18.103 Fall 2013

1 Fourier Integrals of finite measures.

Denote the space of finite, positive, measures on $\mathbb{R}$ by

$$M_+(\mathbb{R}) = \{\mu : \mu \text{ is a positive measure on } \mathbb{R}; \quad \mu(\mathbb{R}) < \infty\}$$

Proposition 1 *For $\mu \in M_+(\mathbb{R})$, we define the Fourier transform by*

$$\mathcal{F}(\mu) = \hat{\mu}(\xi) = \int_{\mathbb{R}} e^{-ix\xi}\, d\mu(x)$$

Then

$$\mathcal{F} : M_+(\mathbb{R}) \to C_b(\mathbb{R})$$

where $C_b(\mathbb{R})$ is the space of bounded continuous functions.

Proof. We have

$$|\hat{\mu}(\xi)| \le \left| \int_{\mathbb{R}} e^{-ix\xi}\, d\mu(x) \right| \le \int_{\mathbb{R}} |e^{-ix\xi}|\, d\mu(x) = \mu(\mathbb{R}) < \infty$$

Furthermore, since $\mu(\mathbb{R}) < \infty$, the function $g(x) \equiv 1$ is integrable, and serves as a majorant, $|e^{-ix\xi}| \le 1 = g(x)$. By the dominated convergence theorem, $\hat{\mu}(\xi)$ is continuous.

Theorem 1 *(Uniqueness) Let $\mu \in M_+(\mathbb{R})$. Then μ is uniquely determined by $\hat{\mu}$.*

Proof. We show that if $\hat{\mu}_1 = \hat{\mu}_2$, then $\mu_1 = \mu_2$. Let $\varphi \in \mathcal{S}$. Then

$$\int_{\mathbb{R}} \hat{\varphi}(y) d\mu(y) = \int_{\mathbb{R}} \varphi(x) \hat{\mu}(x)\, dx \tag{1}$$

Indeed, by Fubini's theorem

$$\int_{\mathbb{R}} \hat{\varphi}(y) d\mu(y) = \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-ixy} \varphi(x)\, dx d\mu(y) = \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-ixy} d\mu(y) \varphi(x)\, dx = \int_{\mathbb{R}} \hat{\mu}(y) \varphi(x)\, dx$$

If $\hat{\mu}_1(y) = \hat{\mu}_2(y)$, then by (1),

$$\int_{\mathbb{R}} \hat{\varphi}(y) d\mu_1(y) = \int_{\mathbb{R}} \hat{\varphi}(y) d\mu_2(y)$$

for every $\varphi \in \mathcal{S}$. Since the Fourier transform is invertible on $\mathcal{S}$, we can also write this as

$$\int_{\mathbb{R}} \varphi(y) d\mu_1(y) = \int_{\mathbb{R}} \varphi(y) d\mu_2(y) \quad \text{for all} \quad \varphi \in \mathcal{S}$$

Choose $\varphi_\epsilon \in \mathcal{S}(\mathbb{R})$ such that

$$1_{[a,b]} \leq \varphi_\epsilon \leq 1_{[a-\epsilon, b+\epsilon]}$$

By the dominated convergence theorem

$$\mu_1([a,b]) = \lim_{\epsilon \to 0} \int_{\mathbb{R}} \varphi_\epsilon d\mu_1 = \lim_{\epsilon \to 0} \int_{\mathbb{R}} \varphi_\epsilon d\mu_2 = \mu_2([a,b])$$

This shows that μ_1 and μ_2 are the same.

Weak convergence of measures. Let μ_j and μ denote positive measures with finite total mass on $\mathbb{R}$. We say that μ_j tends weakly to μ if

$$\lim_{j \to \infty} \int_{\mathbb{R}} \varphi d\mu_j = \int_{\mathbb{R}} \varphi d\mu$$

for all $\varphi \in C_0(\mathbb{R})$.

