LECTURE 18: THE FOURIER TRANSFORM FOR $\mathbf{M}_1(\mathbb{R}^N)$

In many applications, it is important to know the relationship between the weak convergence of measures and convergence of their Fourier transforms, which are often called *characteristic functions* in the probability literature.

Theorem 18.1. *Given* $\{\mu_n : n \geq 1\} \cup \{\mu\} \subseteq \mathbf{M}_1(\mathbb{R}^N)$, $\mu_n \xrightarrow{\text{w}} \mu$ *if and only if* $\hat{\mu}_n(\boldsymbol{\xi}) \longrightarrow \hat{\mu}(\boldsymbol{\xi})$ *for each* $\boldsymbol{\xi} \in \mathbb{R}^N$. *In fact, if* $\mu_n \xrightarrow{\text{w}} \mu$, *then* $\hat{\mu}_n \longrightarrow \hat{\mu}$ *uniformly on compact subsets.*

Proof. Suppose that $\hat{\mu}_n \longrightarrow \hat{\mu}$ pointwise. Then, by Parseval's identity and Lebesgue's dominated convergence theorem, for each $\varphi \in \mathscr{S}(\mathbb{R}^N;\mathbb{C})$,

$$(2\pi)^N \langle \varphi, \mu_n \rangle = \int \hat{\varphi}(\boldsymbol{\xi})\hat{\mu}_n(-\boldsymbol{\xi})\, d\boldsymbol{\xi} \longrightarrow \int \hat{\varphi}(\boldsymbol{\xi})\hat{\mu}(-\boldsymbol{\xi})\, d\boldsymbol{\xi} = (2\pi)^N \langle \varphi, \mu \rangle,$$

and so, by Theorem 17.3, $\mu_n \xrightarrow{\text{w}} \mu$.

Now suppose that $\mu_n \xrightarrow{\text{w}} \mu$ and that $\boldsymbol{\xi}_n \longrightarrow \boldsymbol{\xi}$ in $\mathbb{R}^N$. Then the functions $\varphi_n(\mathbf{x}) = e^{\imath(\boldsymbol{\xi}_n,\mathbf{x})_{\mathbb{R}^N}}$ converge uniformly on compact subsets to the function $\varphi(\mathbf{x}) = e^{\imath(\boldsymbol{\xi},\mathbf{x})}$, and therefore, by Theorem 17.4, $\hat{\mu}_n(\boldsymbol{\xi}_n) \longrightarrow \hat{\mu}(\boldsymbol{\xi})$. Hence $\hat{\mu}_n \longrightarrow \mu$ uniformly on compact subsets. □

Undoubtedly the most famous application of Theorem 18.1 is to the derivation of the Central Limit Theorem in probability theory. The C.L.T. states that if $\{\mathbf{X}_n : n \geq 1\}$ is a sequence of mutually independent, uniformly square integrable random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ have the properties that their expected value is $\mathbf{0}$ and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \mathbb{E}\big[(\boldsymbol{\xi}, \mathbf{X}_m)^2_{\mathbb{R}^N}\big] = |\boldsymbol{\xi}|^2$$

for all $\boldsymbol{\xi} \in \mathbb{R}^N$, then the distribution σ_n of

$$\frac{\sum_{m=1}^{n} \mathbf{X}_m}{n^{\frac{1}{2}}}$$

converges weakly to γ^N, where $\gamma(dx) = (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2}}\, dx$ is the standard Gaussian measure on $\mathbb{R}$. To phrase this in analytic terms, let μ_m be the distribution of $\mathbf{X}_m$. Then the distribution of $\sum_{m=1}^{n} \mathbf{X}_m$ is the measure $\mu_1 * \cdots * \mu_n$, and so

$$\hat{\sigma}_n(\boldsymbol{\xi}) = \prod_{m=1}^{n} \hat{\mu}_m\big(\tfrac{\boldsymbol{\xi}}{n^{\frac{1}{2}}}\big)$$

is the Fourier transform of the distribution of $\frac{1}{n^{\frac{1}{2}}} \sum_{m=1}^{n} \mathbf{X}_m$. Next note that, by Taylor's theorem,

