

# BERNSTEIN POLYNOMIALS AND BROWNIAN MOTION

EMMANUEL KOWALSKI

## 1. INTRODUCTION.

One of the greatest pleasures in mathematics is the surprising connections that often appear between apparently disconnected ideas and theories. Some particularly striking instances exist in the interaction between probability theory and analysis. One of the simplest is the elegant proof of the Weierstrass approximation theorem by S. Bernstein [2]: on the surface, this states that if  $f : [0, 1] \rightarrow \mathbf{C}$  is a continuous function, then the sequence of *Bernstein polynomials*

$$B_n(f)(x) = \sum_{j=0}^n \binom{n}{j} f\left(\frac{j}{n}\right) x^j (1-x)^{n-j}$$

converges uniformly to  $f$  on  $[0, 1]$ . However, the structure and origin of the polynomials become much clearer when given a probabilistic interpretation. We recall this argument in section 3, stating here only that the heart of this connection is the formula

$$B_n(f)(x) = E\left(f\left(\frac{S_n}{n}\right)\right), \quad (1.1)$$

where  $S_n = X_1 + \cdots + X_n$  for any  $n$ -uple  $(X_i)$  of independent random variables with Bernoulli law such that

$$P(X_i = 1) = x, \quad P(X_i = 0) = 1 - x.$$

On a much deeper level lies Brownian motion, which can be described informally as “random” paths in Euclidean space  $\mathbf{R}^d$  starting from the origin. We will recall the precise definition of this concept in the case  $d = 1$  in section 4, and the reader with no previous experience is also encouraged to read [3], which contains a very readable informal description. In the author’s view, there are few mathematical facts more amazing than the connection between plane Brownian motion and, for instance, the theory of harmonic functions. As an example, there is the beautiful formula (discovered by Kakutani [7])

$$f(z) = \int_{\Omega} f_0(z + B(\tau)) dP$$

for the solution of the Dirichlet problem: to find the function  $f$  harmonic in a bounded, open, connected plane set  $U$  with smooth boundary such that  $f(z) = f_0(z)$  for all  $z$  on the boundary  $\partial U$ , where  $f_0 : \partial U \rightarrow \mathbf{R}$  is a given (continuous) function. Here  $B(t)$  signifies a two-dimensional Brownian motion starting at 0 defined on a probability space  $\Omega$  with probability measure  $P$ , and  $\tau$  is the exit time from  $U$  for  $B$  started at  $z$  (i.e.,

$$\tau = \inf\{t : z + B(t) \in \partial U\};$$

for a proof, see [8, chap. 4, sec. 2.B]).

Although not directly relevant to what follows, it might be useful to point out as well the discovery of the (still mysterious) link between the Riemann zeta-function  $\zeta(s)$  (and its generalizations) and random matrices as another striking example where probability and number theory have become intertwined. Roughly speaking, it is expected—with overwhelming numerical evidence—that suitably normalized spacings of zeros of  $\zeta(s)$  are distributed asymptotically “high in the critical strip” (i.e., when the imaginary part goes to infinity) like the normalized spacings of eigenvalues of large unitary matrices selected at random with respect to the natural probability measures on groups of unitary matrices. A shade differently, but conveying the same kind of intuitive idea, it is expected (and confirmed in various ways) that one can “model”  $\zeta(1/2 + it)$  for large  $|t|$  by values of characteristic polynomials of “random” unitary matrices of size roughly  $\log t/2\pi$ . We refer readers to [9] for a survey of this very lively topic and to [10] for some of the remarkable conjectures concerning the zeta-function that arise in this manner.

Since Brownian motion (now one-dimensional) is a way of speaking of a “random” continuous function on  $[0, +\infty)$ , it seems natural to bring together the simple Bernstein polynomials and the more delicate Brownian motion: namely, for  $n \geq 1$  and for  $x$  in  $[0, 1]$ , we consider the random variable given by evaluating the Bernstein polynomial for the Brownian motion at  $x$ :

$$\mathfrak{B}_n(x) = \sum_{j=0}^n \binom{n}{j} B\left(\frac{j}{n}\right) x^j (1-x)^{n-j}. \quad (1.2)$$

Now three different questions (at least) present themselves:

- Forgetting that we know that Brownian motion exists, can we use polynomial approximations to prove its existence? The idea is to consider a sequence  $\mathcal{B}_n$  of random polynomials of the type

$$\mathcal{B}_n(x) = \sum_{j=0}^n \binom{n}{j} X_{nj} x^j (1-x)^{n-j} \quad (1.3)$$

for some suitable random variables  $X_{nj}$  and to prove that  $\mathcal{B}_n$  converges (again in a suitable sense) to a Brownian motion.

- Or, since we know that Brownian motion exists, can one investigate its properties by means of the Bernstein motions  $\mathfrak{B}_n$  and Bernstein’s convergence theorem?

- Or, again assuming that Brownian motion exists, what are the properties of the random polynomials of the type (1.2)?

We will consider simple instances of all three questions. First, we provide a construction of Brownian motion (in the guise of the Wiener measure on the vector space  $C([0, 1], \mathbf{R})$  of real-valued continuous functions on the closed interval  $[0, 1]$ ) from random Bernstein polynomials. The complete proof is not simpler than known ones but (to the author, at least) furnishes a very convincing plausibility argument for the existence of Brownian motion that requires nothing more than the very basics of probability theory.

Second, we use results concerning Bernstein polynomials to prove that almost surely the Brownian path  $t \mapsto B(t)$  is not differentiable outside a set of Lebesgue measure 0 in  $[0, 1]$ . This is, of course, weaker than the theorem of Paley-Wiener-Zygmund [13], which asserts that the exceptional set is almost surely empty, but the proof is particularly simple. Moreover, it is in fact independent of the construction of Brownian motion previously mentioned.

Finally, simple facts about the oscillations of a Brownian path allow us to prove some results about the zeros of random Bernstein polynomials.

In all of this we give complete rigorous proofs, but we also try to highlight the ideas (often quite intuitive) behind the arguments used. Everything should be understandable to a graduate student (not studying probability) or, with a reasonable amount of suspension of disbelief, to an undergraduate who has had a first course in probability theory. Such a reader may wish to start by reading the very nice informal description given by Billingsley [3] and can safely skip all the technical mentions of measurability (so any set and any function can be considered as measurable and all  $\sigma$ -algebras on a set  $\Omega$  are just the entire power set of subsets of  $\Omega$ ).

**Notation.** A triple  $(\Omega, \Sigma, P)$  always signifies a probability space, with  $P$  a probability measure on the  $\sigma$ -algebra  $\Sigma$  of subsets of  $\Omega$ . For a random variable  $X$ ,  $V(X) = E((X - E(X))^2)$  is its variance. Also, as usual, we use  $x \wedge y$  to denote  $\min(x, y)$ .

## 2. REVIEW OF FACTS FROM PROBABILITY.

We list here the various definitions and basic results from probability theory needed in the rest of the paper, with a few exceptions that are deeper results and will be stated only when needed in later sections. As a reference, we use [15], but the material can be found in any textbook on probability theory.

First, for the study of Bernstein polynomials we use the most basic properties of Bernoulli random walks. Let  $x$  in  $[0, 1]$  and  $n \geq 1$  be given, and let  $X_1, \dots, X_n$  be independent random variables on a given probability space  $(\Omega, \Sigma, P)$  taking only the values 0 and 1 according to the same Bernoulli law

$$P(X_i = 1) = x, \quad P(X_i = 0) = 1 - x.$$

The associated ( $n$ -step) random walk is the sequence of sums  $S_i = X_1 + \dots + X_i$  ( $0 \leq i \leq n$ ), and the “empirical” average  $T_n = S_n/n$  is often what we consider. Note that

$$E(X_i) = x, \quad V(X_i) = x(1 - x).$$

Hence

$$E(T_n) = E\left(\frac{S_n}{n}\right) = x, \quad V(T_n) = V\left(\frac{S_n}{n}\right) = \frac{x(1 - x)}{n}, \quad (2.1)$$

the latter because of the additivity of the variance for independent random variables.

We also use the fourth moment of  $T_n - x$ , for which we have the following result:

**Lemma 2.1.** *The estimate*

$$E((T_n - x)^4) = E\left(\left(\frac{S_n}{n} - x\right)^4\right) \leq \frac{1}{n^2}$$

holds for  $n \geq 1$  and for all  $x$  in  $[0, 1]$ .