Exercise. Show that if

$$\lim_{j \to \infty} \int_{\mathbb{R}} \varphi d\mu_j = \int_{\mathbb{R}} \varphi d\mu$$

for all $\varphi \in C_0^\infty(\mathbb{R})$, then $\mu_j \to \mu$ weakly. (The reason is that $C_0^\infty(\mathbb{R})$ is dense in $C_0(\mathbb{R})$ in the uniform norm, i. e. the $L^\infty(\mathbb{R})$ norm. In particular, functions $\varphi \in \mathcal{S}(\mathbb{R})$ sufffice.)

We will establish weak convergence by finding a sufficient condition for it in terms of Fourier transforms. (Note that we established the analogous lemma for Fourier series when we proved Weyl's equidistribution theorem.)

Proposition 2 *If μ_j and μ belong to $M^+(\mathbb{R})$, and for each $\xi \in \mathbb{R}$,*

$$\lim_{j \to \infty} \hat{\mu}_j(\xi) = \hat{\mu}(\xi)$$

then

$$\lim_{j \to \infty} \int_{\mathbb{R}} f d\mu_j = \int_{\mathbb{R}} f d\mu$$

for all $f \in \mathcal{S}(\mathbb{R})$.

Proof. We have in particular, $\mu_j(\mathbb{R}) = \hat{\mu}_j(0) \to \hat{\mu}(0)$ so that the sequence $\mu_j(\mathbb{R})$ is bounded. (In all of our applications we have probability measures, $\mu_j(\mathbb{R}) = 1$, so this step is not needed.) Thus

$$\sup_j |\hat{\mu}_j(\xi)| = \sup_j \mu_j(\mathbb{R}) \leq C < \infty$$

For $\varphi \in \mathcal{S}$, $|\varphi(\xi)\hat{\mu}_j(\xi)| \leq C|\varphi(\xi)|$ is an integrable majorant, so the dominated convergence theorem gives

$$\lim_{j\to\infty} \int_{\mathbb{R}} \varphi(\xi)\hat{\mu}_j(\xi)\, d\xi = \int_{\mathbb{R}} \varphi(\xi)\hat{\mu}(\xi)\, d\xi$$

But by Fubini's theorem this is the same as

$$\lim_{j\to\infty} \int_{\mathbb{R}} \hat{\varphi}(y) d\mu_j(y) = \int_{\mathbb{R}} \hat{\varphi}(y)) d\mu(y)$$

But the set of all $\hat{\varphi}$ such that $\varphi \in \mathcal{S}$ is all of $\mathcal{S}$, so we have the desired weak convergence of μ_j to μ.

In the applications we want more explicit kind of convergence, namely convergence of $\mu_j(I)$ for intervals I. This nearly works. What can go wrong is illustrated using sequences of point masses,

$$\delta_a(x) = \delta(x - a)$$

This measure is defined by $\delta_a(E) = 1$ if $a \in E$ and 0 if $a \notin E$. The sequence $\delta_{1/n}$ tends weakly to δ_0 as $n \to \infty$, but

$$\lim_{j\to\infty} \delta_{1/n}(I) = 1 \neq 0 = \delta_0(I), \qquad I = (0, 1).$$

A *continuous measure* is a measure μ such that $\mu(a) = 0$ for any single point a. A measure that has only point masses is called atomic, and a continuous measure is sometimes called non-atomic. On your homework you show how to split a measure into atomic and continuous parts. The discontinuities arise at at most countably many points. A continuous measure is characterized by the fact that $\mu((a, b)$ is continuous as a function of a and b. Equivalently, $\mu((a, b)) = \mu([a, b])$ for any $a \leq b$.

