TMA4170: Corona Week 3

Fourier transform applications

Miłosz Krupski

Based on
– P. Billingsley *Probability and Measure, Third Edition*, Wiley 1995,
– M. A. Pinsky, *Introduction toFourier Analysisand Wavelets*, AMS, 2009,
– E. M. Stein & R. Shakarchi, *Fourier Analysis, An Introduction*, Princetons 2003,
– J. Dziubański's lecture notes

1. Central limit theorem

1.1. Preliminaries.

Definition 1.1. Suppose $(\mathbb{R}, \mathrm{Bor}(\mathbb{R}), \mu)$ is a measure space such that $\mu(\mathbb{R}) = 1$ (we say that μ is a probability measure). Let

$$\phi_\mu(\xi) = \int_{\mathbb{R}} e^{i\xi x}\,\mu(dx) = \int_{\mathbb{R}} \cos(\xi x)\,\mu(dx) + i\int_{\mathbb{R}} \sin(\xi x)\,\mu(dx).$$

We call ϕ_μ the **characteristic function** of the measure μ.

Notice that the characteristic function is nothing else but the inverse Fourier transform, up to scaling by 2π – and it so happens in probability theory, that it would most often be an irrelevant constant.

More precisely, recall that for every non-negative function $f \in L^1(\mathbb{R}, d\lambda)$ we may define a measure μ by

$$\mu(A) = \int_A f\,d\lambda,$$

and then (up to scaling)

$$\phi_\mu(\xi) = \int_{\mathbb{R}} e^{i\xi x}\,\mu(dx) = \int_{\mathbb{R}} e^{i\xi x} f(x)\,dx = \check{f}(\xi),$$

hence the definition is indeed just an extension of the inverse Fourier transform we are familiar with.

Remark 1.2. *Notice that since μ is a finite measure, we have*

$$\int_{\mathbb{R}} |e^{ix\xi}|\,\mu(dx) = \mu(\mathbb{R}) = 1,$$

hence the integral describing ϕ_μ is always well-defined. In fact, we can show that it is a bounded continuous function (see Problem 1). However, for a general measure, the Riemann-Lebesgue lemma is no longer true (what is the characteristic function of δ_0?).

For general finite Borel measures μ_1, μ_2 on $\mathbb{R}$ we may also define the operation of convolution, the result of which is again a finite Borel measure

$$(\mu_1 * \mu_2)(A) = \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{1}_A(x+y)\,\mu_1(dx)\,\mu_2(dy).$$

Notice that if both measures are defined by integrable functions, i.e. $\mu_1(A) = \int_A f(x)\,dx$ and $\mu_2(A) = \int_A g(x)\,dx$ then

$$\begin{aligned}(\mu_1 * \mu_2)(A) &= \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{1}_A(x+y) f(x) g(y)\,dx\,dy \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{1}_A(x) f(x-y) g(y)\,dy\,dx = \int_A\int_{\mathbb{R}} f(x-y)g(y)\,dy\,dx \\ &= \int_A (f * g)(x)\,dx.\end{aligned}$$

This means that the measure $\mu_1 * \mu_2$ is defined by the integrable funtion $f * g$, as we would expect.

If both μ_1 and μ_2 are probability measures then $\mu_1 * \mu_2$ is also a probability measure.

Just as well, some of the fundamental properties of the Fourier transform are still valid for measures, which we may formulate using the notion of characteristic functions. The following proposition is one of them.

PROPOSITION 1.3. *If ϕ_{μ_1} and ϕ_{μ_2} are the characteristic functions of probability measures μ_1 and μ_2 then $\phi_{\mu_1 * \mu_2} = \phi_{\mu_1} \cdot \phi_{\mu_2}$.*

1.2. Inversion and the Uniqueness Theorem. A characteristic function ϕ uniquely determines the measure μ it comes from. This observation builds on the discussion about inversion of the Fourier transform from the previous set of notes (most importantly, recall that we cannot use the inverse transform directly).

THEOREM 1.4. *If the probability Borel measure μ on $\mathbb{R}$ has characteristic function ϕ, and if $\mu(\{a\}) = \mu(\{b\}) = 0$, then*

$$\mu\big([a,b)\big) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-i\xi a} - e^{-i\xi b}}{i\xi} \phi(\xi)\, d\xi. \tag{1}$$

Distinct measures cannot have the same characteristic function.

