MTH 664 - Lectures 23 & 24

Yevgeniy Kovchegov
Oregon State University
Topics:

• The Fourier transform.

• Plancharel Theorem.

• The Uniqueness Theorem.

• The Convergence Theorem.

• The Centarl Limit Theorem.
Weak convergence in distribution.

We say that probability distributions \( \{\mu_n\}_{n=1,2,...} \) converge weakly if

\[
\lim_{n \to \infty} \int f(x) \, d\mu_n(x) = \int f(x) \, d\mu(x)
\]

for any bounded continuous function \( f(x) \).

If \( X_n \) are random variables distributed according to \( \mu_n \), and if \( X \) has distribution \( \mu \), then

\[
\lim_{n \to \infty} E[f(X_n)] = E[f(X)],
\]

and we say that \( X_n \) converges to \( X \) weakly or in distribution.

**Theorem.** Probability distributions \( \mu_n \) converge weakly to \( \mu \) if and only if

\[
\lim_{n \to \infty} \int f(x) \, d\mu_n(x) = \int f(x) \, d\mu(x)
\]

for all continuous functions \( f(x) \) with compact supports, i.e. \( \forall f \in C_c(\mathbb{R}^d) \).
Fourier transform.

If $X$ is a random variable with distribution $\mu$, then the Fourier transform of $\mu$ is defined as

$$\hat{\mu}(t) = \int_{\mathbb{R}} e^{itx} d\mu(x) = E[e^{itX}]$$

The function $\hat{\mu}(t)$ is called the characteristic function of $X$.

The Fourier transform of a Lebesgue-integrable function $f(x)$ is defined as

$$\hat{f}(t) = \int_{\mathbb{R}} e^{itx} f(x) dx$$
Fourier transform.

Observe that \( \hat{\mu}(t) = E[\cos(tX)] + iE[\sin(tX)] \).

Properties:

• If \( X_1, X_2, \ldots \) are independent random variables with distributions \( \mu_1, \mu_2, \ldots \mu_n \), then the characteristic function of the sum \( S_n = X_1 + \ldots + X_n \) is equal to the product of the corresponding characteristic functions,

\[
\hat{\mu}(t) = E[e^{itS_n}] = \hat{\mu}_1(t) \cdot \ldots \cdot \hat{\mu}_n(t)
\]

• If \( Z \) is a standard normal random variable, then

\[
\hat{\mu}(t) = E[e^{itZ}] = e^{-t^2/2}
\]

The above is proved via complex integration.
Fejér’s Theorem.

For $\epsilon > 0$ let

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

Fejér’s Theorem. $\forall f \in C_c(\mathbb{R})$ (continuous with compact support), $f \ast \phi_\epsilon(x)$ is infinitely differentiable whose $k$-th derivative equals $f \ast \phi_\epsilon^{(k)}(x)$, and

$$\lim_{\epsilon \to 0^+} \sup_{x \in \mathbb{R}} |f \ast \phi_\epsilon(x) - f(x)| = 0$$

Proof. In order to show the uniform convergence of $f \ast \phi_\epsilon(x)$ to $f(x)$ we observe for any given $x \in \mathbb{R}$,

$$f \ast \phi_\epsilon(x) = E[f(x - \epsilon Z)]$$

where $Z$ is a standard normal random variable.

Now, since $\forall f \in C_c(\mathbb{R})$, it is a bounded and uniformly continuous function. Thus, $|f(x)| < M$ ($\forall x \in \mathbb{R}$), and $\forall \delta > 0$ $\exists \epsilon > 0$ such that

$$|f(x) - f(y)| < \delta \quad \text{whenever} \quad |x - y| < \sqrt{\epsilon}$$
Fejér’s Theorem.

Proof. (continued) In order to show the uniform convergence of $f * \phi_\epsilon(x)$ to $f(x)$ we observe for any given $x \in \mathbb{R}$,

$$f * \phi_\epsilon(x) = E[f(x - \epsilon Z)],$$

where $Z$ is a standard normal random variable.