*Proof.* This is well known (it is the basis of Cantelli's "easy" proof of the strong law of large numbers for identically distributed random variables with finite fourth moment; see, for example, [15, Theorem 1, p. 363]). After expanding

$$(T_n - x)^4 = \left(\frac{X_1 - x + \cdots + X_n - x}{n}\right)^4$$

and taking the expectation, among the  $n^4$  terms

$$\frac{1}{n^4} E((X_{i_1} - x)(X_{i_2} - x)(X_{i_3} - x)(X_{i_4} - x))$$

all those with three or four distinct indices vanish because of the independence assumption. For instance (with  $\tilde{X}_i = X_i - x$ ),

$$E((X_1 - x)(X_2 - x)(X_3 - x)(X_4 - x)) = E(\tilde{X}_1) \cdots E(\tilde{X}_4) = 0,$$

$$E((X_1 - x)^2(X_2 - x)(X_3 - x)) = E(\tilde{X}_1^2)E(\tilde{X}_2)E(\tilde{X}_3) = 0.$$

There remain only the  $n$  terms of the type

$$\frac{1}{n^4} E((X_i - x)^4) = \frac{x^4(1-x) + x(1-x)^4}{n^4} \leq \frac{1}{12n^4},$$

and the  $6n(n-1)/2$  terms of the type

$$\frac{1}{n^4} E((X_i - x)^2(X_j - x)^2) = \frac{x^2(1-x)^2}{n^4} \leq \frac{1}{16n^4}$$

with  $i \neq j$ , so we find that

$$E((T_n - x)^4) \leq \frac{n}{12n^4} + \frac{6n(n-1)}{32n^2} \leq \frac{1}{n^2}.$$

□

We also consider  $d$ -dimensional variants of the Bernoulli random walk in which each  $X_i$  is a vector  $X_{ij}$  ( $1 \leq j \leq d$ ) of Bernoulli random variables with

$$P(X_{ij} = 1) = x_j, \quad P(X_{ij} = 0) = 1 - x_j$$

for some vector  $x = (x_1, \dots, x_d)$  in  $[0, 1]^d$ . We assume then that the  $(X_{ij})$  are *all* independent.

In the study of Bernstein polynomials,  $x$  is in fact the variable of the functions under consideration. All random variables depend on it, although we suppress this fact in the notation (but notice that (2.1) recovers  $x$  naturally from  $S_n$  or  $T_n$ ). This is already apparent in formula (1.1) (there is no  $x$  on the right-hand side!); this may seem confusing, but we will explain this formula again carefully in the proof of Theorem 3.1.

We recall Chebychev's inequality [15, chap. 1, sec. 5.1] for a random variable  $X$  having finite variance  $V(X)$ : for any  $\delta > 0$

$$P(|X| \geq \delta) \leq \frac{V(X)}{\delta^2}. \tag{2.2}$$

Also, we have the easy half of the Borel-Cantelli lemma (see [15, chap. 2, sec. 10.3] for the full statement): if  $(A_n)$  is a sequence of events such that the series

$$\sum_{n \geq 1} P(A_n)$$

converges, then almost surely any  $\omega$  in  $\Omega$  belongs to at most finitely many  $A_n$  (i.e.,

$$P\left(\bigcap_{N \geq 1} \bigcup_{n \geq N} A_n\right) = 0.)$$

Our second main probabilistic topic is the elementary theory of Gaussian random variables and vectors (see, for example, [15, chap. 2, sec. 13]). Recall that a real-valued random variable  $X$  is a *centered Gaussian random variable* with variance  $\sigma^2$  ( $\sigma > 0$ ) if its distribution law  $X(P)$  is given by

$$X(P) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}} dt. \quad (2.3)$$

Recall that this means simply that for any (measurable) subset  $B$  of  $\mathbf{R}$  (for example, an interval  $B = [a, b]$ ), we have

$$P(X \in B) = \frac{1}{\sigma\sqrt{2\pi}} \int_B e^{-\frac{t^2}{2\sigma^2}} dt.$$

It follows that

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbf{R}} t e^{-\frac{t^2}{2\sigma^2}} dt = 0, \quad V(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbf{R}} t^2 e^{-\frac{t^2}{2\sigma^2}} dt = \sigma^2.$$

Moreover, a simple computation using (2.3) shows that for  $k = 0, 1, 2, \dots$  the moment  $E(X^k)$  exists and, in particular, that

$$E(X^4) = 3V(X)^2 = 3\sigma^4. \quad (2.4)$$

By the correspondence between probability laws and their characteristic functions, an equivalent condition for  $X$  to be Gaussian is that for each real  $t$  one has

$$E(e^{itX}) = e^{-\sigma^2 t^2/2}.$$

This approach by Fourier theory leads to an easy definition of a *centered Gaussian random vector*  $(X_1, \dots, X_m)$ : this is a vector of real-valued random variables such that there exists a positive quadratic form  $Q(t_1, \dots, t_m)$  for which

$$E(e^{i(t_1 X_1 + \dots + t_m X_m)}) = e^{-Q(t_1, \dots, t_m)/2}$$

for all  $(t_1, \dots, t_m)$  in  $\mathbf{R}^m$ . The coefficients of the quadratic form are related to the Gaussian vector as follows: if

$$Q(t) = \sum_{1 \leq i, j \leq m} a_{ij} t_i t_j,$$

then the *covariance matrix*  $A = (a_{ij})$  is determined by

$$E(X_i X_j) = a_{ij}.$$

As before, the joint law  $(X_1, \dots, X_n)(P)$  of the vector is defined as a measure on  $\mathbf{R}^m$  that permits one to compute  $P((X_1, \dots, X_m) \in B)$  for any measurable subset  $B$  of  $\mathbf{R}^m$ . If  $Q$  is positive definite, one shows that

$$P((X_1, \dots, X_m) \in B) = \frac{1}{\sqrt{|\det(A)|}(2\pi)^{m/2}} \int_B e^{-\sum_{i,j} b_{ij} t_i t_j} dt_1 \dots dt_m,$$

where  $(b_{ij})$  is the inverse of  $A$ .

Exploiting the form of the characteristic functions, one demonstrates easily that the coordinates of a centered Gaussian random vector  $(X_1, \dots, X_m)$  are independent if and only if they are uncorrelated. This means that the vector  $(X_1, \dots, X_m)$  is Gaussian with diagonal covariance matrix, or in other words,  $E(X_i X_j) = 0$  if  $i \neq j$ .

Also, one sees without difficulty that if  $X = (X_1, \dots, X_m)$  is a centered Gaussian random vector and  $u : \mathbf{R}^m \rightarrow \mathbf{R}^n$  is a linear map, then the  $n$ -dimensional vector  $u(X)$  is again a centered Gaussian vector. In particular, a linear combination of the coordinates of  $X$ ,

$$X = \sum_{i=1}^m \alpha_i X_i \quad (\alpha_i \in \mathbf{R}),$$

is a centered Gaussian random variable.

As a final generalization, if  $T$  is an arbitrary set, then a mapping on  $T$  of the sort  $t \mapsto X_t$ , where  $X_t$  is a random variable for each  $t$ , is a *Gaussian stochastic process parameterized by  $T$*  (often  $T = [0, 1]$  or  $[0, +\infty)$ ) if for each finite subset  $t_1, \dots, t_m$  of  $T$  the vector  $(X_{t_1}, \dots, X_{t_m})$  is a centered Gaussian random vector. The covariance function of the stochastic process  $X = (X_t)$  is the function

$$g(s, t) = E(X_s X_t)$$

defined for  $(s, t)$  in  $T^2$ .

Besides the definition and characterization of independence, we will appeal to the following simple lemma, which makes rigorous the quite intuitive fact that the values of a sequence of Gaussian random variables with increasing variance (i.e., more and more “spread out”) can form a convergent sequence only with probability zero.

**Lemma 2.2.** *Let  $(\Omega, \Sigma, P)$  be a probability space, and let  $(X_n)$  be a sequence of centered Gaussian random variables such that  $\sigma_n^2 = V(X_n) \rightarrow +\infty$ . Then  $(X_n)$  is almost surely unbounded.*

*Proof.* If  $Y$  is a centered Gaussian random variable with variance  $\sigma = 1$  and if  $\varepsilon > 0$ , we have by (2.3)

$$P(|Y_n| < \varepsilon) = \frac{1}{\sqrt{2\pi}} \int_{-\varepsilon}^{\varepsilon} e^{-t^2/2} dt \leq \varepsilon.$$

Since  $Y_n = X_n/\sigma_n$  has variance 1, we infer that

$$P(|X_n| < M) = P\left(|Y_n| < \frac{M}{\sigma_n}\right) \leq \frac{M}{\sigma_n}$$

for any  $n$  and arbitrary positive  $M$ , whence

$$\lim_{n \rightarrow \infty} P(|X_n| < M) = 0 \quad (M > 0). \quad (2.5)$$

So, in a sense,  $(X_n)$  is unbounded in probability, and it should come as no surprise that there exists a subsequence of  $(X_n)$  that is unbounded almost surely. Indeed, taking  $M = k$ , an integer, there exists  $N_k$  such that

$$P(|X_n| < k) < 2^{-k} \quad (n \geq N_k),$$

which ensures that

$$\sum_k P(|X_{N_k}| < k) < +\infty.$$

Invoking the easy half of the Borel-Cantelli lemma, we conclude that with probability one the event  $\{|X_{N_k}| < k\}$  occurs only finitely often. Thus the subsequence  $(X_{N_k})$  is almost surely unbounded, as then is  $(X_n)$ .  $\square$

Alternatively, one can establish Lemma 2.2 by applying to  $Y_n = X_n^{-1}$  (with  $Y_n = 0$  if  $X_n = 0$ , which only occurs with probability zero) the standard fact (see, for example, [15, chap. 2, sec. 10.5]) that a sequence converging in probability has a subsequence converging almost surely, since  $Y_n$  converges to 0 in probability by (2.5).

