The Fourier Operator and the Characteristic Function

8/8/18

Anubhav Nanavaty

1 Overview

The Characteristic function is an important concept in probability theory, because it significantly simplifies otherwise cumbersome calculations. In particular, the characteristic function derives its useful properties from those of the Fourier Operator. The first part of the lecture serves to define the Fourier Operator $\mathcal{F}$, determine in what context(s) can we invert the operator, and observe some of its properties. In the second part of the lecture, I define the characteristic function using the Fourier Operator, and demonstrate its usefulness in probability theory by proving some corollaries and lemmas.

2 The Fourier Operator on the Schwartz Space

We begin by observing how the Fourier Operator behaves on a space of C^∞ functions whose derivatives decay very rapidly to $\pm\infty$.

Definition 2.1. The *Schwartz Space*, denoted $\mathcal{S}(\mathbb{R})$, is the set of $C^\infty(\mathbb{R})$ functions such that for any $\alpha, \beta \in \mathbb{N}$

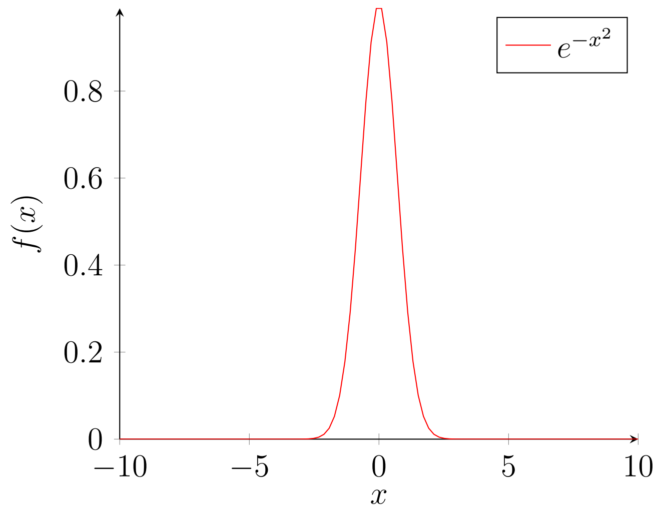
$$\lim_{|x|\to\infty} x^\alpha \frac{d^\beta}{dx^\beta} f(x) = 0$$

In other words, if a function is in the Schwartz space, then given any polynomial p, we can show that the function and its derivatives decay faster than p. A next question might be, what sort of functions live in $\mathcal{S}(\mathbb{R})$?

Example 2.2. The function $f(x) = e^{-ax^2}$, $a \in \mathbb{R}$, is an element of $\mathcal{S}(\mathbb{R})$.

Proof. Due to negative exponentiation, $\lim_{|x|\to\infty} f(x) = 0$. Moreover, given any $\alpha \in \mathbb{N}$, it makes sense that e^{-ax^2} grows at a rate significantly faster than x^α, so $\lim_{|x|\to\infty} x^\alpha e^{-ax^2} = 0$. Since $f(x) \in C^\infty(\mathbb{R})$, taking derivatives, we see that $\frac{d^\beta}{dx^\beta} f(x) = \sum_{n=1}^{2^\beta - 1} x^{\alpha_n} f(x)$, where $\alpha_i \in \mathbb{N}$. Since each term in the sum decays due to what we showed earlier, the entire sum will decay no matter the values of α_i. Further, if we multiply each term by some x^α, our previous claims allows to conclude that the sum will decay for large $|x|$, giving us that $f(x)$ is in the

Schwartz space. Below is a picture:



$\square$

As we can see from the example, functions on the Schwartz space "look almost like" bump functions. This

Let us look at how the Fourier Operator behaves on this space.

Definition 2.3. The *Fourier Operator* is defined as $\mathcal{F} : \mathcal{S}(\mathbb{R}) \to L_1(\mathbb{R})$ such that

$$\mathcal{F}(f)(\xi) = \int_{\mathbb{R}} e^{-2\pi i y \xi} f(y) dy \equiv \hat{f}(\xi)$$