Proposition 3 *If*

$$\lim_{j\to\infty} \int_{\mathbb{R}} f d\mu_j = \int_{\mathbb{R}} f d\mu$$

for all $f \in C_0^\infty(\mathbb{R})$, *then*

$$\limsup_{j\to\infty} \mu_j((a, b)) \leq \mu([a, b])$$

and

$$\liminf_{j\to\infty} \mu_j((a,b)) \geq \mu((a,b))$$

In particular if μ is continuous then

$$\mu_j((a,b)) \to \mu((a,b)) = \mu([a,b])$$

Proof. Let $\varphi \in \mathcal{S}$ be such that $1_{(a,b)} \leq \varphi \leq 1_{(a-\epsilon,b+\epsilon)}$, then

$$\limsup_{j\to\infty} \mu_j((a,b)) \leq \limsup_{j\to\infty} \int_{\mathbb{R}} \varphi\, d\mu_j = \int_{\mathbb{R}} \varphi\, d\mu \leq \mu((a-\epsilon, b+\epsilon))$$

Taking the limit as $\epsilon \to 0$ gives the upper bound. The lower bound is similar using $\varphi \in C_0^\infty(\mathbb{R})$ such that $1_{(a+\epsilon,b-\epsilon)} \leq \varphi \leq 1_{(a,b)}$. It is important to note that the sequence μ_j can have all the atomic parts it wants. Only the target measure μ needs to be continuous at the endpoints. This is crucial to the applications.

2 An application to the central limit theorem.

Recall that in probability theory, we are concerned with measures μ_X on $\mathbb{R}$ that represent the probability distribution of a random variable X satisfying

$$P(a < X < b) = \mu_X((a,b)) = \mathbb{E}(1_{(a,b)}(X)) = \int_{\mathbb{R}} 1_{(a,b)} d\mu_X$$

and, more generally, for any Borel function φ on $\mathbb{R}$ that is μ_f-integrable,

$$\int_{\mathbb{R}} \varphi(x) d\mu_f(x) = \mathbb{E}(\varphi(X)) = \int_M \varphi(X) d\nu$$

For example, all the Rademacher functions R_n all have the same distribution, value 1 with probability 1/2 and value -1 with probability 1/2. The corresponding measure is denoted

$$\mu = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$$

$$\mathbb{E}(\varphi(R_n)) = \int_{\mathbb{R}} \varphi(x) d\mu(x) = \frac{1}{2}\varphi(1) + \frac{1}{2}\varphi(-1)$$

The mean and variance of X is defined by

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\, d\mu_X(x); \quad \mathrm{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2\, d\mu_X(x)$$

For $X = R_n$ the mean is 0 and the variance is

$$\frac{1}{2}(1-0)^2 + \frac{1}{2}(1-0)^2 = 1$$

The *characteristic function* χ_X of a random variable X is defined by

$$\chi_X(t) = \mathbb{E}(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} d\mu_X(x)$$

For example, the characteristic function of each R_n is

$$\chi(t) = (1/2)e^{it} + (1/2)e^{-it} = \cos(t)$$

In general, up to a sign, the characteristic function is the Fourier transform of the distribution of a random variable X,

$$\chi_X(t) = \widehat{\mu_X}(-t)$$

The notion of weak convergence is central to probability theory. Weak convergence of the distributional measures is the same as

$$\mathbb{E}(\varphi(X_j)) \to \mathbb{E}(\varphi(X))$$

for all $\varphi \in C_0(\mathbb{R})$ (or equivalently for all $\varphi \in C_0^\infty(\mathbb{R})$). In probability theory it's also convenient to know (and not hard to show) that one can broaden the class of test functions from $\varphi \in C_0(\mathbb{R})$ to $\varphi \in C_b(\mathbb{R})$, the space of bounded continuous functions.

The word used by probabilists for weak convergence is *convergence in law.* They often prefer an equivalent, but more concrete, formulation using test functions of the form $\varphi = 1_I$, where $I \subset \mathbb{R}$ is an interval. We have

$$\mathbb{E}(X \in I) = \mathbb{E}(1_I(X)) = \int_I d\mu_X = \mu_X(I)$$

The discontinuity of 1_I at its endpoints requires us to change the statement slightly, as we explained earlier and now express using notations from probability theory.

