Laplace and the origin of the Ornstein–Uhlenbeck process

MARTIN JACOBSEN

Institute of Mathematical Statistics, University of Copenhagen, 5 Universitetsparken, 2100 Copenhagen Ø, Denmark

It is shown that a second-order partial differential equation, found by Laplace in 1810, is the Fokker–Planck equation for a one-dimensional Ornstein–Uhlenbeck process. It is argued that Laplace's reasoning, although not rigorous, can be entirely justified by using the modern theory of weak convergence of stochastic processes. The solutions to the differential equation found by Laplace and others, using expansions in terms of Hermite polynomials, are discussed.

Keywords: Bernoulli–Laplace urn model; diffusion process; Fokker–Planck equation; Hermite polynomials; weak convergence

1. Introduction

The backward and forward (or Fokker–Planck) differential equations are some of the basic tools used in the study of diffusion processes. For instance, it is well known that the densities of the transition probabilities for a diffusion satisfy these second-order partial differential equations under quite general conditions.

Major contributions to the (semigroup) theory of diffusions, including the backward and forward equations, were made by Kolmogorov, Feller and Dynkin, but examples of the diffusion differential equations appear earlier, notably in the work of Bachelier (1906; 1912) in connection with random walks and Brownian motion. By far the earliest occurrence of such an equation is, however, in the work of Laplace (1810–11; 1812), and it is this equation, derived by Laplace by approximating with a sequence of urn models first proposed by Daniel Bernoulli (1770), which is the topic of the present paper.

The argument given by Laplace (or rather, the lack of same) has been strongly criticized by several authors, Todhunter (1865) in particular; see Section 3 below. We shall argue, however, that despite the lack of detail and rigour in Laplace's approach, his formidable intuition has led him to a differential equation which is entirely justifiable, and is in fact the Fokker–Planck equation for a one-dimensional Ornstein–Uhlenbeck process, which appears as the weak limit of the Bernoulli–Laplace urn models.

It is surprising that this does not seem to have been noticed before. Bachelier (1906) is possibly the first to write down the transition density of an Ornstein–Uhlenbeck process, and he does refer back to Laplace, while Markov (1915) establishes a connection between the urn models and the Ornstein–Uhlenbeck process through the convergence of moments

– but both papers are too early to be able to place the results in the context of the theory of stochastic processes. (A fairly recent discussion of Markov's results and of the contemporary paper by Steklov (1915) is Ondar (1970), a paper that we have not seen.) Certainly, in the surveys by Molina (1930) and Sheynin (1976–7) of Laplace's work there is no mention of a connection between urn schemes and diffusions other than that found by Markov.

In Section 2 we briefly survey the history of and set down some standard facts about the Ornstein–Uhlenbeck process. Section 3 is devoted to the Bernoulli–Laplace urn model and Laplace's controversial differential equation, while in Section 4 we discuss the solutions to the differential equation that were found by Laplace and investigated more thoroughly by Steklov. The solutions are given as expansions in terms of Hermite polynomials, and we show how the solutions given in this form agree with the forms readily available from Markov process theory.

For overviews of Laplace's work in probability theory, the reader is referred to the papers by Molina (1930) and Sheynin (1976–7) already mentioned. Also, Hald (1990) treats various aspects of Laplace's work and in a forthcoming book on the history of mathematical statistics will present an extensive discussion of the basic contributions to probability and statistics made by Laplace.

2. The Ornstein–Uhlenbeck process

In their famous paper, Ornstein and Uhlenbeck (1930) studied a free particle in Brownian motion, moving in a rarefied gas and affected by a friction force proportional to the pressure. In order to understand the displacement process $x(t)$ of the particle, they investigated the velocity process $u(t) = x'(t)$, which is now known as the (one-dimensional) Ornstein–Uhlenbeck process.

The work of Ornstein and Uhlenbeck continues Einstein's fundamental work on Brownian motion itself (see Einstein 1956), but also owes much to the work of Smoluchowski (1915), who, using reasoning from molecular kinetics, derived the Fokker–Planck equation for the Ornstein–Uhlenbeck process and also determined the transition density (see (2) below).

The paper by Doob (1942) contains a precise mathematical definition of $u(t)$ as a continuous time-homogeneous Markov process, characterized (see Doob 1942, Theorem 1.1) as being, apart from white noise, the only stationary Gaussian process in continuous time which is also a Markov process. (Note that Doob's terminology is 'temporally homogeneous'. We use 'time-homogeneous' to describe a Markov process with transitions from time s to time $s + t$ that depend on t only, and 'stationary' to describe a process that is strictly stationary, i.e. with distributional properties that remain unchanged under translations of time.)

According to Doob, the process $u = (u(t))_{-\infty < t < \infty}$ is a stationary Gaussian process with continuous sample paths, defined by the constant expectation and variance functions

$$\mathrm{E}\,u(t) = m, \quad \mathrm{E}(u(t) - m)^2 = \sigma_0^2 \qquad (t \in \mathbb{R})$$

and the covariance function

$$\mathrm{E}[(u(t) - m)(u(s) - m)] = \sigma_0^2\,\mathrm{e}^{-\beta|t-s|} \qquad (s, t \in \mathbb{R}).$$

Here $m \in \mathbb{R}$ is arbitrary, while $\sigma_0^2 > 0$, and, as is essential for obtaining a stationary process, $\beta > 0$.

In the case where $m = 0$, it follows that the transition probabilities for u are given as follows (Doob 1942, p. 352): for any s and any $t > 0$, the conditional distribution of $u(s + t)$ given $u(s) = a$ is Gaussian with mean $a\,\mathrm{e}^{-\beta t}$ and variance $\sigma_0^2(1 - \mathrm{e}^{-2\beta t})$.