Proposition 2.4. *The Operator is well-defined.*

Proof. We first show that $S(\mathbb{R}) \subset L_1(\mathbb{R})$. For any $f \in \mathcal{S}(\mathbb{R})$, we observe that since $\lim_{|x|\to\infty} x^2 f(x) = 0$ and f is continuous on $\mathbb{R}$, there exists a ball $B_0(\delta) \subset \mathbb{R}$ such that $x^2 f(x) < 1$ for $x \in \mathbb{R} \setminus B_0(\delta)$. Further, f is bounded on $B_0(\delta)$ due to continuity by a constant M. Therefore,

$$\int_{\mathbb{R}} f dx = \int_{B_0(\delta)} f dx + \int_{\mathbb{R}\setminus B_0(\delta)} f dx \leq M + \int_{\mathbb{R}\setminus B_0(\delta)} \frac{1}{x^2} \leq M - 2(\frac{1}{\delta}) < \infty$$

Proving that $f \in L^1(\mathbb{R})$. Looking at the Fourier Transform, we see that Therefore,

$$\Big| \int_{\mathbb{R}} e^{-2\pi i y \xi} f(y) dy \Big| \leq \int_{\mathbb{R}} |e^{-2\pi i y \xi} f(y)| dy \leq ||f||_{L_1}$$

This tells us that $\hat{f} \in L_1(\mathbb{R})$. □

We can go even further to state the following:

Corollary 2.5. *$\mathcal{S}(\mathbb{R}) \subset L^p(\mathbb{R})$, for all $p \geq 1$.*

Proof. The proof uses the exact same method as that of Proposition 2.4, but instead we use the fact that $\lim_{|x|\to\infty} x^{2p} f(x) = 0$ for $p \in [1, \infty)$. For $p = \infty$ we can just use the fact that $\lim_{|x|\to\infty} f(x) = 0$ and invoke $f \in C(\mathbb{R})$ to state that f is bounded. □

Fourier transform of any Schwartz function, $\hat{f}(\xi)$ is now shown well defined for all $\xi \in \mathbb{R}$. We now prove some important properties of $\mathcal{F}$:

Proposition 2.6. [1] *For $\mathcal{F} : \mathcal{S}(\mathbb{R}) \to L_1(\mathbb{R})$.*

1. $\mathcal{F}(f')(\xi) = 2\pi i \xi \hat{f}(\xi)$
2. $\xi^\alpha \hat{f}(\xi) = \frac{1}{(2\pi i)^\alpha} \mathcal{F}(f^{(\alpha)})(\xi)$
3. $\frac{d}{d\xi} \hat{f}(\xi) = -2\pi i \ \widehat{x f(x)}$

Proof. 1. Using the integration by parts formula, if we set $u(x) = e^{-2\pi i \xi x}$, $dv = f'(x)dx$, then:

$$\begin{aligned} \mathcal{F}(f')(\xi) &= \int_{\mathbb{R}} e^{-2\pi i \xi x} f'(x) dx = \int_{-\infty}^{\infty} u dv \\ &= e^{-2\pi i \xi x} f(x) \Big|_{-\infty}^{\infty} + 2\pi i \xi \int_{\mathbb{R}} e^{-2\pi i \xi x} f(x) dx \\ &= 2\pi i \xi \hat{f}(\xi) \end{aligned}$$

The first term of the second line vanishes due to $\lim_{|x|\to\infty} f(x) = 0$. This proves the first equality.

2. Using the first equality, we can see that

$$\mathcal{F}(f^{(\alpha)})(\xi) = 2\pi i \xi \mathcal{F}(f^{(\alpha-1)})(\xi) = (2\pi i \xi)^2 \mathcal{F}(f^{(\alpha-2)})(\xi) = \cdots = (2\pi i \xi)^\alpha \mathcal{F}(f)(\xi)$$

[1] Note that $f^{(\alpha)}$ means that α derivative of f. Note that $\mathcal{F}(f) = \hat{f}$, and different notation was used to avoid any confusion regarding taking the derivative.

3. We can calculate:

$$\begin{aligned}\frac{d}{d\xi}\hat{f}(\xi) &= \frac{d}{d\xi}\int_{\mathbb{R}} e^{-2\pi i\xi x} f(x)dx \\ &= \int_{\mathbb{R}} \frac{d}{d\xi} e^{-2\pi i\xi x} f(x)dx \\ &= \int_{\mathbb{R}} -2\pi i e^{-2\pi i\xi x} x f(x)dx = -2\pi i\ \widehat{xf(x)}\end{aligned}$$