Finally, let  $Y$  be a metric space, and let  $\mathcal{B}$  be the Borel  $\sigma$ -algebra on  $Y$  (i.e., the smallest  $\sigma$ -algebra containing the open sets of the metric topology on  $Y$ ). One says that a sequence  $(P_n)$  of probability measures on  $(Y, \mathcal{B})$  (respectively, a sequence  $(X_n)$  of measurable maps  $X_n : \Omega \rightarrow Y$ ) *converges in distribution* to a probability measure  $Q$  on  $(Y, \mathcal{B})$  if the sequence of measures  $(P_n)$  (respectively, the sequence  $(X_n(P))$  of image measures) converges weakly to  $Q$ . This means that for any bounded continuous function  $f$  on  $Y$  we have

$$\lim_{n \rightarrow \infty} \int_Y f(y) dP_n(y) = \int_Y f(y) dQ(y)$$

or

$$\lim_{n \rightarrow \infty} \int_Y f(y) dX_n(P)(y) = \int_Y f(y) dQ(y).$$

Equivalently, and more intuitively, the  $P_n$ -probability of a subset  $A$  of  $Y$  converges to its  $Q$ -probability for all “reasonable” sets  $A$ . To be precise, convergence in distribution holds if and only if for each Borel subset  $A$  of  $Y$  whose boundary  $\partial A$  has  $Q$ -measure zero it is true that

$$\lim_{n \rightarrow \infty} P_n(A) = Q(A)$$

(see [15, chap. 3, sec. 1.2]).

The Lévy criterion (see [15, chap. 3, sec. 3]) states that if  $Y = \mathbf{R}^d$  with  $d \geq 1$ , then a sequence of  $d$ -dimensional random vectors  $X_n = (X_{n1}, \dots, X_{nd})$  converges in distribution to some  $Q$  if the characteristic functions

$$\varphi_n(t_1, \dots, t_d) = E(e^{i(t_1 X_{n1} + \dots + t_d X_{nd})})$$

converge pointwise to a function  $\varphi$  on  $\mathbf{R}^d$  that is continuous at 0. In particular, it follows that a sequence of centered Gaussian random vectors converges in distribution if and only if the covariance coefficients  $E(X_{n,i} X_{n,j})$  converge to real numbers  $a_{ij}$  ( $1 \leq i, j \leq d$ ) and the limit distribution  $\varphi$  is also the distribution function of a centered Gaussian random vector with covariance matrix  $(a_{ij})$ .

## 3. PRELIMINARIES ABOUT BERNSTEIN POLYNOMIALS.

In this section we summarize the facts about Bernstein polynomials that we need in the sequel. For completeness, we include proofs of all of them, albeit proofs in a probabilistic style.

We first recall the proof of Bernstein's theorem, generalized to functions on  $[0, 1]^d$  (the case  $d = 2$  will be required later).

**Theorem 3.1 (Bernstein).** *Let  $d$  be a positive integer. If  $f : [0, 1]^d \rightarrow \mathbf{C}$  is a continuous function and if*

$$B_n(f)(x) = \sum_{0 \leq i_1, \dots, i_d \leq n} f\left(\frac{i_1}{n}, \dots, \frac{i_d}{n}\right) \prod_{k=1}^d \binom{n}{i_k} x_k^{i_k} (1 - x_k)^{n-i_k} \quad (3.1)$$

for  $n = 1, 2, \dots$  and  $x = (x_1, \dots, x_d)$  in  $[0, 1]^d$ , then  $B_n(f) \rightarrow f$  uniformly on  $[0, 1]^d$ .

*Proof.* Consider a positive integer  $n$  and  $x = (x_1, \dots, x_d)$  in  $[0, 1]^d$ . Let  $X_i = (X_{i1}, \dots, X_{id})$  ( $1 \leq i \leq n$ ) be a sequence of random vectors defined on some probability space such that the  $X_{ij}$  are all independent and, as in the beginning of the previous section, each  $X_{ij}$  takes only the values 0 and 1 according to the Bernoulli law for which

$$P(X_{ij} = 1) = x_j, \quad P(X_{ij} = 0) = 1 - x_j,$$

thus putting our point  $x$  in the picture.

Generalizing (1.1) to dimension  $d$  we claim that

$$B_n(f)(x) = E\left(f\left(\frac{S_n}{n}\right)\right). \quad (3.2)$$

where still  $S_n = X_1 + \dots + X_n$ .

This formula indicates why the theorem holds: by the law of large numbers, it is usually the case that  $S_n/n$  is close to the average  $(x_1, \dots, x_d)$ , and therefore (by continuity)  $f(S_n/n)$  should be close to the constant  $f(x)$ . We infer that the expectation  $B_n(f)(x)$  is close to  $E(f(x)) = f(x)$ .

To make this argument precise, we start by proving (3.2). Writing  $S_n = (S_{n1}, \dots, S_{nd})$ , notice that  $S_{nj}/n$ , as a sum of  $n$  terms each of which is either 0 or 1, can take only values of the form  $i_j/n$  with  $0 \leq i_j \leq n$ . Moreover by counting the number of possible ways of doing this, we see that  $S_{nj}/n$  takes the value  $i_j/n$  with probability

$$P(S_{nj}/n = i_j/n) = \binom{n}{i_j} x_j^{i_j} (1 - x_j)^{n-i_j},$$

so expanding  $E(f(S_n/n))$  according to the possible values of  $S_n/n$  and using the independence of the various coordinate vectors, we get (3.2).

Now we put

$$\sigma_{n,j}^2 = V(S_{nj}) = n^{-2} x_j (1 - x_j) \leq \frac{1}{4n^2}.$$

Using obvious notation for a multidimensional expectation, we have  $E(S_n) = nx = n(x_1, \dots, x_d)$ , whence

$$B_n(f)(x) - f(x) = E\left(f\left(\frac{S_n}{n}\right) - f(x)\right) = E\left(f\left(\frac{S_n}{n}\right) - f\left(E\left(\frac{S_n}{n}\right)\right)\right).$$

Set  $\|x\| = \max\{|x_i|\}$  for any  $x$  in  $\mathbf{R}^d$ . By Chebychev's inequality (2.2), applied componentwise, and independence, we obtain for any  $\delta > 0$  the upper bound

$$P\left(\left\|\frac{S_n}{n} - x\right\| \geq \delta\right) \leq \sum_{1 \leq j \leq d} P\left(\left|\frac{S_{n,j}}{n} - x_j\right| \geq \delta\right) \leq \frac{d\sigma_{n,j}^2}{\delta^2} \leq \frac{d}{4n^2\delta^2}.$$

Now let  $\varepsilon (> 0)$  be given. By the uniform continuity of  $f$  on the compact set  $[0, 1]^d$ , there exists a positive  $\delta$  such that  $|f(x) - f(y)| < \varepsilon$  whenever  $x$  and  $y$  in  $[0, 1]^d$  have  $\|x - y\| < \delta$ . Then

$$\int_{\{\|S_n/n - x\| < \delta\}} \left|f\left(\frac{S_n}{n}\right) - f(x)\right| dP \leq \varepsilon,$$

so

$$\begin{aligned} \int_{\{\|S_n/n - x\| \geq \delta\}} \left|f\left(\frac{S_n}{n}\right) - f(x)\right| dP &\leq 2 \max_{x \in [0, 1]^d} |f(x)| P\left(\left\|\frac{S_n}{n} - x\right\| \geq \delta\right) \\ &\leq \frac{2d}{4n^2\delta^2} \max_{x \in [0, 1]^d} |f(x)|. \end{aligned}$$

Clearly, taken together the foregoing bits of information imply that for all sufficiently large  $n$

$$\left|E\left(f\left(\frac{S_n}{n}\right) - f\left(E\left(\frac{S_n}{n}\right)\right)\right)\right| \leq 2\varepsilon$$

holds independently of  $x$ , hence the result.  $\square$

Another simple property of Bernstein polynomials that we will require is the form of their derivatives. We consider only functions of one or two variables. The result (which is simply a direct computation) reads as follows:

**Lemma 3.2.** (1) *If  $f : [0, 1] \rightarrow \mathbf{C}$  is a continuous function and  $B_n(f)$  is the  $n$ th Bernstein polynomial of  $f$ , then*

$$B_n(f)'(x) = \sum_{j=0}^{n-1} \binom{n-1}{j} n \left(f\left(\frac{j+1}{n}\right) - f\left(\frac{j}{n}\right)\right) x^j (1-x)^{n-1-j}. \quad (3.3)$$

(2) *If  $f : [0, 1]^2 \rightarrow \mathbf{C}$  is a continuous function and  $B_n(f)$  is the  $n$ th Bernstein polynomial of  $f$ , then*

$$\begin{aligned} \partial_x B_n(f)(x, y) &= \sum_{j=0}^{n-1} \sum_{k=0}^n \binom{n-1}{j} \binom{n}{k} n \left(f\left(\frac{j+1}{n}, \frac{k}{n}\right) - f\left(\frac{j}{n}, \frac{k}{n}\right)\right) \\ &\quad x^j (1-x)^{n-1-j} y^k (1-y)^{n-k}, \end{aligned}$$

and symmetrically for  $\partial_y B_n(f)$ .