$$\hat{\mu}_m\big(\tfrac{\boldsymbol{\xi}}{n^{\frac{1}{2}}}\big)) = 1 + \tfrac{\imath}{n^{\frac{1}{2}}} \int (\boldsymbol{\xi}, \mathbf{x})_{\mathbb{R}^N}\, \mu_m(d\mathbf{x}) - \tfrac{1}{2n} \int (\boldsymbol{\xi}, \mathbf{x})^2\, \mu_m(d\mathbf{x}) + o_m\big(\tfrac{1}{n}\big),$$

where, because the $\mathbf{X}_m$'s are uniformly square integrable,

$$\lim_{n\to\infty} n \sup_{m\geq 1} o_m\big(\tfrac{1}{n}\big) = 0.$$

Hence, because the $\mathbf{X}_m$ have expected value $\mathbf{0}$ and

$$\lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^{n}\int(\boldsymbol{\xi},\mathbf{x})^2_{\mathbb{R}^N}\,\mu_m(d\mathbf{x})=|\boldsymbol{\xi}|^2,$$

one has that

$$\hat{\sigma}_n(\boldsymbol{\xi})=\prod_{m=1}^{n}\left(1-\frac{1}{2n}\int(\boldsymbol{\xi},\mathbf{x})^2_{\mathbb{R}^N}\,\mu_m(d\mathbf{x})+o_m\big(\tfrac{1}{n}\big)\right)\longrightarrow e^{-\frac{|\boldsymbol{\xi}|^2}{2}}=\widehat{\gamma^N}(\boldsymbol{\xi}).$$

In spite of Theorem 18.1, it is *not* true that a sequence of probability measures converges weakly just because their Fourier transform converge pointwise. The reason why is that if the sequence converges weakly, then it is relatively compact and therefore must be tight. The following theorem of P. Lévy shows how one can use Fourier transforms to test for tightness.

Theorem 18.2. (**Lévy's Continuity Theorem**) *If $A\subseteq\mathbf{M}_1(\mathbb{R}^N)$, then A is tight if and only if for each $\epsilon>0$ there exists an $r>0$ such that*

$$\sup_{\substack{\mu\in A\\ |\boldsymbol{\xi}|\le r}}\big|1-\hat{\mu}(\boldsymbol{\xi})\big|\le\epsilon. \tag{18.1}$$

Hence, $\{\mu_n:\,n\ge1\}\subseteq\mathbf{M}_1(\mathbb{R}^N)$ is weakly convergent in $\mathbf{M}_1(\mathbb{R}^N)$ if and only if $\hat{\mu}_n$ converges uniformly in a neighborhood of $\mathbf{0}$, in which case there is a $\mu\in\mathbf{M}_1(\mathbb{R}^N)$ to which $\{\mu_n:\,n\ge1\}$ is converging weakly.

Proof. Assume that A is tight and therefore relatively compact. To see that (18.1) hold, suppose it did not. Then there would be an $\epsilon>0$ such that, for each $n\ge1$, $\big|1-\mu_n(\boldsymbol{\xi}_n)\big|\ge\epsilon$ for some $\mu_n\in A$ and $\boldsymbol{\xi}\in B\big(\mathbf{0},\frac{1}{n}\big)$, and, because A is relatively compact, we can choose these μ_n so that they converge weakly to some $\mu\in\mathbf{M}_1(\mathbb{R}^N)$. But then there would exist an $m\ge1$ for which

$$|1-\hat{\mu}(\boldsymbol{\xi})|\vee\big|\hat{\mu}_n(\boldsymbol{\xi})-\hat{\mu}(\boldsymbol{\xi})\big|<\frac{\epsilon}{2}$$

when $n\ge m$ and when $|\boldsymbol{\xi}|\le\frac{1}{n}$, which would lead to the contradiction that $\big|1-\hat{\mu}_n(\boldsymbol{\xi}_n)\big|\ge\epsilon$.