Now, since $\forall f \in C_c(\mathbb{R})$, it is a bounded and uniformly continuous function. Thus, $|f(x)| < M$ ($\forall x \in \mathbb{R}$), and $\forall \delta > 0 \exists \epsilon > 0$ such that $|f(x) - f(y)| < \delta$ whenever $|x - y| < \sqrt{\epsilon}$

Hence,

$$|f * \phi_\epsilon(x) - f(x)| \leq E\left[|f(x - \epsilon Z) - f(x)| \cdot 1_{|Z| < 1/\sqrt{\epsilon}}\right] + 2M \cdot P(|Z| \geq 1/\sqrt{\epsilon})$$

$$\leq \delta + 2M \cdot P(|Z| \geq 1/\sqrt{\epsilon})$$

and

$$\limsup_{\epsilon \to 0+} \sup_{x \in \mathbb{R}} |f * \phi_\epsilon(x) - f(x)| \leq \delta$$

for any choice of $\delta > 0$. \qed
The Plancharel’s Theorem.

If $\mu$ is a probability distribution (or any finite measure on $\mathbb{R}$), and if $f(x)$ is a Lebesgue-integrable function, then

$$\int_{\mathbb{R}} f \ast \phi_\epsilon(x) \, d\mu(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\epsilon^2 t^2/2} \hat{f}(t) \overline{\hat{\mu}(t)} \, dt$$

If in addition $f \in C_c(\mathbb{R})$, then

$$\int_{\mathbb{R}} f(x) \, d\mu(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(t) \overline{\hat{\mu}(t)} \, dt$$

Proof. By the Fubini-Tonelli theorem,

$$\frac{1}{2\pi} \int_{\mathbb{R}} e^{-\epsilon^2 t^2/2} \hat{f}(t) \overline{\hat{\mu}(t)} \, dt = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\epsilon^2 t^2/2} \left( \int_{\mathbb{R}} e^{itx} f(y) \, dy \cdot \int_{\mathbb{R}} e^{-itx} d\mu(x) \right) \, dt$$

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} \left( \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\epsilon^2 t^2/2} e^{it(y-x)} \, dt \right) f(y) \, dy \, d\mu(x) = \int_{\mathbb{R}} \int_{\mathbb{R}} \phi_\epsilon(y-x) f(y) \, dy \, d\mu(x)$$

Thus

$$\int_{\mathbb{R}} f \ast \phi_\epsilon(x) \, d\mu(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\epsilon^2 t^2/2} \hat{f}(t) \overline{\hat{\mu}(t)} \, dt$$
The Plancharel’s Theorem.

If $\mu$ is a probability distribution (or any finite measure on $\mathbb{R}$), and if $f(x)$ is a Lebesgue-integrable function, then

$$\int_{\mathbb{R}} f \ast \phi_\epsilon (x) \, d\mu(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\epsilon^2 t^2 / 2} \hat{f}(t) \overline{\hat{\mu}(t)} \, dt$$

If in addition $f \in C_c(\mathbb{R})$, then

$$\int_{\mathbb{R}} f(x) \, d\mu(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(t) \overline{\hat{\mu}(t)} \, dt$$

**Proof.** (continued) We proved

$$\int_{\mathbb{R}} f \ast \phi_\epsilon (x) \, d\mu(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\epsilon^2 t^2 / 2} \hat{f}(t) \overline{\hat{\mu}(t)} \, dt$$

Now, let $\epsilon \downarrow 0$, and use Fejér’s Theorem to conclude

$$\int_{\mathbb{R}} f(x) \, d\mu(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(t) \overline{\hat{\mu}(t)} \, dt$$

$\square$
The Uniqueness Theorem.

The Uniqueness Theorem. Suppose $\mu$ and $\nu$ are two probability distributions (in general, two finite measures) on $(\mathbb{R}, \mathcal{B})$. If $\hat{\mu} = \hat{\nu}$, then $\mu = \nu$.

Proof. By the Planchareel’s Theorem, if $\hat{\mu} = \hat{\nu}$, then for any $f \in C_c(\mathbb{R})$,

$$\int_{\mathbb{R}} f(x) \, d\mu(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(t) \, \hat{\mu}(t) \, dt = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(t) \, \hat{\nu}(t) \, dt = \int_{\mathbb{R}} f(x) \, d\nu(x)$$

Choose a sequence $f_k \in C_c(\mathbb{R})$ such that $f_k \downarrow 1_{[a,b]}$.

