

STATS310A - Lecture 17

Persi Diaconis
Scribed by Michael Howes

11/18/21

Contents

1	Characteristic functions	1
1.1	Definitions	1
1.2	Properties	1
1.3	Motivation	2
2	Other modes of convergence	5
3	Cantor's diagonalization argument	6

1 Characteristic functions

Today we will be discussing the use of Fourier analysis in probability. In this area, the key objects of study are characteristic functions. The plan is to first define characteristic functions and state some of their properties. We will then use characteristic functions in some interesting examples and then we will prove the properties we used.

1.1 Definitions

Definition 1. Let μ be a probability measure on \mathbb{R} the *characteristic function* of μ is the function $\phi_\mu : \mathbb{R} \rightarrow \mathbb{C}$ given by

$$\phi_\mu(t) = \int_{\mathbb{R}} e^{itx} \mu(dx) = \int_{\mathbb{R}} \cos(tx) \mu(dx) + i \int_{\mathbb{R}} \sin(tx) \mu(dx).$$

Definition 2. If X is a random variable on \mathbb{R} with distribution μ , the *characteristic function* of X is defined to be $\phi_X = \phi_\mu$.

Note that if $X \sim \mu$, then

$$\phi_X(t) = \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX)] + i\mathbb{E}[\sin(tX)].$$

We will regularly interchange ϕ_μ and ϕ_X .

1.2 Properties

Some properties of characteristic functions are

- (1) The integral $\phi_X(t)$ is well defined for all t and X .
- (2) The function $|\phi_X|$ is bounded by 1 and $\phi_X(0) = 1$.

- (3) The function ϕ_X is continuous.

Proof. For all $t, h \in \mathbb{R}$ we have

$$\begin{aligned} |\phi_X(t+h) - \phi_X(t)| &\leq \mathbb{E} \left[\left| e^{i(t+h)X} - e^{itX} \right| \right] \\ &= \mathbb{E} \left[\left| e^{itX} \right| \left| e^{ihX} - 1 \right| \right] \\ &= \mathbb{E} \left[\left| e^{ihX} - 1 \right| \right]. \end{aligned}$$

By the dominated convergence theorem,

$$\mathbb{E} \left[\left| e^{ihX} - 1 \right| \right] \xrightarrow{h \rightarrow 0} 0.$$

Thus ϕ_X is in fact uniformly continuous. □

- (4) If X and Y are independent random variables, then

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t).$$

Equivalently, if μ and ν are two probability measures on \mathbb{R} , then

$$\phi_{\mu * \nu}(t) = \phi_\mu(t)\phi_\nu(t),$$

where $\mu * \nu$ is the convolution of μ and ν .

- (5) If the k^{th} moment $\mathbb{E}[X^k]$ exists, then ϕ is k times differentiable at 0 and

$$\mathbb{E}[X^k] = i^k \phi_X^{(k)}(0).$$

The converse is not necessarily true. The differentiability of ϕ_X does not imply the existence of moments.

- (6) If $\phi_\mu(t) = \phi_\nu(t)$ for all t , then $\mu = \nu$. This is called the uniqueness theorem.
 (7) Let μ_n be a sequence of measures, then $\mu_n \Rightarrow \mu$ if and only if $\phi_{\mu_n}(t) \rightarrow \phi_\mu(t)$ for all t . This is called the continuity theorem.

1.3 Motivation

Regarding characteristic functions, one might ask “what’s it about?”. One answer is that we defined $\mu_n \Rightarrow \mu$ by $\int f d\mu_n \rightarrow \int f d\mu$ for all continuous bounded f . The continuity theorem says we only need to consider the very special functions $f_t(x) = e^{itx}$. The usefulness of this can be seen in the following example.

Example 1. Let’s return to our card guessing ESP example. Suppose we have cards $\{1, 2, \dots, n\}$ which are shuffled and turned over one at a time. We have seen that with complete feedback there is probability $\frac{1}{n-j+1}$ of correctly guessing the j^{th} card. We previously considered examples where guessing card j correctly results in a constant reward. We will now consider what happens if the reward for guess card j is proportional to difficulty of guessing card j .

