Applications of Fourier transform

A. Eremenko

April 16, 2021

1. Heat equation on the line.

$$u_t = k u_{xx}, \quad u(x,0) = f(x). \tag{1}$$

Let us assume that f and $x \mapsto u(x,t)$ tend to 0 for $x \to \pm\infty$ sufficiently fast so that we can take Fourier transforms in the variable x. Then we obtain

$$\hat{u}_t = -ks^2 \hat{u}, \quad \hat{u}(s,0) = \hat{f}(s).$$

(Differentiation with respect to t can be performed under the integral sign). This is the initial value problem for a first order linear ODE whose solution is

$$u(s,t) = \hat{f}(s) e^{-ks^2 t}.$$

Since the inverse Fourier transform of a product is a convolution, we obtain the solution in the form

$$u(x,t) = K(x,t) \star f(x),$$

where $K(x,t)$ is the inverse Fourier transform of $e^{-ks^2 t}$. Using Example 2 (formula (5)) from the previous lecture "Fourier Transform" with $a = 1/(2kt)$, we obtain

$$K(x,t) = \frac{1}{2\sqrt{\pi k t}} e^{-\frac{x^2}{4kt}}. \tag{2}$$

This is called the *heat kernel*. So the solution of our problem is

$$u(x,t) = \frac{1}{2\sqrt{\pi k t}} \int_{-\infty}^{\infty} e^{-\frac{(x-y)^2}{4kt}} f(y) dy = \frac{1}{2\sqrt{\pi k t}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{4kt}} f(x-y) dy.$$

Notice that this solution makes sense under a very mild restrictions on f: because of the factor e^{-y^2}, the integral will be convergent for functions of sufficiently slow growth, even if Fourier transform is not defined for them. On the other hand, once this formula is written, it is easy to verify that it indeed solves our problem.

Indeed, the heat kernel itself does satisfy the heat equation, which is verified by differentiation (do this!), so its convolution with any function also satisfies heat equation.

Now the heat kernel has these three properties:

(i) $K(x,t) > 0$

(ii) $\int_{-\infty}^{\infty} K(x,t)dx = 1$ for all $t > 0$, (check this!)

(iii) For every $\epsilon > 0$ we have $K(x,t) \to 0$ as $t \to 0$, *uniformly* for $|x| \geq \epsilon$. Check this!

Any function $K(x,t)$ with these properties is called a *positive kernel*, and it is easy to see that for every continuous $f \in L^1$ we have

$$(K(.,t) \star f)(x) \to f(x), \quad t \to 0.$$

So we obtained a solution of the initial value problem for the heat equation on the line for a large class of initial conditions.

Actually this solution is unique under a mild restriction of the growth at infinity, namely that $u(x,t)$ as a function of x has slower growth than e^{x^2}.

2. Heat equation on half-line with zero boundary condition.

Problem. Solve the heat equation (1) on the half-line $x > 0$ with the boundary condition

$$u(0,t) = 0, \quad t > 0, \tag{3}$$

and constant initial condition

$$u(x,0) = f(x). \tag{4}$$

Then $f(x) = \text{const}$, this is problem occurs as the flat Earth approximation of the problem of cooling the Earth, solved by Kelvin: x is the depth, and the temperature is supposed to depend on depth and time only. The surface temperature is constant and we may take it as zero. (The constant temperature is a crude approximation, of course).

The question is how to reduce this problem to a problem on the whole line which we just solved. The answer is *consider the odd extension of the initial condition!*.

The heat kernel $K(x,.)$ is even (as a function of x), and a convolution of an even function and odd function is odd: if we suppose that K is even and f is odd, then the convolution

$$u(x) := \int K(x-y)f(y)dy$$

will be odd. Indeed

$$\begin{aligned} u(-x) &= \int K(-x-y)f(y)dy = \int K(x+y)f(y)dy \\ &= \int K(x-u)f(-u)du = -\int K(x-u)f(u)du = -u(x). \end{aligned}$$

Since every odd function is zero at zero, we can replace the boundary condition by $u(x,0) = \tilde{f}(x)$ where $\tilde{f}$ is the odd extension of f, and apply the solution for the whole real line obtained in the previous section. For example, when f is constant, say $f(x) = T$, then $u + T$ will have boundary values 0 for negative x and $2T$ for positive x, and the formula from section 1 gives:

$$u(x,t) = \frac{T}{\sqrt{\pi kt}} \int_0^\infty e^{-\frac{(x-y)^2}{4kt}} dy - T. \tag{5}$$

So we obtained a solution.