□

These properties are interesting, and upon careful observance, we see that if f is Schwartz, then so must $\hat{f}$. Therefore, $\mathcal{F}$ maps $\mathcal{S}(\mathbb{R})$ into itself. Due to a lot of technicality and a limited amount of time, I will state without proof that in fact $\mathcal{F} : \mathcal{S} \to \mathcal{S}$ is bijective and an also a linear automorphism if $\mathcal{S}$ is seen as a group with the convolution operation. However, we will prove that $\mathcal{F}$ is invertible. Before we do so, we must prove a lemma:

Lemma 2.7. *For all* $\epsilon > 0$

$$\widehat{e^{-\pi\epsilon^2 x^2}}(\xi) = \epsilon^{-1} e^{-\pi\frac{\xi^2}{\epsilon^2}}$$

Proof. By direct computation

$$\begin{aligned}\widehat{e^{-\pi\epsilon^2 x^2}}(\xi) &= \int_{\mathbb{R}} e^{-2\pi i x\xi} e^{-\pi(\epsilon x)^2} dx \\ &= \int_{\mathbb{R}} e^{-\pi(\epsilon x)^2 - 2\pi i x\xi} \\ &= \int_{\mathbb{R}} e^{-(\frac{\sqrt{\pi}\xi}{\epsilon} + \sqrt{\pi}\epsilon x)^2 - \frac{\pi\xi^2}{\epsilon^2}} dx \\ &= e^{-\pi\xi^2} \int_{\mathbb{R}} e^{(\frac{\sqrt{\pi}\xi}{\epsilon} + \sqrt{\pi}\epsilon x)^2} dx \\ &= e^{-\pi\xi^2} \frac{1}{\epsilon\sqrt{\pi}} \int_{\mathbb{R}} e^{-u^2} du = \epsilon^{-1} e^{-\pi\xi^2}\end{aligned}$$

The last step was from a result we computed during the complex analysis module (i.e. that the Gaussian integral is equal to $\sqrt{\pi}$). □

Before we begin proving the Fourier Inversion Formula, we must first define a concept important to the proof:

Definition 2.8. An *Approximation to the Identity* is a family of $L_1(\mathbb{R})$ functions $\{F_\epsilon(x)\}$, $\epsilon \in \mathbb{R}^+$ such that

1. $\int_{\mathbb{R}} F_\epsilon = 1$ for all ϵ
2. $|F_\epsilon(x)| \leq A\epsilon^{-1}$
3. $|F_\epsilon(x)| \leq \frac{A\epsilon}{x^2}$

These approximations derive their name from the following theorem, which I will state without proving:

Theorem 2.9. *If $\{F_\epsilon\}$ is an approximation of the identity, and $f \in L_1(\mathbb{R}) \cap C(\mathbb{R})$, then for every $x \in \mathbb{R}$,*

$$\lim_{\epsilon \to 0} F_\epsilon * f(x) = f(x)$$

I would like to state the following proposition, which can be straightforwardly verified:

Proposition 2.10. $\{G_\epsilon\}_{\epsilon>0}$ *such that*

$$G_\epsilon = \epsilon^{-1} e^{-\pi\epsilon^{-2}(x)^2}$$

is an approximation of the identity.

We covered this theorem and did exercises like this in 209; one can look in Stein and Shakarchi Chapter 3. On to the main theorem:

Theorem 2.11 (The Fourier Inversion Formula)**.** $\mathcal{F}^{-1}$ *exists, and is given by*

$$\mathcal{F}^{-1}(\hat{f})(x) = \int_{\mathbb{R}} e^{2\pi i x\xi} \hat{f}(\xi) d\xi$$

Proof. Observe that if we define

$$F_\epsilon(\xi) = e^{2\pi i\xi} e^{-\pi\epsilon^2\xi^2} \hat{f}(\xi)$$

It is clear that $\lim_{\epsilon \to 0} F_\epsilon(\xi) = e^{2\pi i\xi} \hat{f}(\xi)$. Further, for all $\epsilon > 0$,

$$|F_\epsilon(\xi)| \leq |e^{2\pi i\xi} e^{-\pi\epsilon^2\xi^2} \hat{f}(\xi)| \leq |\hat{f}|$$

We have shown already that $\hat{f} \in L_1$, so by the dominated convergence theorem:

$$\lim_{\epsilon \to 0} \int_{\mathbb{R}} e^{2\pi i x\xi} e^{-\pi\epsilon^2\xi^2} \hat{f}(\xi) d\xi = \int_{\mathbb{R}} e^{2\pi i x\xi} \hat{f}(\xi) d\xi$$

