SEVENTH LECTURE

I will now sketch a proof -- one which does not really satisfy me entirely. It's a rather ugly thing, because one has to compute too much. One feels that one shouldn't have to compute anything at all. Such a statement ought to be provable by "pure thought." I would also like to call your attention to the fact that we have not accomplished very much, except to find an interesting way of writing the solution. We will later see a similar treatment for certain parabolic equations -- with the difference that in that case one can actually use the new probabilistic form to draw significant analytical conclusions.

The calculations which we are going to perform will not be wholly wasted, because we will need to do similar things later on. I will only prove the statement (26) for functions φ which are reasonably "decent." In particular, I will assume that my function can be written as a Fourier integral:

$$\varphi(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\xi x}\,\Phi(\xi)\,d\xi \tag{29}$$

Now if you substitute this in the statement we are trying to prove you get:

$$F(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Phi(\xi)\,e^{-i\xi x}\left\langle \cos\left(\xi v\int_0^t (-1)^{N(\tau)}\,d\tau\right)\right\rangle d\xi \tag{30}$$

Now you look at this cosine and expand it in a power series:

$$1-\frac{(\xi V)^2}{2!}\left\langle\left(\int_0^t(-1)^{N(\tau)}d\tau\right)^2\right\rangle+\frac{(\xi V)^4}{4!}\left\langle\left(\int_0^t(-1)^{N(\tau)}d\tau\right)^4\right\rangle+\cdots \tag{31}$$

These averages, as you very well know, are called moments -- they are the moments of this strange "randomized" time. In this case, only the even moments enter. However, let me just show you first how one calculates the first moment. This can actually be done immediately because you can interchange the averaging and integration:

$$\left\langle\int_0^t(-1)^{N(\tau)}d\tau\right\rangle=\int_0^t\left\langle(-1)^{N(\tau)}\right\rangle d\tau \tag{32}$$

(This can be easily justified. The averaging is also an integration, an integration over the space of all functions $N(t)$. So it is merely a question of interchanging the order of integration.) Now the path is clear, because I know what the distribution of $N(\tau)$ is. It is given by (21). Then, from the definition of an average -- or mathematical expectation, if you wish -- we find that:

$$\left\langle(-1)^{N(\tau)}\right\rangle=\sum_{k=0}^{\infty}(-1)^k\cdot e^{-a\tau}\frac{(a\tau)^k}{k!} \tag{33}$$

This is an easy series, isn't it? If I pull $e^{-a\tau}$ out then what I have left is just the series for $e^{-a\tau}$. So the whole business is just $e^{-2a\tau}$. And so we can find the first moment:

$$\mu_1(t) = \left\langle \int_0^t (-1)^{N(\tau)} d\tau \right\rangle = \int_0^t e^{-2a\tau} d\tau \qquad (34)$$

It is only a little bit more complicated to calculate the second moment. It's a very common trick, which is used over and over again, to write the square of an integral as a double integral by introducing two variables. Let me do that:

$$\left\langle \left(\int_0^t (-1)^{N(\tau)} d\tau \right)^2 \right\rangle = \left\langle \int_0^t \int_0^t (-1)^{N(\tau_1)} (-1)^{N(\tau_2)} d\tau_1 d\tau_2 \right\rangle \qquad (35)$$

The expression on the right is an integral over a square. Since it is completely symmetrical in the two variables τ_1 and τ_2, I can integrate over only half the square and multiply by two.

$$2! \left\langle \iint\limits_{0 \leqslant \tau_1 \leqslant \tau_2 \leqslant t} (-1)^{N(\tau_1)} (-1)^{N(\tau_2)} d\tau_1 d\tau_2 \right\rangle = 2! \iint\limits_{0 \leqslant \tau_1 \leqslant \tau_2 \leqslant t} \left\langle (-1)^{N(\tau_1)+N(\tau_2)} \right\rangle d\tau_1 d\tau_2$$

There is a reason why I put $2!$ rather than simply 2: I am anticipating the results for the higher moments. And now you perform a very simple trick. You merely write $N(\tau_2) = N(\tau_1) + \left[N(\tau_2) - N(\tau_1)\right]$. There's a point in writing it this way, because I have separated $N(\tau_2)$ into a sum of two things which are independent. In the case of the higher moments you do the same thing, only there will be more terms in the decomposition. Making now this substitution we have:

$$2!\iint\limits_{0\leqslant\tau_1\leqslant\tau_2\leqslant t}\left\langle(-1)^{2N(\tau_1)+N(\tau_2)-N(\tau_1)}\right\rangle d\tau_1\,d\tau_2 \tag{36}$$