To determine the age of Earth, we compute the gradient $u_x(0,t)$, and since this gradient can be measured, we can determine the age t. Differentiating and substituting $x = 0$, we obtain

$$u_x(0,t) = \frac{T}{\sqrt{\pi kt}}.$$

See the handout "Age of the Earth" for a further discussion.

3. Heat equation on a half-line with arbitrary boundary condition.

In the previous section we solved heat equation on the half-line $x > 0$ with initial condition $u(x,0) = 1, \ x > 0$ and the boundary condition $u(0,t) =$

$0,\ t > 0$. It is given by formula (5) which we write as

$$u(x,t) = 2\int_0^\infty K(x-y,t)dy - 1,$$

where $K(x,t)$ is the heat kernel given by formula (2).

Now it is easy to obtain a solution of the heat equation with initial and boundary conditions $v(x,0) = 0, x > 0;\ v(0,t) = 1$. it is simply

$$v(x,t) = 1 - u(x,t) = 2 - 2\int_0^\infty K(x-y,t)dt.$$

Now we make an important remark: for every $x \neq 0$, the function $t \mapsto K(x,t)$ has limit 0 as $t \to 0$, and moreover, all partial derivatives with respect to t have limit 0, when $t \to 0$. See the discussion of such functions in "Fourier transform", in the Example in section 3. This means that we can extend the function $K(x,t)$ to $t < 0$ by defining $K(x,t) = 0$ for $t < 0$, and the extended function will satisfy the heat equation for all $x > 0$ and all t. So our function $v(x,t)$ also satisfies heat equation for $x > 0$ and all t, and it equals 0 for $t \leq 0$.

Let us denote by $H_s(t)$ the *Heaviside function,*

$$H(t) = \begin{cases} 1, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

We can solve now the heat equation for $x > 0$, $-\infty < t < \infty$ with the boundary condition $u(0,t) = H(t-s)$:

$$v_s(x,t) = 2 - 2\int_0^\infty K(x-y,t-s)dy, \quad v_s(0,t) = H(t-s), \tag{6}$$

where we use the convention that $K(x,t) = 0$ for $t < 0$. Once we can solve the problem whose boundary condition is Heaviside function, we can also solve it with any linear combination of Heaviside functions. And every reasonable function of t can be approximated by a linear combination of Heaviside functions: suppose for example that $f(t)$ is continuous and has bounded support which is contained in the positive ray. Then the linear combinations

$$\sum_{j=0}^{n} \left(f(s_{j+1}) - f(s_j)\right) H(t - s_j)$$

approximate f uniformly when the partition $s_0 < s_1 < \ldots < s_{n+1}$ of the support of f is fine enough, and $f(s_0) = f(s_{n+1}) = 0$. We can write this as

$$\sum_{j=0}^{n} \frac{f(s_{j+1}) - f(s_j)}{s_{j+1} - s_j} H(t - s_j)(s_{j+1} - s_j),$$

which is the integral sum of an integral, and we obtain the representation

$$f(s) = \int_0^\infty H(t - s) f'(s) ds,$$

in the form of convolution with the Heaviside function.

Since we solved the heat equation with boundary condition in the form of a Heaviside function, we can now use the superposition principle, and solve it with arbitrary boundary function:

$$u(x, t) = \int_0^\infty v_s(x, t) f'(s) ds,$$

where v_s is the solution (6). Integrating this by parts, we obtain a convolution formula representing the solution in terms of boundary condition $f(t)$:

$$u(x, t) = -\int_0^\infty f(s) \frac{d}{ds} v_s(x, t) ds.$$

It remains to compute the derivative of the explicit function v_s. The computation is simplified by the fact that it satisfies the heat equation. By differentiating (6) and using the heat equation $K_t = kK_{xx}$ we obtain:

