How Einstein did Brownian motion in 1905

Peter Haggstrom
mathsatbondibeach@gmail.com
https://gotohaggstrom.com

September 1, 2020

1 Background

Albert Einstein received the Nobel Prize in 1921 not for his work on Special Relativity but for his his work on the photoelectric effect. Neverthless, his 1905 work on Browian motion (contemporaneous with his work on Special Relativity and the photoelectric effect) was immensely influential. Einstein's original paper is reproduced in [1], pages 12-17. This Dover edition contains a couple of typos which I have corrected. It is a short derivation of the diffusion equation for such a surprisingly deep result. Jean Perrin used Einstein's work in his Nobel Prize winning experiments which led to an estimate of Avogadro's number. The details of how he did so are set out in [2].

It must be borne in mind that in 1905 the atomistic theory of matter was not universally accepted. Ernst Mach was one well known physicist who wasn't a believer. Wilhelm Ostwald, the Nobel Prize winning chemist, was a fervent opponent of the atomic hypothesis but apparently was ultimately won over by Perrin's experiments.

Einstein's 1905 derivation of the equation for Brownian motion is a model of deceptive simplicity, indeed, C W Gardiner says that "Einstein's reasoning is very clear and elegant". ([3], page 3). While Einstein's approach seems remarkably simple, I think it requires quite a lot of thought to be convincing and in fact Langevin came up with an approach a few years after Einstein which he enthusiastically described as "infinitely more simple" ([3], pages 6-7). Note here I am only dealing with the diffusion aspect of Einstein's analysis of Brownian motion.

Einstein's anaysis proceeds as follows:

" We will turn now to a closer consideration of the irregular movemnets which arise from thermal molecular movements, and give rise to the diffusion investigated in the last paragraph.

Evidently it must be assumed that each single particle executes a movement which is independent of the movement of all other particles; the movements of one and the same particle after different intervals of time must be considered as mutually independent processes, so long as we think of these intervals of time being chosen not too small.

We will introduce a time interval τ in our discussion, which is to be very small compared with the observed interval of time, but, nevertheless, of such a magnitude that the movements executed by a particle in two consecutive intervals of time τ are to be considered as mutually independent phenomena.

Suppose there are altogether n suspended particles in a liquid. In an interval of time τ the x-coordinates of the single particles will increase by Δ, where Δ has a different value (positive or negative) for each particle. For the value of Δ a certain probability law will hold; the number dn of the particles which experience in the time interval τ a displacement which lies between Δ and $\Delta + d\Delta$, will be expressed by an equation of the form:

$$dn = n\,\phi(\Delta)\,d\Delta \tag{1}$$

where

$$\int_{-\infty}^{\infty} \phi(\Delta)\,d\Delta = 1 \tag{2}$$

and ϕ only differs from zero for very small values of Δ and fulfils the condition:

$$\phi(\Delta) = \phi(-\Delta) \tag{3}$$

We will investigate now how the coefficient of diffusion depends on ϕ, confining ourselves again to the case where the number ν of the particles per unit volume is dependent only on x and t.

Putting for the number of particles per unit volume $\nu = f(x,t)$, we will calculate the distribution of the particles at a time $t+\tau$ from the distribution at time t. From the definition of the function $\phi(\Delta)$, there is easily obtained the number of the particles which are located at the time $t+\tau$ between two planes perpendicular to the x-axis with abscissae x and $x + dx$. We get:

$$f(x, t+\tau)\,dx = dx \int_{\Delta=-\infty}^{\Delta=\infty} f(x+\Delta, t)\,\phi(\Delta)\,d\Delta \tag{4}$$

Now since τ is very small, we can put:

$$f(x, t+\tau) = f(x,t) + \tau \frac{\partial f}{\partial t} \tag{5}$$

Further, we can expand $f(x+\Delta, t)$ in terms of Δ:

$$f(x+\Delta, t) = f(x,t) + \Delta \frac{\partial f(x,t)}{\partial x} + \frac{\Delta^2}{2!} \frac{\partial^2 f(x,t)}{\partial x^2} + \dots \tag{6}$$

We can bring this expansion under the integral sign, since only very small values of Δ contribute anything to the latter:

$$f + \frac{\partial f}{\partial t} \tau = f \int_{-\infty}^{\infty} \phi(\Delta)\, d\Delta + \frac{\partial f}{\partial x} \int_{-\infty}^{\infty} \phi(\Delta)\, d\Delta + \frac{\partial^2 f}{\partial x^2} \int_{-\infty}^{\infty} \frac{\Delta^2}{2!} \phi(\Delta)\, d\Delta + \dots \tag{7}$$

(The Dover edition incorrectly has $\frac{\partial x}{\partial f}$ in the second term on the RHS of (7))

On the right hand side of (7) the second, fourth etc terms vanish since $\phi(x) = \phi(-x)$; whilst of the first, third, fifth etc terms, every succeeding term is very small compared with the preceding. Bearing in mind that $\int_{-\infty}^{\infty} \phi(\Delta)\, \Delta = 1$ and putting

$$\frac{1}{\tau} \int_{-\infty}^{\infty} \frac{\Delta^2}{2} \phi(\Delta)\, d\Delta = D \tag{8}$$

and taking into consideration only the first and third terms on the right hand side, we get from this equation

$$\frac{\partial f}{\partial t} = D \frac{\partial^2 f}{\partial x^2} \tag{9}$$

This is the well known differential equation for diffusion, and we recognise that D is the coefficient of diffusion.