Now this becomes greatly simplified, because $2N(\tau_1)$, whatever else it is, is an even number. Therefore,

$$\left\langle(-1)^{2N(\tau_1)+N(\tau_2)-N(\tau_1)}\right\rangle=\left\langle(-1)^{N(\tau_2)-N(\tau_1)}\right\rangle \tag{37}$$

which, from the meaning of the average, is nothing more than:

$$\sum_{k=0}^{\infty}(-1)^k\cdot e^{-a(\tau_2-\tau_1)}\frac{[a(\tau_2-\tau_1)]^k}{k!}=e^{-2a(\tau_2-\tau_1)} \tag{38}$$

Consequently, the second moment finally becomes:

$$\mu_2(t)=\left\langle\left(\int_0^t(-1)^{N(\tau)}d\tau\right)^2\right\rangle=2!\int_0^t d\tau_2\int_0^{\tau_2}d\tau_1\,e^{-2a(\tau_2-\tau_1)} \tag{39}$$

Now this last is just the integral of a convolution. Why does the convolution come in? Precisely because of the decomposition we made of $N(\tau_2)$ into two independent parts. It shouldn't take much imagination to see that this will also happen when you go to the higher moments. Indeed, this happens in all stochastic processes with independent increments. You always have such convolutions coming in. The natural thing to do, then, is to take the Laplace transform. Things will be much simpler, because the Laplace transform of a convolution is the product of transforms. Now I can write (39) in the form of a double convolution by introducing the Heaviside function:

$$\Delta(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 & \text{if } t > 0 \end{cases} \tag{40}$$

It becomes just:

$$\mu_2(t) = 2! \int_0^\infty \Delta(t-\tau_2) \int_0^\infty \Delta(\tau_2-\tau_1) e^{-2a\tau_1} d\tau_1 = \Delta^* \Delta^* e^{-2at} \tag{41}$$

Then we get the Laplace transform immediately:

$$\int_0^\infty e^{-st} \frac{\mu_2(t)}{2!} dt = \frac{1}{s} \cdot \frac{1}{s} \cdot \frac{1}{s+2a} \tag{42}$$

The general formula turns out to be different for n even and odd. It's a nice exercise for you to verify that:

$$\int_0^\infty e^{-st} \frac{\mu_n(t)}{n!} dt = \begin{cases} \dfrac{1}{s^{\frac{n+1}{2}}} \cdot \dfrac{1}{(s+2a)^{\frac{n+1}{2}}}, & \text{for } n \text{ odd} \\[2ex] \dfrac{1}{s^{\frac{n}{2}+1}} \cdot \dfrac{1}{(s+2a)^{n/2}}, & \text{for } n \text{ even} \end{cases} \tag{43}$$

Now the moments are rather messy, because they are inverse Laplace transforms of these things. That suggests that rather than to work with $F(x,t)$ itself, we may better work with its Laplace transform. So let us calculate this Laplace transform.

$$\int_0^{\infty} e^{-st} F(x,t)\,dt = \frac{1}{2\pi s}\int_{-\infty}^{\infty} \Phi(\xi)\, e^{-i\xi x} \left(\sum_{n=0}^{\infty} \left[\frac{-(\xi v)^2}{s(s+2a)} \right]^n \right) d\xi \qquad (44)$$

We have here made use of formula (30) and the moments we have calculated. The series can be summed. It's a very well-known one. You then get an extremely simple formula:

$$\int_0^{\infty} e^{-st} F(x,t)\,dt = \frac{1}{2\pi s}\int_{-\infty}^{\infty} \Phi(\xi)\, e^{-i\xi x} \left[1 + \frac{(\xi v)^2}{s(s+2a)} \right]^{-1} d\xi \qquad (45)$$

And now it is easy to see, at least formally, what transpires. The Laplace transform of the telegrapher's equation is

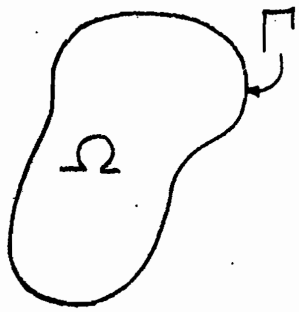
$$v \frac{d^2 f}{dx^2} - \frac{s(s+2a)}{v} f - \frac{s+2a}{v} F(x,0) = 0 \qquad (46)$$

This form of the equation, as a matter of fact, is often used in solving the telegrapher's equation. But now you can directly verify that $f(x,s)$, the Laplace transform of my function $F(x,t)$, satisfies this one. You simply substitute it in. It follows then that my function $F(x,t)$ satisfies the telegrapher's equation.