$$\begin{aligned} \frac{d}{ds} v_s(x, t) &= -2 \frac{d}{ds} \int_0^\infty K(x - y, t - s) dy = 2 \int_0^\infty K_t(x - y, t - s) dy \\ &= 2k \int_0^\infty K_{xx}(x - y, t - s) dy = 2k K_x(x, t - s), \end{aligned}$$

and simple explicit differentiation of the heat kernel shows that

$$L(x, t) := 2k K_x(x, t) = -\frac{x}{2\sqrt{\pi k}} t^{-3/2} e^{-\frac{x^2}{4kt}},$$

and we obtain the final result

$$u(x, t) = (L \star f)(x, t) = \frac{x}{2\sqrt{\pi k}} \int_0^t (t - s)^{-3/2} \exp\left(-\frac{x^2}{4k(t - s)}\right) f(s) ds. \quad (7)$$

Notice that integral from 0 to ∞ is actually from 0 to t since $K(x, t-s) = 0$ for $s > t$. Function $L(x,t)$ is called the *lateral heat kernel*.

Once the formula for a solution is found, one can prove it without any mentioning of its derivation:

Exercise 1. Show that:

a) The kernel

$$L(x,t) = \frac{x}{2\sqrt{\pi k}} t^{-3/2} e^{-\frac{x^2}{2kt}},$$

as a function of t is a positive kernel in the sense defined in Section 1, and

b) It satisfies the heat equation $L_t = kL_{xx}$ for $x > 0,\ t > 0$ and is infinitely differentiable when extended by 0 for $t < 0$.

c) Conclude that (7) is a solution of the heat equation with the boundary function $u(x,t) = f(t)$ and initial condition $u(x,0) = 0$.

See the handout "Transatlantic Cable" for a discussion of this solution.

4. Laplace equation in a half-plane.

Consider the Dirichlet problem

$$u_{xx} + u_{yy} = 0, \quad -\infty < x < \infty, \quad y > 0,$$

with the boundary condition

$$u(x,0) = f(x),$$

and some condition at ∞, for example that u tends to zero sufficiently fast so that Fourier transform can be applied.

Then Fourier transform with respect to x gives

$$-s^2 \hat{u}(s,y) + \hat{u}_{yy} = 0, \quad \hat{u}(s,0) = \hat{f}(s).$$

This is an second order linear ODE whose general solution is

$$C_1(s)e^{sy} + C_2(s)e^{-sy}.$$

One boundary condition $C_1(s) + C_2(s) = \hat{f}(s)$ is not sufficient, but the simplest way to satisfy it[1] is to take $C_1(s) = 0$ for $s > 0$ and $C_2(s) = 0$ for $s < 0$,

[1] Of course, this is not rigorous. But once this correct formula is guessed, it can be proved without any reference to Fourier transform. Uniqueness of solution is a separate problem. In fact it has unique *bounded* solution, but we do not prove this result here.

This gives

$$\hat{u}(s,y) = \hat{f}(s)e^{-y|s|}.$$

Since this is a product, the original u must be a convolution of f with the inverse transform of $e^{-y|s|}$. The last transform we know from Example 3 of the previous lecture. It is

$$P(x,y) = \frac{y}{\pi(x^2+y^2)}.$$

This is called the *Poisson kernel* for the upper half-plane. So we obtain

$$u(x,y) = (P \star f)(x,y) = \frac{y}{\pi}\int_{-\infty}^{\infty} \frac{f(t)}{(x-t)^2+y^2}dt,$$

which is called the *Poisson formula* for the upper half-plane.

This formula has an appealing geometric interpretation. Let us set first $f(x) = 1,\ x < 0$ and $f(x) = 0, x > 0$. Then the Poisson integral can be evaluated:

$$\frac{y}{\pi}\int_{-\infty}^{0} \frac{dt}{(x-t)^2+y^2} = \frac{1}{\pi}\arctan\frac{y}{x},$$

in other words, $u(x,y)$ in this case is just the polar angle of the point (x,y), divided by π. Check that the boundary values for $y=0$ are correct!

More generally, if $f(x) = 1,\ x \in (a,b)$ and 0 otherwise, we conclude that $u(x,y)$ is the angle under which the interval (a,b) is seen from the point (x,y), divided by π.

