Lecture 20. Weak convergence and characteristic functions, continued.

We'll say that a family $\mathfrak{M} = \{\mu\}$ of probability distributions is *weakly precompact* if for every sequence $\mu_1, \mu_2, ..., \mu_n, ...$ of distributions belonging to $\mathfrak{M}$ there exists a subsequence $\mu_{n_1}, \mu_{n_2}, ..., \mu_{n_k}, ..., n_1 < n_2 < ... < n_k < ...$, and a probability distribution ν such that

$$(w) \lim_{k \to \infty} \mu_{n_k} = \nu \tag{20.1}$$

(we would call this family *compact* if it converged necessarily to a distribution also belonging to $\mathfrak{M}$).

Examples: Of course, every *finite* family of distributions is weakly precompact: in this case in a sequence $\mu_1, \mu_2, ..., \mu_n, ...$ some distribution μ will be repeated infinitely many times, and this subsequence converges to μ. Also every weakly convergent series of distributions is precompact. Another example: The family $\mathfrak{N}_{CD}$ of all normal distributions with parameters (a, b), $|a| \le C$, $b \le D$, is weakly precompact: from every sequence (a_n, b_n), $|a_n| \le C$, $b \le D$, we can extract a subsequence (a_{n_k}, b_{n_k}) converging to some (a_∞, b_∞), and by what was said about our example in Lecture 17, the corresponding normal distributions converge weakly to the normal distribution with parameters (a_∞, b_∞).

On the other hand, the family of normal distributions with parameters $(0, n)$ is *not* precompact: if it were, the sequence of distribution functions $F_{n_k}(x)$ would converge, by Theorem 18.2, to some *distribution function* $G(x)$ at all points at which it is continuous; while in fact it converges (for all x) to $1/2$, which is *not* a distribution function.

Theorem 20.1. *A family $\mathfrak{M}$ of probability distributions is weakly precompact if and only if for every positive ε there exists a $C < \infty$ such that for every $\mu \in \mathfrak{M}$*

$$\mu[-C, C] > 1 - \varepsilon \qquad \text{(or, which is the same, } \mu([-C, C]^c) < \varepsilon). \tag{20.2}$$

Proof. Let us prove the "only if" part first: suppose that *not* for every positive ε there exists a $C < \infty$ such that for every $\mu \in \mathfrak{M}$ (20.2) is satisfied. This means that there *exists* a positive ε such that *for every* C there *exists* a $\mu = \mu_C \in \mathfrak{M}$ such that

$$\mu[-C, C] \le 1 - \varepsilon. \tag{20.3}$$

Let us take such an $\varepsilon > 0$.

Let us choose a sequence of distributions $\mu_n \in \mathfrak{M}$ (some of distributions in this sequence may coincide). Namely, we take $C = n$; and we take as μ_n an arbitrary distribution belonging to $\mathfrak{M}$ for which

$$\mu_n[-n, n] \le 1 - \varepsilon. \tag{20.4}$$

No subsequence of this sequence can converge to any probability distribution ν. Indeed, if $\mu_{n_k} \to_w \nu$, the corresponding distribution functions $F_{n_k}(x) = \mu_{n_k}(-\infty, x]$ converge as $k \to \infty$ to $G(x) = \nu(-\infty, x]$ at all points x at which G is continuous (for x in a dense set K). We have:

$$\mu_{n_k}[-n_k, n_k] \le 1 - \varepsilon. \tag{20.5}$$

Now let $a < b$ be arbitrary real numbers. Because K is dense, there exist $a' < a$ and $b' > b$ such that $a', b' \in K$. For sufficiently large k we have $n_k > b'$, $-n_k < a'$; and then we have:

$$F_{n_k}(b') - F_{n_k}(a') = \mu_{n_k}(a', b'] \le \mu_{n_k}[-n_k, n_k] \le 1 - \varepsilon. \tag{20.6}$$

Taking $k \to \infty$, we obtain:

$$G(b') - G(a') \le 1 - \varepsilon; \tag{20.7}$$

and because $a > a'$, $b < b'$,

$$G(b) - G(a) \le 1 - \varepsilon. \tag{20.8}$$

So the difference of the values of the distribution function $G(x)$ at two *arbitrary* points a and b is not greater than $1 - \varepsilon$; which contradicts $\lim_{x\to\infty} G(x) = 1$, $\lim_{x\to-\infty} G(x) = 0$.