PROOF. Assume the inversion formula (1) holds. Let μ and ν share the same characteristic function. Notice there may only be a countable number of points such that such that $\mu(\{a\}) \neq 0$ or $\nu(\{a\}) \neq 0$. For any interval which doesn't have one of its ends at one of such points we have $\mu[a,b) = \nu[a,b)$. Such intervals form a ring generating $\mathrm{Bor}(\mathbb{R})$, hence $\mu = \nu$.

Denote by I_T the quantity inside the limit in (1). By changing the order of integrals[1] we get

$$I_T = \frac{1}{2\pi} \int_{\mathbb{R}} \int_{-T}^{T} \frac{e^{i\xi(x-a)} - e^{i\xi(x-b)}}{i\xi}\, dt\, \mu(dx).$$

Let

$$S(T) = \int_0^T \frac{\sin t}{t}\, dt.$$

We can show that

$$\lim_{T\to\infty} S(T) = \frac{\pi}{2}.$$

[1]It is outside the scope of this course, but one always needs to be extremely careful with changing the order of integrals: look up the Fubini-Tonelli theorem.

Moreover, if

$$\operatorname{sgn}\theta = \begin{cases} -1 & \text{if } \theta < 0, \\ 0 & \text{if } \theta = 0, \\ +1 & \text{if } \theta > 0, \end{cases}$$

then

$$\int_0^T \frac{\sin \theta t}{t}\,dt = \operatorname{sgn}\theta S\bigl(T|\theta|\bigr).$$

Using the Euler formulas and the fact that $\sin s$ and $\cos s$ are odd and even, respectively, we thus obtain

$$I_T = \int_{\mathbb{R}} \left[\frac{\operatorname{sgn}(x-a)}{\pi} S\bigl(T|x-a|\bigr) - \frac{\operatorname{sgn}(x-b)}{\pi} S\bigl(T|x-b|\bigr) \right] \mu(dx).$$

The integrand here is bounded and converges as $T \to \infty$ to the function

$$\psi_{a,b}(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{1}{2} & \text{if } x = a \\ 1 & \text{if } a < x < b \\ \frac{1}{2} & \text{if } x = b \\ 0 & \text{if } b < x \end{cases}$$

Thus, using the Lebesgue dominated convergence theorem, we obtain $I_T \to \int \psi_{a,b}\,d\mu = \mu\bigl([a,b)\bigr)$, if only $\mu(\{a\}) = \mu(\{b\}) = 0$. □

1.3. The Continuity Theorem. Because of the results presented in this section, characteristic functions can also be used to facilitate studying limits of measures – but first we need to introduce a notion of such limits, called (most often) **weak convergence**.

Definition 1.5. We say that a sequence of probability Borel measures μ_n converges weakly to the measure μ if for every bounded continuous function g we have

$$\lim_{n\to\infty} \int_{\mathbb{R}} g\,d\mu_n = \int_{\mathbb{R}} g\,d\mu.$$

We denote $\mu_n \Rightarrow \mu$.

Theorem 1.6. *The following are equivalent*

(1) μ_n *converge weakly to* μ*;*
(2) *for every closed set* A *we have*

$$\limsup_{n\to\infty} \mu_n(A) \leqslant \mu(A);$$

(3) *for every open set* A *we have*

$$\liminf_{n\to\infty} \mu_n(A) \geqslant \mu(A);$$

(4) *for every Borel set A with $\mu(\operatorname{bd} A) = 0$ we have*

$$\lim_{n\to\infty} \mu_n(A) = \mu(A).$$

The proof is not very complicated, but unfortunately it is outside the scope of this course. For more details, you can read Chapters 25 and 26 in P. Billingsley *Probability and Measure, Third Edition*, Wiley 1995.

The following theorem is called the **continuity theorem** and it is an essential tool in probability theory, making it one of the most important applications of the Fourier transform.

THEOREM 1.7. *Let μ_n, μ be probability measures with characteristic functions ϕ_n, ϕ. Then $\mu_n \Rightarrow \mu$ if and only if $\phi(\xi) \to \phi(\xi)$ for every ξ.*

PROOF. Notice that $x \mapsto e^{ix\xi}$ is a bounded continuous function for every ξ. It therefore follows directly from definition of weak convergence that if $\mu_n \Rightarrow \mu$ then $\phi(\xi) \to \phi(\xi)$ for every ξ.