Another important consideration can be related to this method of development. We have assumed that the single particles are all referred to the same coordinate system. But this is unnecessary, since the movements of the single particles are mutually independent. We will now refer the motion of each particle to a coordinate systenm whose origin coincides at the time $t = 0$ with the position of the centre of gravity of the particles in question; with this difference, that $f(x,t)\, dx$ now gives the number of particles whose x coordinate has increased between the time $t = 0$ and the time $t = t$, by a quantity which lies between x and $x + dx$. In this case also the function f must satisfy, in its changes, the equation (I). Further, we must evidently have for $x \neq 0$ and $t = 0$:

$$f(x,t) = 0 \tag{10}$$

and

$$\int_{-\infty}^{\infty} f(x,t)\, dx = n \tag{11}$$

NOTE: The reference in the translation to equation (I) is to an immediately earlier paper titled "On the movement of small particles suspended in a stationary liquid demanded by the molecular kinetic theory of heat" - see [1], pages 1-12. The equation at issue is at page 10 and is:

$$-K\nu + \frac{RT}{N}\frac{\partial \nu}{\partial x} = 0 \tag{12}$$

R is the ideal gas constant, T is temperature and N is the number of molecules in a gram-molecule.

The problem, which accords with the problem of the diffusion outwards from a point (ignoring possibilities of exchange between the diffusing particles) is now mathematically completely defined; the solution is:

$$f(x,t) = \frac{n}{\sqrt{4\pi D}} \frac{e^{\frac{-x^2}{4Dt}}}{\sqrt{t}} \tag{13}$$

The probable distribution of the resulting displacements in a given time t is therefore the same as that of fortuitous error, which was to be expected. But it is significant how the constants in the exponential term are related to the coefficients of diffusion. We will now calculate with the help of this equation the displacement λ_x in the direction of the X axis which a particle experiences on an average, or - more accurately expressed - the square root of the arithmetic mean of the squares of displacements in the direction of the X axis; it is

$$\lambda_x = \sqrt{\bar{x}^2} = \sqrt{2Dt} \tag{14}$$

The mean displacement is therefore proportional to the square root of time. It can be easily shown that the square root of the mean of the squares of the total displacements of the particles has the value $\lambda_x\sqrt{3}$. "

2 Analysis of Einstein's approach

Given the mutual independence assumptions and symmetry assumptions ($\phi(\Delta) = \phi(-\Delta)$) it is not surprising at one level that a function with a Gaussian type of structure emerges. It is not a Gaussian but it has similarities. If you fix the time interval in (13) you get a Gaussian. Maxwell had already shown how his velocity distribution law - a Gaussian function - arose from assumptions of independence and radial symmetry (see [4]). Note that in (11) t is unbound and if in (13) we integrate over space we are simply integrating a Gaussian by treating t as fixed, and with the substitution $u = \frac{x}{\sqrt{4Dt}}$, we get:

$$\begin{aligned} I = & \int_{-\infty}^{\infty} \frac{n}{\sqrt{4\pi D}} \frac{e^{\frac{-x^2}{4Dt}}}{\sqrt{t}}\, dx \\ = & \frac{n}{\sqrt{4\pi D}} \int_{-\infty}^{\infty} \frac{e^{-u^2}}{\sqrt{t}} \sqrt{4Dt}\, du \\ = & \frac{n}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-u^2}\, du \\ = & \frac{n}{\sqrt{\pi}} \times \sqrt{\pi} \\ = & n \end{aligned} \tag{15}$$

This is a comforting result. The number of particles is spatially conserved over all time.

The first point to note is that the relationship between (1) and (2) is essentially tautologous. Einstein is saying that there is some probability distribution which accounts for the behaviour of the particles, hence (2) must hold for a valid probability distribution. If you integrate both sides of (1) you get:

$$n = \int_{-\infty}^{\infty} dn = \int_{0}^{n} dn = \int_{-\infty}^{\infty} n\, \phi(\Delta)\, d\Delta = n \int_{-\infty}^{\infty} \phi(\Delta)\, d\Delta = n \tag{16}$$

Equation (1) is uncontroversial unless you want to dispute the assumption of mutual independence. The editor of [1], R Furth makes the point ([1], page 97) that Einstein's assumption is not previously established and he notes that L S Ornsetin developed a formula without such assumptions, but when the time interval is sufficiently large the formula agrees with Einstein's approach.

If you start out with n particles in the short time τ the number of particles that will lie in the interval $(\Delta, \Delta + d\Delta)$ is n weighted by the probability of occupying that differential region ie $n\, \phi(\Delta)\, d\Delta$.

What really drives Einstein's whole analysis is the step in (4). Gardiner has this to say about equation (4) (see [3], page 5):

“ The Chapman-Kolmogorov Equation occurs as Einstein's equation (4). It states that the probability of a particle being at a point x at time $t+\tau$ is given by the sum of all possible “pushes” Δ from positions $x+\Delta$, multiplied by the probability of being at $x+\Delta$ at time t. This assumption is based on the independence of the push Δ of any previous history of the motion: it is only necessary to know the initial position of the particle at time t - not at any previous time. This is the Markov postulate and the Chapman Kolmogorov equation, of which (4) is a special form, is the central dynamical equation to all Markov processes.”