Then by the Monotone Convergence Theorem,

$$\mu([a,b]) = \nu([a,b])$$

Therefore $\mu(B) = \nu(B)$ for all Borel $B \in \mathcal{B}$. 

\[\square\]
The Convergence Theorem.

Suppose $\mu, \mu_1, \mu_2, \ldots$ are probability distributions on $(\mathbb{R}, \mathcal{B})$. If $\lim_{n \to \infty} \hat{\mu}_n = \hat{\mu}$ pointwise, then $\mu_n$ converges weakly to $\mu$, i.e. $\mu_n \Rightarrow \mu$.

Proof. It is sufficient to prove that for any $f \in C_c(\mathbb{R})$,

$$\lim_{n \to \infty} \int f(x) \, d\mu_n(x) = \int f(x) \, d\mu(x)$$

By the Fejér’s Theorem, $\forall \delta > 0$, $\exists \epsilon > 0$ such that

$$\sup_{x \in \mathbb{R}} |f * \phi_{\epsilon}(x) - f(x)| \leq \delta$$

Thus, by the triangle inequality,

$$\left| \int f(x) \, d\mu_n(x) - \int f(x) \, d\mu(x) \right| \leq 2\delta + \left| \int f * \phi_{\epsilon}(x) \, d\mu_n(x) - \int f * \phi_{\epsilon}(x) \, d\mu(x) \right|$$

$$= 2\delta + \left| \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\epsilon^2 t^2/2} \hat{f}(t) \left( \hat{\mu}_n(t) - \hat{\mu}(t) \right) dt \right|$$
The Convergence Theorem.

Proof. It is sufficient to prove that for any $f \in C_c(\mathbb{R})$, 
\[
\lim_{n \to \infty} \int f(x) \, d\mu_n(x) = \int f(x) \, d\mu(x)
\]
By the Fejér’s Theorem, $\forall \delta > 0, \exists \epsilon > 0$ such that 
\[
\sup_{x \in \mathbb{R}} \left| f \ast \phi_\epsilon(x) - f(x) \right| \leq \delta
\]
Thus, by the triangle inequality, 
\[
\left| \int f(x) \, d\mu_n(x) - \int f(x) \, d\mu(x) \right| \leq 2\delta + \left| \int f \ast \phi_\epsilon(x) \, d\mu_n(x) - \int f \ast \phi_\epsilon(x) \, d\mu(x) \right|
\]
\[
= 2\delta + \left| \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\epsilon^2 t^2 / 2} \hat{f}(t) \left( \hat{\mu}_n(t) - \hat{\mu}(t) \right) \, dt \right|
\]
Notice that if $f \in C_c(\mathbb{R})$, then $\hat{f}$ is uniformly bounded by $\|f\|_{L^1}$, and therefore 
\[
\limsup_{n \to \infty} \left| \int f(x) \, d\mu_n(x) - \int f(x) \, d\mu(x) \right| \leq 2\delta
\]
by DCT.
The Central Limit Theorem (CLT).

The Central Limit Theorem (CLT). Suppose $X_1, X_2, \ldots$ are i.i.d. random variables with mean $E[X_j] = \rho < \infty$ and variance $Var(X_j) = \sigma^2 < \infty$. Let $S_n = X_1 + \ldots + X_n$. Then the distribution $\nu_n$ of

$$\frac{S_n - n\rho}{\sigma\sqrt{n}}$$

converges weakly to the standard normal distribution $\nu$, i.e.

$$\nu_n \Rightarrow \nu.$$

Proof. W.l.o.g. assume $\rho = 0$ and $\sigma = 1$ for all $j$, as otherwise we can consider $\tilde{X}_j = \frac{X_j - \rho}{\sigma}$. Then, we need to show that the distribution $\nu_n$ of $\frac{S_n}{\sqrt{n}}$ converges weakly to the standard normal distribution $\nu$.