*Proof.* For  $d = 1$  we have

$$\begin{aligned} B_n(f)'(x) &= \sum_{j=0}^n \binom{n}{j} f\left(\frac{j}{n}\right) \{jx^{j-1}(1-x)^{n-j} + (n-j)x^j(1-x)^{n-1-j}\} \\ &= \sum_{j=1}^n j \binom{n}{j} f\left(\frac{j}{n}\right) x^{j-1}(1-x)^{n-j} - \sum_{j=0}^{n-1} (n-j) \binom{n}{j} f\left(\frac{j}{n}\right) x^j(1-x)^{n-1-j}. \end{aligned} \quad (3.4)$$

Changing  $j$  to  $j-1$  in the first sum and noting that  $j \binom{n}{j} = n \binom{n-1}{j-1}$  and  $(n-j) \binom{n}{j} = n \binom{n-1}{j}$ , we regroup to get

$$B_n(f)'(x) = \sum_{j=0}^{n-1} \binom{n-1}{j} n \left( f\left(\frac{j+1}{n}\right) - f\left(\frac{j}{n}\right) \right) x^j (1-x)^{n-1-j},$$

as required. For the case  $d = 2$  we observe that

$$B_n(f)(x, y) = \sum_{j=0}^n \binom{n}{j} G_n\left(\frac{j}{n}, y\right) x^j (1-x)^{n-j}$$

with  $G_n(x, y) = B_n(g)(y)$ , the Bernstein polynomial of the function of  $y$  given by  $g(y) = f(x, y)$  for fixed  $x$ . From the case  $d = 1$  we learn that

$$\partial_x B_n(f)(x, y) = \sum_{j=0}^{n-1} \binom{n-1}{j} n \left\{ G_n\left(\frac{j+1}{n}, y\right) - G_n\left(\frac{j}{n}, y\right) \right\} x^j (1-x)^{n-j},$$

which when expanded again gives the desired result. The case of  $\partial_y B_n(f)$  is, of course, similar.  $\square$

The last property is a fairly striking illustration of the “quality” of the approximation to a function given by its Bernstein polynomials. Recall (one of the classic counterexamples of “rigorous” analysis) that for a sequence of smooth functions  $f_n$  converging uniformly (say on  $[0, 1]$ ) to a smooth function  $f$  there is in general no reason that the sequence of derivatives  $f'_n$  should converge or, if it does converge, bear any relation to  $f'$ . However, Lorentz proved (see [11, Theorem 1.8.1]) that this holds for the sequence of Bernstein polynomials in the following local form:

**Lemma 3.3.** *If  $f : [0, 1] \rightarrow \mathbf{C}$  is a continuous function and  $x_0$  is a point of  $[0, 1]$  such that the derivative  $f'(x_0)$  exists then*

$$\lim_{n \rightarrow \infty} B_n(f)'(x_0) = f'(x_0).$$

We will use the contrapositive form of Lemma 3.3 to study the points of differentiability of Brownian paths: if for some  $x_0$  the sequence of derivatives of the Bernstein polynomials of a function  $f$  at  $x_0$  does not converge, then it must be the case that  $f$  is not differentiable at  $x_0$ . For completeness, we include a proof of the lemma that is merely the probabilistic rewriting (for stylistic consistency) of the argument in [11].

*Proof.* First the cases  $x_0 = 0$  and  $x_0 = 1$  follow immediately from

$$B_n(f)'(0) = \frac{f(\frac{1}{n}) - f(0)}{\frac{1}{n}}, \quad B_n(f)'(1) = \frac{f(1) - f(1 - \frac{1}{n})}{\frac{1}{n}},$$

(evaluate (3.3) at 0 or 1). We proceed assuming that  $x_0(1 - x_0) \neq 0$ , which in the probabilistic interpretation means that  $V(S_n) = nx_0(1 - x_0) \neq 0$ .

Writing  $f(x) = f(x_0) + (x - x_0)f'(x_0) + g(x)$  with  $g$  differentiable at  $x_0$  and  $g'(x_0) = 0$ , we see that it is enough to show that  $B_n(g)'(x_0) \rightarrow 0$  as  $n \rightarrow \infty$ . We can write  $g(x) = (x - x_0)\varepsilon(x)$ , where  $\lim_{x \rightarrow x_0} \varepsilon(x) = 0$ . We claim that the following probabilistic formula holds, with  $S_n$  and  $T_n$  as in (2.1): for  $x$  in  $(0, 1)$

$$B_n(g)'(x) = \frac{E\{(T_n - x)g(T_n)\}}{V(T_n)} \quad (3.5)$$

whence for  $x = x_0$

$$B_n(g)'(x_0) = \frac{E\{g(T_n)(T_n - x_0)\}}{V(T_n)} = \frac{E\{\varepsilon(T_n)(T_n - x_0)^2\}}{V(T_n)}. \quad (3.6)$$

To obtain (3.5), we go back to (3.4), now applied to  $g$ . Instead of passing to the next line, we write

$$\begin{aligned} B_n(g)'(x) &= \frac{1}{x(1-x)} \sum_{j=0}^n \binom{n}{j} g\left(\frac{j}{n}\right) (j - nx) x^j (1-x)^{n-j} \\ &= \frac{n}{x(1-x)} \sum_{j=0}^n \binom{n}{j} g\left(\frac{j}{n}\right) \left(\frac{j}{n} - x\right) x^j (1-x)^{n-j}. \end{aligned}$$

After expanding  $g(\frac{j}{n})(\frac{j}{n} - x) = \frac{j}{n}g(\frac{j}{n}) - xg(\frac{j}{n})$  we recognize the expression

$$B_n(g)'(x) = \frac{n}{x(1-x)} (B_n(g_1)(x) - xB_n(g_2)(x))$$

in which the Bernstein polynomials of  $g_1(y) = yg(y)$  and  $g_2(y) = xg(y)$  occur. Writing both in expectation form (3.2) gives (3.5), since

$$V(T_n) = \frac{x(1-x)}{n}.$$

Now we use (3.6) to prove the lemma. Let  $\varepsilon > 0$  be given. There exists  $\delta > 0$  such that when  $|x - x_0| < \delta$  we have  $|\varepsilon(x)| < \varepsilon$ . As in Bernstein's theorem, we estimate the right-hand side of (3.6) separately on  $\{|T_n - x_0| < \delta\}$  and  $\{|T_n - x_0| \geq \delta\}$ . In the first set,  $|\varepsilon(T_n)| < \varepsilon$ , hence

$$\left| \int_{\{|T_n - x_0| < \delta\}} \varepsilon(T_n)(T_n - x_0)^2 dP \right| \leq \varepsilon V(T_n).$$

On the second set, we write  $M = \max |\varepsilon(x)|$  and invoke Cauchy's inequality to conclude that

$$\left| \int_{\{|T_n - x_0| \geq \delta\}} \varepsilon(T_n)(T_n - x_0)^2 dP \right|^2 \leq M^2 P(|T_n - x_0| \geq \delta) E((T_n - x_0)^4).$$

By Lemma 2.1, we have

$$E((T_n - x_0)^4) \leq \frac{1}{n^2}$$

for all  $x$  in  $[0, 1]$ . Applying Chebychev's inequality (2.2) and the formulas in (2.1), we obtain

$$|B_n(g)'(x_0)| \leq \varepsilon + \frac{M}{x_0(1-x_0)} \left( \frac{x_0(1-x_0)}{n\delta^2} \right)^{1/2}.$$

For all sufficiently large  $n$  the last expression does not exceed  $2\varepsilon$ , which gives the desired conclusion  $B_n(g)'(x_0) \rightarrow 0$ .  $\square$

#### 4. BROWNIAN MOTION FROM BERNSTEIN POLYNOMIALS.

The name ‘‘Brownian motion’’ refers to Robert Brown’s studies in 1827–28 of the erratic movement of small particles found in grains of pollen (and then in various other organic and inorganic substances) under microscopic observation (see [4], and also chapters 2 to 4 of [12] for comments and quotes about the early history). A mathematical description of this type of phenomenon was obtained by L. Bachelier in 1900 in a very early work about mathematical finance that was not recognized until much later (see [1]), and then in physics by A. Einstein (see [6]) and M. Smoluchowski in 1905 in support of the molecular theory of matter.<sup>1</sup>

These descriptions were largely heuristic. The modern definition reads as follows. For a given probability space  $(\Omega, \Sigma, P)$ , a (one-dimensional) Brownian motion  $B$  is a collection of random variables  $B(x)$  ( $0 \leq x < \infty$ ) that have the following properties:

- (i)  $B(0) = 0$  almost surely;
- (ii) if  $0 \leq x_1 < x_2 < \dots < x_n$ , then the vector  $(B(x_1), \dots, B(x_n))$  is a centered Gaussian random vector (i.e.,  $B$  is a Gaussian process);
- (iii) the covariance function  $E(B(x)B(y))$  is given for  $x$  and  $y$  in  $[0, \infty)$  by

$$E(B(x)B(y)) = x \wedge y; \tag{4.1}$$

- (iv) almost surely, the path  $x \mapsto B(x)$  (called a *sample path*) is a continuous function on  $[0, +\infty)$ .

