MATHEMATICS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903