This is not really a very nice proof. It is very inelegant -- although perhaps one should never speak of elegance among people who are engaged at least part of their time in applied work. The lack of elegance here is that in simply verifying this formula you are essentially solving the equation. It's sort of cheap to simply verify something by brute force. It would be much nicer if one could see it directly. But I don't want you to take me seriously. I don't want to stop doing things because they don't adhere to certain principles of elegance. Boltzmann used to say, when he was criticized that his work was inelegant, that elegance should be left to shoemakers and tailors. Perhaps this is really true. But this proof is a little bit aesthetically dissatisfying, I would say.

This same proof goes also for a higher number of dimensions. Again it's simply a matter of writing the Laplace transform and verifying the same formula. There are certain disappointments in connection with this. Because really one learns comparatively little. At least I haven't learned anything really startling by doing it this way. The situation changes radically if you go to other differential equations, those of parabolic and elliptic type. When you study them from the point of view of stochastic problems from which they arose, then a considerable amount of new knowledge and a new approach results.

Let me first give you one more example of the great advantage which can accrue if one looks at the same thing from a different point of view. I will consider a very classical problem: The asymptotic behavior of eigenvalues of the Laplacian. You have some region Ω with a boundary Γ . These are in the plane, let us say. Everything is assumed to be suitably smooth so that classical analysis is applicable. The problem we are considering can then be stated:



$$\frac{1}{2}\Delta u + \lambda u = 0 \quad \text{in } \Omega \tag{47}$$

$$u = 0 \quad \text{on } \Gamma \tag{48}$$

You will recognize this as the vibrating membrane problem. I now consider the eigenvalues in their increasing order: $\lambda_1, \lambda_2, \lambda_3, \ldots$. And I denote by $A(\lambda)$ the number of eigenvalues less than λ. In the plane as λ goes to infinity,

$$A(\lambda) \sim \frac{|\Omega|}{2\pi}\lambda \tag{49}$$

The symbol $|\Omega|$ means the area of the region. In three dimensions it becomes the volume and the lambda gets raised to the power 2/3. The curious thing is that the constant depends only on the area, not on the shape at all.

There is a very heroic story connected with it. Let me tell you the story. This theorem, in three-dimensional space, was conjectured for the first time by the great H. A. Lorentz in 1908. This was during a meething in Goettingen devoted to the new quantum theory -- it wasn't quantum mechanics yet. There was, by that time, a new theory of specific heats which Debye proposed. It extended the older theory of Einstein and was one of the first great triumphs of the new quantum theory. By playing around with Debye's theory ou can deduce some properties of the heat content. Now it's perfectly natural hat the heat content should be proportional to the volume. And certainly it annot depend on the shape. So there was the conjecture. Three years later, n 1911, the late Herman Weyl proved this. In so doing he made a tremendously nteresting contribution to this branch of mathematics. It is in this connec-ion that he first introduced the famous variational characterization of the igenvalues, the so-called minimax characterization. From there on the method vas applied very successfully to different problems in many different ways. Veyl's original proof was not difficult, but not entirely "understandable." Although you understand the steps you still don't quite know what makes it tick.

I will now give you a proof which has the advantage that one understands very well how it works. In fact, it makes the theorem appear relatively superficial. The way it is stated now it has a certain appearance of depth. But we will look at it from a somewhat different point of view -- the point of view connected with probability and diffusion. What I'm going to give you will not be a complete proof, because there are several delicate points which have to be justified. This can be done and it has been done.