By the Convergence Theorem, it is sufficient to show that

$$\lim_{n \to \infty} \hat{\nu}_n(t) = \hat{\nu}(t) = e^{-t^2/2} \quad \text{pointwise} \quad \forall x \in \mathbb{R}.$$

Let $\mu$ denote the distribution of $X_j$. Then $\hat{\mu}(t) = E[e^{itX_j}]$ is its characteristic function.
The Central Limit Theorem (CLT).

Proof. (continued) W.l.o.g. assume $\rho = 0$ and $\sigma = 1$ for all $j$, as otherwise we can consider $\tilde{X}_j = \frac{X_j - \rho}{\sigma}$. Then, we need to show that the distribution $\nu_n$ of $\frac{S_n}{\sqrt{n}}$ converges weakly to the standard normal distribution $\nu$.

By the Convergence Theorem, it is sufficient to show that

$$\lim_{n \to \infty} \hat{\nu}_n(t) = \hat{\nu}(t) = e^{-t^2/2} \quad \text{pointwise} \quad \forall x \in \mathbb{R}.$$ 

Let $\mu$ denote the distribution of $X_j$. Then $\hat{\mu}(t) = E[e^{itX_j}]$ is its characteristic function.

Now, $e^{is} = 1 + is - \frac{1}{2}s^2 + R(s)$, where $|R(s)| \leq \min(|s|^3, s^2)$, and

$$\hat{\mu}(t) = E \left[ 1 + itX_j - \frac{t^2}{2}X_j^2 + R(t \cdot X_j) \right] = 1 - \frac{t^2}{2} + E[R(t \cdot X_j)]$$

$$\hat{\nu}_n(t) = E[e^{itS_n/\sqrt{n}}] = \left( \hat{\mu}(t/\sqrt{n}) \right)^n = \left( 1 - \frac{t^2}{2n} + E \left[ R(t \cdot X_j/\sqrt{n}) \right] \right)^n$$
The Central Limit Theorem (CLT).

**Proof.** (continued) By the Convergence Theorem, it is sufficient to show that

$$\lim_{n \to \infty} \hat{\nu}_n(t) = \hat{\nu}(t) = e^{-t^2/2} \quad \text{pointwise \ } \forall x \in \mathbb{R}.$$ 

Let $\mu$ denote the distribution of $X_j$. Then $\hat{\mu}(t) = E[e^{itX_j}]$ is its characteristic function.

Now, $e^{is} = 1 + is - \frac{1}{2}s^2 + R(s)$, where $|R(s)| \leq \min \left( |s|^3, s^2 \right)$, and

$$\hat{\mu}(t) = E \left[ 1 + itX_j - \frac{t^2}{2}X_j^2 + R(t \cdot X_j) \right] = 1 - \frac{t^2}{2} + E[R(t \cdot X_j)]$$

$$\hat{\nu}_n(t) = E \left[ e^{i\frac{tS_n}{\sqrt{n}}} \right] = \left( \hat{\mu} \left( \frac{t}{\sqrt{n}} \right) \right)^n = \left( 1 - \frac{t^2}{2n} + E \left[ R \left( \frac{t \cdot X_j}{\sqrt{n}} \right) \right] \right)^n$$

$$\ln \hat{\nu}_n(t) = n \cdot \ln \left( 1 - \frac{t^2}{2n} + E \left[ R \left( \frac{t \cdot X_j}{\sqrt{n}} \right) \right] \right),$$

where

$$\left| E \left[ R \left( \frac{t \cdot X_j}{\sqrt{n}} \right) \right] \right| \leq E \left[ \min \left( \frac{t^3 |X_j|^3}{n \sqrt{n}}, \frac{t^2 X_j^2}{n} \right) \right] = o \left( \frac{1}{n} \right)$$

for a fixed $t$, as $n \to \infty$. 
The Central Limit Theorem (CLT).