Now assume that (18.1) holds. To show that A must be tight, begin by observing that

$$|1-\hat{\mu}(\boldsymbol{\xi})|\ge\int\big(1-\cos(\boldsymbol{\xi},\mathbf{y})_{\mathbb{R}^N}\big)\,\mu(d\mathbf{y}).$$

Therefore, if[8] $\mathbf{e}\in\mathbb{S}^{N-1}$, for all $r>0$,

$$\frac{1}{r}\int_0^r\big|1-\hat{\mu}(t\mathbf{e})\big|\,dt\ge\int_{\mathbb{R}^N\setminus\{\mathbf{0}\}}\left(1-\frac{\sin\big(r(\mathbf{e},\mathbf{y})_{\mathbb{R}^N}\big)}{r(\mathbf{e},\mathbf{y})_{\mathbb{R}^N}}\right)\mu(d\mathbf{y}).$$

Now set

$$s(t)=\inf\left\{\frac{\sin\tau}{\tau}\,:\,\tau\ge t\right\}\quad\text{for } t>0.$$

Then $s(t)>0$ for all $t>0$ and, for all $R>0$,

$$\sup_{|\boldsymbol{\xi}|\le r}\big|1-\hat{\mu}(\boldsymbol{\xi})\big|\ge\frac{1}{r}\int_0^r\big|1-\hat{\mu}(t\mathbf{e})\big|\,dt\ge s(rR)\mu\big(\{\mathbf{y}:\,|(\mathbf{e},\mathbf{y})_{\mathbb{R}^N}|\ge R\}\big).$$

[8] $\mathbb{S}^{N-1}$ is the unit sphere in $\mathbb{R}^N$.

Since

$$\mu\big(\{\mathbf{y}:\,|\mathbf{y}|\ge R\}\big) \le N \sup_{\mathbf{e}\in\mathbb{S}^{N-1}} \mu\big(\{\mathbf{y}:\,\big|(\mathbf{e},\mathbf{y})_{\mathbb{R}^N}\big| \ge N^{-\frac{1}{2}}R\}\big),$$

we have the estimate

$$\mu\big(\{\mathbf{y}:\,|\mathbf{y}|\ge R\}\big) \le \frac{N}{s(rN^{-\frac{1}{2}}R)} \sup_{|\boldsymbol{\xi}|\le r}\big|1-\hat{\mu}(\boldsymbol{\xi})\big|. \tag{18.2}$$

In particular, (18.1) implies that, for each $\epsilon > 0$, there is an $R > 0$ such that

$$\sup_{\mu\in A}\mu\big(\{\mathbf{y}:\,|\mathbf{y}|\ge R\}\big) \le \epsilon.$$

□

Bochner found an interesting characterization of characteristic functions, one which is intimately related to Lévy's Continuity Theorem. To describe his result, say that a function $f:\mathbb{R}^N \longrightarrow \mathbb{C}$ is *non-negative definite* if the matrix

$$\big(\big(f(\boldsymbol{\xi}_j-\boldsymbol{\xi}_k)\big)\big)_{1\le j,k\le n}$$

is non-negative definite for all $n\ge 2$ and $\boldsymbol{\xi}_1,\dots,\boldsymbol{\xi}_n\in\mathbb{R}^N$, which is equivalent to saying

$$\sum_{j,k=1}^{n} f(\boldsymbol{\xi}_j-\boldsymbol{\xi}_k)\alpha_j\overline{\alpha_k}\ge 0$$

for all $\alpha_1,\dots,\alpha_n\in\mathbb{C}$.