Let $g \in \mathcal{S}(\mathbb{R})$. Then we have

$$\int_{\mathbb{R}} \breve{g}(\xi)\phi_n(\xi)\, d\xi = \int_{\mathbb{R}} g(x)\, \mu_n(dx),$$

which, like the case in $L^1(\mathbb{R})$, can be proved by extending the same formula known for μ_n in the Schwartz class.

Letting $n \to \infty$, the Lebesgue dominated convergence theorem implies that

$$\begin{aligned}\lim_{n\to\infty} \int_{\mathbb{R}} g(x)\, \mu_n(dx) &= \lim_{n\to\infty} \int_{\mathbb{R}} \psi(\xi)\phi_n(\xi)\, d\xi \\ &= \int_{\mathbb{R}} \psi(\xi)\phi(\xi)\, d\xi = \int_{\mathbb{R}} g(x)\, \mu(dx).\end{aligned}$$

For an interval (a, b), let $g^{\pm} \in \mathcal{S}(\mathbb{R})$ be such that $g^{-} < \mathbb{1}_{(a,b)} < g^{+}$. Then

$$\begin{aligned}\limsup_{n\to\infty} \mu_n\big((a,b)\big) &\leqslant \limsup_{n\to\infty} \int_{\mathbb{R}} g^{+}(x)\, \mu_n(dx) = \int_{\mathbb{R}} g^{+}(x)\, \mu(dx), \\ \liminf_{n\to\infty} \mu_n\big((a,b)\big) &\geqslant \liminf_{n\to\infty} \int_{\mathbb{R}} g^{-}(x)\, \mu_n(dx) = \int_{\mathbb{R}} g^{-}(x)\, \mu(dx).\end{aligned}$$

Now we let $g^{+} \downarrow \mathbb{1}_{(a,b)}$ and $g^{-} \uparrow \mathbb{1}_{(a,b)}$ to conclude that

$$\mu((a,b)) \leqslant \liminf \mu_n((a,b)) \leqslant \limsup \mu_n((a,b)) \leqslant \mu([a,b]).$$

If $\mu(\{a\}) = \mu(\{b\}) = 0$, then the extreme values are equal, so that the limit exists as required. □

1.4. Central limit theorem, analytical version. Another application of the Fourier transform is the central limit theorem. Formulated in the language of analysis, it is not very complicated. We are going to discuss its proabilistic formulation as well in the following sections. Notice how it depends on the special properties of the heat kernel (the Gaussian function) in connection with the Fourier transform.

THEOREM 1.8. *Suppose that* μ *is a Borel measure on* $\mathbb{R}$ *such that* $\mu(\mathbb{R}) = 1$ *and in addition*

$$\int_{\mathbb{R}} x\,\mu(dx) = 0 \quad \textit{and} \quad \int_{\mathbb{R}} x^2\,\mu(dx) = \sigma^2 < \infty$$

Then for every interval A *(or every Borel set* A*) we have*

$$\lim_{n\to\infty} \big(\underbrace{\mu * \mu * \cdots * \mu}_{n\ \textit{times}}\big)(\sqrt{n}A) = \frac{1}{\sigma\sqrt{2\pi}} \int_A e^{-x^2/\sigma^2}\,dx.$$

Notice that the type of convergence that we see in the formulation of the central limit theorem is exactly the weak convergence of measures. We call the measure on the right hand side the **Gaussian measure** (with mean 0 and variance σ^2).

PROOF. We compute the characteristic function of the convolution in question. This is $[\phi_\mu(\xi/\sqrt{n})]^n$. The Taylor expansion of this function at $\xi = 0$ is

$$\phi_\mu(\xi) = 1 - \frac{\xi^2\sigma^2}{2} + o(\xi^2).$$

The characteristic function of the Gaussian measure with mean zero and variance σ^2 has the same Taylor expansion.

Notice that if a, b are complex numbers with $|a| \leqslant 1$, $|b| \leqslant 1$, then $|a^n - b^n| \leqslant n|a-b|$, hence we can write

$$\Big|[\phi_\mu(\xi/\sqrt{n})]^n - e^{-\xi^2/2\sigma^2}\Big| \leqslant n\Big|[\phi_\mu(\xi/\sqrt{n})] - e^{-\xi^2/2n\sigma^2}\Big|.$$

But from the Taylor expansions

$$\phi_\mu(\xi/\sqrt{n}) - e^{-\xi^2/2\sigma^2} = o(1/n),$$

which proves that $\lim_{n\to\infty}[\phi_\mu(\xi/\sqrt{n})]^n = e^{-\xi^2/2\sigma^2}$. The conclusion now follows from the continuity theorem (and the uniqueness of the inversion). ☐

1.5. Independence. In order to discuss probability theory, we have to adjust our language.

Let $(\Omega, \Sigma, \mathcal{P})$ be a probability space. Every set $A \in \Sigma$ is called an **event**. Every measurable function $X : \Omega \to \mathbb{R}$ (or $\mathbb{C}$) is called a **random variable**.