From the first three properties and the fact that centered Gaussian random variables are independent if and only if they are orthogonal, one deduces the additional property:

- (v) the process has independent increments (i.e., whenever  $x_1 < \dots < x_n$ , the random variables  $B(x_1)$ ,  $B(x_2) - B(x_1)$ ,  $\dots$ ,  $B(x_n) - B(x_{n-1})$  are independent (Gaussian) random variables).

(For instance with  $n = 2$  and  $x_1 < x_2$ , we compute

$$E(B(x_1)(B(x_2) - B(x_1))) = E(B(x_1)B(x_2)) - E(B(x_1)^2) = x_1 - x_1 = 0$$

by invoking (4.1).)

To a mathematician not previously acquainted with the theory, the existence of Brownian motion should by no means be obvious. To someone with a sufficient background in measure theory, it may indeed be ‘‘obvious’’ that the first three conditions can be met through some abstract construction. However, the continuity of sample paths is not at all clear.

---

<sup>1</sup> Although Brownian motion would appear to be a wonderful subject for a book geared to a general audience, there does not seem to be one available.

The first construction of Brownian motion was due to Wiener [16]. Many other constructions followed, notably that of P. Lévy; a number of them are described in [8, chap 2., secs. 2–4], for instance. However the following construction, based on Bernstein polynomials, seems to be new, although the method closely resembles the one based on the limit in distribution of scaled random walks (see [3, secs. 2, 3] for an introduction).

Our idea is to construct  $B$  on  $[0, 1]$  as a limit of its Bernstein polynomials, described a priori. Indeed, the  $n$ th Bernstein polynomial of  $B(x)$  involves only  $B(j/n)$  with  $0 \leq j \leq n$ , and from property (iii) we know that  $B(j/n)$  must be a centered Gaussian random variable with variance  $V(B(j/n)) = j/n$ . Moreover,

$$B\left(\frac{j}{n}\right) = \sum_{1 \leq k \leq j} \left( B\left(\frac{k}{n}\right) - B\left(\frac{k-1}{n}\right) \right) = Y_{n1} + \cdots + Y_{nj},$$

where  $Y_{nj} = B(j/n) - B((j-1)/n)$ , so (by property (iv))  $(Y_{nj})_{1 \leq j \leq n}$  is a vector of independent centered Gaussian random variables with variance  $V(Y_{jn}) = 1/n$ .

We now suppose given for each positive integer  $n$  a vector  $(Y_{nj})_{1 \leq j \leq n}$  of independent centered Gaussian random variables with variance  $1/n$ , and we define

$$X_{n0} = 0, \quad X_{nj} = Y_{n1} + \cdots + Y_{nj} \quad (1 \leq j \leq n) \quad (4.2)$$

$$\mathcal{B}_n(x) = \sum_{j=0}^n \binom{n}{j} X_{nj} x^j (1-x)^{n-j}. \quad (4.3)$$

As a finite linear combination of Gaussian random variables,  $\mathcal{B}_n(x)$  is a Gaussian random variable for each  $n$ , so  $x \mapsto \mathcal{B}_n(x)$  is a Gaussian process.

The first step in deciding whether these *Bernstein motions* converge somehow to a model of Brownian motion is to compute the covariance function  $E(\mathcal{B}_n(x)\mathcal{B}_n(y))$  (since those are centered Gaussian random variables, this will identify the law of  $\mathcal{B}_n$ ).

**Lemma 4.1.** *The covariance function of  $\mathcal{B}_n$  satisfies*

$$E(\mathcal{B}_n(x)\mathcal{B}_n(y)) = Q_n(x, y), \quad (4.4)$$

where  $Q_n = B_n(\wedge)$  is the  $n$ th Bernstein polynomial of the continuous function  $\wedge : [0, 1]^2 \rightarrow \mathbf{R}$  given by  $\wedge(x, y) = x \wedge y$ .

Recall that the Bernstein polynomial of a function of several variables is defined in (3.1).

*Proof.* From (4.3) we see that

$$E(\mathcal{B}_n(x)\mathcal{B}_n(y)) = \sum_j \sum_k \binom{n}{j} \binom{n}{k} x^j (1-x)^{n-j} y^k (1-y)^{n-k} E(X_{nj}X_{nk}) \quad (4.5)$$

and that

$$E(X_{nj}X_{nk}) = \sum_{1 \leq j' \leq j} \sum_{1 \leq k' \leq k} E(Y_{nj'}Y_{nk'}). \quad (4.6)$$

But

$$E(Y_{nj'}Y_{nk'}) = \frac{\delta(j', k')}{n},$$

since  $Y_{nj'}$  and  $Y_{nk'}$  are independent when  $j' \neq k'$  and  $Y_{nj}$  has variance  $1/n$ . Thus, after the smoke has cleared only the diagonal terms with  $j' = k'$  remain in (4.6), and they can occur only if  $j' \leq j$  and  $j' \leq k$  (i.e., if  $j' \leq j \wedge k$ ). Therefore

$$E(X_{nj}X_{nk}) = \frac{j \wedge k}{n} = \frac{j}{n} \wedge \frac{k}{n},$$

which, when inserted into (4.5) gives precisely (4.4) in light of (3.1).  $\square$

**Corollary 4.2.** (1) *For each  $x$  in  $[0, 1]$  the sequence of random variables  $(\mathcal{B}_n(x))$  converges in distribution to a centered Gaussian random variable with variance  $x$ .*

(2) *For each  $(x, y)$  in  $[0, 1]^2$  the sequence of vectors  $(\mathcal{B}_n(x), \mathcal{B}_n(y))$  converges in distribution to a vector  $(B(x), B(y))$  of Gaussian centered random variables with  $E(B(x)B(y)) = x \wedge y$ .*

Corollary 4.2 follows immediately from Lemma 4.1 and Theorem 3.1. (Recall from section 2 that a sequence  $(X_n)$  of centered Gaussian random vectors will converge in distribution if their covariance matrices converge.) From this corollary we see that if we temporarily denote by  $B(x)$  any random variable with the same distribution as the limit of  $\mathcal{B}_n(x)$ , the resulting process is a Gaussian process satisfying all conditions of a Brownian motion (on  $[0, 1]$ ) except the (almost sure) continuity of sample paths.

This last condition is naturally more subtle, because in defining  $\mathcal{B}_n$  as we do, there is no a priori relation whatsoever between the values of  $X_{n,j}$  and  $X_{m,k}$  for  $n \neq m$ , contrary to the heuristic argument that  $X_{n,j}$  should represent  $B(j/n)$ . One could try to choose the variables  $Y_{n,j}$  and construct  $X_{n,j}$  in such a way that for almost all fixed samples  $\omega$  in  $\Omega$  the definition  $B(j/n) = X_{n,j}$  is coherent and extends to a continuous function. However, this does not seem to be easy. Instead we consider the limiting distribution of polynomials  $\mathcal{B}_n$  viewed as elements of the space  $C([0, 1], \mathbf{R})$  of real-valued continuous functions on  $[0, 1]$ .

Recall that  $C([0, 1], \mathbf{R})$  is a complete metric space with the distance given by the metric  $\|f\| = \sup |f(x)|$ . In particular, there is the associated Borel  $\sigma$ -algebra  $\mathcal{B}$  of subsets of  $C([0, 1], \mathbf{R})$ . Then, as might be expected, the map

$$\mathcal{B}_n : \Omega \rightarrow C([0, 1], \mathbf{R})$$

is measurable. This is true because  $\mathcal{B}$  is generated by the cylinder sets

$$C_{x,A} = \{f \in C([0, 1], \mathbf{R}) : f(x) \in A\}$$

for  $x$  in  $[0, 1]$  and  $A$  a Borel subset of  $\mathbf{R}$  (see [8, chap. 2, sec.4, Problem 4.2]), and  $\mathcal{B}_n^{-1}(C_{x,A}) = \{\omega : |\mathcal{B}_n(\omega, x)| \in A\}$ , which is measurable by definition for fixed  $x$ .