5. Wave equation on the whole line.

$$u_{tt} = c^2 u_{xx},$$

and let us take the initial conditions

$$u(x,0) = f(x), \quad u_t(x,0) = g(x), \quad -\infty < x < \infty.$$

Applying Fourier's transform, we obtain

$$\hat{u}_{tt} = -c^2 s^2 \hat{u}, \quad \hat{u}(s,0) = \hat{f}(s), \quad \hat{u}_t(s,0) = \hat{g}(s).$$

The general solution is

$$C_1(s)\cos(cst) + C_2(s)\sin(cst),$$

and the first initial conditions imply $C_1(s) = \hat{f}s,\ C_2(s) = \hat{g}(s)/(cs)$. So we solved the problem in the sense that we found the Fourier transform of the solution

$$\hat{u}(s,t) = \hat{f}(s)\cos(cst) + \hat{g}(s)\frac{\sin(cst)}{cs}. \tag{8}$$

To obtain an explicit formula of convolution type, we would like to know inverse transforms of $\cos(cst)$ and $\sin(cst)/(cs)$. There is no problem with the second one, this is essentially Example 1 in "Fourier Transform", and it is

$$t\chi_{ct}(x) = \begin{cases} (2c)^{-1}, & |x| < ct, \\ 0 & \text{otherwise.} \end{cases}$$

So we obtain the solution

$$u(x,t) = \frac{1}{2c}\int_{x-ct}^{x+ct} g(u)du,$$

which coincides with the sccond part of d'Alembert's formula for zero initial position.

Now what about the boundary condition where the shape of the srting is prescribed? In following previous steps we run into a trouble that the inversion formula cannot be applied to $\cos(cst)$, since the integral diverges. So one needs a slightly different argument.

The solution with zero initial velocity $g = 0$ can be written as the inverse Fourier transform

$$u(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(s)\cos(cst)e^{isx}ds, \tag{9}$$

and the integral is convergent if for example $f \in \mathscr{S}$, the Schwartz space; since in this case also $\hat{f} \in \mathscr{S}$. Now we write

$$\cos(cst) = \frac{1}{2}\left(e^{icst} + e^{-icst}\right).$$

Plugging this to (9) we obtain one half of the sum of two integrals, the first one is the inverse transform of $\hat{f}(s)e^{icst}$ and the second in the inverse transform of $\hat{f}(s)e^{-icst}$. Using Property 2 of Fourier transform (multiplication of transform corresponds to the shift of the original), we recover the first part of d'Alambert's formula:

$$u(x,t) = \left(f(x-ct) + f(x+ct)\right)/2.$$

Notice that this is not a convolution of f with any function, though it 'looks like one".

This suggest the idea that perhaps all notions of Fourier transform and convolution must be generalized. Indeed this is so, but to do this one needs to generalize the notion of function itself! So that in particular such formulas as the first part of d'Alembert formula can be written as convolutions with "something".

6. Signal processing.

The most conspicuous thing in the real world which is modeled by the real line is time. In engineering, especially electric engineering, one considers "signals" which can be thought as functions of time. Many functions of time can be represented by Fourier inversion formula

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(s) e^{isx} dx.$$

This means that f is a superposition, or a linear combination of simple harmonic oscillations $x \mapsto e^{isx}$ of angular frequency s. The set of frequencies which are involved in the representation of f, in other words the *support* of $\hat{f}$ is called the spectrum of f. (Some authors prefer to call $\hat{f}$ itself the spectrum, but we will not do this.) It is important that f and $\hat{f}$ completely determine each other, so engineers are talking about "time representation" $f(t)$ and "frequency representation" $\hat{f}$ of the same "signal".

A "system" is a device which transforms signals: $g = A(f)$, where g and f are functions of time, and A is an operator, some device which transforms signals f (inputs) into signals g (outputs). Common properties of the operator A are:

a) linearity, and

b) time invariance, that is if $f = A(g)$ then $f(t-a) = A(g(t-a))$. This means that is if input is shifted in time, then the output is shifted in time by the same amount.