Theorem 18.3. *A function $f:\mathbb{R}^N\longrightarrow\mathbb{C}$ is a characteristic function if and only if f is continuous, $f(0)=1$, and f is non-negative definite.*

Proof. Assume that $f=\hat{\mu}$ for some $\mu\in M_1(\mathbb{R}^N)$. Then it is obvious that f is continuous and that $f(0)=1$. To see that it is non-negative definite, observe that

$$\begin{aligned}\sum_{j,k=1}^{n} f(\boldsymbol{\xi}_j-\boldsymbol{\xi}_k)\alpha_j\overline{\alpha_k} &= \int\left(\sum_{j,k=1}^{n} e^{i(\boldsymbol{\xi}_j-\boldsymbol{\xi}_k,\mathbf{x})_{\mathbb{R}^N}}\alpha_j\overline{\alpha_k}\right)\mu(d\mathbf{x})\\ &= \int\left|\sum_{j,k=1}^{n} e^{i\boldsymbol{\xi}_j x}\alpha_j\right|^2\mu(d\mathbf{x})\ge 0.\end{aligned}$$

Now assume that f is a continuous, non-negative definite function with $f(0)=1$. Because

$$A\equiv\begin{pmatrix}1 & f(\boldsymbol{\xi})\\ f(-\boldsymbol{\xi}) & 1\end{pmatrix}$$

is non-negative definite, $\mathfrak{Im}\big(f(\boldsymbol{\xi})+f(-\boldsymbol{\xi})\big)$ and $\mathfrak{Im}\big(if(\boldsymbol{\xi})-if(-\boldsymbol{\xi})\big)$ are both 0, and therefore $f(\boldsymbol{\xi})=\overline{f(-\boldsymbol{\xi})}$. Thus A is Hermitian, and because it is non-negative definite, $1-|f(\boldsymbol{\xi})|^2\ge 0$. Therefore $|f(\boldsymbol{\xi})|\le 1$. Next, let $\psi\in\mathscr{S}(\mathbb{R}^N;\mathbb{R})$, and use Riemann approximations to see that

$$\iint f(\boldsymbol{\xi}-\boldsymbol{\eta})\hat{\psi}(\boldsymbol{\xi})\overline{\hat{\psi}(\boldsymbol{\eta})}\,d\xi d\boldsymbol{\eta}\ge 0.$$

Assume for the moment that $f\in L^1(\lambda_{\mathbb{R}^N};\mathbb{C})$, and set

$$g(\mathbf{x})=(2\pi)^{-N}\int e^{-i(\boldsymbol{\xi},\mathbf{x})_{\mathbb{R}^N}}f(\boldsymbol{\xi})\,d\boldsymbol{\xi}.$$

By Parseval's identity, Fubini's Theorem and the fact that $\overline{\hat{\psi}(\boldsymbol{\xi})} = \hat{\psi}(-\boldsymbol{\xi})$,

$$(2\pi)^N \int g(\mathbf{x})\psi(\mathbf{x})^2\,d\mathbf{x} = \int f(\boldsymbol{\xi})\widehat{\psi^2}(-\boldsymbol{\xi})\,d\boldsymbol{\xi} = \int f(\boldsymbol{\xi})\big(\hat{\psi} * \hat{\psi}\big)(-\boldsymbol{\xi})\,d\boldsymbol{\xi}$$

$$= \iint f(\boldsymbol{\xi})\overline{\hat{\psi}(\boldsymbol{\xi}+\boldsymbol{\eta})}\hat{\psi}(\boldsymbol{\eta})\,d\boldsymbol{\xi}d\boldsymbol{\eta} = \iint f(\boldsymbol{\xi}-\boldsymbol{\eta})\overline{\hat{\psi}(\boldsymbol{\xi})}\hat{\psi}(\boldsymbol{\eta})\,d\boldsymbol{\xi}d\boldsymbol{\eta} \geq 0.$$

Hence, since g is continuous, it follows that $g \geq 0$. In addition, $f = \hat{g}$ and so $\int g(\mathbf{x})\,d\mathbf{x} = f(0) = 1$ and f is the Fourier transform of the probability measure $d\mu = g\,d\lambda_{\mathbb{R}^N}$.