We use the maps  $\mathcal{B}_n$  to define image measures  $W_n = P(\mathcal{B}_n)$  on  $C([0, 1], \mathbf{R})$ . In other words, for a (Borel) set  $X$  of continuous functions  $W_n(X) = P(Y)$ , where  $Y$  is the set of “events” in  $\Omega$  for which the corresponding  $n$ th Bernstein polynomial is in  $X$ . Our model of Brownian motion is then associated with the probability space  $(C([0, 1], \mathbf{R}), \mathcal{B}, W)$  for the probability measure  $W$  on

$C([0, 1], \mathbf{R})$  that is the limit in distribution of the measures  $W_n$ , with the random variable  $B(x)$  being simply the “evaluation at  $x$ ” map on  $C([0, 1], \mathbf{R})$ . The measure  $W$  is called *Wiener measure*.

**Theorem 4.3.** *With  $\mathcal{B}_n$  defined as in (4.3), the sequence  $(W_n)$  of probability measures on  $C([0, 1], \mathbf{R})$  converges in distribution to a probability measure  $W$  such that on the probability space  $(C([0, 1], \mathbf{R}), \mathcal{B}, W)$  the stochastic process defined by the random variables  $B(x)$ , where  $B(x)(f) = f(x)$  for  $f$  in  $C([0, 1], \mathbf{R})$ , is a Brownian motion restricted to  $[0, 1]$ .*

Note that passing from  $B(x)$  for  $x$  in  $[0, 1]$  to a proper Brownian motion defined whenever  $0 \leq x < +\infty$  is not difficult (see [8, chap. 2, Corollary 3.4]).

We will use the following application of Skorohod’s criterion for convergence of probability measures in distribution:

**Theorem 4.4.** *Let  $(\mathcal{B}_n)$  be a sequence of continuous Gaussian processes on  $[0, 1]$  with the property that for each finite set  $t_1 < t_2 < \dots < t_m$  in  $[0, 1]$  the Gaussian random vectors  $(\mathcal{B}_n(t_1), \dots, \mathcal{B}_n(t_m))$  converge in distribution. Suppose that  $\mathcal{B}_n(0) = 0$  almost surely and that*

$$E((\mathcal{B}_n(x) - \mathcal{B}_n(y))^4) \leq C|x - y|^2 \quad (4.7)$$

holds on  $[0, 1]^2$  for some absolute constant  $C$ . Then the associated sequence of probability measures  $W_n = P(\mathcal{B}_n)$  on  $C([0, 1], \mathbf{R})$  converges in distribution to a measure  $W$  with the following property: for each  $x$  in  $[0, 1]$  the sequence  $\mathcal{B}_n(x)$  converges in distribution to the random variable  $f \mapsto f(x)$  on the probability space  $C([0, 1], \mathbf{R})$  equipped with the measure  $W$ .

*Proof.* This is a special case of [8, chap. 2, Problem 4.11, Theorem 4.15], or [14, chap. 13, sec. 1.8].  $\square$

*Proof of Theorem 4.3.* Because of the previous results, it suffices to check (4.7) for the Bernstein processes  $\mathcal{B}_n$ . In fact, we verify (4.7) with the constant  $C = 12$ . For this, notice that  $\mathcal{B}_n(x) - \mathcal{B}_n(y)$  is a Gaussian centered random variable with variance

$$H_n(x, y) = E((\mathcal{B}_n(x) - \mathcal{B}_n(y))^2) = Q_n(x, x) + Q_n(y, y) - 2Q_n(x, y)$$

(Lemma 4.1). As for any centered Gaussian, the fourth moment is given in terms of the variance by (2.4), so

$$E((\mathcal{B}_n(x) - \mathcal{B}_n(y))^4) = 3H_n(x, y)^2.$$

To estimate  $H_n(x, y)$ , we write

$$H_n(x, y) = Q_n(x, x) - Q_n(x, y) + Q_n(y, y) - Q_n(x, y).$$

For the last two terms we obtain

$$Q_n(y, y) - Q_n(x, y) = \int_x^y \partial_x Q_n(t, y) dt.$$

We then appeal to Lemma 3.2 to compute the partial derivative  $\partial_x Q_n$ . Noting that

$$\left| \frac{x_1 \wedge y - x_2 \wedge y}{x_1 - x_2} \right| \leq 1$$

for all  $x_1, x_2$ , and  $y$  in  $[0, 1]$  (look at the slopes of the graph of  $x \mapsto x \wedge y$ ), we derive from the lemma and the binomial theorem the estimate

$$|\partial_x Q_n(t, y)| \leq \sum_{j=0}^{n-1} \sum_{k=0}^n \binom{n-1}{j} \binom{n}{k} t^j (1-t)^{n-1-j} y^k (1-y)^{n-k} = 1.$$

It follows that

$$|Q_n(y, y) - Q_n(x, y)| \leq |x - y|.$$

Symmetrically, we have

$$|Q_n(x, x) - Q_n(x, y)| \leq |x - y|.$$

Finally, we arrive at the inequalities  $0 \leq H_n(x, y) \leq 2|x - y|$  and

$$E((\mathcal{B}_n(x) - \mathcal{B}_n(y))^4) = 3H_n(x, y)^2 \leq 12|x - y|^2.$$

□

**Remark 4.5.** This construction of Brownian motion is essentially theoretical and not really suitable for numerical modeling purposes, since to compute a sample  $\mathcal{B}_n(x)$  effectively, one must either compute very large binomial coefficients (e.g.,  $\binom{1000}{500} > 10^{299}$ ) to expand the Bernstein polynomial or compute many approximate values of the binomial distribution for various parameters. Figure 1 shows a sample path computed by the latter method with  $n = 5000$ .

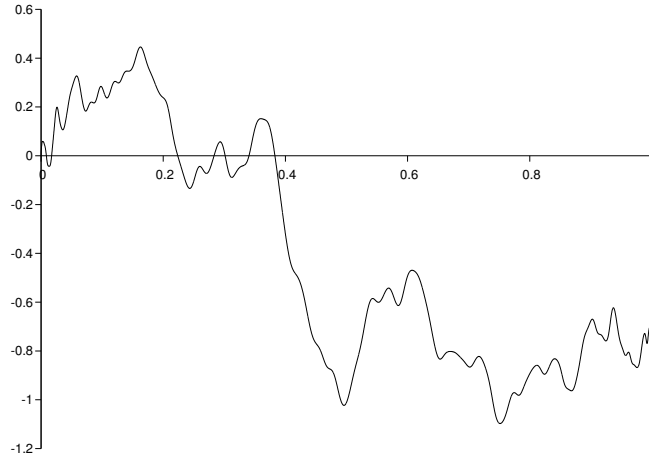


FIGURE 1. A Bernstein polynomial with  $n = 5000$ .

## 5. NONDIFFERENTIABILITY OF BROWNIAN PATHS.

We now adopt the point of view that we know that the mathematical model of Brownian motion exists — for instance (but not necessarily) as derived from Bernstein polynomials — and we seek further properties of the sample paths (i.e., of the functions  $x \mapsto B(x)$ ). In particular, we ask: Are these functions differentiable? This is a classical question, but we approach it via the Bernstein polynomials.

Define the  $n$ th Bernstein process  $\mathfrak{B}_n$  associated with  $B$  by taking the  $n$ th Bernstein polynomial of the function  $x \mapsto B(x)$  on  $[0, 1]$ :

$$\mathfrak{B}_n(x) = \sum_{j=0}^n \binom{n}{j} B\left(\frac{j}{n}\right) x^j (1-x)^{n-j},$$

which is almost surely the  $n$ th Bernstein polynomial of a continuous function. Note that this definition is indeed independent from that of the  $\mathcal{B}_n$  of the previous section, although of course  $\mathfrak{B}_n$  and  $\mathcal{B}_n$  share the same distribution.

On the basis of section 3, we see that almost surely  $\mathfrak{B}_n(x)$  converges uniformly to  $B(x)$ , and by Lemma 3.3 we find that the convergence of the derivatives can detect differentiability of  $B(x)$  at a point. Using this we prove in a quite elementary fashion that for a fixed  $x_0$  in  $[0, 1]$  the path  $x \mapsto B(x)$  is almost surely nondifferentiable at  $x_0$ .

By the remark immediately following the statement of Lemma 3.3, it suffices to show that almost surely the sequence  $(\mathfrak{B}'_n(x_0))$  is not convergent. Since from the first formula in Lemma 3.2,  $\mathfrak{B}'_n(x_0)$  is again a centered Gaussian random variable, we start once more by computing its variance. We have

$$\mathfrak{B}'_n(x) = n \sum_{j=0}^{n-1} \binom{n-1}{j} Y_{nj} x^j (1-x)^{n-1-j}, \quad (5.1)$$

where  $Y_{nj} = B((j+1)/n) - B(j/n)$ . By property (v) of Brownian motion, the  $Y_{n,j}$  are independent centered Gaussian random variables with variance  $1/n$ , so the corresponding orthogonality gives the variance

$$V(\mathfrak{B}'_n(x)) = n \sum_{j=0}^{n-1} \binom{n-1}{j}^2 x^{2j} (1-x)^{2(n-1-j)}.$$

We denote this by  $nR_{n-1}(x)$ . Note that  $R_n(x)$  is essentially yet another Bernstein polynomial: it is the  $n$ th Bernstein polynomial associated with the discontinuous function  $\delta(x, y)$  on  $[0, 1]^2$  that is the characteristic function of the diagonal  $\{(x, y) : x = y\}$ , evaluated at  $(x, x)$ . So Bernstein's theorem does not apply to determine the behavior of  $R_n(x_0)$  as  $n \rightarrow \infty$ , but the presence of the factor  $n$  leads one to believe that  $nR_{n-1}(x_0)$  should be large. If so, then it would follow from the fairly intuitive Lemma 2.2 that the sequence  $\mathfrak{B}'_n(x_0)$  were almost surely unbounded, giving the desired result.