Having defined u, Doob proceeds to define the displacement process $x = (x(t))_{t \in \mathbb{R}}$ introduced by Ornstein and Uhlenbeck (1930) as

$$x(t) - x(0) = \int_0^t u(s)\,\mathrm{d}s$$

and shows, in particular, that the variance of $x(t) - x(0)$ is given by the expression originally found by Ornstein and Uhlenbeck.

From now on the time axis will be $\mathbb{R}_0 = [0, \infty[$ rather than $\mathbb{R}$, and we write $u = (u(t))_{t \geq 0}$. The Ornstein–Uhlenbeck process (with $m = 0$) may then be described (Doob 1942, p. 358) as the solution to the stochastic differential equation

$$\mathrm{d}u(t) = -\beta u(t)\,\mathrm{d}t + \sigma\,\mathrm{d}B(t), \tag{1}$$

where $B = (B(t))_{t \geq 0}$ is a standard one-dimensional Brownian motion and $\sigma = \sqrt{2\beta}\sigma_0$. The solution to (1) is completely specified once its initial value $u(0)$ is given, the only requirement being that $u(0)$ should be stochastically independent of B. We may then write for any $s \geq 0$, $t > 0$, $a, v \in \mathbb{R}$,

$$P(u(s + t) \leq v \mid u(s) = a) = \int_{-\infty}^v p_t(a, \mu)\,\mathrm{d}\mu,$$

with $p_t(a, \mu)$ the Gaussian transition density

$$p_t(a, \mu) = \frac{1}{\sqrt{2\pi\sigma^2\dfrac{1 - \mathrm{e}^{-2\beta t}}{2\beta}}}\exp\left(-\frac{1}{2\sigma^2\dfrac{1 - \mathrm{e}^{-2\beta t}}{2\beta}}(\mu - a\,\mathrm{e}^{-\beta t})^2\right). \tag{2}$$

If, in particular, $u(0)$ is Gaussian with mean 0 and variance $\sigma^2/2\beta = \sigma_0^2$, u is stationary.

It was noted by Hostinský (1932, p. 52), that this type of expression goes back at least to Smoluchowski (1915). The expression appears, however, also in the two contemporary papers by Markov (1915, p. 103) and Steklov (1915, p. 1536) and even earlier in the work of Bachelier (1906, p. 273).

From (1) it follows in particular that the infinitesimal generator A for the transition semigroup of the Ornstein–Uhlenbeck process u has the form

$$Af(\mu) = -\beta f'(\mu) + \tfrac{1}{2}\sigma^2 f''(\mu) \tag{3}$$

for f twice continuously differentiable with f and Af bounded. From the general theory of

Markov processes (see, for example, Dynkin 1965, or Feller 1971) it then also follows that the transition densities satisfy the backward Feller–Kolmogorov differential equations, and, as will be much more important to us, the forward or Fokker–Planck equations

$$\frac{\partial}{\partial t}\,p_t(a,\mu) = -\frac{\partial}{\partial\mu}\big(-\beta\mu p_t(a,\mu)\big) + \tfrac{1}{2}\sigma^2\frac{\partial^2}{\partial\mu^2}\,p_t(a,\mu)$$

$$= \beta p_t(a,\mu) + \beta\mu\frac{\partial}{\partial\mu}\,p_t(a,\mu) + \tfrac{1}{2}\sigma^2\frac{\partial^2}{\partial\mu^2}\,p_t(a,\mu). \tag{4}$$

This expression holds for all $a \in \mathbb{R}$ on the domain $(t,\mu) \in\,]0,\infty[\times\mathbb{R}$. Of course, as $t \to 0$, $p_t(a,\mu)$ has a singularity, but for $g,h : \mathbb{R} \to \mathbb{R}$ bounded and continuous it is true that

$$\lim_{t\to 0}\int_{-\infty}^{\infty} p_t(a,\mu)g(\mu)\,\mathrm{d}\mu = g(a) \qquad (a \in \mathbb{R}) \tag{5}$$

$$\lim_{t\to 0}\int_{-\infty}^{\infty} p_t(a,\mu)h(a)\,\mathrm{d}a = h(\mu) \qquad (\mu \in \mathbb{R}). \tag{6}$$

3. The Bernoulli–Laplace urn model

The main topic of this paper is Laplace's thorough discussion of the urn model introduced by Daniel Bernoulli (1770). This part of Laplace's work appears in Book II of his *Théorie Analytique des Probabilités* (Laplace 1812) and also in the earlier memoir (Laplace 1810–11).

Laplace considers the following special case of the Bernoulli model: two urns, A and B, each contain n balls, with n of the $2n$ balls white and n black. A ball is drawn at random from each urn and then the ball that came from A is placed in urn B, and the ball that came from B is placed in urn A. The problem, as originally posed by Bernoulli, is then to find the distribution of the number of white balls in urn A after r draws.

Denoting by $z_{x,r}$ the probability that there are precisely x white balls in A after r draws, Laplace derives the partial second-order difference equation

$$z_{x,r+1} = \left(\frac{x+1}{n}\right)^2 z_{x+1,r} + 2\frac{x}{n}\left(1 - \frac{x}{n}\right)z_{x,r} + \left(1 - \frac{x-1}{n}\right)^2 z_{x-1,r}. \tag{7}$$

It is characteristic (see the discussion in Hald 1990, Section 23.3) that Laplace, in order to solve a concrete problem, first finds a difference equation for the function z of interest. He would then normally proceed to determine the generating function for z, which in the case of (7) is too difficult, so instead he uses a different approach: he defines a new space variable μ and a new time variable r' by

$$x = \tfrac{1}{2}(n + \mu\sqrt{n}), \qquad r = nr', \tag{8}$$

and claims that $U = U(\mu, r') := z_{x,r}$ satisfies the second-order partial differential equation

$$\frac{\partial U}{\partial r'} = 2U + 2\mu\frac{\partial U}{\partial \mu} + \frac{\partial^2 U}{\partial \mu^2}, \tag{9}$$

which, as is important to note, is of the form of equation (4) with $\beta = 2$ and $\sigma^2 = 2$.