We will look at this theorem in a different way, in what I think is the proper way. We regard the equation (47) as arising not from the vibrating membrane or anything of that sort, but from the diffusion equation:

$$\frac{\partial P}{\partial t} = \frac{1}{2}\Delta P \qquad (50)$$

You know perfectly well how this goes. I will, of course, require that

$$P(x,y,t) = 0 \quad \text{on } \Gamma \qquad (51)$$

and, in addition, I want to make the following assumption:

$$P(x,y,t) \longrightarrow \delta(x-x_0)\delta(y-y_0) \text{ as } t \longrightarrow 0 \tag{52}$$

So, if you look on $P(x,y,t)$ as being the concentration of some diffusing stuff, it's initially all concentrated at (x_0, y_0).

If you take the equation (50) subject to the initial condition (52) and the boundary condition (51), then it is well known from entirely classical stuff how to write at least a formal solution. You simply make a separation of variables and find that you can write

$$P(x_0,y_0 \mid x,y,t) = \sum_{j=1}^{\infty} e^{-\lambda_j t} \varphi_j(x,y)\varphi_j(x_0,y_0) \tag{53}$$

The $\varphi_j(x,y)$ are the normalized eigenfunctions of the operator $\frac{1}{2}\Delta$; the λ_j are the eigenvalues, the same ones as before. This is one of the most standard results in classical mathematical physics.

Now let us try to interpret this $P(x_0,y_0 \mid x,y,t)$. The condition (51) means that Γ is an absorbing barrier -- any of this diffusing stuff which reaches it is eaten up. So it's perfectly normal that $P(x_0,y_0 \mid x,y,t)$, the concentration, is going to approach zero as you approach the boundary. Initially all the mass is at the point (x_0,y_0), but as time goes on it diffuses out and gets absorbed.

Now suppose that t is very small. Place yourself in the position of the diffusing stuff -- you are there together with all the other particles that are going to diffuse. Now you are going to move for a very short time.

In such a short time you don't know what will happen on the boundary, because you haven't had a chance to discover what horrible disasters are going to befall you. The knowledge of your fate, that you are going to be eaten at the boundary, has not yet reached you. The smaller t is, the less knowledge of this you possess. Consequently, for small t, a good approximation ought to be given by the solution of the same diffusion equation without regard to the boundary. Of course, it's an intuitive principle, this principle of not feeling the boundary. But it leads me to think that early in the game the solution will be that of the unrestricted diffusion problem:

$$\frac{1}{2\pi t} e^{-\frac{\left[(x-x_0)^2+(y-y_0)^2\right]}{2t}} \tag{54}$$

which is perfectly well known.

If we are so courageous to think that this is a good approximation for small t, then perhaps it also holds when $x = x_0$ and $y = y_0$. It then becomes $1/2\pi t$. But then that gets us, using (53),

$$\sum_j e^{-\lambda_j t} \varphi_j^2(x_0, y_0) \sim \frac{1}{2\pi t} \quad \text{as } t \to 0 \tag{55}$$

Now integrate both sides of this asymptotic equality over the region. Using the fact that the eigenfunctions were normalized -- the integral of $\varphi_j^2(x,y)$ is one -- you get:

$$\sum_j e^{-\lambda_j t} \sim \frac{|\Omega|}{2\pi t} \tag{56}$$

And now we are almost through. All we have to do is know a little bit of mathematics. Because I can write the sum as a Stieltjes integral, a very convenient way of writing it:

$$\sum_j e^{-\lambda_j t} = \int_0^\infty e^{-\lambda t}\, dA(\lambda) \tag{57}$$

$\frac{1}{t}$ can also be written as a similar integral, so you get from (56):

$$\int_0^\infty e^{-\lambda t}\, dA(\lambda) \sim \frac{|\Omega|}{2\pi} \int_0^\infty e^{-\lambda t}\, d\lambda \tag{58}$$

which is true as t goes to zero. Now you know about this function $A(\lambda)$ that it is non-decreasing. After all, the bigger λ is, the more λ_j's you include. Now there is a theorem that says that from such an asymptotic equality, for t going to zero, there results another asymptotic equality for λ going to infinity:

$$A(\lambda) \sim \frac{|\Omega|}{2\pi}\lambda \quad \text{as } \lambda \longrightarrow \infty \tag{59}$$

This theorem is known as the Hardy-Littlewood-Karamata Tauberian theorem. It produces the result (59) we were after.