Under some mild restrictions on the space of functions, every linear time stationary transformation is performed by convolution with some function h which characterizes the device, that is

$$g = A(f) = h \star f.$$

Applying the Fourier transform we obtain

$$\hat{g} = H\hat{f},$$

where $H = \hat{h}$ is called the transfer function, so in the "frequency domain" our device works by just multiplying the input on H.

Of many mathematical problems arising with this model, we address only one: how can a signal be discretized. Since digital processing works with discrete signals (sequences) this is an important question for communication technology.

It turns out that signals with *bounded spectrum* can be sampled at certain sequences of time, and this sampling does not lead to the loss of information: the signal can be recovered from these samplings. The precise statement is the following:

Sampling Theorem. *Let f be a function in $L^2(\mathbf{R})$, whose Fourier transform is supported on an interval $|s| \leq \omega$. Then the sequence $(f(\pi n/\omega))_{n=-\infty}^{\infty}$ completely determines f, and in fact f can be recovered by the formula*

$$f(t) = \sum_{-\infty}^{\infty} f\left(\frac{\pi n}{\omega}\right) \frac{\sin(\omega t - n\pi)}{\omega t - n\pi}. \tag{10}$$

The series converges in L^2.

Proof. Since $\hat{f}$ is supported on $[-\omega, \omega]$, we can expand it into a Fourier series

$$\hat{f}(s) = \sum_{-\infty}^{\infty} c_n e^{i\pi ns/\omega}, \quad |s| \leq \omega, \tag{11}$$

where the Fourier coefficients are

$$c_n = \frac{1}{2\omega}\int_{-\omega}^{\omega} \hat{f}(s) e^{-i\pi ns/\omega} ds = \frac{1}{2\omega}\int_{-\infty}^{\infty} \hat{f}(s) e^{-i\pi ns/\omega} = \frac{\pi}{\omega} f(-\pi n/\omega), \tag{12}$$

by the Fourier inversion formula. On the other hand, again by the inversion formula, and using (11), (12):

$$\begin{aligned} f(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(s) e^{ist} dt = \frac{1}{2\omega}\int_{-\omega}^{\omega} \sum_{-\infty}^{\infty} f(-\pi n/\omega) e^{-i\pi ns/\omega} e^{ist/\omega} ds \\ &= \sum_{-\infty}^{\infty} f(-\pi n/\omega) \frac{1}{2\omega}\int_{-\omega}^{\omega} e^{is(\omega t + \pi n)/\omega} ds, \end{aligned}$$

and it remains to evaluate the elementary integral

$$\frac{1}{2\omega}\int_{-\omega}^{\omega} e^{is(\omega t+\pi n)/\omega} ds = \frac{\sin(\omega t+\pi n)}{\omega t+\pi n}.$$

To obtain (10) one switches $n \mapsto -n$.

7. Probability.

This is one of the main applications of Fourier transform. I will not discuss modern rigorous foundations of probability, which are actually irrelevant for our purposes.

A *random variable* is a numerical quantity X which “depends on chance”. For example, the number of deaths in the US on a given day. Or the number of moleculas in a volume of a gas whose speed belongs to a given interval.

A real random variable is completely characterized by its *distribution function*

$$F_X(t) = \mathbf{P}(X \leq t), \quad -\infty < t < \infty.$$

Here $\mathbf{P}(.)$ is the probability of an event related to X. Probability is a number in $[0,1]$. So

$$\mathbf{P}(a < X \leq b) = F_X(b) - F_X(a).$$

The distribution function is increasing,

$$\lim_{t\to-\infty} F_X(t) = 0 \quad \text{and} \quad \lim_{t\to+\infty} F_X(t) = 1.$$

It is also continuous from the right,

$$\lim_{t\to t_0+} F_X(t) = F_X(t_0).$$

These three properties characterize distribution functions. If $F'(t) = f_X(t)$ exists and is integrable, then f_X is called the *density* of the distribution. In this case the probabilities of events related to X are given by the formula

$$\mathbf{P}(a < X < b) = \int_a^b f_X(t)dt, \tag{13}$$

and in particular $\mathbf{P}(X = t) = 0$ for every t. A random variable whose distribution has a density is called *continuous*. The density must satisfy

$$f_X(t) \geq 0 \quad \text{and} \quad \int_{-\infty}^{\infty} f_X(t)dt = 1,$$

and any function with these properties is a density of some continuous random variable.