Every real valued random variable has a corresponding **distribution**, which a probability measure on $\mathbb{R}$, defined by

$$\mu(A) = \mathcal{P}\Big(\big\{X^{-1}(A)\big\}\Big).$$

By $E[X]$ we denote the **expectation**, which is simply the integral

$$E[X] = \int_\Omega X \, d\mathcal{P} = \int_\mathbb{R} x \, \mu(dx).$$

In the same way we may define the **variance** $\mathrm{Var}(X) = E[X^2] - E[X]^2$, where $E[X^2] = \int_\mathbb{R} x^2 \, \mu(dx)$.

If μ is the distribution of the random variable X then we also call ϕ_μ the characteristic function of X and we have

$$\phi_\mu(\xi) = E[e^{i\xi X}] = \int_\mathbb{R} e^{ix\xi} \, \mu(dx).$$

We may denote $\phi_X = \phi_\mu$.

Definition 1.9. We say that events $A_1, \ldots A_n \in \Sigma$ are **mutually independent** if for every $I \subseteq \{1, \ldots, n\}$ we have

$$\mathcal{P}\left(\bigcap_{k \in I} A_k\right) = \prod_{k \in I} \mathcal{P}(A_k)$$

Definition 1.10. By F_X we denote the **cumulative distribution function** of a random variable X, i.e.

$$F_X(t) = \mathcal{P}\big(\{X \leqslant t\}\big).$$

For a vector of random variables $(X_1, \ldots, X_n)$ we denote

$$F_{(X_1, \ldots, X_n)}(x_1, \ldots, x_n) = \mathcal{P}\Big(\{X_1 \leqslant x_1\} \cap \ldots \cap \{X_n \leqslant x_n\}\Big).$$

Definition 1.11. We say that X_1 and X_2 are **independent random variables** if

$$F_{(X_1, X_2)}(x_1, x_2) = F_{X_1}(x_1) F_{X_2}(x_2).$$

This can be extended to vectors of random variables of any size, following the preceding definitions, as well as independence between vectors of random variables.

Proposition 1.12. *If X_1 and X_2 are independent random variables then*

$$E[X_1 X_2] = E[X_1] E[X_2].$$

This property implies an important relation for characteristic functions. If $Y_j = \cos(\xi X_j)$ and $Z_j = \sin(\xi X_j)$ for $j = 1, 2$, then (Y_1, Z_1)

and (Y_2, Z_2) are independent We thus have

$$\phi_{X_1}(\xi)\phi_{X_2}(\xi) = (E[Y_1] + iE[Z_1])(E[Y_2] + iE[Z_2])$$
$$= E[Y_1]E[Y_2] - E[Z_1]E[Z_2] + i(E[Y_1]E[Z_2] + E[Z_1]E[Y_2])$$
$$= E[Y_1Y_2 - Z_1Z_2 + i(Y_1Z_2 + Z_1Y_2)]$$
$$= E[e^{i\xi(X_1+X_2)}] = \phi_{(X_1+X_2)}(\xi).$$

Similarly, if $X_1, \ldots, X_n$ are independent, then

$$\phi_{(X_1+\ldots+X_n)}(\xi) = E\Big[e^{i\xi\sum_{k=1}^n X_k}\Big] = \prod_{k=1}^n E[e^{i\xi X_k}]$$
$$= \phi_{X_1}(\xi) \cdot \ldots \cdot \phi_{X_n}(\xi).$$

1.6. The central limit theorem, probabilistic version. We can now reformulate the cenral limit theorem in the language of probability theory. Notice that it is in fact the same theorem as we already proved.