There seems to be no doubt that with (9), Laplace presented the first ever differential equation of the type now known from the probabilistic theory of diffusion processes. He does not give an argument that leads from (7) to (9), but only hints that one should use expansions of z, written by him as identities in the following fashion:

$$z_{x+1,r} = z_{x,r} + \frac{\partial z_{x,r}}{\partial x} + \frac{1}{2}\frac{\partial^2 z_{x,r}}{\partial x^2}$$

$$z_{x-1,r} = z_{x,r} - \frac{\partial z_{x,r}}{\partial x} + \frac{1}{2}\frac{\partial^2 z_{x,r}}{\partial x^2}$$

$$z_{x,r+1} = z_{x,r} + \frac{\partial z_{x,r}}{\partial r},$$

ignoring terms of order n^{-2}. (In fact, as will be shown below, terms of order $o(n^{-1})$ should be ignored. The expansions are in terms of powers of $n^{-1/2}$, but if they are carried far enough, one finds that the terms of order $n^{-3/2}$ cancel and, hence, for these expansions the first term to be ignored is of order n^{-2}).

Several authors have reacted strongly to Laplace's line of reasoning at this point. Todhunter (1865, §999) writes that (7) 'is too difficult for exact solution, and so Laplace mutilates it most unsparingly'; Bachelier (1906, p. 275), states that the argument yielding (9) from (7) is badly executed, and that the solution provided by (9) is inexact; Hostinský (1932, p. 50) expresses the opinion that it is difficult to follow the calculations done by Laplace; and Fréchet (1938, p. 13) says that Laplace treats only a special case of the Bernoulli model, and that his reasoning is lacking in rigour. Both Hostinský and Fréchet admit, however, that Laplace's argument is of interest since it leads to a second-order partial differential equation obtained much later by Smoluchowski (1915) in his study of (physical) diffusion. Both authors refer to Markov (1915) as the best source for a precise proof leading from (7) to (9).

Molina (1930, 1936) reproduces (9) with the comment that its solution approximates that of (7). He has an interesting reference to Lotka (1956, pp. 30–31), where the urn model is used to illustrate the concept of irreversibility in molecular kinetics.

It is the main thesis of the present paper, that not only is Laplace (and Markov) absolutely right, but also that the appearance of a Fokker–Planck equation for an Ornstein–Uhlenbeck process can be entirely justified by referring to the modern theory of weak convergence of stochastic processes. Since the step from (7) to (9) thus becomes crucial, we shall reproduce in some detail the argument given by Markov (1915).

Markov considers a more general urn model with n balls in urn A, n_1 balls in urn B and $(n + n_1)p$ white and $(n + n_1)q$ black balls in all, where $0 < p < 1$, $q = 1 - p$. This model

gives the partial difference equation (Markov 1915, p. 88)

$$z_{x,r+1} = \frac{x+1}{n} \frac{n_1 q - np + x + 1}{n_1} z_{x+1,r} + \frac{n-x+1}{n} \frac{(n+n_1)p - x + 1}{n_1} z_{x-1,r}$$
$$+ \left\{ \frac{x}{n} \frac{(n+n_1)p - x}{n_1} + \frac{n-x}{n} \frac{n_1 q - np + x}{n_1} \right\} z_{x,r}, \tag{10}$$

and, for the corresponding Markov chain, Markov proceeds to find the stationary initial distribution, which is hypergeometric. He changes variables from (x,r) to (μ, ρ) (ρ corresponding to Laplace's r'; see (8)) using the formulae

$$x = np + \mu \frac{1}{\Delta \mu}, \qquad r\left(\frac{1}{n} + \frac{1}{n_1}\right) = 2\rho$$

where

$$\Delta \mu = \sqrt{\frac{n+n_1}{2pqnn_1}}.$$

He then demands that $n_1 \alpha = n$, where $\alpha > 0$, is a constant (so $\Delta \mu = \sqrt{(1+\alpha)/2pqn}$), writes

$$U(\mu, \rho) = z_{x,r}$$

and expresses (10) in terms of U, μ and ρ instead of z, x and r, using the approximations

$$z_{x+1,r} = U(\mu + \Delta \mu, \rho) \approx U + \Delta \mu \frac{\partial U}{\partial \mu} + \tfrac{1}{2}(\Delta \mu)^2 \frac{\partial^2 U}{\partial \mu^2}$$

$$z_{x-1,r} = U(\mu - \Delta \mu, \rho) \approx U - \Delta \mu \frac{\partial U}{\partial \mu} + \tfrac{1}{2}(\Delta \mu)^2 \frac{\partial^2 U}{\partial \mu^2}$$

$$z_{x,r+1} = U\left(\mu, \rho + \frac{1+\alpha}{2n}\right) \approx U + \frac{1+\alpha}{2n} \frac{\partial U}{\partial \rho},$$

which are of course precise to the order $o(n^{-1})$ as $n \to \infty$, when U is smooth enough and $\mu \in \mathbb{R}$ is fixed. In the resulting version of (10) thus obtained by Markov, all terms of order n^0, $n^{-\frac{1}{2}}$ cancel and, equating the coefficients of order n^{-1}, one arrives at (9) exactly, irrespective of the values of p and α (which shows that the change-of-variable formulae used by Markov have been chosen with great care in order for him to arrive at precisely the Fokker–Planck equation given by Laplace).