On the other hand, if $p_0 = \mathbf{P}(X = t_0) > 0$, then F_X has a jump at t_0 of magnitude p_0. So if X takes only a discrete sequence of values $\ldots < t_k < t_{k+1} < \ldots$ with probabilities p_k, then

$$F_X(t) = \sum_{k: t_k \leq t} p_k. \tag{14}$$

Such distributions are called *discrete*. It is convenient to unite these two cases and write

$$\mathbf{P}(a < X \leq b) = \int_a^b dF_X(t),$$

which gives (13) in the continuous case, and (14) in the discrete case. This is called the Stieltjes integral. In general, if g is a continuous function, and F is increasing and continuous from the right, then the integral

$$\int g(x) dF(x)$$

is defined as the limit of integral sums

$$\sum g(x_k)(F(x_{k+1}) - F(x_k)),$$

when partitions $\ldots < x_k < x_{k+1} < \ldots$ become finer and finer. So if F is continuously differentiable, we obtain in the limit

$$\int g(x) F'(x) dx,$$

and if F is discrete, we obtain the sum

$$\sum g(x_k)(F(x_k + 0) - F(x_k - 0)).$$

This permits a uniform treatement of discrete and continuous case, and permits to consider mixed cases when the distribution is neither continuous nor discrete.

The quantity

$$EX := \int_{-\infty}^{\infty} t\, dF_X(t) = \int_{-\infty}^{\infty} t f_X(t) dt$$

is called the *expectation* or the average of the random variable X. You may think of it in terms of a gain in a game of chance: suppose your gain is X. In the discrete case, you gain t_k with probability p_k. Then in a long series of games you expect the average gain given by the formula

$$\sum_k t_k p_k = \int_{-\infty}^{\infty} t\, dF_X(t).$$

Expectation of a sum of two random variables is the sum of their expectations:

$$E(X+Y) = E(X) + E(Y).$$

If g is a continuous function, and X is a random variable, we can consider the new random variable $g(X)$ and its expectation will be

$$E(g(X)) = \int_{-\infty}^{\infty} g(t) dF_X(t).$$

Suppose now that we have two random variables X and Y. They are called *independent*[2] if the joint probabilities of events are products:

$$P(X \leq x, Y \leq y) = \mathbf{P}(X \leq y)\mathbf{P}(Y \leq Y) = F_X(x)F_Y(y),$$

for all x, y. Let us find the distribution function of the *sum of two independent random variables*. The event $X + Y \leq t$ occurs when $X \leq t - y$ *and* $Y \leq y$ for some y, so

$$F_{X+Y}(t) = \mathbf{P}(X+Y \leq t) = \int_{-\infty}^{\infty} \mathbf{P}(X \leq t-y)\mathbf{P}(Y \leq y) dy = \int_{-\infty}^{\infty} F_X(t-y)\, dF_Y(y),$$

and this is called the "convolution of distribution functions". So *the distribution function of a sum of independent random variables is the convolution of their distribution functions.* If the densities exist, we have a similar rule for the densities:

$$f_{X+Y} = f_X \star f_Y,$$

the density of the sum of independent random variables is the convolution of their densities. Fourier transform of the distribution function

$$\phi_X(s) = E(e^{-isX}) = \int_{-\infty}^{\infty} e^{-isx} dF_X(x) = \int_{-\infty}^{\infty} e^{-isx} f_X(x) dx$$

[2] This is just a mathematical definition. We skip the complicated discussion about real world phenomena which this definition reflects.

is called the *characteristic function* of the random variable X; it always exists (why?) and completely characterizes the random variable. From the fact that Fourier transform sends convolutions to products, we conclude that *the characteristic function of the sum of two independent random variables is the product of their characteristic functions.*

Example 1. Improper random variable: $\mathbf{P}(X = a) = 1$ for some real a. Then $E(X) = a$ and $\phi_X(s) = e^{-ias}$.