Using our formula for $\hat{f}(\xi)$, and the fact that $\left| \int_{\mathbb{R}} e^{2\pi i x\xi} \hat{f}(\xi) d\xi \right| \leq ||\hat{f}||_{L_1}$ We can substitute and invoke Fubini's theorem to get that

$$\begin{aligned}
\lim_{\epsilon \to 0} \int_{\mathbb{R}} e^{2\pi i x\xi} e^{-\pi\epsilon^2\xi^2} \hat{f}(\xi) d\xi &= \lim_{\epsilon \to 0} \int_{\mathbb{R}} e^{2\pi i x\xi} e^{-\pi\epsilon^2\xi^2} d\xi \int_{\mathbb{R}} e^{-2\pi i y\xi} f(y) dy \\
&= \lim_{\epsilon \to 0} \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-2\pi i (y-x)\xi} e^{-\pi\epsilon^2\xi^2} d\xi f(y) dy \\
&= \lim_{\epsilon \to 0} \int_{\mathbb{R}} \epsilon^{-1} e^{-\pi\epsilon^{-2}(y-x)^2} f(y) dy \\
&= \lim_{\epsilon \to 0} (G_\epsilon * f)(x) \\
&= f(x)
\end{aligned}$$

The second to last step was due to the observation that $f \in C(\mathbb{R}) \cap L_1(\mathbb{R})$, along with the fact that $\{G_\epsilon\}$ is an approximation of the identity (which I did not prove, but omitted due to time constraints). This proves that the Fourier inversion formula holds, i.e., that $\mathcal{F}$ is invertible, and

$$\mathcal{F}^{-1}(\hat{f})(x) = \int_{\mathbb{R}} e^{2\pi i x\xi} \hat{f}(\xi) d\xi$$

□

Now that we have shown that $\mathcal{F}$ is bijective, a natural question that arises is: Given that L_2 is Hilbert Space that is important in analysis, how does the L_2 norm of $f \in \mathcal{S}(\mathbb{R})$ compare with that of $\hat{f}$? The following theorem shows us that they are exactly the same!

Theorem 2.12. *(Plancherel's Theorem) Given $f, g \in \mathcal{S}(\mathbb{R})$, $\langle f, g \rangle = \langle \hat{f}, \hat{g} \rangle$, where $\langle \cdot, \cdot \rangle$ is the inner product on $L_2(\mathbb{R})$.*

Proof. We first observe that

$$\begin{aligned}
\int_{\mathbb{R}} f(x)\hat{g}(x) dx &= \int_{\mathbb{R}} f(x) \int_{\mathbb{R}} e^{-2\pi i x\xi} g(\xi) d\xi \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-2\pi i x\xi} f(x) dx g(\xi) d\xi \qquad \text{(Fubini)} \\
&= \int_{\mathbb{R}} \hat{f}(\xi) g(\xi) d\xi
\end{aligned}$$

Then, recalling the formula for the inner product, we see that

$$\begin{aligned}\langle \hat{f}, \hat{g} \rangle = \int_{\mathbb{R}} \hat{f}(\xi)\overline{\hat{g}}(\xi) d\xi &= \int_{\mathbb{R}} \overline{\hat{g}}(\xi)\hat{f}(\xi) \\ &= \int_{\mathbb{R}} \hat{\overline{\hat{g}}}(\xi) f(\xi) d\xi \qquad \text{(from the previous calculation)} \\ &= \int_{\mathbb{R}} \overline{g} f d\xi \\ &= \langle f, g \rangle\end{aligned}$$

The last step was due to the observation that $\hat{\overline{\hat{g}}} = \overline{\hat{\hat{g}}} = \overline{g}$, since conjugates pass through integrals. □

Plancherel's Theorem hints at the fact that the Fourier Operator is in fact a *Unitary* operator on L_2.

3 The Characteristic Function

In Probability Theory, the Characteristic function is essentially the Fourier transform of a the density function of a random variable. For this section, we set the Fourier transform of a function to be

$$\int_{\mathbb{R}} e^{-ixt} f(x) dx = \hat{f}(t)$$

All properties of the previous section still hold for this modified Fourier Operator - but exact results differ by constants.