If, in the model studied by Markov, we take $n_1 = n$, $p = q = \frac{1}{2}$, $\alpha = 1$, we return to the special case investigated by Laplace. It is worth noting that the critical Todhunter (1865) almost arrives at (9), but then ignores the crucial assumption that after the change of variable through (8), μ should be considered a finite quantity. As will be argued below, this assumption has a well-understood analogue if one wishes to prove the central limit theorem, although formally, of course, the range for μ in the nth urn model is the interval $[-\sqrt{n}, \sqrt{n}]$, which becomes unbounded in the limit. (Todhunter seems very concerned about the extreme values $\pm\sqrt{n}$.)

The Laplace urn model is, of course, a Markov chain, time-homogeneous with state space $E_n = \{0, 1, \ldots, n\}$ and transition probability matrix (p_{xy}), where $x, y \in E_n$, given by

$$p_{x,x+1} = \left(1 - \frac{x}{n}\right)^2, \qquad p_{xx} = 2\frac{x}{n}\left(1 - \frac{x}{n}\right), \qquad p_{x,x-1} = \left(\frac{x}{n}\right)^2 \tag{11}$$

with $p_{xy} = 0$ for all other pairs (x, y). The Markov chain $(Z_r^{(n)})_{r \geq 0}$ is completely specified by also giving the distribution of $Z_0^{(n)}$ (the initial distribution), and it is then clear that

$$z_{x,r} := P(Z_r^{(n)} = x)$$

satisfies (7). If we now consider the transformation (8), we obtain a process $u^{(n)} = (u^{(n)}(t))_{t \geq 0}$ in continuous time given by

$$Z_{[nt]}^{(n)} = \tfrac{1}{2}\left(n + u^{(n)}(t)\sqrt{n}\right)$$

where we use the standard notation t for the time variable rather than Laplace's r or Markov's ρ. Then the following result holds:

Theorem. *Let $\mu_0 \in \mathbb{R}$ and, for all sufficiently large n, let the initial value of $Z^{(n)}$ be given by*

$$Z_0^{(n)} = [\tfrac{1}{2}(n + \mu_0\sqrt{n})]. \tag{12}$$

Then, as $n \to \infty$,

$$u^{(n)} \xrightarrow{\mathscr{D}} u \tag{13}$$

where u is the Ornstein–Uhlenbeck process with parameters

$$\beta = 2, \qquad \sigma^2 = 2$$

and initial value $u(0) \equiv \mu_0$. In particular, the transition densities $(p_t(a, \mu))$ for u satisfy Laplace's differential equation (9) with $U(\mu, t) = p_t(a, \mu)$ in the domain $(\mu, t) \in \mathbb{R} \times]0, \infty[$ for any $a \in \mathbb{R}$.

Remarks. In (12) the brackets $[\cdot]$ denote the integer part. The qualification 'n sufficiently large' simply means that n should be so large that the value of $Z_0^{(n)}$ belongs to E_n. The convergence in (13) means convergence in distribution, when $u^{(n)}$ and u are viewed as random variables with values in the Skorohod space $D[0, \infty[$ of right continuous paths with left limits. (See Billingsley 1968 for $D[0, 1]$ and, for example, Ethier and Kurtz 1986 for $D[0, \infty[$). In particular, it follows from (13) that the finite-dimensional distributions converge: for all $N \in \mathbb{N}$ and $0 \leq t_1 < \cdots < t_N$ we have that

$$(u^{(n)}(t_1), \ldots, u^{(n)}(t_N)) \xrightarrow{\mathscr{D}} (u(t_1), \ldots, u(t_N)). \tag{14}$$

The theorem has been stated only for the Laplace urn model, but it holds also for the more general models studied by Markov (1915) if one defines $u^{(n)}$ by $Z_{[nt]}^{(n)} = np + u^{(n)}(t)\sqrt{2pqn/(1 + \alpha)}$ and uses Markov chains $Z^{(n)}$ with transition probabilities adjusted in an obvious manner from (11) to the general case treated by Markov.

We shall not prove the theorem here, but merely indicate why it is true. (A proof may be given using the techniques and results from Ethier and Kurtz 1986.)

The infinitesimal generator for the limit process u is given by (see (3))

$$Af(\mu) = -2\mu f'(\mu) + f''(\mu)$$

(retaining Laplace's notation μ for the space variable); in particular, for sufficiently well-behaved f,

$$\lim_{h \downarrow 0} \frac{1}{h} \left(E[f(u(h)) \mid u(0) = \mu] - f(\mu) \right) = Af(\mu).$$

But here we can approximate the limit on the left by replacing u with $u^{(n)}$, taking $h = 1/n$ and letting $n \to \infty$, keeping μ fixed and thinking of the Markov chain $Z^{(n)}$ as starting from

$$x_n := \tfrac{1}{2}(n + \mu\sqrt{n})$$

and, even though this is not an integer, use the transition probabilities (p_{xy}) as given formally by (11). The result is

$$nE\left[f\left(\frac{1}{\sqrt{n}} (2Z_1^{(n)} - n) \right) \Bigg| Z_0^{(n)} = x_n \right] - f(\mu)$$

$$= n\left[\left(1 - \frac{x_n}{n}\right)^2 f\left(\mu + \frac{2}{\sqrt{n}}\right) + 2\frac{x_n}{n}\left(1 - \frac{x_n}{n}\right)f(\mu) + \left(\frac{x_n}{n}\right)^2 f\left(\mu - \frac{2}{\sqrt{n}}\right) - f(\mu) \right]$$

$$\xrightarrow[n \to \infty]{} Af(\mu)$$

for all f twice continuously differentiable in a neighbourhood of μ.