The (cumulative) distribution function F_X of a random variable X (with probability distribution μ_X) is defined as

$$F_X(a) = \mathbb{E}(X \leq a) = \mu_X((-\infty, a])$$

We say that X_j *converges in law to* X if

$$\lim_{j\to\infty} F_{X_j}(a) = F_X(a)$$

for every a at which $F(x)$ is continuous. It follows that if X_j tends to X in law

$$\lim_{j\to\infty} \mathbb{E}(a < X_j < b) = \mathbb{E}(a < X < b)$$

provided F_X is continuous at a and b. Proposition 3 says that weak convergence of μ_{X_j} to μ_X implies convergence in law of X_j to X. The converse is left as an exercise; we won't use it. (To prove this, first figure out how convergence in law identifies the jumps of the increasing function F_X.)

A gaussian random variable X with mean zero and variance 1 is a variable whose distribution is given by

$$\mathbb{E}(a < X < b) = \int_a^b g(x)\,dx;$$

in which

$$d\mu_X = g(x)dx, \qquad g(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

Thus

$$\mathbb{E}(\varphi(X)) = \int_{-\infty}^{\infty} \varphi(x)g(x)\,dx$$

The characteristic function is

$$\chi(t) = \chi_X(t) = \mathbb{E}(e^{itX}) = \int_{-\infty}^{\infty} e^{ixt}g(x)\,dx = \hat{g}(-t) = e^{-t^2/2} \tag{2}$$

In particular,

$$\mathbb{E}(1) = \int_{-\infty}^{\infty} g(x)\,dx = 1$$

(This is just the required normalization of a probability measure: total mass 1.) Differentiating (2) with respect to t, we get

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\,g(x)\,dx = 0 \quad \text{(mean zero)}$$

Differentiating again, we get

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2\, g(x)\, dx = 1 \quad \text{(variance 1)}$$

We can rescale to get any variance. Let

$$g_\sigma(x) = (1/\sigma)g(x/\sigma), \quad g(x) = g_1(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

The parameter σ is known as the standard deviation and the variance is the square of the standard deviation.

$$\int_{-\infty}^{\infty} x^2\, g_\sigma(x)\, dx = \sigma^2$$

We can also change the mean by translation: If X has distribution $g_\sigma(x - x_0)dx$, then it has mean x_0 and variance σ^2.

To clarify the meaning of the scale factor σ, consider X_σ a gaussian random variable with standard deviation σ, and X_1, a gaussian random variable with standard deviation 1. Then

$$P(a\sigma < X_\sigma - \mathbb{E}(X_\sigma) < b\sigma) = P(a < X_1 - \mathbb{E}X_1 < b)$$

Theorem 2 *(Central Limit Theorem) Let X_1, X_2, ... be independent, identically distributed random variables such that*

$$\mathbb{E}(X_1) = M; \quad \mathbb{E}[(X_1 - M)^2] = \sigma^2; \quad \mathbb{E}(|X_1|^{2+\alpha}) = A < \infty$$

for some $\alpha > 0$. Then

$$\mathbb{E}\left(a < \frac{X_1 + \cdots + X_n - nM}{\sqrt{n}} < b \right) \longrightarrow \int_a^b g_\sigma(x)\, dx$$

Lemma 1

$$e^{ix} = 1 + ix + (ix)^2/2 + R(x)$$

with

$$|R(x)| \le 4 \min(|x|^2, |x|^3) \le 4|x|^{2+\alpha}$$

for all α, $0 \le \alpha \le 1$.