With this in mind, define independent random variables $(X_j)_{j=1}^\infty$ by

$$X_j = \begin{cases} j & \text{with probability } \frac{1}{j}, \\ 0 & \text{with probability } 1 - \frac{1}{j}. \end{cases}$$

Then $\mathbb{E}[X_j] = 1$ and

$$\text{Var}(X_j) = j^2 \left(\frac{1}{j} \left(1 - \frac{1}{j} \right) \right) = j - 1.$$

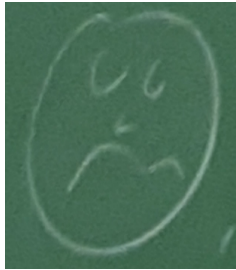
Next define $\tilde{X}_j = X_j - 1 = X_j - \mathbb{E}[X_j]$ so that \tilde{X}_j has mean 0. We can try to apply the central limit theorem to \tilde{X}_j . To do this we need to check Lindeberg's condition. With notation as in the statement of the central limit theorem we have

$$S_n = \sum_{j=1}^n \tilde{X}_j, \quad \sigma_j = \sqrt{\text{Var}(\tilde{X}_j)} = \sqrt{j-1} \quad \text{and} \quad s_n^2 = \sum_{j=1}^n \sigma_j^2 = \binom{n}{2}.$$

For $\varepsilon = \frac{\sqrt{2}}{2}$ we have

$$\begin{aligned} \frac{1}{s_n^2} \sum_{j=1}^n \int_{\{|\tilde{X}_j| > \varepsilon s_n\}} |\tilde{X}_j|^2 d\mathbb{P} &= \frac{2}{n(n-1)} \sum_{j=1}^n \int_{\{|\tilde{X}_j| > \varepsilon \sqrt{n(n-1)/2}\}} |\tilde{X}_j|^2 d\mathbb{P} \\ &= \frac{2}{n(n-1)} \sum_{j=1}^n \int_{\{|\tilde{X}_j| > \frac{1}{2} \sqrt{n(n-1)}\}} |\tilde{X}_j|^2 d\mathbb{P} \\ &\geq \frac{2}{n^2} \sum_{j=\lceil n/2 \rceil}^n \int_{\{|\tilde{X}_j| > \frac{1}{2} n\}} |\tilde{X}_j|^2 d\mathbb{P} \\ &= \frac{2}{n^2} \sum_{j=\lceil n/2 \rceil}^n \frac{1}{j} (j-1)^2 \\ &\not\rightarrow 0. \end{aligned}$$

This Lindeberg's condition does not hold for $\varepsilon = \frac{\sqrt{2}}{2}$ and thus our central limit theorem cannot be used:



But we can still use characteristic functions! To do this we will work with original random variables X_j . Note that

$$\begin{aligned} \mathbb{E}[e^{itX_j}] &= e^{itj} \frac{1}{j} + e^{-it \cdot 0} \left(1 - \frac{1}{j} \right) \\ &= 1 - \frac{1 - e^{itj}}{j}. \end{aligned}$$

Let $S_n = \sum_{j=1}^n X_j$. By independence we have

$$\begin{aligned}\mathbb{E}[e^{itS_n}] &= \prod_{j=1}^n \mathbb{E}[e^{itX_j}] \\ &= \prod_{j=1}^n \left(1 - \frac{1 - e^{itj}}{j}\right) \\ &= \exp\left(\sum_{j=1}^n \log\left(1 - \frac{1 - e^{itj}}{j}\right)\right).\end{aligned}$$

Thus

$$\begin{aligned}\phi_{\frac{S_n}{n}}(t) &= \mathbb{E}\left[e^{it\frac{S_n}{n}}\right] \\ &= \exp\left(\sum_{j=1}^n \log\left(1 - \frac{1 - e^{it\frac{j}{n}}}{j}\right)\right) \\ &\sim \exp\left(-\sum_{j=1}^n \frac{1 - e^{it\frac{j}{n}}}{j}\right),\end{aligned}$$

since for large j , $\frac{1 - e^{it\frac{j}{n}}}{j}$ is close to zero and so we can use the approximation $\log(1 - x) \approx -x$ for small x . For small values of j , we have $e^{it\frac{j}{n}} \approx 1$ and so

$$\log\left(1 - \frac{1 - e^{it\frac{j}{n}}}{j}\right) \approx \log(1) = 0 \approx \frac{1 - e^{-t\frac{j}{n}}}{j}.$$

Note that

$$\sum_{j=1}^n \frac{1 - e^{it\frac{j}{n}}}{j} = \frac{1}{n} \sum_{j=1}^n \frac{e^{it\frac{j}{n}} - 1}{\frac{j}{n}},$$

is a Riemann sum that approximates the integral $\int_0^1 \frac{1 - e^{itx}}{x} dx$. And thus for all $t \in \mathbb{R}$ we have

$$\phi_{\frac{S_n}{n}}(t) \rightarrow \exp\left(-\int_0^1 \frac{1 - e^{itx}}{x} dx\right).$$

Thus $\frac{S_n}{n} \rightarrow F_\infty$ where F_∞ is the distribution satisfying

$$\phi_{F_\infty}(t) = \exp\left(-\int_0^1 \frac{1 - e^{itx}}{x} dx\right).$$

Remark 1. Persi has a story about the distribution F_∞ . There are three classes of distribution which arise as the limit of sums of independent random variables.

