BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 51, Number 3, July 2014, Pages 491–497
S 0273-0979(2013)01432-0
Article electronically published on September 20, 2013

Normal approximations with Malliavin calculus. From Stein's method to universality, by Ivan Nourdin and Giovanni Peccati, Cambridge Tracts in Mathematics, Vol. 192, 2012, xii+239 pp., US \$80.00, ISBN 978-1-107-01777-1

This monograph contains some recent results by the authors and their collaborators on the application of Stein's method combined with Malliavin calculus to the normal approximation for functionals of a Gaussian process. It is addressed to researchers and graduate students in probability and statistics who would like to learn the basis of Gaussian analysis and its application to asymptotic techniques related to normal approximations.

1. Stein's method

The standard normal distribution γ is a probability on the real line with density

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

From the fact that ϕ satisfies the differential equation $\phi'(x) = -x\phi(x)$, it follows that a real-valued random variable N has the normal probability distribution γ, if and only if for every differentiable function $f : \mathbb{R} \to \mathbb{R}$ such that $xf(x)$ and $f'(x)$ are integrable with respect to γ,

$$E[Nf(N)] = E[f'(N)].$$

Given a general random variable F, if the expectation $E[Ff(F)] - E[f'(F)]$ is close to zero for a large class of smooth functions f, then we should be able to conclude that the law of F is close to γ in some sense. This is the heuristics of Stein's method (see [7]). To make this argument rigorous, given a measurable function $h : \mathbb{R} \to \mathbb{R}$ such that $E[|h(N)|] < \infty$, where N has the distribution γ, we introduce the *Stein's equation*

$$h(x) - E[h(N)] = f'(x) - xf(x). \tag{1.1}$$

The function

$$f_h(x) = e^{x^2/2} \int_{-\infty}^{x} [h(y) - E[h(N)]] e^{-y^2/2} dy \tag{1.2}$$

turns out to be the unique solution to equation (1.1) satisfying $\lim_{x\to\pm\infty} e^{-x^2/2} f(x) = 0$. Substituting x by F in equation (1.1) and taking the expectation yields

$$E[h(F)] - E[h(N)] = E[f_h'(F) - Ff_h(F)]. \tag{1.3}$$

In the particular case $h = \mathbf{1}_{(-\infty,x]}$, for some $x \in \mathbb{R}$, one can show from (1.2) that $\|f_h\|_\infty \le \sqrt{2\pi}/4$ and $\|f_h'\|_\infty \le 1$. As a consequence, (1.3) leads to the following inequality for the Kolmogorov distance between the law of F, denoted by $\mathcal{L}(F)$, and the normal distribution γ

$$d_{\mathrm{Kol}}(\mathcal{L}(F), \gamma) = \sup_{x\in\mathbb{R}} |P(F \le x) - P(N \le x)| \le \sup_{h\in\mathcal{F}_{\mathrm{Kol}}} |E[f_h'(F) - Ff_h(F)]|, \tag{1.4}$$

where $\mathcal{F}_{\mathrm{Kol}} = \{f \in \mathcal{C}^1 : \|f\|_\infty \le \sqrt{2\pi}/4, \|f'\|_\infty \le 1\}$. In a similar way, by choosing the class of functions h which are indicators of Borel sets, we obtain the *total variation distance* d_{TV}, and we if choose the class of functions h such that

$|h(x)-h(y)| \le |x-y|$, then we get the Wasserstein distance d_W. Then, the estimate (1.4) holds for these distances, where in the right-hand side $\mathcal{F}_{\text{Kol}}$ is replaced by $\mathcal{F}_{\text{TV}} = \{f \in \mathcal{C}^1 : \|f\|_\infty \le \sqrt{\pi/2}, \|f'\|_\infty \le 2\}$ for the total variation distance and by $\mathcal{F}_{\text{W}} = \{f \in \mathcal{C}^1 : \|f'\|_\infty \le \sqrt{2/\pi}\}$ for the Wasserstein distance.

When the random variable F is a functional of a Gaussian process, one can estimate the right-hand side of equation (1.3) using the integration by parts formula of Malliavin calculus.

2. Malliavin calculus

The Malliavin calculus is a stochastic calculus of variations in a Gaussian space, developed from the probabilistic proof of Hörmander's hypoellipticity theorem by Malliavin in [1]. The main application of this calculus is to establish the regularity of the probability distribution of functionals of an underlying Gaussian process. In this way one can prove the existence and smoothness of the density for solutions to ordinary and partial stochastic differential equations. Basic references for the Malliavin calculus and its applications are the monographs by Malliavin [2] and Nualart [5]. Chapters 1 and 2 of the monograph under review contain the basic elements of Malliavin calculus; in particular, the basic ideas in the one-dimensional case are presented in Chapter 1.