Proof. The fundamental theorem of calculus implies

$$f(1) = f(0) + f'(0) + \frac{1}{2!}f''(0) + \frac{1}{2!}\int_0^1 f'''(t)(1-t)^2\,dt$$

Let $f(t) = e^{itx}$. Then

$$R(x) = \frac{1}{2}\int_0^1 (ix)^3 e^{itx}(1-t)^2\,dt$$

Therefore,

$$|R(x)| \leq \frac{|x|^3}{2}\int_0^1 (1-t)^2\,dt = |x|^3/6$$

for all $|x| \leq 1$. On the other hand, for $|x| \geq 1$,

$$|R(x)| = |1 + ix + (ix)^2/2 - e^{ix}| \leq 2 + |x| + x^2/2 \leq 4|x|^2$$

Replacing X_j with $X_j - M$ we may assume without loss of generality that $M = 0$. We compute the Fourier transform of the measure μ_n defined by

$$\mu_n((a,b)) = P\left(a < \frac{X_1 + \cdots + X_n}{\sqrt{n}} < b\right)$$

$$\hat{\mu}_n(\xi) = \mathbb{E}\left(e^{-i\xi(X_1+\cdots+X_n)/\sqrt{n}}\right) = \prod_{j=1}^{n}\mathbb{E}(e^{-i\xi X_j/\sqrt{n}}) = \left(\mathbb{E}(e^{-i\xi X_1/\sqrt{n}})\right)^n$$

and since $M = 0$,

$$\begin{aligned}\mathbb{E}(e^{-i\xi X_1/\sqrt{n}}) &= \int_{\mathbb{R}} e^{-ix\xi}\,d\mu_1(x) = \int_{\mathbb{R}}[1 - ix\xi/\sqrt{n} + (-ix\xi)^2/2n + R(-x\xi/\sqrt{n})]\,d\mu_1(x) \\ &= 1 - \sigma^2\xi^2/2n + O(|\xi|^{2+\alpha}/n^{1+\alpha/2})\end{aligned}$$

For each fixed ξ we therefore get

$$\lim_{n\to\infty}[1 - \sigma^2\xi^2/2n + O(1/n^{1+\alpha/2})]^n = e^{-\sigma^2\xi^2/2}$$

In other words, for all ξ

$$\lim_{n\to\infty}\hat{\mu}_n(\xi) = \hat{g}_\sigma(\xi)$$

Then we apply Proposition 2 and 3 to finish the proof.

Sums of random variables and convolution.

We can rephrase the central limit theorem in terms of convolution. For independent X_1 and X_2, suppose that $\nu_j = \mu_{X_j}$, then

$$\mathbb{E}(\varphi(X_1 + X_2)) = \int_{\mathbb{R}} \int_{\mathbb{R}} \varphi(x_1 + x_2) d\nu_1(x_1) d\nu_2(x_2)$$

In particular, if we define a measure by

$$\nu(I) = \int_{\mathbb{R}} \int_{\mathbb{R}} 1_I(x_1 + x_2) d\nu_1(x_1) d\nu_2(x_2) \tag{3}$$

for all intervals I, the $\nu = \mu_{X_1+X_2}$, the distribution of the sum $X_1 + X_2$.

Exercise. Show that if $d\nu_j = f_j(x)dx$, with $f_j \in L^1(\mathbb{R})$, then

$$d\nu = g(x)dx, \quad \text{with} \quad g = f_1 * f_2$$

This justifies turning (3) into the *definition* of $\nu = \nu_1 * \nu_2$.

Exercise. Show that under this definition,

$$\widehat{\nu_1 * \nu_2}(\xi) = \hat{\nu}_1 \hat{\nu}_2 \tag{4}$$

For a sequence of independent random variables, X_j,

$$\mu_{X_1+\cdots+X_n} = \mu_{X_1} * \mu_{X_2} * \cdots * \mu_{X_n}$$

For instance, if $\mu_{X_1} = (1/2)(\delta_{-1} + \delta_1)$, and the variables are i. i. d., then

$$\mu_{X_1+\cdots+X_n} = 2^{-n} \sum_{k=0}^{n} \binom{n}{k} \delta_{-n+2k}$$

Exercise. Express the central limit theorem (in this special case) in terms of this measure, and find the scaling under which this measure tends weakly to the gaussian distribution $g_1(x)dx$.