Let $\mathcal{N}$ denote a random variable with the standard normal distribution:

$$P(\mathcal{N} \in A) = \frac{1}{\sqrt{2\pi}} \int_A e^{-x^2/2}\,dx.$$

Theorem 1.13. *Suppose that X_n is a sequence of independent random variables having the same distribution with mean c and finite positive variance σ^2. If $S_n = X_1 + \cdots + X_n$, then*

$$\frac{S_n - nc}{\sigma\sqrt{n}} \Rightarrow \mathcal{N}.$$

2. Radon transform

Consider a flat section of an object (e.g. an internal organ inside a human) being scanned by an X-ray beam (see Figure 1)

Let I_{in} be the intensity of the beam before entering the object and d be the distance travelled. If the material is homogeneous and ρ is the absorption rate (perhaps related to the density, or water content, or something else that physics and biology can tell us), then the output intensity is given by I_{out} in the following relation

$$I_{out} = I_{in}e^{-d\rho},$$

However, if the absorption rate ρ is variable, then

$$I_{out} = I_{in}e^{-\int_L \rho(\tau)\,d\tau},$$

where the integral is taken along the path of the beam L inside the object. If we artificially put $\rho = 0$ everywhere on the outside, then L may be considered to simply be any given straight line in $\mathbb{R}^2$, since such a change doesn't affect the value of the integral.

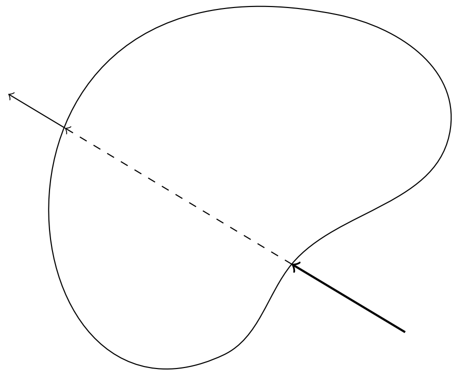

FIGURE 1. An X-ray beam passing through an object.

Now imagine a CAT[2] scanner, able to send X-ray beams in all directions from every point around the section of an organ and then take a measurement on the opposite side (see Figure 2).

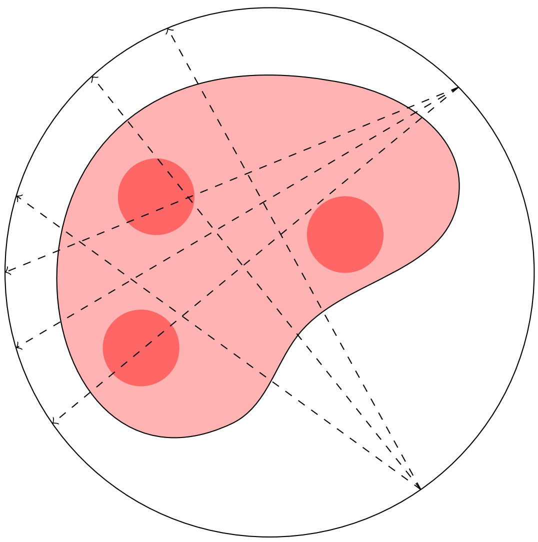

FIGURE 2. A very simple model of a CAT scanner.

To model this situation, we may define the **Radon transform**

$$(\mathcal{R}\rho)(L) = \int_L \rho(\tau)\, d\tau = \log(I_{in}/I_{out}),$$

[2]Computed Axial Tomography; many other instruments are based on the same principle – in electron microscopy, seismology, etc.

which, for a given function ρ, assigns a value to each line L. Since we are most interested in objects that can be scanned from all directions on the outside, we could assume that ρ has a compact support. However, it is slightly more general, and more convenient, to assume $\rho \in \mathcal{S}(\mathbb{R}^2)$, which stands for the Schwartz class on $\mathbb{R}^2$ (defined in an analogous way to the one-dimensional case). It contains all smooth, compactly supported functions.

Suppose we know the value of the measurement (like the CAT scan does), along each line L, but the function ρ is otherwise unknown – and it is our goal is to recover this function (since it may give us a picture of a damaged kidney, or something like this).

First we parametrize the set of lines on the plane $\mathbb{R}^2$. By $L(t, \alpha)$ we denote the line perpendicular to the vector $(\cos\alpha, \sin\alpha)$ and passing through the point $(t\cos\alpha, t\sin\alpha)$. It may also be described by the equation $x\cos\alpha + y\sin\alpha = t$ or the parametrization

$$\mathbb{R} \ni u \mapsto (t\cos\alpha - u\sin\alpha, t\sin\alpha + u\cos\alpha) \in L$$

Thanks to this, way we may define $(\mathcal{R}\rho)(t, \alpha)$, where $t \in \mathbb{R}$, $\alpha \in [0, \pi)$, by the formula

$$\begin{aligned}(\mathcal{R}\rho)(t, \alpha) &= \int_{L(t,\alpha)} \rho(\tau)\, d\tau \\ &= \int_{\mathbb{R}} \rho(t\cos\alpha + u\sin\alpha, t\sin\alpha - u\cos\alpha)\, du.\end{aligned}$$