Now to the "if" part. Suppose that for every positive ε there exists a $C < \infty$ such that for every $\mu \in \mathfrak{M}$ (20.2) is satisfied. Let us take the function $G(x)$ whose existence is stated in Theorem 19.1. It is nondecreasing and right-continuous. Let us prove that

$$\lim_{x\to\infty} G(x) = 1, \qquad \lim_{x\to-\infty} G(x) = 0. \tag{20.9}$$

Let us prove, say, the first equality here. This equality means that for every $\varepsilon > 0$ there exists a D such that for every $x \ge D$

$$G(x) \ge 1 - \varepsilon. \tag{20.10}$$

Let us take a C so that (20.2) holds for all $\mu \in \mathfrak{M}$. This C may not belong to the set K of continuity points of G; but K is dense, so we can take a $D > C$ such that D belongs to K. We have for this D:

$$F_{n_k}(D) = \mu_{n_k}(-\infty,\, D] \ge \mu_{n_k}[-C,\, C] > 1 - \varepsilon; \tag{20.11}$$

and

$$G(D) = \lim_{k\to\infty} F_{n_k}(D) \ge 1 - \varepsilon. \tag{20.12}$$

We prove that $\lim_{x\to-\infty} G(x) = 0$ using the inequality $\mu([-C,\, C]^c) = \mu\big((-\infty,\, -C) \cup (C,\, \infty)\big) < \varepsilon$ in (20.2).

Now by Theorem 5.5 there exists a distribution ν having $G(x)$ as its distribution function; and by Theorem 18.1 we have $\mu_{n_k} \to_w \nu$ $(k \to \infty)$.

The statement of Theorem 20.1 holds also in the multidimensional case: a family $\mathfrak{M}$ of distributions in $\mathbb{R}^n$ is weakly precompact if and only if for every positive ε there exists a C such that $\mu\{\boldsymbol{x} \in \mathbb{R}^n \colon |\boldsymbol{x}| \le C\} > 1 - \varepsilon$ for every $\mu \in \mathfrak{M}$. The proof is mostly the same but much longer because we have to consider parallelepipeds instead of intervals, continuity in n variables instead of one, etc.

It turns out that the corresponding theorem is also true for distributions on a metric space X (with its Borel σ-algebra) – if this metric space is *complete* and *separable*; instead of intervals $[-C,\, C]$ or spheres $\mu\{\boldsymbol{x} \in \mathbb{R}^n \colon |\boldsymbol{x}| \le C\}$ we take *compact sets*.

Now to characteristic functions. First a simple result.

Theorem 20.2. *If* $\mu_n \to_w \nu$ $(n \to \infty)$, *then the corresponding characteristic functions converge at every point*:

$$\lim_{n\to\infty} f_{\mu_n}(t) = f_\nu(t), \qquad t \in \mathbb{R}^1. \tag{20.13}$$

Proof. The characteristic functions corresponding to distributions are expressed as integrals with respect to these measures; so we have to prove that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} e^{itx}\, \mu_n(dx) = \int_{-\infty}^{\infty} e^{itx}\, \nu(dx). \tag{20.14}$$

But for every fixed real t the function $g_t(x) = e^{itx}$ is bounded and continuous on $(-\infty, \infty)$, so (20.14) follows immediately from the definition of weak convergence.

Theorem 20.3 (the order of exposing the material and the numbering of the theorems is different here from what was in the lecture). *Let* $\mu_1, \mu_2, \ldots, \mu_n, \ldots$ *be a sequence of probability distributions*; *and let the limit*

$$\lim_{n\to\infty} f_{\mu_n}(t) = g(t) \tag{20.15}$$

exist for every $t \in \mathbb{R}^1$.

If the family of distributions $\mathfrak{M} = \{\mu_n, n = 1, 2, 3, \ldots\}$ *is weakly precompact*, *then there exists a probability distribution* ν *such that*

$$\mu_n \to_w \nu \qquad (n \to \infty) \tag{20.16}$$

(and, by Theorem 20.2, the limiting characteristic function is the characteristic function corresponding to this limiting distribution: $\lim_{n\to\infty} f_{\mu_n}(t) = f_\nu(t) = g(t)$).