The Chapman-Kolmogorov equation is explained in detail in Appendix 1. It is a statement about probabilities whereas equation (4) is actually a statement about a “volume” (in Einstein's case a degenerate volume ie a length) multiplied by a length giving a pure number. The RHS is the integral of a probabilistic component with a degenerate “volume”. Thus (4) is indeed a special “species” of Chapman-Kolmogorov equations since the RHS is not a sum or integral of the product of two probabilities.

In effect Einstein is saying that:

$$f(x,t+\tau) = \int_{\Delta=-\infty}^{\Delta=+\infty} f(x+\Delta,t)\,\phi(\Delta)\,d\Delta \tag{17}$$

If you imagine fixing a spatial point at a time $t+\tau$ where τ is very small in comparison to t and movements of the particle in consecutive intervals of time t are mutually independent phenomena, the “volume” of particles captured in the interval $(x, x+dx)$ in the time interval $(t, t+\tau)$ is the sum (because of the independence assumption) of the symmetric displacements about x of the “volumetric” function $f(x,t)$ (number of particles per unit volume) weighted by the probability.

Note that there is no τ on the RHS of (17). This small time displacement has in a sense been integrated out over the spatial dimension.

So integrating from $\Delta = -\infty$ to $\Delta = +\infty$ gives the total number of particles per unit volume at $f(x,t+\tau)$. Given that τ is very small the following approximation applies (Einstein writes it as an equality without an error term but we know what he means):

$$f(x,t+\tau) = f(x,t) + \tau\,\frac{\partial f(x,t)}{\partial t} \tag{18}$$

Next Einstein expands $f(x+\Delta,t)$ in powers of Δ using Taylor's Theorem:

$$f(x+\Delta,t)) = f(x,t) + \Delta\,\frac{\partial f(x,t)}{\partial x} + \frac{\Delta^2}{2!}\,\frac{\partial^2 f(x,t)}{\partial x^2} + \ldots \tag{19}$$

What he does next is deliciously straightforward:

$$
\begin{aligned}
f(x,t+\tau) = f(x,t) + \tau\,\frac{\partial f(x,t)}{\partial t} \\
= \int_{\Delta=-\infty}^{\Delta=+\infty} f(x+\Delta,t)\,\phi(\Delta)\,d\Delta \\
= \int_{\Delta=-\infty}^{\Delta=+\infty} \left\{ f(x,t) + \Delta\,\frac{\partial f(x,t)}{\partial x} + \frac{\Delta^2}{2!}\,\frac{\partial^2 f(x,t)}{\partial x^2} + \dots \right\} \phi(\Delta)\,d\Delta \\
= f(x,t) \int_{\Delta=-\infty}^{\Delta=+\infty} \phi(\Delta)\,d\Delta + \frac{\partial f(x,t)}{\partial x} \int_{\Delta=-\infty}^{\Delta=+\infty} \Delta\,\phi(\Delta)\,d\Delta + \frac{\partial^2 f(x,t)}{\partial x^2} \int_{\Delta=-\infty}^{\Delta=+\infty} \frac{\Delta^2}{2!}\,\phi(\Delta)\,d\Delta + \dots \\
= f(x,t) + \frac{\partial f(x,t)}{\partial x} \int_{\Delta=-\infty}^{\Delta=+\infty} \Delta\,\phi(\Delta)\,d\Delta \;+\; \frac{\partial^2 f(x,t)}{\partial x^2} \int_{\Delta=-\infty}^{\Delta=+\infty} \frac{\Delta^2}{2!}\,\phi(\Delta)\,d\Delta + \dots
\end{aligned}
\tag{20}
$$

We have used on the RHS of (20) the facts that $\int_{-\infty}^{\infty} \phi(\Delta)\,d\Delta = 1$ and because of the assumption $\phi(\Delta) = \phi(-\Delta)$ it follows that odd terms such as $\frac{\partial f(x,t)}{\partial x} \int_{\Delta=-\infty}^{\Delta=+\infty} \Delta\,\phi(\Delta)\,d\Delta$ and $\frac{\partial^3 f(x,t)}{\partial x^3} \int_{\Delta=-\infty}^{\Delta=+\infty} \frac{\Delta^3}{3!}\,\phi(\Delta)\,d\Delta$ etc are all zero. If you cannot immediately see this symmetry, just split the integrals up into the two intervals $(-\infty,0]$ and $[0,\infty)$ and make the substitution $\Delta \to -\Delta$ for the integral on $(-\infty,0]$. It becomes the negative of the integral on $[0,\infty)$ so the end result of the sum of the two integrals is zero.

Einstein asserts without proof that of the first, third, fifth etc terms on the RHS of (20), every succeeding term is very small compared with the preceding, Thus:

$$
f(x,t) > \frac{\partial^2 f(x,t)}{\partial x^2} \int_{\Delta=-\infty}^{\Delta=+\infty} \frac{\Delta^2}{2!}\,\phi(\Delta)\,d\Delta \text{ etc} \tag{21}
$$

This is a non-trivial assertion since we have no information about the rates of decay of the derivatives (which are are assumed to exist). How is it justified? The assertion is that the products are decreasing. The first step in the argument is that (19) can be brought under the integral sign "since only very small values of Δ contribute anything to the latter." In other words, the probability of large values of Δ is very small so that when we integrate in (17) it is only the small values of Δ that contribute significantly to the integral. Einstein assumed that "ϕ only differs from zero for very small values of Δ "-ie compact support. This is physically believable since it would be unlikely in a very small time scale to see a very large displacement. Indeed, all the empirical work on Brownian motion justifies that belief. Einstein is also implicitly saying that the derivatives are all bounded. Let us probe this in a bit more detail.

