Lecture 17. More about characteristic functions. Weak convergence of distributions.

Now let us consider the relation between characteristic functions and the moments.

Theorem 17.1. *Let* ξ *be a real-valued random variable with finite expectation:* $E|\xi| < \infty$. *Then the characteristic function* $f(t) = f_\xi(t)$ *is differentiable, and*

$$f'(0) = i \cdot E\xi. \tag{17.1}$$

Proof. We have:

$$f'(0) = \lim_{h\to 0} \frac{f(h) - f(0)}{h} = \lim_{h\to 0} \frac{Ee^{ih\xi} - Ee^{i0\xi}}{h} = \lim_{h\to 0} E\frac{e^{ih\xi} - e^{i0\xi}}{h}. \tag{17.2}$$

As $h \to 0$, the ratio $\frac{e^{ih\xi} - e^{i0\xi}}{h}$ converges, for every $\omega \in \Omega$, to the derivative of the function $e^{it\xi}$ at $t = 0$, which is equal to $i\xi$. All random variables $\frac{e^{ih\xi} - e^{i0\xi}}{h}$, for every $h \neq 0$ and every $\omega \in \Omega$, are dominated by the random variable $|\xi|$:

$$\left|\frac{e^{ih\xi} - e^{i0\xi}}{h}\right| \leq |\xi| \tag{17.3}$$

($|e^{ih\xi} - e^{i0\xi}|$ is the distance between the two points $e^{ih\xi}$ and $e^{i0\xi}$ on the unit circle in the complex plane along the straight line, while $|h||\xi|$ is the distance between them along the arc of the circle); so by the dominated-convergence theorem (Theorem 14.3) we have:

$$f'(0) = \lim_{h\to 0} E\frac{e^{ih\xi} - e^{i0\xi}}{h} = E \lim_{h\to 0} \frac{e^{ih\xi} - e^{i0\xi}}{h} = Ei \cdot \xi = i \cdot E\xi. \tag{17.4}$$

Just the same way we prove also that for every $t \in \mathbb{R}^1$

$$f'(t) = i \cdot E(\xi \cdot e^{it\xi}. \tag{17.5}$$

Theorem 17.2. *Suppose* $E|\xi|^k < \infty$. *Then the characteristic function is* k *times differentiable, and*

$$f^{(k)}(0) = i^k \cdot E\xi^k. \tag{17.6}$$

Proof. Differentiating formula (17.5), we get:

$$f''(t) = \lim_{h\to 0} \frac{i \cdot E(\xi e^{i(t+h)\xi}) - i \cdot E(\xi e^{it\xi}}{h}) = \lim_{h\to 0} i \cdot E\frac{\xi e^{i(t+h)\xi} - \xi e^{it\xi}}{h}; \tag{17.7}$$

$$\left|\frac{\xi e^{i(t+h)\xi} - \xi e^{it\xi}}{h}\right| \leq |\xi|^2, \tag{17.8}$$

and by the dominated-convergence theorem

$$f''(t) = E \lim_{h\to 0} i \cdot \frac{\xi e^{i(t+h)\xi} - \xi e^{it\xi}}{h} = E(i^2 \xi^2 e^{it\xi}); \tag{17.9}$$

etc. Taking these derivatives at $t = 0$, we get (17.6).

For ξ having the normal distribution with parameters (a, b) we have $f(t) = e^{-bt^2/2+iat}$,

$$f'(t) = e^{-bt^2/2+iat} \cdot (-bt + ia), \qquad f''(t) = e^{-bt^2/2+iat} \cdot [(-bt + ia)^2 - b], \tag{17.10}$$

$$E\xi = a, \quad E\xi^2 = a^2 + b, \qquad \mathrm{Var}(\xi) = E\xi^2 - (E\xi)^2 = b. \tag{17.11}$$

So the meaning of the two parameters of the one-dimensional normal distribution is the expectation and the variance.

Theorem 17.3. *If the characteristic function is twice differentiable at* $t = 0$, *then the second moment* $E\xi^2$ *is finite* (and, by Theorem 17.2, $E\xi^2 = -f''(0)$).