Consider a centered Gaussian family of random variables $X = \{X(h) : h \in \mathfrak{H}\}$, defined in a probability space $(\Omega, \mathcal{F}, P)$ and indexed by a real separable Hilbert space $\mathfrak{H}$, such that $E[X(h)X(g)] = \langle h, g\rangle_{\mathfrak{H}}$ for every $h, g \in \mathfrak{H}$. The family X is called an *isonormal* Gaussian process. Particular examples are $\mathfrak{H} = \mathbb{R}^d$, where X is just a d-dimensional random vector and $\mathfrak{H} = L^2([0,T]))$, where $X(h) = \int_0^T h(t)dB(t)$, with B a Brownian motion in the time interval $[0,T]$. We will assume that the σ-field $\mathcal{F}$ is generated by X.

The fundamental operators in Malliavin calculus are the derivative operator D, its adjoint δ called the divergence operator, and the generator of the Ornstein–Uhlenbeck semigroup, denoted by L. The derivative operator D behaves as an infinite-dimensional gradient and is defined by $D(X(h)) = h$ for any $h \in \mathfrak{H}$. More generally, for any cylindrical and smooth random variable of the form $F = g(X(h_1), \dots, X(h_m))$, where $h_i \in \mathfrak{H}$ and $g : \mathbb{R}^m \to \mathbb{R}$ is an infinitely differentiable function with bounded partial derivatives, DF is the $\mathfrak{H}$-valued random variable given by the chain rule,

$$DF = \sum_{i=1}^m \frac{\partial g}{\partial x_i}(X(h_1), \dots, X(h_m))h_i.$$

The operator D can be extended to the Sobolev space $\mathbb{D}^{1,2} \subset L^2(\Omega; \mathfrak{H})$ of random variables such that $E[F^2] + E[\|DF\|_{\mathfrak{H}}^2] < \infty$.

The divergence operator δ is the adjoint of D. It is an unbounded operator in $L^2(\Omega; \mathfrak{H})$ whose domain is the set of elements $u \in L^2(\Omega; \mathfrak{H})$ such that $|E[\langle u, DF\rangle_{\mathfrak{H}}]| \le c_u \|F\|_{L^2(\Omega)}$ for every $F \in \mathbb{D}^{1,2}$. Then, for u in the domain of δ, $\delta(u)$ is the square integrable random variable defined by the duality relationship

$$E[\langle u, DF\rangle_{\mathfrak{H}}] = E[\delta(u)F], \tag{2.1}$$

for any $F \in \mathbb{D}^{1,2}$. To introduce the operator L, given $F \in L^2(\Omega)$ consider the orthogonal expansion $F = E[F] + \sum_{p=1}^\infty J_p(F)$, where J_p is the projection on the pth Wiener chaos. Then $LF = \sum_{p=1}^\infty -pJ_p(F)$, provided that this sum converges in $L^2(\Omega)$. A fundamental result that connects the three operators is that F belongs to the domain of L if and only if $F \in \mathbb{D}^{1,2}$ and DF belongs to the domain of δ, and in this case,

$$LF = -\delta(DF). \tag{2.2}$$

The operator L^{-1} defined by $L^{-1}F = \sum_{p=1}^{\infty} -\frac{1}{p} J_p(F)$ is the pseudo-inverse of L. The following integration by parts formula is the key ingredient in the applications of Malliavin calculus to estimate the right-hand side of equation (1.3).

Theorem 2.1. *Let $F \in \mathbb{D}^{1,2}$ be such that $E[F] = 0$, and let $f : \mathbb{R} \to \mathbb{R}$ be a $\mathcal{C}^1$ function with a bounded derivative. Then*

$$E[Ff(F)] = E[f'(F)\langle DF, -DL^{-1}F\rangle_{\mathfrak{H}}]. \tag{2.3}$$

Proof. Taking into account that $E[F] = 0$ and using (2.2), we obtain $F = LL^{-1}F = -\delta(DL^{-1}F)$. Then, the result follows from the duality relationship (2.1)

$$\begin{aligned} E[Ff(F)] &= -E[f(F)\delta(DL^{-1}F)] = E[\langle D(f(F)), -DL^{-1}F\rangle_{\mathfrak{H}}] \\ &= E[f'(F)\langle DF, -DL^{-1}F\rangle_{\mathfrak{H}}]. \end{aligned}$$