Let's start with the second derivative. It is a standard exercise in calculus that:

$$\frac{\partial^2 f(x,t)}{\partial x^2} = \lim_{h\to 0} \frac{f(x+2h) - 2f(x+h) + f(x)}{h^2} \tag{22}$$

For those who want the gory details see Appendix 2.

Now let's estimate the integral in (21) as follows. We will ignore values of Δ greater than some very small value Δ_s consistent with Einstein's comment:

$$\begin{aligned}\int_{\Delta=-\infty}^{\Delta=+\infty} \frac{\Delta^2}{2!}\,\phi(\Delta)\,d\Delta \approx & \int_{\Delta=-\Delta_s}^{\Delta=+\Delta_s} \frac{\Delta^2}{2!}\,\phi(\Delta)\,d\Delta \\ =& 2\int_{0}^{\Delta=+\Delta_s} \frac{\Delta^2}{2!}\,\phi(\Delta)\,d\Delta \\ \le& B\int_{0}^{\Delta=+\Delta_s} \Delta^2\,d\Delta \\ =& \frac{1}{3}B\Delta_s^3\end{aligned} \tag{23}$$

where $\phi(\Delta) \le B$ for some constant B, since the probability density is bounded.

Putting this all together in (21) with $h = \Delta_s$ we have:

$$\begin{aligned}f(x,t) - \frac{\partial^2 f(x,t)}{\partial x^2}\int_{\Delta=-\infty}^{\Delta=+\infty} \frac{\Delta^2}{2!}\,\phi(\Delta)\,d\Delta \approx & f(x,t) - \frac{1}{3}B\Delta_s^3\Big[\frac{f(x+2\Delta_s) - 2f(x+\Delta_s) + f(x)}{\Delta_s^2}\Big] \\ =& f(x,t) - \frac{1}{3}B\Delta_s\Big[f(x+2\Delta_s) - 2f(x+\Delta_s) + f(x)\Big] \\ =& f(x,t)\underbrace{[1 - \frac{1}{3}B\Delta_s]}_{\to 1} - \underbrace{\frac{1}{3}B\Delta_s\Big[f(x+2\Delta_s) - 2f(x+\Delta_s)\Big]}_{\to 0} \\ \to& f(x,t) > 0\end{aligned} \tag{24}$$

So the qualitative argument does make sense.

What Einstein ultimately gets from (20) is this:

$$f(x,t) + \tau\frac{\partial f(x,t)}{\partial t} = f(x,t) + \frac{\partial^2 f(x,t)}{\partial x^2}\int_{\Delta=-\infty}^{\Delta=+\infty} \frac{\Delta^2}{2!}\,\phi(\Delta)\,d\Delta \tag{25}$$

Einstein defines the diffusion coefficient as:

$$D = \frac{1}{\tau} \int_{\Delta=-\infty}^{\Delta=+\infty} \frac{\Delta^2}{2!}\, \phi(\Delta)\, d\Delta \tag{26}$$

So from (25) and (26) he gets:

$$\frac{\partial f(x,t)}{\partial x} = D\, \frac{\partial^2 f(x,t)}{\partial x^2} \tag{27}$$

To solve the parabolic partial differential equation in (27) some initial conditions are needed. Bearing in mind that (27) is the PDE for the heat equation we might expect some sinusoidal components in the solution if the model is that of temperature distribution in a ring. For instance, the solution to the heat equation:

$$\frac{\partial u(x,t)}{\partial t} = \frac{\partial^2 u(x,t)}{\partial x^2} \text{ with initial condition } u(x,0) = g(x) \tag{28}$$

is:

$u(x,t) = (g * \mathcal{H}_t)(x)$

where $\mathcal{H}_(x)$ is the heat kernel on the circle given by $\mathcal{H}_t(x) = \sum_{n=-\infty}^{n=\infty} e^{-4\pi^2 n^2 t}\, e^{2\pi\, inx}$

(29)

See the discussion in [6], Chapter 4. However, when the model is that of temperature distribution in a rod (which is analogous to the set up for Einstein's Brownian motion solution) we get a heat kernel of the following form (with diffusion coefficient $D = 1$):

$$\mathcal{H}_t(x) = \frac{1}{\sqrt{4\pi t}} e^{-\frac{x^2}{4t}} \tag{30}$$

Again, see the discussion in [6], Chapter 5.

Different initial conditions for (27) will give rise to different types of solutions. Einstein notes that his derivation assumed that single particles are all referred to the same coordinate system but this is unnecessary "since the movements of the single particles are mutually independent". He then assumes that the motion of each particle is referenced to the "coordinate system whose origin coincides at the time $t = 0$ with the position of the centre of gravity of the particles in questions.”

Given this assumption ”$f(x,t)\,dx$ now gives the number of particles whose x coordinate has increased between the time $t=0$ and time $t=t$, by a quantity which lies between x and $x+dx$.”