Proof. (continued) $e^{is} = 1 + is - \frac{1}{2}s^2 + R(s)$, where $|R(s)| \leq \min \left( |s|^3, s^2 \right)$, and

$$\hat{\mu}(t) = E \left[ 1 + itX_j - \frac{t^2}{2}X_j^2 + R(t \cdot X_j) \right] = 1 - \frac{t^2}{2} + E[R(t \cdot X_j)]$$

$$\hat{\nu}_n(t) = E \left[ e^{\frac{its_n}{\sqrt{n}}} \right] = \left( \hat{\mu}(t/\sqrt{n}) \right)^n = \left( 1 - \frac{t^2}{2n} + E \left[ R(t \cdot X_j/\sqrt{n}) \right] \right)^n$$

$$\ln \hat{\nu}_n(t) = n \cdot \ln \left( 1 - \frac{t^2}{2n} + E \left[ R(t \cdot X_j/\sqrt{n}) \right] \right),$$

where

$$\left| E \left[ R(t \cdot X_j/\sqrt{n}) \right] \right| \leq E \left[ \min \left( \frac{t^3|X_j|^3}{n\sqrt{n}}, \frac{t^2X_j^2}{n} \right) \right] = o \left( \frac{1}{n} \right)$$

for a fixed $t$, as $n \to \infty$ [to be shown on next slide].

Therefore,

$$\ln \hat{\nu}_n(t) = n \cdot \left( - \frac{t^2}{2n} + o \left( \frac{1}{n} \right) \right) \longrightarrow - \frac{t^2}{2}$$

and

$$\hat{\nu}_n(t) \longrightarrow e^{-t^2/2}$$
The Central Limit Theorem (CLT).

Proof. (continued) We need to show $E \left[ \min \left( \frac{t^3 |X_j|^3}{n \sqrt{n}}, \frac{t^2 X_j^2}{n} \right) \right] = o \left( \frac{1}{n} \right)$ for any fixed $t$, as $n \to \infty$. Equivalently, we need to establish

$$E \left[ \min \left( \frac{t^3 |X_j|^3}{\sqrt{n}}, t^2 X_j^2 \right) \right] \to 0 \quad \text{as} \quad n \to \infty$$

For any given $K > 0$ and $n \geq K^2 t^2$,

$$0 \leq \min \left( \frac{t^3 |x|^3}{\sqrt{n}}, t^2 x^2 \right) = 1_{|x| \leq \sqrt{n}/t} \cdot \frac{t^3 |x|^3}{\sqrt{n}} + 1_{|x| > \sqrt{n}/t} \cdot t^2 x^2 \leq 1_{|x| \leq K} \cdot \frac{t^3 |x|^3}{\sqrt{n}} + 1_{|x| > K} \cdot t^2 x^2$$

and

$$E \left[ \min \left( \frac{t^3 |X_j|^3}{\sqrt{n}}, t^2 X_j^2 \right) \right] \leq E \left[ 1_{|X_j| \leq K} \cdot \frac{t^3 |X_j|^3}{\sqrt{n}} \right] + E \left[ 1_{|X_j| > K} \cdot t^2 X_j^2 \right]$$

Therefore, $\forall K > 0$,

$$\limsup_{n \to \infty} E \left[ \min \left( \frac{t^3 |X_j|^3}{\sqrt{n}}, t^2 X_j^2 \right) \right] \leq E \left[ 1_{|X_j| > K} \cdot t^2 X_j^2 \right],$$

where $E \left[ 1_{|X_j| > K} \cdot t^2 X_j^2 \right] \to 0$ by the DCT as $K \to \infty$. $\square$
The Central Limit Theorem (CLT).

The Central Limit Theorem (Lindeberg’s version). For each $n \in \mathbb{N}$, let

$$X_{n,1}, X_{n,2}, \ldots, X_{n,k_n}$$

be independent random variables satisfying

$$E[X_{n,j}] = 0, \quad Var(X_{n,j}) = \sigma^2_{n,j} < \infty, \quad \sigma^2_{n,1} + \ldots + \sigma^2_{n,k_n} = 1,$$

and, for every $\epsilon > 0$,

$$\lim_{n \to \infty} \sum_{j=1}^{k_n} E\left[ X_{n,j}^2 \cdot 1_{|X_{n,j}| > \epsilon} \right] = 0,$$

the Lindeberg’s condition.

Then the distribution of $S_n = X_{n,1} + \ldots + X_{n,k_n}$ converges weakly to the standard normal distribution.