Proof. If a function $f(t)$ is twice differentiable at some point t_0, its second derivative at this point can be found as

$$\lim_{h\to 0} \frac{f(t_0 + h) - 2f(t_0) + f(t_0 - h)}{h^2}. \tag{17.12}$$

This is a fact from Analysis, and my general policy is not to give proofs of such results; but here is the proof:

Let us write the Taylor formula with terms up to the second derivative:

$$f(t_0 + h) = f(t_0) + f'(t_0) \cdot h + \tfrac{1}{2} f''(t_0) \cdot h^2 + o(h^2) \qquad (h \to 0). \tag{17.13}$$

Here $o(h^2)$ denotes a function that goes to 0 faster than h^2 as $h \to 0$; that is,

$$\lim_{h\to 0} \frac{f(t_0 + h) - f(t_0) - f'(t_0) \cdot h - \frac{1}{2} f''(t_0) \cdot h^2}{h} = 0. \tag{17.14}$$

The equality (17.14) can be obtained by using l'Hôpital's rule a couple of times.

Of course, equality (17.13) can also be written for $-h$ instead of h:

$$f(t_0 - h) = f(t_0) - f'(t_0) \cdot h + \tfrac{1}{2} f''(t_0) \cdot h^2 + o(h^2) \qquad (h \to 0). \tag{17.15}$$

Putting this together with (17.13) in (17.12), we obtain that the limit (17.12) is equal to $f''(t_0)$.

Taking $t_0 = 0$, we have:

$$f''(0) = \lim_{h\to 0} \frac{f(h) - 2f(0) + f(-h)}{h^2} = \lim_{h\to 0} \frac{Ee^{ih\xi} - 2 + Ee^{-ih\xi}}{h^2} = -2 \lim_{h\to 0} E \frac{1 - \cos h\xi}{h^2}. \tag{17.16}$$

By Fatou's Lemma (Theorem 14.5), we have:

$$E \lim_{h\to 0} \frac{1-\cos h\xi}{h^2} = E \lim_{h\to 0} \frac{1-\cos h\xi}{h^2} = \frac{\xi^2}{2} \\ \le \lim_{h\to 0} E \frac{1-\cos h\xi}{h^2} = \lim_{h\to 0} E \frac{1-\cos h\xi}{h^2} = -\frac{f''(0)}{2} < \infty. \tag{17.17}$$

And now we apply Theorem 17.2 with $k = 2$.

So, for $k = 2$ twice-differentiability of the characteristic function f_ξ at zero is equivalent to the existence of the moment $E\xi^2$. Is this so for $k \ne 2$: does $E|\xi|^k < \infty$ follow from the characteristic function being k times differentiable? I am giving a non-obligatory problem **33*** (with $k = 1$).

The multidimensional version of Theorem 17.2:

Theorem 17.4. *Let* ξ_1, ..., ξ_n *be random variables, and* $E(|\xi_1|^{j_1} \cdot ... \cdot |\xi_n|^{j_n}) < \infty$ *for* $0 \le j_1 \le k_1$, ..., $0 \le j_n \le k_n$. *Then*

$$\frac{\partial^{k_1+...+k_n} f}{(\partial t_1)^{k_1} ... (\partial t_n)^{k_n}}(0, ..., 0) = i^{k_1+...+k_n} \cdot E(\xi_1^{k_1} \cdot ... \cdot \xi_n^{k_n}). \tag{17.18}$$

Applying this to a random vector $\boldsymbol{\xi}$ having the normal distribution with parameters $(\boldsymbol{a}, B)$:

$$\frac{\partial \exp\{-\frac{1}{2}\sum_{k,\,l=1}^n b_{kl}t_kt_l + i\sum_{k=1}^n a_kt_k\}}{\partial t_j} \\ = \exp\{-\tfrac{1}{2}\sum_{k,\,l=1}^n b_{kl}t_kt_l + i\sum_{k=1}^n a_kt_k\} \cdot (-\sum_{k=1}^n b_{jk}t_k + ia_j) \tag{17.19}$$