□

3. Normal approximations

Suppose that F is a random variable defined in the probability space $(\Omega, \mathcal{F}, P)$ associated with an isonormal Gaussian process X. We assume that $F \in \mathbb{D}^{1,2}$, $E[F] = 0$ and $E[F^2] = 1$. Substituting equation (2.3) into the right-hand side of (1.3) yields

$$|E[h(F)] - E[h(N)]| \le \|f_h'\|_\infty E[|1 - \langle DF, -DL^{-1}F\rangle_{\mathfrak{H}}|],$$

which leads to the inequality

$$d_i(\mathcal{L}(F), \gamma) \le C_i E[|1 - \langle DF, -DL^{-1}F\rangle_{\mathfrak{H}}|] \tag{3.1}$$

for $i = 1, 2, 3$, where $d_1 = d_{\mathrm{Kol}}$ is the Kolmogorov distance, $d_2 = d_{\mathrm{TV}}$ is the total variation distance, $d_3 = d_{\mathrm{W}}$ is the Wasserstein distance, and $C_1 = 1$, $C_2 = 2$, and $C_3 = \sqrt{2/\pi}$. In the first two cases one has to assume that F has a density.

Fix $q \ge 2$, and suppose that the random variable F belongs to the qth Wiener chaos. That is, F is a generalized multiple stochastic integral of order q. In that case, $L^{-1}F = -\frac{1}{q}F$, and, therefore, $\langle DF, -DL^{-1}F\rangle_{\mathfrak{H}}) = \frac{1}{q}\|DF\|_{\mathfrak{H}}^2$. On the other hand, $E[\|DF\|_{\mathfrak{H}}^2] = q$. Thus,

$$E[|1 - \langle DF, -DL^{-1}F\rangle_{\mathfrak{H}}|] \le \frac{1}{q}\sqrt{\mathrm{Var}\left(\|DF\|_{\mathfrak{H}}^2\right)},$$

and we obtain the inequalities

$$d_i(\mathcal{L}(F), \gamma) \le C_i/q\sqrt{\mathrm{Var}\left(\|DF\|_{\mathfrak{H}}^2\right)}. \tag{3.2}$$

Using Wiener chaos expansions and product formulas for multiple stochastic integrals, one can show the following general formula for a random variable on the qth Wiener chaos:

$$\mathrm{Var}\left(\|DF\|_{\mathfrak{H}}^2\right) \le \frac{q-1}{3q}\left(E[F^4] - 3\right) \le (q-1)\mathrm{Var}\left(\|DF\|_{\mathfrak{H}}^2\right).$$

Thus, in the right-hand side of (3.2) we can replace the variance of the square norm of the derivative by $E[F^4] - 3$ (which is nonnegative!) Moreover, these inequalities can be extended to the case $E[F^2] = \sigma^2 > 0$. In this case a factor σ^{-2} appears for

$i = 1, 2$, a factor σ^{-1} appears in the case $i = 3$, and $E[F^4] - 3$ has to be replaced by $E[F^4] - 3\sigma^4$.

An immediate consequence of these bounds is a quantitative and direct proof of the so-called *fourth-moment theorem* (see Nualart and Peccati [6]), which represents a drastic simplification of the method of moments and cumulants to prove convergence to the normal distribution. This theorem says that for a sequence $F_n = I_q(f_n)$, $n \geq 1$, of random variables in the qth chaos of X (where f_n belongs to the symmetric tensor product $\mathfrak{H}^{\odot q}$) such that $\lim_{n\to\infty} E[F_n^2] = \sigma^2 > 0$, the following assertions are equivalent:

(i) F_n converges in distribution to $N(0, \sigma^2)$.
(ii) $E[F_n^4] \to 3\sigma^4$ as n tends to infinity.
(ii) Var $\left(\|DF\|^2_{\mathfrak{H}}\right) \to 0$ as n tends to infinity.
(iv) $\|f_n \otimes_r f_n\|_{\mathfrak{H}^{\otimes 2(q-r)}} \to 0$ as n tends to infinity, where $f_n \otimes_r f_n$ is the contraction of r indices between the elements $f_n \in \mathfrak{H}^{\odot q}$.