To remove the assumption that f is integrable, set $g_t(\mathbf{x}) = (2\pi t)^{-\frac{N}{2}}e^{-\frac{|\mathbf{x}|^2}{2t}}$ and define $\gamma_t(d\mathbf{x}) = g_t(\mathbf{x})\,d\mathbf{x}$. Then $\widehat{\gamma_t}(\boldsymbol{\xi}) = e^{-\frac{t|\boldsymbol{\xi}|^2}{2}}$ and therefore $f_t \equiv \widehat{\gamma_t}f$ is a continuous, $\lambda_{\mathbb{R}^N}$-integrable function that is 1 at 0. To see that f_t is non-negative definite, note that

$$\sum_{j,k=1}^n f_t(\boldsymbol{\xi}_j - \boldsymbol{\xi}_k)\alpha_j\overline{\alpha_k} = \sum_{j,k=1}^n f(\boldsymbol{\xi}_j - \boldsymbol{\xi}_k)\alpha_j\overline{\alpha_k}\int e^{i(\boldsymbol{\xi}_j-\boldsymbol{\xi}_k,\mathbf{x})_{\mathbb{R}^N}}\gamma_t(d\mathbf{x})$$

$$= \int\left(\sum_{j,k=1}^n f(\boldsymbol{\xi}_j - \boldsymbol{\xi}_k)\big(\alpha_j e^{i(\boldsymbol{\xi}_j,\mathbf{x})_{\mathbb{R}^N}}\big)\big(\overline{\alpha_k e^{i(\boldsymbol{\xi}_k,\mathbf{x})_{\mathbb{R}^N}}}\big)\right)\gamma_t(d\mathbf{x}) \geq 0.$$

Thus $f_t = \widehat{\mu_t}$ for some $\mu_t \in M_1(\mathbb{R}^N)$, and so, since $f_t \longrightarrow f$ uniformly on compact subsets, Lévy's Continuity Theorem implies that μ_t tends weakly to a $\mu \in M_1(\mathbb{R}^N)$ for which $f = \hat{\mu}$. □

Because it is difficult to check whether a function is non-negative definite, it is the more or less trivial necessity part of Bochner's Theorem that turns out in practice to be more useful than the sufficiency conditions.

Exercise 18.4. Given $f \in C_{\mathrm{b}}(\mathbb{R}^N;\mathbb{C})$ with $f(0) = 1$, define the quadratic form

$$(\varphi,\psi)_f = \iint_{\mathbb{R}^N\times\mathbb{R}^N}\varphi(\boldsymbol{\xi})f(\boldsymbol{\xi}-\boldsymbol{\eta})\overline{\psi(\boldsymbol{\eta})}\,d\boldsymbol{\xi}d\boldsymbol{\eta}$$

for $\varphi,\psi \in \mathscr{S}(\mathbb{R}^N;\mathbb{C})$. Show that this quadratic form is an inner product (i.e., $(\varphi,\varphi)_f \geq 0$) if and only if f is a characteristic function. Further, if $f = \hat{\mu}$, show that $(\varphi,\psi)_f = (\hat{\varphi},\hat{\psi})_{L^2(\mu;\mathbb{C})}$ and therefore that $(\,\cdot\,,\,\cdot\,)_f$ is a Hilbert inner product (i.e., $(\varphi,\varphi)_f = 0 \implies \varphi = 0$) if and only if $\mu(G) > 0$ for all non-empty open sets G.

Exercise 18.5. Here are some interesting facts about characteristic functions.

(**i**) It is easy to check that if $\mu \in \mathbf{M}_1(\mathbb{R}^N)$, then

$$|\hat{\mu}(\boldsymbol{\eta}) - \hat{\mu}(\boldsymbol{\xi})|^2 \leq 2\mathfrak{Re}\big(1 - \hat{\mu}(\boldsymbol{\eta}-\boldsymbol{\xi})\big),$$

and so, by Theorem 18.3, one sees that if f is a continuous, non-negative definite function for which $f(\mathbf{0}) = 1$, then $|f(\boldsymbol{\xi})| \leq 1$ and $|f(\boldsymbol{\eta})-f(\boldsymbol{\xi})|^2 \leq 2\big(1-\mathfrak{Re}f(\boldsymbol{\eta}-\boldsymbol{\xi})\big)$. Show that these inequalities hold even if one drops the continuity assumption.