($\frac{1}{2}$ disappears, because for $k \ne l$ we have to take into account the summands with $k = j$, and those with $l = j$; and for $l = k$ we differentiate t_j^2),

$$\frac{\partial^2 \exp\{-\frac{1}{2}\sum_{k,\,l=1}^n b_{kl}t_kt_l + i\sum_{k=1}^n a_kt_k\}}{\partial t_j \partial t_s} = \exp\{-\tfrac{1}{2}\sum_{k,\,l=1}^n b_{kl}t_kt_l + i\sum_{k=1}^n a_kt_k\} \times \\ \times \big[(-\sum_{k=1}^n b_{jk}t_k + ia_j) \cdot (-\sum_{l=1}^n b_{sl}t_s + ia_s) - b_{js}\big], \tag{17.20}$$

$$E\xi_j = \frac{1}{i} \cdot ia_j = a_j, \qquad E(\xi_j\xi_s) = -\frac{\partial^2 f}{\partial t_j \partial t_s}(0, ..., 0) = a_ja_s + b_{js}, \tag{17.21}$$

$$\mathrm{Cov}(\xi_j, \xi_s) = E(\xi_j - E\xi_j)(\xi_s - E\xi_s) = E(\xi_j\xi_s) - E\xi_j \cdot E\xi_s = b_{js}. \tag{17.22}$$

That is, the first parameter, $\boldsymbol{a}$, is the vector of expectations, and the second, B, is *the matrix of covariances* (the covariance matrix).

Just as sometimes we have to consider convergence of sequences of random variables – in fact, different types of convergence – sometimes we have to consider convergence of sequences of *distributions*. In this case too, we can consider different types of convergence of random variables; here we are going to consider only one type of convergence: *weak convergence* of distributions.

Let μ_1, μ_2, ..., μ_n, ... be a sequence of distributions. We say that this sequence converges weakly as $n \to \infty$ to a distribution ν (notation: $\mu_n \to_w \nu$, or: $(w) \lim_{n\to\infty} \mu_n = \nu$) if for every bounded continuous function $f(x)$

$$\lim_{n\to\infty} \int f(x)\ \mu_n(dx) = \int f(x)\ \nu(dx). \tag{17.23}$$

Integrals over what?? As a matter of fact, we have introduced a convention: if a sum, or an integral, or whatever (say, a product) is taken, without any mention of the set over which the variable (x in the present case) runs, we take it over *all* possible values of the variable.

Note that we haven't specified in what space we consider our distributions.

If it is the real line, we take the integral over $\mathbb{R}^1$, i.e. from $-\infty$ to ∞; if it is n-dimensional distributions we are talking about (distributions of n-vectors), we take the integrals over $\mathbb{R}^n$. But also we can consider weak convergence of distributions in an arbitrary metric (or topological) space X (with its Borel σ-algebra $\mathcal{B}_X$): in this case too it makes sense considering continuity of functions $f(x)$ (and their *boundedness* makes sense in an *arbitrary* space X).

Most of what follows holds for wide classes of metric spaces; but for simplicity and concreteness we'll restrict ourselves to distributions on the real line (with going to $\mathbb{R}^n$ only to say that in the n-dimensional space this is exactly the same).

Example 17.1. The binomial distribution with parameters (n, p), where n is a natural number, and $0 \le p \le 1$, is, by definition, the discrete distribution concentrated at the points 0, 1, 2, ..., k, ..., $n-1$, n, with

$$\mu\{k\} = \binom{n}{k}\ p^k (1-p)^{n-k}, \qquad k = 0, 1, 2, ..., n-1, n, \tag{17.24}$$

where $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$ is the number of combinations of size k out of n. (Such will be the distribution of the number of successes in n Bernoulli trials with probability of success in one trial equal to p; i.e., the distribution of the sum of n independent random variables ξ_1, ..., ξ_n with $P\{\xi_i = 1\} = p$, $P\{\xi_i = 0\} = 1 - p\}$.)

Let μ_n, for every natural n, be the binominal distribution with parameters (n, p_n), where $p_n \to 0$ $(n \to \infty)$, $\lim_{n\to\infty} n \cdot p_n = a \in [0, \infty)$. We have for every integer $k \ge 0$:

$$\begin{aligned}\lim_{n\to\infty} \mu_n\{k\} &= \lim_{n\to\infty} \binom{n}{k}\ p_n^k (1-p_n)^{n-k} \\ &= \lim_{n\to\infty} \frac{n\cdot(n-1)\cdot\ldots\cdot(n-k+1)}{k!} \cdot p^n (1-p_n)^{n-k} \\ &= \frac{1}{k!} \prod_{j=0}^{k-1} [(n-j)p_n] \cdot \frac{(1-p_n)^n}{(1-p_n)^k} = \frac{1}{k!} \cdot a^k \cdot \frac{e^{-a}}{1^k} = \frac{a^k e^{-a}}{k!}.\end{aligned} \tag{17.25}$$

We see that this is the probability mass function of the Poisson distribution with parameter a.