- (1) *Stable laws* are limits $\frac{S_n - a_n}{b_n} \Rightarrow F$ where $S_n = X_1 + X_2 + \dots + X_n$ where $\{X_j\}$ are independent and identically distributed.
- (2) *Infinitely divisible laws* are limits $\frac{S_n - a_n}{b_n} \Rightarrow F$ where $S_n = X_{n,1} + X_{n,2} + \dots + X_{n,k_n}$ and $\{X_{n,j}\}$ is a triangular array with independent rows.
- (3) *Class L laws* are limits $\frac{S_n - a_n}{b_n} \Rightarrow F$ where $S_n = X_1 + \dots + X_n$ and $\{X_j\}$ are independent.

Note that (1) \subseteq (3) \subseteq (2). The distribution F_∞ from the previous example is a class L law. There was much debate amongst probabilists about whether all class L laws are unimodal. Proofs with mistakes were published, F_∞ was incorrectly claimed to be bimodal but eventually it was proved that all class L laws are unimodal. The paper that settled the score was read very closely.

2 Other modes of convergence

In addition to convergence in distribution $F_n \Rightarrow F$, we also have the following two forms of convergence.

Definition 3. Let X_n, X be random variables on a probability space $(\Omega, \mathbb{P}, \mathcal{F})$. We say that X_n *converges in probability* to X if for every $\varepsilon > 0$,

$$\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0.$$

We will denote convergence in probability by $X_n \xrightarrow{\mathbb{P}} X$.

Definition 4. Again let X_n, X be random variables on a probability space $(\Omega, \mathbb{P}, \mathcal{F})$. We say that X_n *converges almost surely* to X if

$$\mathbb{P}(\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}) = 1.$$

We will denote almost sure convergence by $X_n \xrightarrow{a.s.} X$.

Note that almost sure convergence is equivalent to for all $\varepsilon > 0$

$$\mathbb{P}(|X_n - X| > \varepsilon \text{ i.o.}) = 0,$$

which shows that almost sure convergence implies convergence in probability. The converse is not true as the following examples show.

Example 2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the interval $[0, 1]$ with the Borel σ -algebra and Lebesgue measure. Define

$$\begin{aligned} X_1 &= \delta_{[0, 1/2]}, \quad X_2 = \delta_{[1/2, 1]} \\ X_3 &= \delta_{[0, 1/4]}, \quad X_4 = \delta_{[1/4, 1/2]}, \quad X_5 = \delta_{[1/2, 3/4]}, \quad X_6 = \delta_{[3/4, 1]}, \\ X_7 &= \delta_{[0, 1/8]}, \quad \dots, \\ &\dots \end{aligned}$$

Then $\mathbb{P}(|X_n| > \varepsilon) \rightarrow 0$ for all $\varepsilon > 0$ since $\mathbb{P}(|X_n| > 0) \leq 2^{-\lfloor \log_2(n) \rfloor}$. However for every $\omega \in \Omega$, $X_n(\omega) = 1$ infinitely often. Thus $X_n \xrightarrow{\mathbb{P}} 0$ but X_n does not converge almost surely to 0.

Example 3. Define

$$f(x) = \begin{cases} \frac{\log(2) \log(|x|+1)}{x^2 \log(|x|)^2} & \text{if } |x| > 2, \\ 0 & \text{if } |x| \leq 2. \end{cases}$$

The function f is a probability density with respect to Lebesgue measure on \mathbb{R} . Let X_1, X_2, \dots be an i.i.d. sequence with density f . Define $S_n = \sum_{j=1}^n X_j$, then $\frac{S_n}{n} \xrightarrow{\mathbb{P}} 0$ but $\mathbb{E}[X_1]$ is not defined and $\frac{S_n}{n}$ does not converge to 0 almost surely. Thus this is an example when the weak law of large numbers holds but the strong law does not hold. See these [lecture notes](#) (the function F there is the CDF of a random variable with density f).

In general convergence in distribution is weaker than convergence in probability but if X_n converges in distribution to a constant c , then X_n converges in probability to c . Another result relating the different modes of convergence is the following.

Theorem 1 (Slutsky). *Let X_n, Y_n, Z be random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that $X_n \Rightarrow Z$ and $X_n - Y_n \xrightarrow{\mathbb{P}} 0$, then*

$$Y_n \Rightarrow Z.$$

We have already implicitly used this result many times when manipulating limits as in Example 1 of this lecture.