Proof. According to the definition of precompactness, there exists a subsequence μ_{n_k}, $k = 1, 2, 3, \ldots$, and a probability distribution ν such that

$$\mu_{n_k} \to_w \nu \qquad (k \to \infty). \tag{20.17}$$

We want to prove that *the whole sequence* converges to the same limit. Suppose that this is not so: (20.16) does not hold. What does it mean?

Convergence (20.16) means that for every bounded continuous function $g(x)$ we have $\int_{-\infty}^{\infty} h(x)\, \mu_n(dx) \to \int_{-\infty}^{\infty} h(x)\, \nu(dx)$ as $n \to \infty$ (in Lecture 17 we denoted the "test function" with $f(x)$, but here the letter f is used for characteristic functions); or, in detail: for every bounded continuous $h(x)$ and every positive ε there exists a natural n_0 such that for every $n \ge n_0$ we have $\left|\int_{-\infty}^{\infty} h(x)\, \mu_n(dx) - \int_{-\infty}^{\infty} h(x)\, \nu(dx)\right| < \varepsilon$.

The negation of this statement is as follows: there *exist* a bounded continuous function $h(x)$ and a positive ε such that *for every* natural n_0 there *exists* an $n \geq n_0$ such that

$$\Big|\int_{-\infty}^{\infty} h(x)\ \mu_n(dx) - \int_{-\infty}^{\infty} h(x)\ \nu(dx)\Big| \geq \varepsilon. \tag{20.18}$$

Let us fix this function $h(x)$ and the positive ε. We are going to construct a certain sequence $n'_k \to \infty$ (different from the sequence n_k for which $\mu_{n_k} \to_w \nu$ as $k \to \infty$). Take $n_0 = 1$. There exists a natural n such that (20.18) holds; let us denote this n as n'_1. Then we take $n_0 = n'_1 + 1$; there exists an $n'_2 \geq n'_1 + 1$ (which can be rewritten as $n'_2 > n'_1$) such that (20.18) holds for $n = n'_2$. Proceeding like this, we obtain a sequence of natural numbers $n'_1 < n'_2 < n'_3 < ... < n'_k < ...$ such that

$$\Big|\int_{-\infty}^{\infty} h(x)\ \mu_{n'_k}(dx) - \int_{-\infty}^{\infty} h(x)\ \nu(dx)\Big| \geq \varepsilon, \qquad k = 1, 2, 3, \dots . \tag{20.19}$$

Because of precompactness, we can extract from this sequence a subsequence n'_{k_j} such that

$$\mu_{n'_{k_j}} \to_w \nu' \qquad (j \to \infty), \tag{20.20}$$

where ν' is some probability distribution. By Theorem 22.1??, we get

$$\lim_{j\to\infty} f_{\mu_{n'_{k_j}}}(t) = f_{\nu'}(t), \qquad t \in \mathbb{R}^1. \tag{20.21}$$

But $f_{\mu_{n'_{k_j}}}(t)$ is a subsequence of the sequence $f_{\mu_n}(t)$, which converges to $g(t)$. So we have

$$f_{\nu'}(t) = \lim_{j\to\infty} f_{\mu_{n'_{k_j}}}(t) = g(t) = f_\nu(t), \qquad t \in \mathbb{R}^1. \tag{20.22}$$

By the uniqueness theorem (Theorem 15.1) we have $\nu' = \nu$.

It follows from (20.19) that

$$\Big|\int_{-\infty}^{\infty} h(x)\ \mu_{n'_{k_j}}(dx) - \int_{-\infty}^{\infty} h(x)\ \nu(dx)\Big| \geq \varepsilon, \qquad j = 1,\ 2,\ 3, ..., \tag{20.23}$$

which we can (because $\nu = \nu'$) rewrite as

$$\Big|\int_{-\infty}^{\infty} h(x)\ \mu_{n'_{k_j}}(dx) - \int_{-\infty}^{\infty} h(x)\ \nu'(dx)\Big| \geq \varepsilon, \qquad j = 1, 2, 3, \dots . \tag{20.24}$$

But this clearly contradicts (20.20).

The source of the contradiction is that we supposed that the whole sequence $\mu_n \not\to_w \nu$. So this is impossible, and our theorem is proved.