We can also consider the class of signed measures Borel measures $M(\mathbb{R})$, real-valued functions μ on Borel sets of $\mathbb{R}$ such that

$$\|\mu\| = \sup\{\sum_{j=1}^{\infty} |\mu(I_j)| : \bigcup I_j = \mathbb{R}, \ \text{(disjoint intervals)}\} < \infty$$

The norm $\|\mu\|$ represents the total mass of μ (also called the total variation). A signed measure can always be written as

$$\mu = \mu_+ - \mu_-$$

for two finite, positive measures $\mu_\pm$ such that, in addition, there are Borel sets $E_\pm$ such that $\mu_\pm(E_\pm) = \mu_\pm(\mathbb{R})$ and $\mu_+(E_-) = \mu_-(E_+) = 0$. (In this case we say μ_+ and μ_- are mutually singular, and write $\mu_+ \perp \mu_-$. This specifies the measures uniquely.) Using this decomposition, we can define $\mu * \nu$ for signed measures using linearity and (4), and it remains the case that Fourier transform of the convolution is the product of the Fourier transforms.

3 $\mathcal{S}'(\mathbb{R})$, the class of tempered distributions

Define a metric on the Schwartz class

$$\mathcal{S}(\mathbb{R}) = \{\varphi \in C^\infty(\mathbb{R}) : \|\varphi\|_{j,k} < \infty,\ j,\ k = 0,\ 1,\ 2,\ \dots\}$$

with seminorms[1]

$$\|\varphi\|_{j,k} = \sup_{x\in\mathbb{R}} |x^j \varphi^{(k)}(x)|$$

by

$$\operatorname{dist}(f, g) = \sum_{j,k=0}^{\infty} 2^{-j-k} \min(\|f-g\|_{j,k},\ 1)$$

We take the minimum with 1 so that the sum is finite. The seminorm property is used to confirm the triangle inequality so that this is a metric.

The dual space to $\mathcal{S}(\mathbb{R})$, denoted $\mathcal{S}'(\mathbb{R})$ is the set of continuous, linear functions $T : \mathcal{S} \to \mathbb{C}$. (The point of defining the metric was to say in what sense T is continuous.) This class is also known as the space of tempered distributions. (More generally, the class of distributions, are dual to the smaller function space $C_0^\infty(\mathbb{R})$. Because the tempered distributions have to act on Schwartz class functions, they have better behavior at infinity than general distributions. The more general distributions don't have well defined Fourier transforms because they can grow too fast at infinity.)

The most useful notion of convergence in $\mathcal{S}'$ is the same weak convergence we already discussed. We say that T_j tends to T if

$$\lim_{j\to\infty} T_j(\varphi) = T(\varphi)$$

[1]A seminorm satisfies $\|f\| \ge 0$, $\|f+g\| \le \|f\| + \|g\|$ and $\|cf\| = |c|\|f\|$, but not necessarily the last axiom: $\|f\| = 0$ need not imply that $f = 0$.

for all $\varphi \in \mathcal{S}$. (The technical reason why this is a useful kind of limit is that the existence of the limit point by point for each $\varphi \in \mathcal{S}(\mathbb{R})$ suffices in order that the limiting linear operator T be continuous in the $\mathcal{S}(\mathbb{R})$ metric.)

If $u(x)$ is a bounded measurable function or $u \in L^p(\mathbb{R})$, then we define

$$T_u(\varphi) = \int_{\mathbb{R}} u(x)\varphi(x)\,dx$$

for any $\varphi \in \mathcal{S}$. In this way we identify every such function u with a distribution. Measures are also identified with

$$T_\mu(\varphi) = \int_{\mathbb{R}} \varphi\,d\mu$$

Any linear continuous operation on $\mathcal{S}$ defines a corresponding (dual) operation on $\mathcal{S}'$. For example, we have calculated for $u \in \mathcal{S}$ and for $\varphi \in \mathcal{S}$ that

$$\int_R u\hat{\varphi}\,dx = \int_{\mathbb{R}} \hat{u}\varphi\,dx$$

This explains the *definition* of the Fourier transform of any $T \in \mathcal{S}'(\mathbb{R})$, namely