Thus the change of variable in space and time used by Laplace (see (8)) is the correct one that gives weak convergence of his urn scheme to the Ornstein–Uhlenbeck process. This does not mean that Laplace discovered this process: he was only interested in finding the probabilities $z_{x,r}$ or good approximations thereof, and because of the difficult nature of the combinatorics involved this led him to the difference equation (7) in space and time, and by a flash of genius to the partial differential equation (9), which we now know describes the limiting Ornstein–Uhlenbeck process. The connection between the findings made by Laplace and the weak convergence theorem given above is that Laplace came close to obtaining weak convergence of the one-dimensional distributions ($N = 1$ in (14)).

We have not been able to find the weak convergence result in the literature. For example, in his two-volume classic Feller mentions the Bernoulli–Laplace urn model briefly in vol. 1 (Feller 1957, Chapter XV, Exercise 10), and the Ornstein–Uhlenbeck process in vol. 2 (Feller 1971, pp. 335–336), but does not make the connection. However, the one-dimensional convergence suggested by Laplace's work is established more firmly by Markov (1915) who, in his generalization of Laplace's model uses recursion formulae to find explicit expressions for the moments $E(Z_r^{(n)})^k$ (in terms of the initial values $Z_0^{(n)}$ of the chains) and then shows that the moments $E[(u^{(n)}(t))^k]$ converge as $n \to \infty$. More specifically, if in the limit $u^{(n)}(0)$ is $N(a, \tfrac{1}{2}(1 + l))$ (Gaussian with mean a and variance $\tfrac{1}{2}(1 + l)$) where $l > -1$, Markov shows that for all $k \in \mathbb{N}$, $E(u^{(n)}(t))^k$ converges to the kth

moment of the distribution $N(ae^{-2t}, \frac{1}{2}(1 + le^{-4t}))$, which is in fact the correct distribution of $u(t)$ when $u(0) \sim N(a, \frac{1}{2}(1 + l))$, as we shall see in the next section.

It must be noted that Markov (1915, p. 103), observes that if $t > 0$, the Gaussian distribution of $u(t)$ is also well defined if $l = -1$ (which would correspond to $u(0)$ being degenerate at a), and he writes down the density

$$p_t(a, \mu) = \frac{1}{\sqrt{\pi(1 - e^{-4t})}} \exp\left(-\frac{(\mu - ae^{-2t})^2}{1 - e^{-4t}}\right). \tag{15}$$

This is in fact the correct transition density for the Ornstein–Uhlenbeck process u and one of the first instances where it appears in the literature.

With the focus on Ornstein–Uhlenbeck processes it is natural to ask why (9), the first second-order partial differential equation associated with a diffusion process ever to appear, comes so much earlier (about a century) than the corresponding equation for Brownian motion. In fact, the reasoning that led Laplace to (9) might just as well have yielded the Brownian motion equation, as we shall see shortly. One possible explanation why Laplace did not find this equation could be that, since his main concern was to find the probabilities $z_{x,r}$, which in the random walk/Brownian motion case are simple binomial probabilities, in the latter case there was no need to proceed via partial difference equations. It seems likely, however, that Laplace's knowledge about the central limit theorem and the normalizations needed to obtain convergence there inspired him to the crucial normalization (8) needed for the urn models to converge!

To see how Laplace might have deduced the Fokker–Planck equation for Brownian motion, consider a Bernoulli random walk as follows (see Feller 1957, Chapter XIV, Section 6): Let $Y_1, Y_2, \ldots$ be i.i.d. with $P(Y_1 = 1) = P(Y_1 = -1) = \frac{1}{2}$. Define $S_0 \equiv 0$, $S_r = Y_1 + \ldots + Y_r$ and write $z_{x,r} = P(S_r = x)$. Then

$$z_{x,r+1} = \tfrac{1}{2} z_{x+1,r} + \tfrac{1}{2} z_{x-1,r}. \tag{16}$$

Introducing the new variables μ, r' defined by $x = \mu\sqrt{n}$, $r = nr'$, defining $U(\mu, r') = z_{x,r}$ and expressing (16) in terms of μ, r' and U, one finds, arguing as Markov, that in the limit as $n \to \infty$,

$$\frac{\partial U}{\partial r'} = \frac{1}{2} \frac{\partial^2 U}{\partial \mu^2}, \tag{17}$$

the Fokker–Planck equation for standard Brownian motion.

Before leaving this section we shall comment briefly on the paper by Bachelier (1906, p. 273), since this may well mark the first appearance of the expression (2) for the Ornstein–Uhlenbeck transition density. (It must also be noted that Feller 1957, in the footnote on p. 323, credits Bachelier with the discovery of the connection between random walks and diffusions – equations (16) and (17). Notice that Feller refers to Bachelier's work as important but 'heuristic'.) Bachelier studies the problem of finding the probability that a player, after a certain number of games, has lost a certain amount. In the case of Bernoulli trials he gives the De Moivre–Laplace limit approximation to the binomial probabilities and then goes on to discuss the case where the games are no longer independent

(*probabilités connexes* is the term used by Bachelier), and it is here that he presents the Ornstein–Uhlenbeck density, assuming that the loss at any game is proportional to the total loss already incurred. He then goes through the Bernoulli–Laplace urn model with the critical comments on Laplace's work quoted above, but ends up by essentially showing that the densities (2) solve (4) (and hence, in special cases (9))!