For random variables which are not necessarily in a fixed chaos, one can show the estimate

$$E[|1 - \langle DF, -DL^{-1}F\rangle_{\mathfrak{H}}|] \leq \frac{3}{2} E[\|D^2F \otimes_1 D^2F\|^2_{\mathfrak{H}^{\otimes 2}}]^{1/4} E[\|DF\|^4_{\mathfrak{H}}]^{1/4},$$

and the term that provides an estimate of the error in the normal approximation is $E[\|D^2F \otimes_1 D^2F\|^2_{\mathfrak{H}^{\otimes 2}}]^{1/4}$.

The Stein's method described above and its connection with Malliavin calculus are developed in Chapters 3 and 5 of the book. The corresponding multivariate extension of these techniques is explained in Chapters 4 and 6. In the case of a centered d-dimensional Gaussian random vector N whose covariance matrix C is positive definite and $h : \mathbb{R}^d \to \mathbb{R}$ is such that $E[|h(N)|] < \infty$, Stein's equation is the partial differential equation

$$h(x) - E[h(N)] = \langle C, \mathrm{Hess} f(x)\rangle_{\mathrm{HS}} - \langle x, \nabla f(x)\rangle_{\mathbb{R}^d}.$$

If h is Lipschitz, a solution to this equation is given by

$$f_h(x) = \int_0^\infty E[h(N) - h(e^{-t}x + \sqrt{1 - e^{-2t}}N)]dt.$$

Then, by the same methodology as in the one-dimensional case, one can obtain bounds for the Wasserstein distance between the law of a general random vector F and the law of N.

4. Applications

Chapters 7 through 9 contain a series of applications of the general results on normal approximations. First, Chapter 7 deals with the *Breuer–Major Central Limit Theorem.* The goal is to establish the normal approximation for sequences of the form

$$V_n := \frac{1}{\sqrt{n}} \sum_{i=1}^n f(X_i),$$

where X_i is a centered stationary Gaussian sequence with unit variance. The basic assumption is that $f \in L^2(\gamma)$, has an expansion into a series of Hermite polynomials of the form $f(x) = \sum_{q=d}^\infty a_q H_q(x)$. The number $d \geq 1$ is called the Hermite rank of f. The main result says that if $\sum_{v\in\mathbb{Z}} |\rho(v)|^d < \infty$, where $\rho(v) = E(X_1 X_{1+v})$, then V_n converges to a normal distribution $N(0, \sigma^2)$, where $\sigma^2 = \sum_{q=d}^\infty q! a_q^2 \sum_{v\in\mathbb{Z}} \rho(v)^q$.

This is proved in the book using chaos expansions and the above fourth-moment theorem.

An example of such a stationary sequence is given by the increments of a fractional Brownian motion with Hurst parameter H, that is, $X_k = B^H_{k+1} - B^H_k$, where B^H is a centered Gaussian process with covariance

$$E[B^H_t B^H_s] = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right).$$

If $f = H_q$ is an Hermite polynomial of degree $q \geq 2$, the central limit theorem for the sequence V_n holds if $0 < H < 1 - \frac{1}{2q}$. When $H = 1 - \frac{1}{2q}$, there is also a normal approximation but with a logarithmic normalization. In the quadratic case, where $f(x) = x^2 - 1$ and $H \leq 3/4$, using the techniques based on Stein's method and Malliavin calculus, one can obtain precise rates of convergence for the total variation distance between the law of $V_n/\sqrt{E(V_n^2)}$ and γ.

In Chapter 8 the authors apply the integration by parts technique of Malliavin calculus to the explicit computation of the cumulants associated with regular functionals of an underlying Gaussian process.

Consider a sequence of centered random variables $\{F_n : n \geq 1\}$ with unit variance, converging in law to the normal distribution γ. The purpose of Chapter 9 is to establish an exact asymptotic expression for the sequence $P(F_n \leq z) - P(N \leq z)$, when $z \in \mathbb{R}$ is fixed. Suppose that $F_n \in \mathbb{D}^{1,2}$, and let

$$\varphi(n) = \sqrt{\mathrm{Var}[\langle DF_n, -DL^{-1}F_n\rangle_{\mathfrak{H}}]}.$$

Then it is proved that

$$\frac{P(F_n \leq z) - P(N \leq z)}{\varphi(n)} \to \frac{\rho}{3}\Phi^{(3)}(z)$$

as n tends to infinity, where $\Phi(z) = P(N \leq z)$, assuming that $\varphi(n)$ converges to zero and is strictly positive for n large enough, the random vector

$$(F_n, (1 - \langle DF_n, -DL^{-1}F_n\rangle_{\mathfrak{H}})/\varphi(n))$$

converges in law to a centered two-dimensional Gaussian vector with covariance matrix $\left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$, and the law of F_n is absolutely continuous with respect to the Lebesgue measure. Some connections of this result with Edgeworth expansions and particular examples are discussed in this chapter.