$$\hat{T}(\varphi) = T(\hat{\varphi})$$

For a measure, $\mu \in M^+(\mathbb{R})$ we could also define the Fourier transform by

$$\hat{\mu}(\xi) = \int_{\mathbb{R}} e^{-ix\xi}\,d\mu(x)$$

The two definitions are consistent. The proof is left as an exercise. But let's illustrate it by comparing the two definitions in one case, namely $\mu_0 = \delta$, the delta function (unit mass at 0). The correponding tempered distribution $T_0 \in \mathcal{S}'(\mathbb{R})$ is given by $T_0(\varphi) = \varphi(0)$. Then

$$\hat{\mu}_0(\xi) = \int_{\mathbb{R}} e^{-ix\xi}\,d\mu_0(x) = e^{-i0\xi} = 1$$

On the other hand,

$$\hat{T}_0(\varphi) = T_0(\hat{\varphi}) = \hat{\varphi}(0) = \int_{\mathbb{R}} \varphi(x)\,dx = \int_{\mathbb{R}} \varphi(x)u_0(x)\,dx$$

with $u_0(x) = 1$. In other words, the definitions are consistent, yielding $\hat{\mu}_0 \equiv 1$.

The idea of duality also leads to the definition of the derivative of a tempered distribution, namely,

$$T'(\varphi) = -T(\varphi')$$

This is motivated by the formula (in the special case $u \in \mathcal{S}$, $\varphi \in \mathcal{S}$,

$$\int_{\mathbb{R}} u'\varphi \, dx = -\int_{\mathbb{R}} u\varphi' \, dx,$$

which follows from integration by parts (or the product rule $(u\varphi)' = u'\varphi + u\varphi'$).

Exercise. Find $\widehat{T'}(\xi)$ directly from the definition in the case $T = \delta$. More examples and formulas are on PS11.

Fourier inversion on $\mathcal{S}'(\mathbb{R})$. The Fourier transform is a continuous linear mapping from $\mathcal{S}(\mathbb{R})$ to $\mathcal{S}(\mathbb{R})$ and it inverse is also a continuous mapping from $\mathcal{S}(\mathbb{R})$ to $\mathcal{S}(\mathbb{R})$ with the formula

$$\check{\varphi}(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \varphi(\xi)e^{ix\xi}d\xi$$

Proposition 4 *If $T \in \mathcal{S}'(\mathbb{R})$, and define S by*

$$S(\varphi) = T(\check{\varphi})$$

Then $S \in \mathcal{S}'(\mathbb{R})$ and $\hat{S} = T$. In other words, the mapping $T \mapsto \check{T}$ defined by

$$\check{T}(\varphi) = T(\check{\varphi})$$

inverts the Fourier transform on $\mathcal{S}'(\mathbb{R})$.

This proposition is a corollary of Fourier inversion on $\mathcal{S}(\mathbb{R})$ and the definition of Fourier transform on $\mathcal{S}'(\mathbb{R})$. Let $T \in \mathcal{S}'(\mathbb{R})$. The inverse Fourier tranform mapping $\varphi \mapsto \check{\varphi}$ is a continuous linear mapping from $\mathcal{S}(\mathbb{R}) \to \mathcal{S}(\mathbb{R})$, so S defined by

$$S(\varphi) = T(\check{\varphi})$$

is a continuous linear mapping from $\mathcal{S}(\mathbb{R}) \to \mathbb{C}$. In order to show that $\hat{S} = T$, recall that the Fourier inversion formula on $\mathcal{S}(\mathbb{R})$ says that for all $\eta \in \mathcal{S}(\mathbb{R})$, if $\varphi = \hat{\eta}$, then $\check{\varphi} = \eta$. Thus for all $\eta \in \mathcal{S}(\mathbb{R})$ (and denoting $\varphi = \hat{\eta}$)

$$\hat{S}(\eta) = S(\hat{\eta}) = S(\varphi) = T(\check{\varphi}) = T(\eta)$$