4. The solution of Laplace's differential equation

Having derived equation (9), Laplace devotes his main efforts to solving it, and thereby solving the original problem of finding the probabilities $z_{x,r}$, at least for large values of n, the number of white (or black) balls. He writes the solution in the form

$$U(\mu, r') = \int_{-\infty}^{\infty} \varphi(t, r') \, \mathrm{e}^{-\mu t} \, \mathrm{d}t \tag{18}$$

and finds that φ is the solution of a first-order partial differential equation, and that

$$\varphi(t, r') = \mathrm{e}^{\frac{1}{4}t^2} \psi(t \, \mathrm{e}^{-2r'})$$

with ψ an arbitrary function. He then makes the change of variable $t = 2\mu + 2\mathrm{i}s$ in the integral (18), and arrives at

$$U(\mu, r') = \mathrm{e}^{-\mu^2} \int_{-\infty}^{\infty} \mathrm{d}s \, \mathrm{e}^{-s^2} G\!\left(\frac{s - \mathrm{i}\mu}{\mathrm{e}^{2r'}}\right) \tag{19}$$

where $G(z) = \psi(-\frac{1}{2}\mathrm{i}z)$ (Laplace denotes G by Γ), having thus replaced the unpleasant factor $\mathrm{e}^{t^2/4}$ in the expression for φ by the much more appealing factor e^{-s^2} in the integrand in (19). Now, substituting t in the expression $\exp(-\mu t + \frac{1}{4}t^2)$ of course gives $\exp(-\mu^2 - s^2)$, but the domain of integration in (19) should not be $\mathbb{R}$ but the line $\mathrm{i}\mathbb{R}$ in the complex plane, where $s \to \mathrm{e}^{-s^2}$ is unbounded. Thus, Laplace's derivation of (19) is imprecise, but as he noted himself, the important point is of course that it is still true that if G is smooth enough and suitably bounded, U given by (19) solves (9), as may be verified directly!

Using a power-series expansion of G, Laplace writes U in the form

$$U(\mu, r') = \sum_{j=0}^{\infty} \frac{H^{(j)} \, \mathrm{e}^{-\mu^2}}{\mathrm{e}^{4jr'}} \int_{-\infty}^{\infty} \mathrm{d}s \, \mathrm{e}^{-s^2} (s - \mathrm{i}\mu)^{2j}$$

$$+ \sum_{j=0}^{\infty} \frac{L^{(j)} \, \mathrm{e}^{-\mu^2}}{\mathrm{e}^{(4j+2)r'}} \int_{-\infty}^{\infty} \mathrm{d}s \, \mathrm{e}^{-s^2} (s - \mathrm{i}\mu)^{2j+1} \tag{20}$$

and observes that the integrals

$$P_{2j}(\mu) := \int_{-\infty}^{\infty} \mathrm{d}s \, \mathrm{e}^{-s^2} (s - \mathrm{i}\mu)^{2j}, \qquad P_{2j+1}(\mu) := \mathrm{i} \int_{-\infty}^{\infty} \mathrm{d}s \, \mathrm{e}^{-s^2} (s - \mathrm{i}\mu)^{2j+1}$$

and both real-valued polynomials such that only terms of even powers of μ appear in P_{2j} and only terms of odd powers of μ appear in P_{2j+1}. He then proceeds to show that

for $n \neq n'$,

$$\int_{-\infty}^{\infty} \mathrm{d}\mu\, \mathrm{e}^{-\mu^2} P_n(\mu) P_{n'}(\mu) = 0,$$

i.e. the polynomials P_n are orthogonal with respect to the Gaussian probability measure

$$\nu(\mathrm{d}\mu) = \frac{1}{\sqrt{\pi}}\, \mathrm{e}^{-\mu^2}\, \mathrm{d}\mu \tag{21}$$

on the real line.

Laplace goes on to find examples of solutions to (9), to which we shall return below, but before that we need briefly to discuss the polynomials P_n.

It is an acknowledged fact that in describing the polynomials P_n, Laplace essentially discovered what are now known as the Hermite polynomials. More precisely, it was noted by Markov (1915, p. 91) that the P_n used by Laplace are proportional to the Hermite polynomials H_n. (For an account of this see also Molina 1930; 1936; Hald 1990). We shall now prove this proportionality, using the original definition (Chebyshev 1859; Hermite 1864)

$$\mathrm{e}^{-\mu^2} H_n(\mu) = \frac{\mathrm{d}^n}{\mathrm{d}\mu^n}\, \mathrm{e}^{-\mu^2}$$

of what both Markov (1915) and Steklov (1915) refer to as the Laplace–Chebyshev–Hermite polynomials. In this form the H_n are given by

$$H_n(\mu) = \sum_{k=0}^{[n/2]} \frac{(-1)^{n-k} n!}{k!\,(n-2k)!} (2\mu)^{n-2k}$$

(see, for example, Hille 1925–6, (11) p. 432; or Szegö 1939, (5.5.4), p. 102). For $n = 2j$ one finds that the integral after the factor $H^{(j)} \exp(-\mu^2 - 4jr')$ in (20) equals

$$P_{2j}(\mu) = \sum_{k=0}^{j} (-1)^{j-k} \Gamma\left(k + \frac{1}{2}\right) \binom{2j}{2k} \mu^{2j-2k}$$

$$= \sum_{k=0}^{j} (-1)^{j-k} \frac{\sqrt{\pi}}{2^{2j}} \frac{(2j)!}{k!\,(2j-2k)!} (2\mu)^{2j-2k}$$

$$= (-1)^j \frac{\sqrt{\pi}}{2^{2j}} H_{2j}(\mu).$$

Similarly, one finds

$$P_{2j+1}(\mu) = (-1)^{j+1} \frac{\sqrt{\pi}}{2^{2j+1}} H_{2j+1}(\mu).$$

Thus, in terms of the polynomials H_n, Laplace's solution to (9) takes the form of the expansion

$$U(\mu, r') = \sum_{n=0}^{\infty} C_n H_n(\mu)\, \mathrm{e}^{-\mu^2 - 2nr'}.$$

A thorough discussion of when this series actually solves (9) is given by Steklov (1915). He writes the expansion as

$$U(\mu, r') = e^{-\mu^2} \sum_{n=0}^{\infty} A_n e^{-2nr'} \varphi_n(\mu), \tag{22}$$

where

$$\varphi_n(\mu) = (2^n n!)^{-\frac{1}{2}} H_n(\mu)$$

are the H_n normalized so as to form a complete orthonormal basis for the space $L^2(\nu)$ of functions, square-integrable with respect to the probability ν. Steklov specifically considers the problem of solving (9) subject to the boundary condition

$$U(\mu, 0) = f(\mu), \tag{23}$$

with f a given function. In this case he shows that the coefficients A_n in (22) are given by

$$A_n = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} f(\mu) \varphi_n(\mu)\, d\mu, \tag{24}$$

i.e. the coefficients in the Gram–Charlier expansion of $\tilde{f}(\mu) = e^{\mu^2} f(\mu)$ in the orthonormal basis (φ_n),

$$\tilde{f} \sim \sum_{n=0}^{\infty} \left(\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} f(\mu) \varphi_n(\mu)\, d\mu \right) \varphi_n.$$