Can it be that the characteristic functions converge, but the sequence of distributions does not have any weak limit? Or the precompactness condition is satisfied automatically if $\lim_{n\to\infty} f_n(t)$ exists for every $t \in \mathbb{R}^1$?

It turns out that there are convergent sequences of characteristic functions such that the corresponding distributions do not converge.

We turn to our old example of μ_n being the normal distribution with parameters $(0,\, n)$. We have:

$$f_{\mu_n}(t) = e^{-nt^2/2}, \tag{20.25}$$

the limit

$$\lim_{n\to\infty} f_{\mu_n}(t) = \begin{cases} 0, & t \neq 0, \\ 1, & t = 0 \end{cases} \tag{20.26}$$

exists for every t, but μ_n has no weak limit.

However the following result holds:

Theorem 20.4. *Let* $\mu_1,\, \mu_2,\, ...,\, \mu_n,\, ...$ *be a sequence of probability distributions*; *let the limit*

$$\lim_{n\to\infty} f_{\mu_n}(t) = g(t) \tag{20.27}$$

exist for every $t \in \mathbb{R}^1$, *and let the function* $g(t)$ *be continuous at* $t = 0$.

Then the sequence of distributions μ_n *converges weakly as* $n \to \infty$ (to a distribution whose characteristic function turns out to be the limiting function $g(t)$).

Proof. First of all, $g(0) = \lim_{n\to\infty} f_{\mu_n}(t) = \lim_{n\to\infty} 1 = 1$. We are going to use the continuity of $g(t)$ at $t = 0$, which means that for every positive ε there exists a positive δ such that for all $t \in (-\,\delta,\, \delta)$

$$|g(t) - g(0)| = |g(t) - 1| < \varepsilon/3. \tag{20.28}$$

The functions $f_{\mu_n}(t)$ are continuous, and therefore Borel measurable. The function $g(t)$, being a pointwise limit of a sequence of measurable functions is also measurable; so we can integrate it:

$$\left|\frac{1}{2\delta}\int_{-\delta}^{\delta} [g(t) - 1]\, dt\right| < \varepsilon/3. \tag{20.29}$$

Now, the functions $f_{\mu_n}(t)$ converge to $g(t)$ as $n \to \infty$, and they are dominated on the interval $(-\,\delta,\, \delta)$ by the constant 1, which is integrable with respect to the Lebesgue measure. So by the dominated convergence theorem we have:

$$\lim_{n\to\infty} \frac{1}{2\delta}\int_{-\delta}^{\delta} f_n(t)\, dt \to \frac{1}{2\delta}\int_{-\delta}^{\delta} g(t)\, dt \qquad (n \to \infty). \tag{20.30}$$

This means that for $n \geq n_0$ we have:

$$\left|\frac{1}{2\delta}\int_{-\delta}^{\delta} f_{\mu_n}(t)\, dt - \frac{1}{2\delta}\int_{-\delta}^{\delta} g(t)\, dt\right| < \varepsilon/3. \tag{20.31}$$

Together with (20.29) this yields

$$\Big|\frac{1}{2\delta}\int_{-\delta}^{\delta}[f_{\mu_n}(t)-1]\;dt\Big| < 2\varepsilon/3. \tag{20.32}$$

Let us remember that the characteristic function is also an integral; so the integral in (20.32) is an iterated integral:

$$\frac{1}{2\delta}\int_{-\delta}^{\delta}\Big[\int_{-\infty}^{\infty} e^{itx}\;\mu_n(dx)-1\Big]\;dt = \frac{1}{2\delta}\int_{-\delta}^{\delta}\Big[\int_{-\infty}^{\infty}(e^{itx}-1)\;\mu_n(dx)\Big]\;dt. \tag{20.33}$$

The function $e^{itx}-1$ is certainly measurable in the variables t, x, it is also bounded ($|e^{itx}-1|\le 2$), and the integral of its absolute value over $(-\delta,\,\delta)\times(-\infty,\,\infty)$ with respect to the measure $\lambda_1\times\mu_n$ is definitely finite. So we can use Fubini's Theorem and change the order of integration:

$$\frac{1}{2\delta}\int_{-\delta}^{\delta}\Big[\int_{-\infty}^{\infty}(e^{itx}-1)\;\mu_n(dx)\Big]\;dt = \int_{-\infty}^{\infty}\Big[\frac{1}{2\delta}\int_{-\delta}^{\delta}(e^{itx}-1)\;dt\Big]\;\mu_n(dx). \tag{20.34}$$

`From this point, not in` Lecture 22??.