The point is thus to get a lower bound for  $R_n(x)$ . A few cases, done by hand or using a computer algebra system, suggest quickly that  $R_n(x)$  for  $x$  in  $[0, 1]$  is minimal for  $x = 1/2$ . But at  $x = 1/2$  we have

$$R_n\left(\frac{1}{2}\right) = 2^{-2n} \sum_{j=0}^n \binom{n}{j}^2 = \frac{\binom{2n}{n}}{2^{2n}} \sim \frac{1}{\sqrt{\pi n}}$$

by Stirling's formula, so we would anticipate  $nR_{n-1}(x_0)$  to be of size on the order of  $\sqrt{n}$  at least.

The inequality

$$R_n(x) \geq R_n\left(\frac{1}{2}\right) \quad (n \geq 1, x \in [0, 1]) \quad (5.2)$$

turns out to be true, but the proof is surprisingly involved. It was provided by T. Rivoal in the following stronger form:

$$R_n(u + \frac{1}{2}) = \sum_{j=0}^n a_{nj} u^{2j}$$

with  $a_{nj} \geq 0$  for all  $n$  and  $j$ . His argument uses an identity of Watson concerning hypergeometric functions. It would be quite interesting to have a more elementary proof of this seemingly innocent proposition. This is the subject of Problem 11155 in the May 2005 issue of this MONTHLY. Perhaps the following probabilistic interpretation can help:  $R_n(x)$  is the probability that a walker starting at the origin returns to the origin after performing a  $2n$ -step Bernoulli random walk with independent steps 1 and  $-1$  having probability  $x$  and  $1-x$ , respectively, for the first  $n$  steps and then probability  $1-x$  and  $x$ , respectively, for the last  $n$  steps.

However, instead of (5.2), we can prove quite easily an inequality that is almost as strong and sufficient for our application.

**Lemma 5.1.** *The estimates*

$$R_n(x) \geq \frac{9}{32} \frac{1}{\sqrt{n}}$$

and

$$V(\mathfrak{B}'_n(x)) = E(\mathfrak{B}'_n(x)^2) \geq \frac{9\sqrt{n}}{32}$$

hold for all  $x$  in  $[0, 1]$ .

*Proof.* Let  $S_n = X_1 + \cdots + X_n$  with  $(X_i)$  the independent random variables with Bernoulli law giving the probabilistic interpretation of  $B_n(x)$ . We can write

$$R_n(x) = \sum_{j=0}^n P(S_n = j)^2.$$

We then estimate as follows: by Chebychev's inequality (2.2), we know that for any positive  $\delta$

$$1 - \frac{1}{4\delta^2 n} \leq P\left(\left|\frac{S_n}{n} - x\right| \leq \delta\right).$$

The right-hand side has the expansion

$$\begin{aligned} P\left(\left|\frac{S_n}{n} - x\right| \leq \delta\right) &= \sum_{\substack{0 \leq j \leq n \\ |j/n - x| \leq \delta}} P(S_n = j) \\ &\leq \left(\sum_{\substack{0 \leq j \leq n \\ |j/n - x| \leq \delta}} 1\right)^{1/2} \left(\sum_{\substack{0 \leq j \leq n \\ |j/n - x| \leq \delta}} P(S_n = j)^2\right)^{1/2} \leq \sqrt{2n\delta} \sqrt{R_n(x)}, \end{aligned}$$

which is justified by an application of Cauchy's inequality and positivity. Hence by taking  $\delta = n^{-1/2}$  we arrive at

$$R_n(x) \geq \frac{\left(1 - \frac{1}{4n\delta^2}\right)^2}{2n\delta} \geq \frac{9}{32} \frac{1}{\sqrt{n}}.$$

This choice is in fact predictable from the central limit theorem, which can be interpreted as implying that with positive probability the  $n$ -step random walk is, uniformly in  $n$ , within a distance of  $\sqrt{n}$  from the average value  $x$ .) Hence we get

$$nR_{n-1}(x) \geq \frac{9}{32} \frac{n}{\sqrt{n-1}} \geq \frac{9\sqrt{n}}{32},$$

as desired.  $\square$

As mentioned earlier, Lemma 5.1 ensures that almost surely  $B(x)$  is non-differentiable at  $x_0$  for any fixed  $x_0$ . A well-known formal argument based on Fubini's theorem shows that this implies that almost surely the set of  $x$  where  $B(x)$  is differentiable is of measure 0. We recall this quickly:<sup>2</sup>

**Proposition 5.2.** *Let  $(\Omega, \Sigma, P)$  be a probability space and let  $B(x)$  be a Brownian motion defined on this space for  $0 \leq x \leq 1$ . Then almost surely the set of  $x_0$  in  $[0, 1]$  such that  $x \mapsto B(x)$  is differentiable at  $x_0$  is of Lebesgue measure zero.*

*Proof.* Consider Brownian motion as a mapping

$$B : \Omega \times [0, 1] \rightarrow \mathbf{R}$$

and observe that the almost sure continuity of paths implies easily that this mapping is measurable for the product  $\sigma$ -algebra on  $\Omega \times [0, 1]$  (see [8, chap 1, Proposition 1.13]), for it is the pointwise limit of the usual simple “vertical” dyadic approximations that are clearly measurable. Let

$$D = \{(\omega, y) \in \Omega \times [0, 1] : x \mapsto B(\omega, x) \text{ is differentiable at } y\},$$

$$E = \{(\omega, y) \in \Omega \times [0, 1] : \lim_{n \rightarrow \infty} \frac{B(\omega, y + 1/n) - B(\omega, y)}{1/n} \text{ exists}\}$$

Note that  $D$  lies in  $E$ . The usual manipulations with countable intersections and unions show that, because of the aforementioned measurability,  $E$  is measurable for the product  $\sigma$ -algebra. By Tonelli's theorem

$$P \times \lambda(E) = \int_{\Omega} \lambda(E(\omega)) dP = \int_{[0,1]} P(E(x)) d\lambda(x),$$

where  $\lambda$  is Lebesgue measure and  $E(x)$  (respectively,  $E(\omega)$ ) are the “slices” of  $E$  for fixed  $x$  or  $\omega$ . From the preceding discussion we infer that  $P(E(x)) = 0$  for all  $x$ . Hence  $P \times \lambda(E) = 0$ , so that we must have  $\lambda(E(\omega)) = 0$  for almost all  $\omega$ . This gives  $\lambda(D(\omega)) = 0$  also, which is the statement of Proposition 5.2.  $\square$

**Remark 5.3.** Figure 2 depicts the graph of the derivative of the Bernstein polynomial in Figure 1.

<sup>2</sup> Again, this is weaker than what was proved by Paley, Wiener, and Zygmund, namely, that this set is empty almost surely.

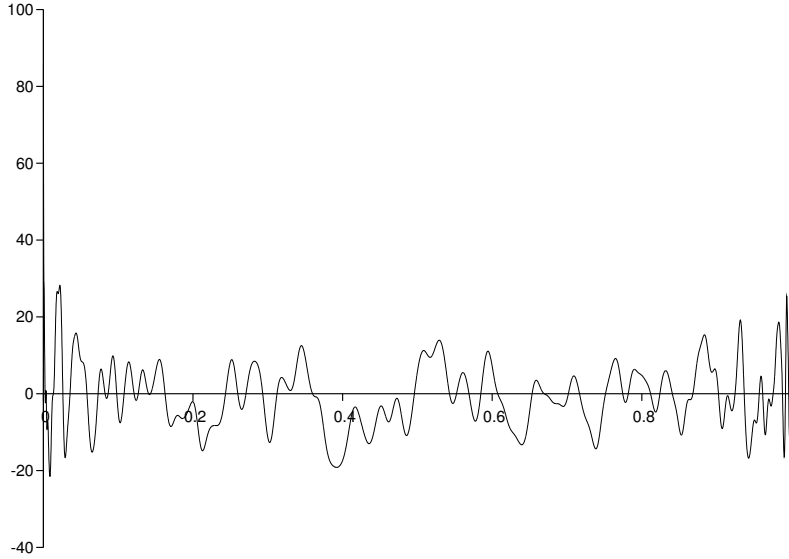


FIGURE 2. The derivative of the Bernstein polynomial in Figure 1.