Proof. Let F_Z be the CDF of Z and let x be a continuity point of F . Given $\delta > 0$ choose $\varepsilon > 0$ so that $|F_Z(y) - F_Z(x)| < \delta$ for all y such that $|y - x| < 2\varepsilon$. Now choose y' and y'' which are continuity points of F_Z and satisfy

$$x - 2\varepsilon < y' < x - \varepsilon,$$

and

$$x + \varepsilon < y'' < x + 2\varepsilon.$$

Suppose that $\omega \in \{Y_n \leq x\}$. If $\omega \notin \{X_n \leq y''\}$, then

$$|X_n(\omega) - Y_n(\omega)| \geq y'' - x > x + \varepsilon - x = \varepsilon.$$

Thus we have the inclusion

$$\{Y_n \leq x\} \subseteq \{X_n \leq y''\} \cup \{|X_n - Y_n| > \varepsilon\}.$$

Similarly suppose that $\omega \in \{X_n \leq y'\}$. If $\omega \notin \{Y_n \leq x\}$, then

$$|X_n(\omega) - Y_n(\omega)| \geq x - y' > x - x + \varepsilon = \varepsilon.$$

And so

$$\{X_n \leq y'\} \subseteq \{Y_n \leq x\} \cup \{|X_n - Y_n| > \varepsilon\}.$$

Thus we have

$$\begin{aligned} \mathbb{P}(X_n \leq y') - \mathbb{P}(|X_n - Y_n| > \varepsilon) &\leq \mathbb{P}(Y_n \leq x) \\ &\leq \mathbb{P}(X_n \leq y'') + \mathbb{P}(|X_n - Y_n| > \varepsilon). \end{aligned}$$

Since $X_n - Y_n \xrightarrow{\mathbb{P}} 0$ and $X_n \Rightarrow F$ we thus have

$$F_Z(y') \leq \liminf_n \mathbb{P}(Y_n \leq x) \leq \overline{\lim}_n \mathbb{P}(Y_n \leq x) \leq F_Z(y'').$$

However $|y' - x|, |y'' - x| \leq 2\varepsilon$ and so we have

$$F_Z(x) - \delta \leq \liminf_n \mathbb{P}(Y_n \leq x) \leq \overline{\lim}_n \mathbb{P}(Y_n \leq x) \leq F_Z(x) + \delta.$$

Thus

$$\lim_n \mathbb{P}(Y_n \leq x) = F_Z(x),$$

and so $Y_n \Rightarrow Z$. □

3 Cantor's diagonalization argument

We will end this lecture by developing one of the tools that will help us prove the uniqueness and continuity theorems for characteristic functions.

Theorem 2. Let $(x_{n,j})_{n,j=1}^\infty$ be a double sequence of real numbers,

$$\begin{array}{cccc} x_{1,1} & x_{1,2} & x_{1,3} & \dots \\ x_{2,1} & x_{2,2} & x_{2,3} & \dots \\ x_{3,1} & x_{3,2} & x_{3,3} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

such that each row $(x_{n,j})_{j=1}^\infty$ is bounded. Then there exists a sequence $(l_n)_{n=1}^\infty$ and an increasing sequence of natural numbers $(j_k)_{k=1}^\infty$ such that for all n

$$\lim_{k \rightarrow \infty} x_{n,j_k} = l_n.$$

Proof. First consider the sequence $(x_{1,j})$. This is a bounded sequence of real numbers. By the Heine-Borel theorem, $(x_{1,j})$ contains a convergent subsequence $(x_{1,j_{1,k}})$ such that

$$\lim_{k \rightarrow \infty} x_{1,j_{1,k}} = l_1.$$

The sequence $(x_{2,j_{1,k}})$ is also a bounded sequence of real numbers. Thus there exists a further subsequence $(x_{2,j_{2,k}})$ such that

$$\lim_{k \rightarrow \infty} x_{2,j_{2,k}} = l_2.$$

We can repeat this with each row which gives us the following. For each $n \geq 2$ we have a sequence $(j_{n,k})$ which is a subsequence of $(j_{n-1,k})$ such that

$$\lim_{k \rightarrow \infty} x_{n,j_{n,k}} = l_n.$$

We next take the diagonal of these sequence of subsequences. That is, define $j_k := j_{k,k}$. For every n , the tail of (j_k) is a subsequence of $(j_{n,k})$ and so $\lim_k x_{n,j_k} = l_n$ for every n . \square