Notice that if $\rho \in \mathcal{S}(\mathbb{R}^2)$ then for a fixed α we have $\mathcal{R}\rho(\,\cdot\,, \alpha) \in \mathcal{S}(\mathbb{R})$ (see Problem 3), hence we may calculate the Fourier transform of the function $\mathcal{R}\rho$ along the t variable

$$\begin{aligned}\widehat{(\mathcal{R}\rho)}(\xi, \alpha) &= \int_{\mathbb{R}} e^{-2\pi i t\xi} (\mathcal{R}\rho)(t, \alpha)\, dt \\ &= \int_{\mathbb{R}} e^{-2\pi i t\xi} \int_{\mathbb{R}} \rho(t\cos\alpha + u\sin\alpha, t\sin\alpha - u\cos\alpha)\, du\, dt \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-2\pi i (x\cos\alpha + y\sin\alpha)\xi} \rho(x, y)\, dx\, dy \\ &= \int_{\mathbb{R}} e^{-2\pi i y\sin\alpha\xi} \int_{\mathbb{R}} e^{-2\pi i x\cos\alpha\xi} \rho(x, y)\, dx\, dy \\ &= \widehat{\rho}(\xi\cos\alpha, \xi\sin\alpha),\end{aligned}$$

where we substituted $\binom{t}{u} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \binom{x}{y}$ (a rotation) and $\widehat{\rho}$ stands for the Fourier transform of ρ taken in variables $x \in \mathbb{R}$ and $y \in \mathbb{R}$ independently.

The Fourier transform may be inverted and then we can switch back to polar coordinates, obtaining

$$\begin{aligned}\rho(x,y) &= \int_{\mathbb{R}}\int_{\mathbb{R}} e^{2\pi i(x\eta+y\delta)}\widehat{\rho}(\eta,\delta)\,d\eta\,d\delta\\ &= \int_{\mathbb{R}}\int_0^{\pi} e^{2\pi i(x\xi\cos\alpha+y\xi\sin\alpha)}\,\widehat{\rho}(\xi\cos\alpha,\xi\sin\alpha)\,d\alpha\,|\xi|\,d\xi\\ &= \int_{\mathbb{R}}\int_0^{\pi}\int_{\mathbb{R}} |\xi| e^{2\pi i\xi\left(x\cos\alpha+y\sin\alpha-t\right)}(\mathcal{R}\rho)(t,\alpha)\,dt\,d\alpha\,d\xi.\end{aligned}$$

In this way recovered the function ρ in terms of the values of the Radon transform $\mathcal{R}\rho$ (we obtained the formula for inverting the Radon transform).

Of course, in real life the CT scan can only take a finite number of measurements. The "discrete Fourier transform" will be the topic of the next set of lecture notes.

Questions:

- Can you calculate the characteristic functions of the Dirac deltas δ_0 and δ_y, $y \in \mathbb{R}$?
- Can you prove Proposition 1.3, at least for $\mu_i(A) = \int_A f_i(x)\,dx$?
- Can you justify why for a probability measure μ the set A such that $\mu(\{a\}) \neq 0$ for $a \in A$ can only be countable?
- Can you define independence for more than two random variables? Notice that you can define pairwise independence or mutual independence.
- Can you prove Proposition 1.12?
- Can you see that both formulations of the central limit theorem are essentially the same?
- Can you try to define the Schwartz class on $\mathbb{R}^2$ or $\mathbb{R}^n$ (remember that the derivatives in different directions may be mixed)?

Problems:

PROBLEM 1. Prove that for a probability measure μ, the characterictic function ϕ_μ is a continuous, bounded function.

PROBLEM 2. Prove that $\lim_{T\to\infty} \int_0^T \frac{\sin t}{t}\,dt = \frac{\pi}{2}$.

PROBLEM 3. Show that if $\rho \in \mathcal{S}(\mathbb{R}^2)$ then for every $\alpha \in [0, \pi)$ we have $\mathcal{R}\rho(\,\cdot\,, \alpha) \in \mathcal{S}(\mathbb{R})$ and

$$\|\mathcal{R}\rho(t, \alpha) - \mathcal{R}\rho(t, \beta)\| \leqslant C|\alpha - \beta|.$$