## 6. ZEROS OF RANDOM BERNSTEIN POLYNOMIALS.

In this section we again consider a Brownian motion and the Bernstein processes  $\mathfrak{B}_n(x)$  as in (1.2), but this time as a family of random polynomials of degree at most  $n$ . As for any family of random polynomials, one of the most natural questions asks for the structure of its zero set. Here is a simple result of this type.

**Remark.** This statement is incorrect, as pointed out by J. Ruf; the problem is that it is not true that Brownian motion has probability zero to be  $\geq 0$  on  $[a, b]$  if  $a \neq 0$  and  $b > a$ . The statement might be true for  $a = 0$ , but the proof doesn't work. We will try to give a correct proof in a subsequent version...

**Proposition 6.1.** *If  $0 \leq a < b \leq 1$  and  $N \geq 1$ , then almost surely the polynomial  $\mathfrak{B}_n(x)$  has at least  $N$  zeros in  $[a, b]$  for all sufficiently large  $n$ .*

*Proof.* We first establish the result for  $N = 1$  and then deduce the general case. Let  $\omega$  in  $\Omega$  be such that the path  $x \mapsto B(\omega, x)$  is continuous and such that there exists a subsequence  $(n_k)$  with the property that  $\mathfrak{B}_{n_k}(x)$  has no zero in  $[a, b]$ . By continuity (of polynomials), this means that  $\mathfrak{B}_{n_k}$  is either positive or negative on  $[a, b]$ . Passing to a further subsequence if necessary, we may assume that either  $\mathfrak{B}_{n_k} > 0$  on  $[a, b]$  for all  $k$  or  $\mathfrak{B}_{n_k} < 0$  on  $[a, b]$  for all  $k$ . Assume the former. Since  $\mathfrak{B}_{n_k}(x)$  converges uniformly to  $B(x)$  on  $[0, 1]$ , it follows that  $B(x) \geq 0$  for all  $x$  in  $[a, b]$ . In the second case, of course, we have  $B(x) \leq 0$  for all  $x$  in  $[a, b]$ . However both cases occur with probability zero by known properties of Brownian paths (see [8, chap. 2, Problem 7.18] applied to  $B(t) - B(a)$  with  $t \leq b - a$ ).

Now consider an arbitrary  $N$ . Choosing (arbitrarily)  $N$  disjoint subintervals  $[a_j, b_j]$  of  $[a, b]$  with  $a_j < b_j$ , we know by the case  $N = 1$  that for  $\omega$  outside a set  $E_j$  of probability 0 the polynomial  $\mathfrak{B}_n(x)$  has a zero in  $[a_j, b_j]$

whenever  $n \geq N(j, \omega)$ . Hence for  $\omega$  not in the union of the  $E_j$ ,  $\mathfrak{B}_n(x)$  has at least  $N$  zeros in  $[a, b]$  as soon as  $n$  is at least  $\max\{N(j, \omega)\}$ .  $\square$

**Remark 6.2.** It would be quite interesting to have quantitative versions of Proposition 6.1. In particular, it is natural to ask whether one can choose  $n$  large enough independently of  $\omega$  so that the conclusion holds? Certainly, putting

$$F_N = \{\omega : \mathfrak{B}_n(\omega, x) \text{ has a zero on } [a, b] \text{ for all } n \geq N\},$$

which is measurable (because the set  $Z$  of continuous functions on  $[0, 1]$  that have zeros is closed in  $C([0, 1], \mathbf{R})$ , being the complement of the obviously open set of functions  $f$  such that  $\min |f| > 0$ ), we have  $F_N \subset F_{N+1}$  and  $\lim P(F_N) = 1$  by Proposition 6.1. Accordingly, for each  $\varepsilon$  ( $0 < \varepsilon < 1$ ) there exists some  $N$  such that with probability greater than  $1 - \varepsilon$  the polynomial  $\mathfrak{B}_n(x)$  has a zero on  $[a, b]$  once  $n \geq N$ .

Another interesting question is: How large must  $n$  be as a function of  $N$ ,  $a$ , and  $b$  so that  $\mathfrak{B}_n$  has at least  $N$  zeros in  $[a, b]$ ? One would hope to link these questions to the finer structure of the zero set  $Z(\omega) = \{x \in [0, 1] : B(\omega, x) = 0\}$ , which is known, for instance, to be almost surely an uncountable set of Lebesgue measure zero but Hausdorff dimension one-half (see [8, chap. 2, sec. 9B] for an introduction to this topic).

As a first step, note that for quite general random functions of the type

$$f(x) = \sum_{i=0}^d a_i f_i(x),$$

where  $a_i$  is a centered Gaussian random variable, Theorem 3.1 in [5] shows that the expected number of zeros of  $f$  in a measurable subset  $I$  of  $\mathbf{R}$  is given by

$$E(\text{Card}\{x \in I : f(x) = 0\}) = \frac{1}{\pi} \int_I \rho(x)^{1/2} dx.$$

Here

$$\rho(x) = \left( \frac{\partial^2}{\partial x \partial y} \log Q \right)(x, x),$$

where  $Q$  is the covariance function  $Q(x, y) = E(f(x)f(y))$  (although it is not defined this way in [5]). With  $f_i(x) = \binom{n}{i} x^i (1-x)^{n-i}$  and  $a_i = B(i/n)$ , we have  $Q(x, y) = Q_n(x, y)$ , the Bernstein polynomial of  $x \wedge y$ , as computed in Lemma 4.1 of section 4.

**ACKNOWLEDGMENTS.** It should be clear that the author is at best an amateur probabilist. It may well be in particular that the ideas exposed here are already known, although a Math Reviews search did not reveal any obvious previous work. But in any event, thinking and writing about these questions was a very pleasant task, bringing back fond memories of learning probability theory and stochastic calculus with A. Goldman in 1990–91. I also thank R. Deville and E. Saint Loubert Bié for some probabilistic help and T. Rivoal for proving the polynomial inequality (5.2).

## REFERENCES

- [1] L. Bachelier, Théorie de la spéculation, *Ann. Sci. École Norm. Sup.* (3) **17** (1900) 21–86; available at <http://www.numdam.org>
- [2] S. Bernstein, Démonstration du théorème de Weierstrass basée sur le calcul des probabilités, *Commun. Soc. Math. Kharkow* (2) **13** (1912–1913) 1–2.
- [3] P. Billingsley, Prime numbers and Brownian motion, *Amer. Math. Monthly* **80** (1973) 1099–1115.
- [4] R. Brown, A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies, in *The Miscellaneous Botanical Works of Robert Brown*, Vol. 1; available at <http://sciweb.nybg.org/science2/pdfs/dws/Brownian.pdf>
- [5] A. Edelman and E. Kostlan, How many zeros of a random polynomial are real, *Bull. Amer. Math. Soc.* **82** (1995) 1–37.
- [6] A. Einstein, Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen (On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat), *Ann. Physik* **17** (1905); available at <http://www.physik.uni-augsburg.de/annalen/history/Einstein-in-AdP.htm>
- [7] S. Kakutani, Two-dimensional brownian motion and harmonic functions, *Proc. Acad. Japan* **20** (1944) 706–714.
- [8] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York, 1988.
- [9] N. Katz and P. Sarnak, Zeroes of zeta functions and symmetry, *Bull. Amer. Math. Soc.* **36** (1999) 1–26.
- [10] J. Keating and N. Snaith, Random matrix theory and  $\zeta(1/2 + it)$ , *Comm. Math. Phys.* **214** (2000) 57–89.
- [11] G. G. Lorentz, *Bernstein polynomials*, University of Toronto Press, Toronto, 1953; 2nd ed., Chelsea, New York, 1986.
- [12] E. Nelson, *Dynamical theories of Brownian motion*, Princeton University Press, Princeton, 1967; also available at <http://www.math.princeton.edu/~nelson/books.html>
- [13] R.E.A.C Paley, N. Wiener and A. Zygmund, Notes on random functions, *Math. Zeitschrift* **37** (1933) 647–668.
- [14] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, 3rd ed., Springer-Verlag, Berlin, 1999.
- [15] A.N. Shiryaev, *Probability*, Springer-Verlag, New York, 1984.
- [16] N. Wiener, Differential space, *J. Math. Phys.* **2** (1923) 131–174.

**EMMANUEL KOWALSKI** studied mathematics in France at the E.N.S Lyon and Institut Fourier (Grenoble), then in the U.S.A at Rutgers University where he obtained his Ph.D. under the supervision of H. Iwaniec in 1998. After a postdoc position at Princeton University and the Institute for Advanced Study, he returned to France, where he is now a professor in Bordeaux. Most of his research is devoted to number theory, in particular analytic number theory, but he keeps from his student years an amateur interest in other topics, especially probability.

UNIVERSITÉ BORDEAUX I - A2X, 351, COURS DE LA LIBÉRATION, 33405 TALENCE CEDEX, FRANCE

*E-mail address:* [emmanuel.kowalski@math.u-bordeaux1.fr](mailto:emmanuel.kowalski@math.u-bordeaux1.fr)