Einstein states that ”we must evidently have for $x \neq 0$ and $t=0$:

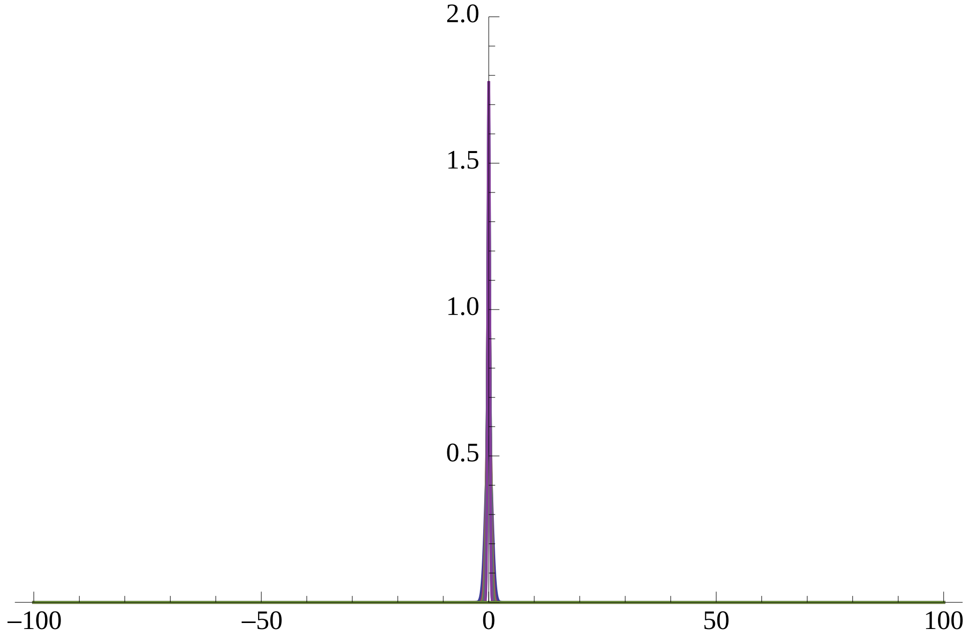
$$f(x,t) = 0 \quad \text{and} \int_{-\infty}^{\infty} f(x,t)\,dx = n\text{”} \tag{31}$$

On this basis he states that the diffusion problem is completely solved and he gives as the solution:

$$f(x,t) = \frac{n\,e^{\frac{-x^2}{4Dt}}}{\sqrt{4\pi Dt}} \tag{32}$$

The solution to the heat PDE was well known at the time of Einstein’s 1905 paper but the innovation was how it figured in a purely particulate process with an underlying probability distribution.

It is useful to look at the initial conditions closely. What they say is that $f(x,0)=0$ and $\int_{-\infty}^{\infty} f(x,t)\,dx = n$ which suggests that at $t=0$ the concentration of the diffusing substance is everywhere zero with the exception of an infinitesimally narrow band around the plane $x=0$. This in turn suggests that at $x=0, t=0$, $f(x,t)$ is the limit of Gaussian functions as the bands become smaller ie it is the Dirac delta “function” (scaled to the number of particles n) which is consistent with the mass concentration at $x=0$ and $t=0$. For instance if we consider Gaussian kernels of the form $\delta_p(x) = \sqrt{\frac{p}{\pi}}\,e^{-px^2}$ we get the following peaking behaviour for $1 \leq p \leq 100$ and $-100 \leq x \leq 100$:



Recall that the Dirac delta "function" has the following properties:

$$\begin{aligned} \int_{-\infty}^{\infty} \delta(t)\, x(t)\, dt =& x(0) \\ \int_{-\infty}^{\infty} \delta(t)\, dt =& 1 \end{aligned} \tag{33}$$

Einstein probably (but I can't be sure) solved (32) by using the method of separation of variables ie assuming $u(x,t) = X(x)\,T(t)$ as that was the classical line of approach during the 19^{th} century. It is improbable that he used Fourier transform techniques. The solution of the heat equation was common knowledge for a physicist of Einstein's calibre. The actual calculations are reasonably intricate and are covered in [7], pages 152-156. In particular, in [7], pages 154-156 the set up is:

$$\frac{\partial u(x,t)}{\partial t} = k \frac{\partial^2 u(x,t)}{\partial x^2} \qquad -\infty < x < \infty, t > 0 \tag{34}$$

$$u(x,0) = f(x) \qquad -\infty < x < \infty \tag{35}$$

and $u(x,t)$ satisfies a boundedness condition $|u(x,t)| < M$.

The solution to (34)-(35) then has this form ([7] page 156):

$$u(x,t) = \frac{1}{2\sqrt{\pi k t}} \int_{-\infty}^{\infty} f(\xi) e^{-\frac{(x-\xi)^2}{4kt}}\, d\xi \qquad t > 0 \tag{36}$$

How do we get (32) out of (36)? The answer revolves around the form of $f(x)$ in (35). If we treat $f(x)$ as a Delta "function" of height n at $x = 0, t = 0$ it will select out the value $\xi = 0$ (see (33)) and the integral on the RHS of (36) becomes:

$$\int_{-\infty}^{\infty} f(\xi) e^{-\frac{(x-\xi)^2}{4kt}}\, d\xi = n e^{-\frac{x^2}{4kt}} \tag{37}$$

With the identification of $k = D$ and noting that $\frac{1}{2\sqrt{\pi k t}} = \frac{1}{\sqrt{4\pi k t}}$ we then get (32).

More detail of how the method of separation of variables is performed in the context of the heat equation can be found in [8].