5. Density estimates

A new technique for deriving Gaussian estimates using Malliavin calculus, introduced by Nourdin and Viens in [4], is presented in Chapter 10. Let $F \in \mathbb{D}^{1,2}$ be such that $E(F) = 0$. The integration by parts formula proved in Theorem 2.1 can also be written as

$$E[Ff(F)] = E[f'(F)g(F)],$$

where $g(x) = E[\langle DF, -DL^{-1}F\rangle_{\mathfrak{H}} | F = x]$. This simple observation leads to a differential equation satisfied by the density $\rho(x)$ of F. Solving this equation leads to the following explicit formula for the density of a random variable F, assuming that $g(F) > 0$ almost surely

$$\rho(x) = \frac{E[|F|]}{2g(x)} \exp\left(-\int_0^x \frac{y\,dy}{g(y)}\right).$$

The above formula holds for all x in the support of the law of F, which is a closed interval containing the origin. In particular, if $0 < c \leq g(x) < C < \infty$ for all $x \in \mathbb{R}$, one can obtain lower and upper Gaussian bounds for the density of F. An important example of an application discussed in the monograph is the case where F is the centered maximum of a finite number of Gaussian random variables, with a nonsingular covariance matrix.

6. Universality

In the last chapter of the book, the authors relate the results and techniques discussed in the monograph with the *universality phenomenon*, according to which the asymptotic behavior of large random systems does not depend on the distribution of its components. The authors consider random homogenous sums defined as multilinear symmetric polynomials of degree $d \leq M$, vanishing on diagonals, on a collection of M independent random variables:

$$Q(f, Y_1, \ldots, Y_M) = \sum_{1 \leq i_1, \ldots, i_d \leq M} f(i_1, \ldots, i_d) Y_{i_1} Y_{i_2} \cdots Y_{i_d}.$$

Given a sequence of m-dimensional vectors of random homogeneous sums of the form

$$(Q_k(f_k^{(n)}, Y_1, \ldots, Y_{M_n}))_{1 \leq k \leq m},$$

it is proved that if the L^2-norm of the functions $f_k^{(n)}$ is uniformly bounded in n, then the convergence to a normal vector when the Y_i have the normal distribution γ is equivalent to the convergence to a normal vector when the Y_i satisfy $E(Y_i) = 0$, $E(Y_i^2) = 1$, and $\max_i E(Y_i^4) < \infty$. The proof is based on the techniques developed in this monograph combined with the extension of the Lindberg principle to the framework of polynomial functionals of sequences of independent random variables done by Mossel, O'Donnel, and Oleszkiewicz in [3].

The book contains many examples and exercises which help the reader understand and assimilate the material. Also bibliographical comments at the end of each chapter provide useful references for further reading.

References

[1] Paul Malliavin, *Stochastic calculus of variation and hypoelliptic operators*, Equations (Res. Inst. Math. Sci., Kyoto Univ., Kyoto, 1976), Wiley, New York, 1978, pp. 195–263. MR536013 (81f:60083)

[2] Paul Malliavin, *Stochastic analysis*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 313, Springer-Verlag, Berlin, 1997. MR1450093 (99b:60073)

[3] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz, *Noise stability of functions with low influences: invariance and optimality*, Ann. of Math. (2) **171** (2010), no. 1, 295–341, DOI 10.4007/annals.2010.171.295. MR2630040 (2012a:60091)

[4] Ivan Nourdin and Frederi G. Viens, *Density formula and concentration inequalities with Malliavin calculus*, Electron. J. Probab. **14** (2009), no. 78, 2287–2309, DOI 10.1214/EJP.v14-707. MR2556018 (2011a:60147)

[5] David Nualart, *The Malliavin calculus and related topics*, 2nd ed., Probability and its Applications (New York), Springer-Verlag, Berlin, 2006. MR2200233 (2006j:60004)

[6] David Nualart and Giovanni Peccati, *Central limit theorems for sequences of multiple stochastic integrals*, Ann. Probab. **33** (2005), no. 1, 177–193, DOI 10.1214/009117904000000621. MR2118863 (2005k:60077)

[7] Charles Stein, *Approximate computation of expectations*, Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7, Institute of Mathematical Statistics, Hayward, CA, 1986. MR882007 (88j:60055)

David Nualart
Kansas University
E-mail address: nualart@math.ku.edu