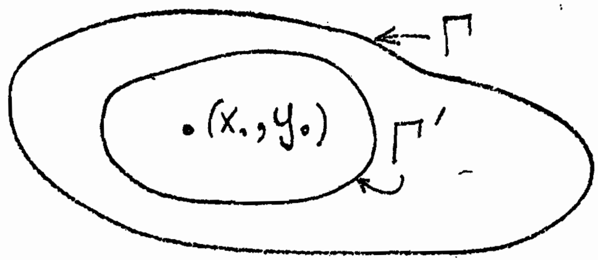
Now this proof has one tremendous advantage. (Actually, it isn't a proof yet because we have done a certain amount of skullduggery.) It is intuitively completely appealing. Moreover, it is unforgettable. The basic principle, this not feeling the boundary for a short time, is only visible if you look upon the problem from the point of view of the diffusion equation. From the wave equation viewpoint there is no such simple interpretation. This principle of not feeling the boundary has by now been exploited many more times in similar connections. Even here it can tell you something more. It tells you that it doesn't matter what boundary condition you put on. The same asymptotic behavior ought to be found for any homogeneous boundary condition. That, of course, is part of Weyl's theorem.

This is an example of the advantage accrued from looking at a mathematical equation in different physical contexts. If you look at it from one point of view it may be much more revealing than if you look at it from another. This is why one should try to formulate even familiar things in as many different ways as possible. You undoubtedly will learn something in the process. Just to finish up this story, let me show you that our principle not only illuminates the result, it also suggests the detailed proof. I claim that immediately I can have an inequality,

$$P(x_0,y_0|x,y,t) \leqslant \frac{1}{2\pi t} e^{-\frac{(x-x_0)^2+(y-y_0)^2}{2t}} \tag{60}$$

From the physics of the problem this is a yelling triviality. Because what does it mean? It means that the concentration when there is an absorbing barrier is less than when there is not.

To find an inequality which goes the other way is only a little bit more difficult. Let us draw around this point (x_0, y_0) another curve, call it Γ' . We make Γ' lie entirely within Γ :



Consider now the same problem with Γ replaced by Γ' . You put the absorbing barrier at Γ' . I will denote the solution by $P'(x_0, y_0 | x, y, t)$. Now I claim that P must be greater than or equal to P' :

$$P'(x_0, y_0 | x, y, t) \leqslant P(x_0, y_0 | x, y, t) \tag{61}$$

Because, after all, the stuff is going to be eaten earlier. In the first problem I still have the possibility of diffusing outside Γ' and then coming back in without having been eaten. These possibilities are denied to me if I have the boundary at Γ' . So P is bigger than P'. What is the obvious thing to do? To select the boundary Γ' so that you can solve for P' . Either a circle or a rectangle will do. So now you have inequalities going both ways, inequalities which certainly hold for all x and y. So you can put $x = x_0$ and $y = y_0$ to get.

$$P'(x, y | x, y, t) \leqslant \sum_j e^{-\lambda_j t} \varphi_j^2(x, y) \leqslant \frac{1}{2\pi t} \tag{62}$$

This inequality is true for every point (x, y) provided I surround it by a certain square. Now think of the whole region as being covered by little squares. At the center of each such square you have the inequality. Then you just integrate out. Of course, to finish this up, you have to know how the left hand side behaves as t goes to zero. But I know the solution explicitly for a square, so it can be simply verified. So this is a very brief sketch of essentially the full proof.

I wish to make one observation. Since one tries to sell the method, one may as well be honest. The reason the proof works so nicely and neatly is because you have all these inequalities which are properties of parabolic equations. One should distinguish between being able to see things clearly and being able to prove them. If you wanted to do the same problem for the biharmonic equation then you could still enunciate the principle of not feeling the boundary. You could still get some answer, but you would not be quite sure how to prove it. There are no corresponding inequalities for the biharmonic heat equation, for instance. In our diffusion problem $P(x_0, y_0 | x, y, t)$ remains always positive. The inequalities are quite obvious, and their proof is intuitively quite clear.

Now this is actually a very interesting question. What equations do and what equations don't have the properties that we need? If you have an elliptic operator Λ of second order then for the corresponding equation $\frac{\partial P}{\partial t} = \Lambda P$ will have all the necessary inequalities. For higher order equations I don't think anybody really knows what to do in this particular way.

But anyway, here is an example of the very classical theorem, done in all the textbooks, which can be treated by a probability approach. Our problem was one in differential equations, a purely mathematical problem. But that it came from a physical situation is not something to be sneezed at.