To use Fourier transforms we need some basic facts about how they are manipulated. In the modern approach to Fourier theory if we assume that the relevant functions decay sufficiently rapidly the proofs of the following facts are essentially straightforward (see [6] for details). For instance if we assume that the functions f inhabit Schwartz space $\mathcal{S}(\mathbb{R})$ the results follow since the functions decay sufficiently rapidly because the functions satisfy:

$$\sup_{x \in \mathbb{R}} |x|^k \, |f^{(l)}(x)| < \infty \text{ for all k,l} \geq 0 \tag{38}$$

Note here that when $l = 0$ (38) is saying that $\sup_{x \in \mathbb{R}} |x|^k \, |f(x)| < \infty$ for all k, which means that $f(x)$ decays at least as fast as e^{-x} - just recall that e^x grows faster than any power of x.

The critical properties of the Fourier transform $\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)\, e^{-2\pi\, i\, \xi\, x}\, dx$ are these:

1. $f(x+h) \to \hat{f}(\xi)\, e^{2\pi\, i\, \xi\, h}$ when $h \in \mathbb{R}$
2. $f(x)\, e^{-2\pi\, i\, x\, h} \to \hat{f}(\xi + h)$ when $h \in \mathbb{R}$

3. $f(\delta x) \to \delta^{-1}\, \hat{f}(\delta^{-1}\, \xi)$ when $\delta > 0$
4. $f'(x) \to 2\pi\, i\, \xi\, \hat{f}(\xi)$
5. $-2\pi\, i\, x\, f(x) \to \frac{d}{d\xi}\, \hat{f}(\xi)$
6. $\widehat{f * g}(\xi) = f(\xi)\, g(\xi)$ where $(f * g)(x) = \int_{-\infty}^{\infty} f(x-u)\, g(u) du$

We start by writing (27) in this form for notational convenience:

$$f_t(x,t) = D\, f_{xx}(x,t) \tag{39}$$

We now use property 4 above to take Fourier transforms with respect to x of both sides of (39):

$$\hat{f}_t(\xi,t) = D\,(2\pi\ i\xi)^2 \hat{f}(\xi,t) = -4D\pi^2\xi^2 \hat{f}(\xi,t) \tag{40}$$

(Note that property 4 is used twice on the RHS of (40))

Thus we have:

$$\hat{f}_t(\xi,t) + 4D\pi^2\xi^2 \hat{f}(\xi,t) = 0 \tag{41}$$

(41) is just a run of the mill linear ordinary differential equation which invites the use of an integrating factor $e^{4D\pi^2\xi^2\, t}$: Thus:

$$\frac{\partial}{\partial t}\{\hat{f}(\xi,t)\ e^{4D\pi^2\xi^2\, t}\} = 0 \tag{42}$$

Solving (42) we see that:

$$\hat{f}(\xi,t)\ e^{4D\pi^2\xi^2\, t} = h(\xi) \implies \hat{f}(\xi,t) = h(\xi)\ e^{-4D\pi^2\xi^2\, t} \tag{43}$$

where $h(\xi)$ is an arbitrary function which is to be determined by the initial conditions.

Recall from (31) that for $x \neq 0$ and $t = 0$:

$$f(x,t) = 0 \quad \text{and} \int_{-\infty}^{\infty} f(x,t)\, dx = n \tag{44}$$

We need to take the Fourier transform represented by $\mathcal{F}$ or $\hat{}$ of the initial conditions in (31) which essentially involves taking the Fourier transform of the Delta "function". The Fourier transform of $\delta(t)$ is simply 1 ie

$$\mathcal{F}(\delta(t)) = \mathbf{1} = \hat{\delta}(\xi) \tag{45}$$

where the bolding is intended to emphasis the "functional" character of the Dirac delta "function".

The required Fourier transform of the initial conditions from (43) is therefore:

$$\hat{f}(\xi, 0) = \hat{h(\xi)} = n\,\hat{\delta}(\xi) \tag{46}$$

Thus we have:

$$\hat{f}(\xi, t) = n\,\hat{\delta}(\xi)\; e^{-4D\pi^2\xi^2\,t} \tag{47}$$

(47) is the product of two Fourier transforms and this is the Fourier transform of the convolution of those functions ie:

$$\hat{f}(\xi, t) = n\;\hat{\delta}(\xi)\;\hat{g}(\xi, t) = n\widehat{(\delta * g)}(\xi, t) \tag{48}$$

where:

$$(p * q)(x) = \int_{-\infty}^{\infty} p(x-y)\,q(y)\,dy \tag{49}$$

When we take the inverse Fourier transform of (48) we get:

$$f(x, t) = \mathcal{F}^{-1}\big\{\widehat{(\delta * g)}(\xi, t)\big\} = \delta(x) * g(x, t) \tag{50}$$

Because the Fourier transform of a Gaussian is a Gaussian ie if $g(x) = e^{-\pi\,x^2}$ then $\hat{g}(\xi) = e^{-\pi\,\xi^2}$, all we have to do is find a suitably scaled $g(x, t)$ to give $e^{-4D\pi^2\xi^2\,t}$. From (45) we know that the Fourier transform of the Dirac delta is $\hat{\delta}(\xi) = \mathbf{1}$. So we have both parts.