The interior integral in (20.34) is evaluated very easily:

$$\frac{1}{2\delta}\int_{-\delta}^{\delta}(e^{itx}-1)\;dt = \begin{cases}\dfrac{e^{i\delta x}-e^{-i\delta x}}{2\delta x}-1=\dfrac{\sin\delta x}{\delta x}-1, & x\ne 0,\\ \qquad 0, & x=0.\end{cases} \tag{20.35}$$

Let us introduce the notation:

$$h(x)=\begin{cases}1-\dfrac{\sin\delta x}{\delta x}, & x\ne 0,\\ \quad 0, & x=0.\end{cases} \tag{20.36}$$

Draw the graph of the function $h(x)$.

We see that $h(x)$ is nonnegative everywhere, and

$$h(x)\ge 1-1/2\pi \qquad \text{for } |x|\ge\pi/\delta. \tag{20.37}$$

Formulas (20.32) – (20.36) mean that for $n\ge n_0$ we have:

$$\int_{-\infty}^{\infty} h(x)\;\mu_n(dx) < \frac{2\varepsilon}{3}. \tag{20.38}$$

Using a Chebyshev inequality, we get:

$$\mu_n(\mathbb{R}^1\setminus[-\pi/\delta,\,\pi/\delta]) \le \frac{\displaystyle\int_{-\infty}^{\infty} h(x)\;\mu_n(dx)}{\inf\{h(x)\colon\, x\notin[-\pi/\delta,\,\pi/\delta]\}} < \frac{2\varepsilon}{3\cdot(1-1/2\pi)} < \varepsilon. \tag{20.39}$$

But this is only for $n \geq n_0$; what about $n = 1, 2, ..., n_0 - 1$? For every fixed n we can find a C_n such that

$$\mu_n(\mathbb{R}^1 \setminus [-C_n, C_n]) < \varepsilon. \tag{20.40}$$

So we take

$$C = \max(\pi/\delta, C_1, C_2, ..., C_{n_0-1}), \tag{20.41}$$

and we have that for every $n = 1, 2, ..., n_0 - 1, n_0, n_0 + 1, ...$

$$\mu_n(\mathbb{R}^1 \setminus [-C, C]) < \varepsilon, \qquad \text{or, which is the same,} \qquad \mu_n[-C, C] > 1 - \varepsilon. \tag{20.42}$$

Remembering that ε was an arbitrary nonnegative number, we see that the family of distributions μ_n is weakly precompact, and we can apply Theorem 20.3.

Theorem 20.2 and Theorem 20.4 (and also Theorems 20.1 and 20.3) are true also in the multidimensional case:

Theorem 20.5. *Let* $\mu_1, \mu_2, ..., \mu_n, ...$ *be probability distributions in the* r*-dimensional space* $\mathbb{R}^r$.

If $\mu_n \to_w \nu$ *as* $n \to \infty$, *where* ν *is some probability distribution in* $\mathbb{R}^r$, *then the corresponding characteristic functions converge at every point*:

$$\lim_{n\to\infty} f_{\mu_n}(\boldsymbol{t}) = f_\nu(\boldsymbol{t}), \qquad \boldsymbol{t} \in \mathbb{R}^r. \tag{20.43}$$

If

$$\lim_{n\to\infty} f_{\mu_n}(\boldsymbol{t}) = g(\boldsymbol{t}) \tag{20.44}$$

for every $\boldsymbol{t} \in \mathbb{R}^r$, *and the function* $g(\boldsymbol{t})$ *is continuous at* $\boldsymbol{t} = \boldsymbol{0}$, *then there exists a weak limit*

$$\nu = (w) \lim_{n\to\infty} \mu_n, \tag{20.45}$$

and its characteristic function is equal to the limiting function $g(\boldsymbol{t})$:

$$f_\nu(\boldsymbol{t}) = \int_{\mathbb{R}^r} e^{i\boldsymbol{t}\cdot\boldsymbol{x}} \, \nu(d\boldsymbol{x}) = g(\boldsymbol{t}). \tag{20.46}$$