Here '$\sim$' signifies that if $\tilde{f} \in L^2(\nu)$ then the series converges to $\tilde{f}$ in $L^2(\nu)$.

Now, using the transition densities (2) with $\beta = 2$, $\sigma^2 = 2$, the same solution may, for suitable f, be expressed as

$$U(\mu, t) = \int_{-\infty}^{\infty} d\mu' f(\mu') p_t(\mu', \mu), \tag{25}$$

where (cf. (15))

$$p_t(\mu', \mu) = \frac{1}{\sqrt{\pi(1 - e^{-4t})}} \exp\left(-\frac{(\mu - \mu' e^{-2t})^2}{1 - e^{-4t}} \right). \tag{26}$$

(To show that (25) solves (9) one merely has to use the fact that $(\mu, t) \to p_t(\mu', \mu)$ is a solution, and to impose conditions that permit changing the order of integration and the differentiations $\partial/\partial t$, $\partial/\partial \mu$, $\partial^2/\partial \mu^2$. Formally, $U(\mu, t)$ given by (25) is defined only for $\mu \in \mathbb{R}$, $t > 0$, but if f is bounded and continuous, U will satisfy the boundary condition (23) in the sense that for all μ, $\lim_{t \to 0} U(\mu, t) = f(\mu)$; see (6).)

We shall see below that the fact that the two expressions (22) and (25) agree amounts to what is essentially a classical expansion of Ornstein–Uhlenbeck transition densities. Before doing that we shall, however, first comment on the examples of solutions to (9) found by Laplace and Steklov.

Laplace considers the special case of solving (9) with boundary condition $U(\mu, 0) = f_L(\mu)$, where

$$f_L(\mu) = \frac{2j}{\sqrt{n\pi}}\, e^{-j^2\mu^2} \tag{27}$$

with $j > 0$. He gives the solution as

$$U(\mu, r') = \frac{2}{\sqrt{n\pi(1 + \mathfrak{C}')}}\, \exp\left(-\frac{\mu^2}{1 + \mathfrak{C}'}\right)$$

where

$$\mathfrak{C}' = \frac{1 - j^2}{j^2\, e^{4r'}}.$$

But if, more generally, following Steklov (1915, (41) and (42), p. 1536), we take $f = f_S$ in (23) as the density of a Gaussian distribution

$$f_S(\mu) = \frac{1}{\sqrt{\pi(1 + l)}}\, e^{-(\mu - a)^2/(1 + l)}$$

where $l > -1$ (as was done by Markov 1915 in his work on convergence of moments; see Section 3 above), it is clear from (25) that $U(\mu, t)$ will be the density of a Gaussian random variable $u(t)$, with $(u(t))_{t \geq 0}$ an Ornstein–Uhlenbeck process where the initial distribution for $u(0)$ has density f_S and, conditionally on $u(0) = \mu'$, $u(t)$ has density $p_t(\mu', \mu)$ given by (26). Thus

$$E[u(t)] = E[e^{-2t}u(0)] = a\, e^{-2t},$$

$$\text{var}[u(t)] = \tfrac{1}{2}(1 - e^{-4t}) + e^{-4t}\frac{1 + l}{2} = \tfrac{1}{2}(1 + l\, e^{-4t})$$

and hence

$$U(\mu, t) = \frac{1}{\sqrt{\pi(1 + l\, e^{-4t})}}\, \exp\left(-\frac{(\mu - a\, e^{-2t})^2}{1 + l\, e^{-4t}}\right)$$

in agreement with the result found by Steklov (1915, (42), p. 1536), who, as already mentioned, based his deduction on the expansion (22), and also in accordance with the moment convergence established by Markov (1915).

Now Laplace's initial condition (27) is, in terms of Steklov's notation,

$$f_L = \frac{2}{\sqrt{n}}\, f_S,$$

with $a = 0$, $j^2 = 1/(1 + l)$. But then

$$\mathfrak{C}' = l\, e^{-4r'}$$

and we see that the solution found by Laplace is the correct one, even though again his reasoning is difficult to follow and perhaps not quite rigorous!

As our final point, we shall show that the fact that the solutions given by (22) (with A_n as in (24)) and (25) agree, amounts to a classical expansion of the transition density (26), namely

$$p_t(\mu',\mu) = \frac{1}{\sqrt{\pi}}\sum_{n=0}^{\infty} e^{-2nt-\mu^2}\varphi_n(\mu')\varphi_n(\mu). \qquad (28)$$

That the series converges to $p_t(\mu',\mu)$ for all $t > 0$, $\mu,\mu' \in \mathbb{R}$ is well known; see Hille (1925–6, (39), p. 439), who traces the formula he gives back to Mehler (1866).