In what follows, t will be treated as a constant and so we want $\hat{g}(\xi) = e^{-4\pi^2 Dt\,\xi^2}$. We let $u = \sqrt{4\pi Dt}\,x = \alpha x$ and using property 3 of Fourier transforms we have:

$$\mathcal{F}(g(u)) = \frac{1}{\alpha}\hat{g}\Big(\frac{1}{\alpha}x\Big) = \frac{1}{\sqrt{4\pi Dt}}e^{-\pi\,\frac{\xi^2}{4\pi Dt}} = \frac{1}{\sqrt{4\pi Dt}}e^{-\frac{\xi^2}{4Dt}} \tag{51}$$

Thus (50) becomes:

$$\begin{aligned} f(x,t) =& \delta(x) * g(x,t) \\ =& n\int_{-\infty}^{\infty}\frac{1}{\sqrt{4\pi Dt}}e^{-\frac{(x-y)^2}{4Dt}}\,\delta(y)\,dy \\ =& \frac{n}{\sqrt{4\pi Dt}}e^{-\frac{x^2}{4Dt}} \end{aligned} \tag{52}$$

The property of the Dirac delta function that is used in the last line of (52) is its "selection" property ie $\int_{-\infty}^{\infty} p(x)\,\delta(x)\,dx = p(0)$

After all of this we need to check that (52) actually solves the PDE in (27):

$$\begin{aligned} \frac{\partial f}{\partial t} =& \frac{n}{\sqrt{4\pi Dt}} \times e^{-\frac{x^2}{4Dt}} \times \frac{-x^2}{4D} \times \frac{-1}{t^2} + \frac{n}{\sqrt{4\pi D}} \times e^{-\frac{x^2}{4Dt}} \times \frac{-1}{2t^{3/2}} \\ =& \frac{n}{\sqrt{4\pi Dt}} \times e^{-\frac{x^2}{4Dt}}\left(\frac{x^2}{4Dt^2} - \frac{1}{2t}\right) \end{aligned} \tag{53}$$

$$\begin{aligned} \frac{\partial f}{\partial x} =& \frac{n}{\sqrt{4\pi Dt}} \times e^{-\frac{x^2}{4Dt}} \times \frac{-2x}{4Dt} \\ =& \frac{-nx}{2Dt\,\sqrt{4\pi Dt}} \times e^{-\frac{x^2}{4Dt}} \end{aligned} \tag{54}$$

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2} =& \frac{-nx}{2Dt\,\sqrt{4\pi Dt}} \times e^{-\frac{x^2}{4Dt}} \times \frac{-2x}{4Dt} - \frac{-n}{2Dt\,\sqrt{4\pi Dt}} \times e^{-\frac{x^2}{4Dt}} \times 1 \\ =& \frac{nx^2}{4D^2t^2\,\sqrt{4\pi Dt}} \times e^{-\frac{x^2}{4Dt}} - \frac{n}{2Dt\,\sqrt{4\pi Dt}} \times e^{-\frac{x^2}{4Dt}} \\ =& \frac{n}{\sqrt{4\pi Dt}}e^{-\frac{x^2}{4Dt}}\left(\frac{x^2}{4D^2t^2} - \frac{1}{2Dt}\right) \end{aligned} \tag{55}$$

Hence:

$$
\begin{aligned}
D\frac{\partial^2 f}{\partial x^2} =& D\frac{n}{\sqrt{4\pi Dt}}e^{-\frac{x^2}{4Dt}}\left(\frac{x^2}{4D^2t^2} - \frac{1}{2Dt}\right) \\
=& \frac{n}{\sqrt{4\pi Dt}}e^{-\frac{x^2}{4Dt}}\left(\frac{x^2}{4Dt^2} - \frac{1}{2t}\right) \\
=& \frac{\partial f}{\partial t}
\end{aligned}
\tag{56}
$$

The initial conditions can only really be understood in terms of Dirac delta "function" terms which suggests that the work of Oliver Heaviside and others had permeated into Germany before 1905 (Dirac is later of course- he was born in 1902). I haven't drilled down into this but Einstein was confident enough that his analysis was not so ropey that people would dismiss it. As I have demonstrated it does hang together. The interesting thing is what is going on is an averaging process known to harmonic analysts in the context of the Laplacian (see [9] for more details).

3 Appendix 1

As a prelimimary to describing the Chapman-Kolmogorov equations, we start with the concept of a Markov chain in a stochastic process which requires that whenever we have a stochastic process which is in state i, then there is a fixed transition probability P_{ij} that its next state will be in j.

The current state at time n is denoted by X_n.

If $A = \{X_0 = i_0, X_1 = i_1 = \cdots = X_{n-1} = i_{n-1}\}$ is the previous history of the Markov chain by the time n, then $\{X_n\}$ has the Markov property it forgets about its past. In other words:

$$
Pr\{X_{n+1} = j | A \cap X_n = i\} = Pr\{X_{n+1} = j | X_n = i\} \tag{57}
$$

Put another way, the next state depends only on the current state and is independent of time.

$\{X_n\}$ is time homogeneous if:

$$Pr\{X_{n+1} = j | A \cap X_n = i\} = Pr\{X_1 = j | X_0 = i\} = P_{ij} \tag{58}$$

ie if the transition probabilities are independent of n.