Because the knowledge of where it comes from gives you a way of approaching the problem. You also see that taking it as coming from one part of physics may be more useful, more illuminating, than taking it as coming from another part of physics. It's a real advantage, both to look at an equation from different points of view and to be aware of the physical interpretation. That is the reason why I believe that if any real breakthrough ever comes in non-linear differential equations, it will come only when a certain amount of physical knowledge will be amassed, so that some intuition will be developed.

The rest of the time I will devote to discussing problems arising from the simplest cases of Brownian motion. I was hoping, when I planned the lectures, to be able to do much more with Brownian motion. But it turns out that one always is too optimistic as to how much time one has. Consequently, I will have to confine myself to the so-called Brownian motion of a free particle. This is already familiar to you in the classical theory of Einstein and Smoluchowski. But I will look at it from a somewhat different point of view, a point of view you might call integration in function spaces.

Now how does the original theory of Einstein or Smoluchowski approach the simplest case of Brownian motion? Let's take a straight line and say to ourselves that we have a Brownian particle starting at $X = X_0$. I assume from the very beginning that, at best, I will be able to predict the probability density $P(X_0|X, t)$. If I multiply by dX then this is simply the probability of finding a particle between X and $X+dX$ at time t if I start from X_0. And now one makes an assumption of the past being independent of the future. Mathematically, it can be formulated as follows:

$$P(X_0|X,t) = \int_{-\infty}^{\infty} dy\, P(X_0|y,t')\, P(y|X,t-t') \qquad (63)$$

To see this, you notice that $P(X_0|X,t)$ is the probability of coming from X_0 to X in time t. But at any intermediate time, say t', I have to be somewhere. Let's say that at time t' I have the position y. So I have made a transition from X_0 to y in the time t'. But then I must make a transition from y to X in the time $t-t'$. What is the probability of this? Well, since the past is independent of the future, the probability is the product of two probabilities. In fact it is just the integrand in equation (63). But since y could be anywhere, since I don't know what it is, I have to integrate over all the possibilities.

This is a very famous equation. It is known under a different name, depending on whether you are a mathematician or a physicist. In the mathematical literature it became known as the Chapman-Kolmogoroff equation, although I do not quite know why Chapman was attached to it. Among the physicists, it is known as the Smoluchowski equation, because he considered it in great detail. It is interesting to remark that a similar equation can also be written in quantum mechanics. Except that it is not anymore the probability density that is involved. Instead it's a complex-valued function known as the probability amplitude. One of the reasons this is so is because in quantum mechanics you cannot make the same argument. You cannot say that if you are at one place at time t and another at time t' then you must have been somewhere in the meantime. This is wrong for a very interesting reason. In quantum mechanics one must always look upon things operationally. I must be able to perform an experiment to find the intermediate position. But if I do that it will change where I end up. Any experiment you can think of will disturb the particle in some way.

Now if you don't make any further assumptions, there are slews of solutions of this equation (63). This is even true in the spatially homogeneous case. This is the case where $P(X_0 \mid X, t)$ depends only on the difference $X - X_0$, not on X and X_0 separately. I can go further even than that. I can consider the symmetric case where $P(X_0 \mid X, t)$ depends only on the absolute value $|X - X_0|$:

$$P(|X-X_0|, t) = \int_{-\infty}^{\infty} P(|X-y|, t')\, P(|y-X_0|, t-t')\, dy \tag{64}$$

The probabilities of going in one direction or the other are completely equal. Even then there are a tremendous number of solutions of this equation. For instance, a somewhat unusual solution, not perhaps known to you all, would be

$$\frac{t}{\pi}\, \frac{1}{t^2 + (X - X_0)^2} \tag{65}$$

If you substitute this into the equation you will verify that it is a solution. And there are many, many others.

But if you assume, in addition to the symmetry, that the second moment is finite, then there is only one solution. Or, rather, one form of a solution. It is namely the Gaussian distribution:

$$\frac{1}{\sqrt{2\pi t}\,\sigma} \cdot e^{-\frac{(X-X_0)^2}{2\sigma^2 t}} \tag{66}$$

σ^2 is the second moment, the mean square displacement:

$$\int_{-\infty}^{\infty}(x-x_0)^2 P(x-x_0,t)\,dx < \infty \qquad (67)$$

It is really the finiteness of this moment that forces the solution (65). You recognize it as the solution of the diffusion equation. I will show you formally how equation (64) can be reduced back to the diffusion equation when the second moment (66) is finite.