Inserting (24) in (22) and changing the order of summation and integration gives

$$U(\mu,t) = \int_{-\infty}^{\infty} d\mu' f(\mu') \frac{1}{\sqrt{\pi}}\sum_{n=0}^{\infty} e^{-2nt-\mu^2}\varphi_n(\mu')\varphi_n(\mu).$$

Demanding that this agree with (22) for a large class of functions f is clearly the same as requiring that, for a given t, μ, (28) must hold for almost all μ' with respect to Lebesgue measure. Since both sides of (28) are continuous in μ', (28) is valid for all $t > 0$, $\mu,\mu' \in \mathbb{R}$.

Acknowledgements

References

Bachelier, L. (1906) Théories des probabilités continues. *J. Math. Pures Appl.* (6), **12**, 259–327.

Bachelier, L. (1912) *Calcul des Probabilités*, Vol. 1. Paris: Gauthier-Villars.

Bernoulli, Daniel (1770) Disquitiones analyticae de novo problemate conjecturali. *Novi Commentarii Academiae Scientiarum Imperialis Petropolitanae*, **XIV**, 3–25.

Billingsley, P. (1968) *Convergence of Probability Measures*. New York: Wiley.

Chebyshev, P.L. (1859) Sur le développement des fonctions à une seule variable. *Bull. Physico-Mathématique de St.-Pétersbourg*, I, 193–200. Reprinted in A. Markoff and N. Somin (eds), *Œuvres de P.L. Tchebysheff*, Vol. 1, pp. 501–508. St Petersburg (1899).

Doob, J.L. (1942) Brownian motion and stochastic equations. *Ann. Math.* (2), **43**, 351–369.

Dynkin, E.B. (1965) *Markov Processes*, Vol. 1. Berlin: Springer-Verlag.

Einstein, A. (1956) *Investigations on the Theory of Brownian Motion*. Edited with notes by R. Fürth. Dover Publications.

Ethier, S.N. and Kurtz, T.G. (1986) *Markov Processes. Characterization and Convergence*. New York: Wiley.

Feller, W. (1957) *An Introduction to Probability Theory and its Applications* (2nd edn), Vol. 1. New York: Wiley.

Feller, W. (1971) *An Introduction to Probability Theory and its Applications* (2nd edn), Vol. 2. New York: Wiley.

Fréchet, M. (1938) *Recherches Théoriques Modernes sur le Calcul des Probabilités. Second Livre. Méthode des Fonctions Arbitraries. Théorie des Événements en Chaîne dans le Cas d'un Nombre Fini d'États Possibles*. Paris: Gauthier-Villars.

Hald, A. (1990) *A History of Probability and Statistics and Their Applications before 1750*. New York: Wiley.

Hermite, C. (1864) Sur un nouveau développement en série des fonctions. *C.R. Acad. Sci. Paris*, **LVIII**(I), 93, 266. Reprinted in E. Picard (ed.), *Œuvres de Charles Hermite*, Vol. 2, pp. 293–308. Paris: Gauthier-Villars (1908).

Hille, E. (1925–6) A class of reciprocal functions. *Ann. Math. (2)*, **27**, 427–464.

Hostinský, B. (1932) Application du calcul des probabilités à la théorie du mouvement Brownien. *Ann. Inst. H. Poincaré*, **3**, 1–74.

Laplace, P.S. (1810–11) Mémoire sur les intégrales définies et leur application aux probabilités, et spécialement à la recherche du milieu qu'il faut choisir entre les résultats des observations. *Mem. Acad. Sci.* (I), XI, Section V. Reprinted in *Œuvres de Laplace*, Vol. 12, pp. 375–387. Paris: Gauthier-Villars (1848).

Laplace, P.S. (1812) *Théorie Analytique des Probabilités. Livre II. Théorie Générale des Probabilités*, Chapter 3. Reprinted in *Œuvres de Laplace* (2nd edn), Vol. 7, pp. 314–322. Paris: Gauthier-Villars (1847); and in *Œuvres de Laplace* (3rd edn), Vol. 7, pp. 292–308. Paris: Gauthier-Villars (1886).

Lotka, A.J. (1956) *Elements of Mathematical Biology*. New York: Dover. First published in 1925 as *Elements in Physical Biology*.

Markov, A.A. (1915) On a problem of Laplace. *Izv. Akad. Nauk.* (6), **2**, 87–104 (in Russian).

Mehler, F.G. (1866) Reihenentwicklungen nach Laplaceschen Funktionen Höherer Ordnung. *Journal für Mathematik*, **66**, 161–176.

Molina, E.C. (1930) The theory of probability: some comments on Laplace's théorie analytique. *Bull. Amer. Math. Soc.*, **36**, 369–392.

Molina, E.C. (1936) A Laplacian expansion for Hermitian–Laplace functions of high order. *Bell System Technical J.*, **XV**, 355–362.

Ondar, H.O. (1970) A certain work of V.A. Steklov on probability theory. In *History and Methodology of Natural Sciences, No. IX, Mechanics, Mathematics*, pp. 262–266. Moscow: Izdat. Moskov. Univ. (in Russian).

Ornstein, L.S. and Uhlenbeck, G.E. (1930) On the theory of Brownian motion. *Phys. Rev.*, **36**, 823–841.

Sheynin, O.O. (1976–7) P.S. Laplace's work on probability theory. *Arch. Hist. Exact Sci.*, **16**, 137–187.

Smoluchowski, M. v. (1915) Über Brownsche Molekularbewegung unter Einwirkung äußerer Kräfte und deren Zusammenhang mit der verallgemeinerten Diffusionsgleichung. *Ann. Physik* (4), **XLVIII**, 1103.

Steklov, V. (1915) On a problem of Laplace. *Izv. Akad. Nauk.* (6), **14**, 1515–1537 (in Russian).

Szegö, G. (1939) *Orthogonal Polynomials.* Amer. Math. Soc. Colloq. Publ. XXIII. New York: American Mathematical Society.

Todhunter, I. (1865) *A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace.* London: Macmillan. Reprinted New York: Chelsea (1949).