Definition 3.1. Given a random variable X, the *Characteristic Function* is defined as $\phi : \mathbb{R} \to \mathbb{C}$ such that

$$\phi_X(t) = \mathbb{E}[e^{iXt}]$$

One can immediately conclude the following:

Proposition 3.2. *1.* $|\phi(t)| = |\mathbb{E}[e^{iXt}]| \leq \mathbb{E}[|e^{iXt}|] = 1$

2. $\lim_{s\to t} \phi(s) = \lim_{s\to t} \mathbb{E}[e^{iXt}] = \mathbb{E}[\lim_{s\to t} e^{iXt}] = \phi(t)$ *by DCT. Hence, ϕ is continuous.*

3. *If $Y = aX + b$, $a, b \in \mathbb{R}$, then* $\phi_Y(t) = \mathbb{E}[e^{2\pi i(aX+b)t}] = e^{ibt}\mathbb{E}[e^{i(aX)t}] = \phi_{aX}(t)$

4. If X has a density function f, then

$$\phi_X(t) = \int_{\mathbb{R}} e^{ixt} f(x)dx = \hat{f}(-t)$$

5. If $X_1, \dots, X_n$ are independent random variables, then

$$\phi_{X_1+\cdots+X_n}(t) = \mathbb{E}\Big[\prod_{k=1}^{n} e^{iX_k t}\Big] = \prod_{k=1}^{n} \mathbb{E}[e^{iX_k t}] = \prod_{k=1}^{n} \phi_{X_k}(t)$$

The last two observations show why the characteristic function is important. It allows us to find the distribution of a sum of independent random variables just by multiplying their characteristic functions together and taking a Fourier transform. At this point, we should check for understanding by seeing an example of a characteristic function.

Example 3.3. The characteristic function of a normal random variable X with mean 0 and variance 1 is $\phi_X(t) = e^{-t^2/2}$

Proof. The probability density of this random variable is given as $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. $f \in \mathcal{S}(\mathbb{R})$ so

$$\begin{aligned}
\phi_X(t) &= \int_{\mathbb{R}} e^{ixt} e^{-x^2/2} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(\frac{t}{\sqrt{2}} - i\frac{x}{\sqrt{2}})^2 - \frac{t^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \int_{\mathbb{R}} e^{-(\frac{t}{\sqrt{2}} - i\frac{x}{\sqrt{2}})^2} dx \\
&= \frac{1}{\sqrt{2\pi}} \sqrt{2\pi} e^{-\frac{t^2}{2}} = e^{-\frac{t^2}{2}}
\end{aligned}$$

□

Example 3.4. The Characteristic function of $\sigma X + \mu$, where X is a normal random variable X with mean 0 and variance 1, is

$$\phi(t) = e^{i\mu t} e^{-\sigma^2 t^2/2}$$

We prove one last Theorem regarding the derivatives of the function - another major reason why the Fourier series are useful.

Theorem 3.5. *If X is a random variable with ϕ as its characteristic function and $\mathbb{E}[|X|] < \infty$, then ϕ is continuously differentiable, and*

$$\phi'(0) = i\mathbb{E}[X]$$

Proof. We would like to use the Dominated Convergence Theorem to prove this. We therefore observe that

$$|e^{i\theta} - 1| \leq \Big| \int_0^\theta ie^{is} ds \Big| \leq \int_0^\theta ds = \theta$$

Now, we can show that

$$\Big| \frac{e^{i(t+\delta)x} - e^{i(t)x}}{\delta} \Big| \leq \frac{|\delta x|}{|\delta|} \leq |x|$$

Now, since $\mathbb{E}[|X|] < \infty$ it must follow that $\int_{\mathbb{R}} x d\mu(x) < \infty$ as well. Therefore, we can use DCT to calculate

$$\begin{aligned}
\phi'(t) = \lim_{\delta \to 0} \frac{\phi(t+\delta) - \phi(t)}{\delta} &= \lim_{\delta \to 0} \int_{\mathbb{R}} \frac{e^{i(t+\delta)x} - e^{i(t)x}}{\delta} d\mu(x) \\
&= \int_{\mathbb{R}} \lim_{\delta \to 0} \frac{e^{i(t+\delta)x} - e^{i(t)x}}{\delta} d\mu(x) \\
&= \int_{\mathbb{R}} ixe^{itx} d\mu(x) \\
\phi'(0) = i\mathbb{E}[0]
\end{aligned}$$

Proving the Theorem □

The ease with which we can take derivatives and add random variables using the characteristic function is the major reason as to why the Fourier Operator is useful for studying probability.