Example 2. Bernoulli's random variable: $\mathbf{P}(X = -1) = p$ and $\mathbf{P}(X = 1) = q$ where $p, q > 0,\ p+q = 1$. This describes the simplest game of chance: coin tossing; it is falls head you win 1, if it falls tail you loose 1. (The coin may by unfair: the probabilities are not necessarily equal). We have $E(X) = q - p$, $\phi_X(s) = pe^{is} + qe^{-is}$. When $p = q = 1/2$ (fair coin) we have $\phi_X(s) = \cos s$.

Example 3. Normal distribution with parameters a and $\sigma > 0$ is a continuous distribution with density

$$f_X(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right).$$

Exercises 2. a) Check that this is indeed a density, that is a positive function whose integral is 1. b) Show that $E(X) = a$. c) Check that it coincides with the *heat kernel* when $a = 0$ and $\sigma = \sqrt{2kt}$. When $(a, \sigma) = (0, 1)$ it is called the *standard normal distribution*, or the Gauss distribution.

An improper random variable can be thought as a non-random variable: it takes a definite value with probability 1. All other random variables have some spread about their average. A convenient measure of this spread is called *variance*. It is defined by the formula

$$\sigma_X^2 = E(X - E(X))^2 = E(X^2) - (E(X))^2.$$

The positive square root σ of the variance σ^2 is called the *standard deviation*.

Exercise 3. Check that the standard deviation of the normal distribution is σ.

Exercise 4. Check that for every random variable $\phi_X(0) = 1$,

$$E(X) = -i\phi'_X(0) \quad \text{and} \quad E(X^2) = -\phi''(0).$$

Exercise 5. Let X be a random variable, $a > 0$ and b is real. Derive from the scaling properties of Fourier transform the property

$$\phi_{aX+b}(s) = e^{-isb}\phi_X(as). \tag{15}$$

Now we state and prove a special case of the fundamental result of Probability theory:

Central Limit Theorem. *Let $X_1, X_2, \ldots$ be a sequence of independent random variables with the same distribution function F_X for which expectation $a = E(X_j)$ and variance $\sigma = \sigma_{X_j}$. Then the distribution function of*

$$\frac{\sum_{n=1}^{n} X_n - na}{\sigma\sqrt{n}} \tag{16}$$

tends to the standard normal distribution with density

$$\frac{1}{\sqrt{2\pi}}e^{-x^2/2}.$$

This theorem explains why normal distribution is ubiquotous. In popular literature it is sometimes called a "bell curve" because of the shape of the graph of its density.

Proof. Random variables $Y_n = (X_n - a)/\sigma$ have zero average and variance 1, and the same characteristic function, ϕ, with

$$\phi(0) = 1, \quad \phi'(0) = 0, \quad \phi''(0) = -1, \tag{17}$$

in view of Exercises 4 and 5. So the characteristic function of (16) is

$$\phi^n(s/\sqrt{n}) \tag{18}$$

since the characteristic function of the sum of independent random variables is he product of their characteristic fuunctions, and by the scaling property in Exercise 5. Let $g = \log\phi$. Then

$$g' = \frac{\phi'}{\phi} \quad \text{and} \quad g'' = \frac{\phi''\phi - (\phi')^2}{\phi^2}.$$

Now we take logarithm of (18) and apply Taylor's formula at 0:

$$n \log \phi(s/\sqrt{n}) = n \left(\log \phi(0) + \frac{\phi'(0)}{\phi(0)} \frac{s}{\sqrt{n}} + \frac{1}{2} \frac{\phi''(0)\phi(0) - (\phi'(0))^2}{\phi^2(0)} \frac{s^2}{n} + \dots \right).$$

Using the values (17) we obtain

$$n \log \phi(s/\sqrt{n}) = -s^2/2 + \dots,$$

where the omitted terms tend to zero as $n \to \infty$. Thus the limit is $-s^2/2$, so the characteristic function of (16) tends to the characteristic function of the stardard normal distribution. To conclude the proof, one needs a continuity property of characteristic functions: that convergence of characteristic functions corresponds to the convergence of distributions. This part of the proof is not difficult, but is not discussed here.

I finish this section with a citation of Henri Poincaré:

"There must be something mysterious about the normal law since mathematicians think that this is a law of nature whereas physicists are convinced that this is a mathematical theorem."

This observation applies to many theorems of probability theory.