Using the result of Problem **38** (or not using it – if you *dis*prove its statement), we can prove that the binomial distribution with parameters $(n,\, p_n)$ converges weakly to the Poisson distribution with parameter a as $n \to \infty$ if $n \cdot p_n \to a$.

Another example: μ_n is the normal distribution with parameters $(a_n,\, b_n)$; suppose $a_n \to a$, $b_n \to b$ as $n \to \infty$. Let us prove that $\mu_n \to_w \nu$ $(n \to \infty)$, where ν is the normal distribution with parameters $(a,\, b)$.

First consider the case of b being positive (according to our new definition, we can also consider the normal distribution with the second parameter equal to 0).

We have to prove that for every bounded continuous function $f(x)$

$$\int_{-\infty} f(x) \cdot \frac{1}{\sqrt{2\pi b_n}}\, e^{-(x-a_n)^2/2b_n}\, dx \to \int_{-\infty} f(x) \cdot \frac{1}{\sqrt{2\pi b}}\, e^{-(x-a)^2/2b}\, dx. \tag{17.26}$$

The integrand in the left-hand side integral converges to that in the right-hand side for every $x \in (-\infty, \infty)$; so to prove (17.26) it is enough to establish that all integrands (say, starting with some $n = n_0$) are dominated by some integrable function. Take some positive $\underline{b} < b = \lim_{n\to\infty} b_n$. There exists a constant C such that for all $n \geq n_0$ and all $x \in (-\infty,\, \infty)$

$$\frac{1}{\sqrt{2\pi b_n}}\, e^{-(x-a_n)^2/2b_n} \leq C \cdot e^{-x^2/\underline{b}} \tag{17.27}$$

(of course we have to *prove* that such a constant C exists, but I omit the proof). Now, all integrands $f(x) \cdot \dfrac{1}{\sqrt{2\pi b_n}}\, e^{-(x-a_n)^2/2b_n}$ with $n \geq n_0$ are dominated in absolute value by the function $\sup_x |f(x)| \cdot C \cdot e^{-x^2/2\underline{b}}$ (remember, the function $f(x)$ was supposed bounded, i. e., $\sup_x |f(x)| < \infty$), and this function is integrable.

Note that we have used only the boundedness of the function $f(x)$, but not its continuity.

Take on the case of $b = 0$ in the following problem:

39 Let μ_n be the normal distribution with parameters (a_n, b_n), $b_n > 0$; suppose $a_n \to a$, $b_n \to 0$ as $n \to \infty$.

Prove that $\mu_n \to_w \delta_a$ $(n \to \infty)$, where δ_a is the distribution concentrated at the point a: $\delta_a(C) = 1$ if $C \ni a$, and $= 0$ if $C \not\ni a$.

Before now we considered only *distributions*: probability measures on $(\mathbb{R}^1, \mathcal{B}^1)$; and did not mention random variables *whose* distributions they are.

Suppose we have a sequence of random variables ξ_1, ξ_2, ..., ξ_n, ..., and another random variable η. By definition, the sequence of their distributions μ_{ξ_1}, μ_{ξ_2}, ..., μ_{ξ_n}, ... converges weakly to the distribution μ_η: $\mu_{\xi_n} \to_w \mu_\eta$ $(n \to \infty)$ if and only if for every bounded continuous function $f(x)$

$$\int_{-\infty}^{\infty} f(x)\ \mu_{\xi_n}(dx) \to \int_{-\infty}^{\infty} f(x)\ \mu_\eta(dx) \qquad (n \to \infty). \tag{17.28}$$

But the integrals here can be rewritten as expectations: $\mu_{\xi_n} \to_w \mu_\eta$ if and only if for every bounded continuous $f(x)$

$$Ef(\xi_n) \to Ef(\eta) \qquad (n \to \infty). \tag{17.29}$$

Theorem 17.5. *If* $\xi_n \to \eta$ *almost surely as* $n \to \infty$, *then the distribution* $\mu_{\xi_n} \to_w \mu_\eta$ $(n \to \infty)$.

Proof. We have to prove (17.29) for every bounded continuous $f(x)$. Since f is continuous, we have, almost surely:

$$f(\xi_n) \to f(\eta) \qquad (n \to \infty). \tag{17.30}$$

All these random variables are dominated by a random variable having a finite expectation, namely with the constant $\sup_x |f(x)| < \infty$; so by the dominated-convergence theorem we have (17.29).