Hint: Use the non-negative definiteness of the matrices

$$\begin{pmatrix} 1 & f(-\boldsymbol{\xi}) \\ f(\boldsymbol{\xi}) & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & f(-\boldsymbol{\xi}) & f(-\boldsymbol{\eta}) \\ f(\boldsymbol{\xi}) & 1 & f(\boldsymbol{\xi}-\boldsymbol{\eta}) \\ f(\boldsymbol{\eta}) & f(\boldsymbol{\eta}-\boldsymbol{\xi}) & 1 \end{pmatrix}$$

to see that $f(-\boldsymbol{\xi}) = \overline{f(\boldsymbol{\xi})}$ and that

$$1 + 2\alpha\bigl(1 - \mathfrak{Re} f(\boldsymbol{\eta} - \boldsymbol{\xi})\bigr) + 2\alpha^2\bigl|f(\boldsymbol{\eta}) - f(\boldsymbol{\xi})\bigr| \geq 0 \text{ for all } \alpha \in \mathbb{R}.$$

(**ii**) Without using Bochner's theorem, show that if f_1 and f_2 are non-negative definite functions, then so are $f_1 f_2$ and, for any $a, b \geq 0$, $af_1 + bf_2$ is also.

Hint: Show that if A and B are non-negative definite, Hermitian $N \times N$ matrices, then $\bigl(\bigl(A_{k,\ell}B_{k,\ell}\bigr)\bigr)_{1\leq k,\ell\leq N}$ is also. One way to see this is to use the fact that B admits a square root.

(**iii**) Suppose that $f : \mathbb{R}^N \longrightarrow \mathbb{C}$ is a function for which $f(\mathbf{0}) = 1$. Show that if $\lim_{|\mathbf{x}|\searrow 0} \frac{1-f(\mathbf{x})}{|\mathbf{x}|^2} = 0$, then f cannot be a characteristic function. In particular, if $\alpha > 2$, then $e^{-|\boldsymbol{\xi}|^\alpha}$ is not a characteristic function.

(**iv**) Given a finite signed Borel measure μ on $\mathbb{R}^N$, define

$$\hat{\mu}(\boldsymbol{\xi}) = \int e^{i(\boldsymbol{\xi},\mathbf{x})_{\mathbb{R}^N}}\, \mu(d\mathbf{x}),$$

and show that $\hat{\mu} = 0$ if and only if $\mu = 0$.

Hint: Use the Hahn Decomposition Theorem to write μ as the difference of two, mutually singular, non-negative Borel measures on $\mathbb{R}^N$.

(**v**) Suppose that $f : \mathbb{R} \longrightarrow \mathbb{C}$ is a non-constant, twice continuously differentiable characteristic function. Show that $f''(0) < 0$ and that $\frac{f''}{f''(0)}$ is again a characteristic function. In addition, show that $\|f'\|_{\mathrm{u}}^2 \vee \|f''\|_{\mathrm{u}} \leq |f''(0)|$ and that $|f(\boldsymbol{\eta}) - f(\boldsymbol{\xi})| \leq |f''(0)|^{\frac{1}{2}}|\boldsymbol{\eta} - \boldsymbol{\xi}|$.

(**vi**) Suppose that $\{\mu_n : n \geq 1\} \subseteq M_1(\mathbb{R})$ and that $f(\boldsymbol{\xi}) = \lim_{n\to\infty} \widehat{\mu_n}(\boldsymbol{\xi})$ exists for each $\boldsymbol{\xi} \in \mathbb{R}$. Show that f is a characteristic function if and only if it is continuous at $\mathbf{0}$, and notice that this provides an alternative proof of Theorem 18.2.

(**vii**) Let $\mu_n \in M_1(\mathbb{R})$ be the measure for which $\frac{d\mu_n}{d\lambda_{\mathbb{R}}} = (2n)^{-1}\mathbf{1}_{[-n,n]}$. Show that $\widehat{\mu_n} \longrightarrow \mathbf{1}_{\{0\}}$ pointwise, and conclude that $\{\mu_n : n \geq 1\}$ has no weak limits. This example demonstrates the essential role that continuity plays in Bochner's and Lévy's theorems.