The probability that a process currently in state i will be in state j after n additional transitions or steps is:

$$P_{ij}^{(n)} = Pr\{X_n = j | X_0 = i\} \text{ for } n, i, j \geq 0 \tag{59}$$

where $P_{ij}^{(1)} = P_{ij}$ and

$$P_{ij}^{(0)} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Chapman-Kolmogorov equations:

$$\boxed{P_{ij}^{(n+m)} = \sum_{k=0}^{\infty} P_{ik}^{(n)} \, P_{kj}^{(m)}} \tag{60}$$

That is going from i to j in $n + m$ steps where there is an intermediate stop in state k after n steps so you have to sum over all possible k. Also note that the form of (60) screams matrix multiplication and that is a good way to remember its structure.
To prove (60) we need to know the following:

$$Pr\{X_{n+m} = j | X_0 = i\} = \sum_{k=0}^{\infty} Pr\{X_{n+m} = j \, \cap \, X_n = k | X_0 = i\} \tag{61}$$

This is a total probability result and be seen by noting that $A = A \cap \Omega = A \cap \big(\cup_{k=0}^{\infty} C_k \big)$ where the C_k are mutually disjoint events such that $\cup_{k=0}^{\infty} C_k = \Omega$. Note that the $A \cap C_k$ are mutually disjoint too and that $A \cap \big(\cup_{k=0}^{\infty} C_k \big) = \cup_{k=0}^{\infty} \big(A \cap C_k \big)$. Equation (34) then follows.

The other critical thing to note is that $Pr\{A \cap C | B\} = Pr\{A | B \cap C\} \, Pr\{C | B\}$. This is proved as follows, with $D = B \cap C$:

$$
\begin{aligned}
Pr\{A \cap C|B\} =& \frac{Pr\{A \cap C \cap B\}}{Pr\{B\}} \\
=& \frac{Pr\{A \cap D\}}{Pr\{B\}} \\
=& \frac{Pr\{A|D\} \times Pr\{D\}}{Pr\{B\}} \\
=& \frac{Pr\{A|B \cap C\} \times Pr\{B \cap C\}}{Pr\{B\}} \\
=& \frac{Pr\{A|B \cap C\} \times Pr\{C \cap B\}}{Pr\{B\}} \\
=& Pr\{A|B \cap C\} \times Pr\{C|B\}
\end{aligned}
\tag{62}
$$

With these preliminaries we can now prove (33):

$$
\begin{aligned}
P_{ij}^{(n+m)} =& Pr\{X_{n+m} = j|X_0 = i\} \text{ by (32)} \\
=& \sum_{k=0}^{\infty} Pr\{X_{n+m} = j \cap X_n = k|X_0 = i\} \text{ by (34)} \\
=& \sum_{k=0}^{\infty} Pr\{X_{n+m} = j|X_0 = i \cap X_n = k\}\, Pr\{X_n = k|X_0 = i\} \text{ by (35)} \\
=& \sum_{k=0}^{\infty} Pr\{X_{n+m} = j|X_n = k\}\, Pr\{X_n = k|X_0 = i\} \text{ by the Markov property} \\
=& P_{ik}^{(n)}\, P_{kj}^{(m)}
\end{aligned}
\tag{63}
$$

4 Appendix 2

In equation (22) I boldly asserted that:

$$
\frac{\partial^2 f(x,t)}{\partial x^2} = \lim_{h \to 0} \frac{f(x+2h) - 2f(x+h) + f(x)}{h^2} \tag{64}
$$

To prove this just let:

$$
F(x,t) = \frac{\partial f(x,t)}{\partial x} = \lim_{h \to 0} \frac{f(x+h,t) - f(x,t)}{h} \tag{65}
$$

So:

$$\begin{aligned}\frac{\partial^2 f(x,t)}{\partial x^2} &= \lim_{h\to 0}\frac{F(x+h,t)-F(x,h)}{h}\\ &= \lim_{h\to 0}\frac{f(x+2h)-f(x+h)-[f(x+h)-f(x)])}{h}\times\frac{1}{h}\\ &= \lim_{h\to 0}\frac{f(x+2h)-2f(x+h)+f(x)])}{h^2}\end{aligned} \tag{66}$$

5 References

[1] Albert Einstein, Investigations on the theory of Brownian Movement , Dover, 1956

[2] Jean Perrin, Brownian Movement and Molecular Reality , Taylor and Francis, London, 1910. This book can be accessed online here http://web.mit.edu/swangroup/footer/perrin_bm.pdf

[3] C W Gardiner, Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences , Third Edition, Springer, 2004

[4] Peter Haggstrom, How Maxwell derived his velocity distribution law, https://gotohaggstrom.com/How%20Maxwell%20derived%20his%20velocity%20distribution%20law_24April2019.pdf

[5] Richard P Feynman, Albert R Hibbs, Quantum Mechanics and Path Integrals, Emended by Daniel F Styer, Dover, 2005.

[6] Elias Stein and Rami Shakarchi, Fourier Analysis: An Introduction", Princeton University Press, 2003.

[7] Ruel V Churchill, Fourier Series and Boundary Value Problems", Second Edition,McGraw Hill, 1963.

[8] Peter Haggstrom, Using the heat equation to motivate the idea of the Fourier transform , https://gotohaggstrom.com/Using%20the%20heat%20equation%20to%20motivate%20the%20idea%20of%20the%20Fourier%20transform.pdf

[9] Peter Haggstrom, Basic Fourier integrals, https://gotohaggstrom.com/Basic%20Fourier%20integrals.pdf

6 History

Created 01 September 2020