Lecture 18. Weak convergence of distributions.

Theorem 17.5, in particular, allows us to prove (17.26) in another, easier way: we take a random variable ζ having the standard normal distribution (i. e., with parameters (0, 1)); then $\xi_n = \sqrt{b_n} \cdot \zeta + a_n$ has the normal distribution with parameters (a_n, b_n), and $\eta = \sqrt{b} \cdot \zeta + a$ the normal distribution with parameters (a, b). So we have to prove:

$$Ef(\sqrt{b_n} \cdot \zeta + a_n) \to Ef(\sqrt{b} \cdot \zeta + a). \tag{18.1}$$

We have for every $\omega \in \Omega$:

$$\xi_n(\omega) = \sqrt{b_n} \cdot \zeta(\omega) + b_n \to \eta(\omega) = \sqrt{b} \cdot \zeta(\omega) + b; \tag{18.2}$$

so by Theorem 17.5 we have $\mu_{\xi_n} \to_w \mu_\eta$.

In our examples we had discrete distributions converging to discrete ones, and continuous to continuous ones; but it is possible for a sequence of discrete distributions to converge to a continuous one (see the next example), and the opposite is also possible (see Problem **39**).

Example: Let μ_n be the mixture of unit measures concentrated at the $n+1$ points 0, $1/n$, $2/n$, ..., k/n, ..., $(n-1)/n$, 1 with weights $1/(n+1)$ each:

$$\mu_n(C) = \sum_{k=0}^{n} \frac{1}{n+1} \cdot \delta_{k/n}(C), \tag{18.3}$$

where

$$\delta_{k/n}(C) = \begin{cases} 1, & C \ni k/n, \\ 0, & C \not\ni k/n. \end{cases} \tag{18.4}$$

Let us prove that the weak limit ν of this sequence of measures (distributions) is the uniform distribution on the interval [0, 1]:

$$\nu(C) = \int_C I_{[0,\,1]}(x)\ dx = \lambda_1(C \cap [0,\, 1]). \tag{18.5}$$

For an arbitrary bounded continuous function $f(x)$ we have:

$$\int_{-\infty}^{\infty} f(x)\ \mu_n(dx) = \sum_{k=0}^{n} f\big(\frac{k}{n}\big) \cdot \frac{1}{n+1}, \tag{18.6}$$

$$\int_{-\infty}^{\infty} f(x)\ \nu(dx) = \int_0^1 f(x)\ dx \tag{18.7}$$

(of course, (18.6) holds not only for continuous bounded functions, but for *all* functions).

The sum (18.6) is the Riemann sum for the integral if we take the partition of the interval [0, 1] by the points $x_k = \dfrac{k}{n+1}$, $k = 0, 1, 2, ..., n-1, n, n+1$, and choose in each interval $[\dfrac{k}{n+1}, \dfrac{k+1}{n+1}]$ the point $\dfrac{k}{n}$ (this point is indeed in this interval: $\dfrac{k}{n+1} \le \dfrac{k}{n} \le \dfrac{k+1}{n+1}$); and we have:

$$\int_{-\infty}^{\infty} f(x)\ \mu_n(dx) = \sum_{k=0}^{n} f(\frac{k}{n}) \cdot [\frac{k+1}{n+1} - \frac{k}{n+1}] \to \text{(Riemann)} \int_0^1 f(x)\ dx. \qquad (18.8)$$

But, as we know, the (proper) Riemann integral, if it exists, coincides with the Lebesgue integral with respect to the Lebesgue measure; so

$$\int_{-\infty}^{\infty} f(x)\ \mu_n(dx) \to \int_{-\infty}^{\infty} f(x)\ \nu(dx) \qquad (n \to \infty) \qquad (18.9)$$

for every bounded continuous function $f(x)$, and $\mu_n \to_w \nu$.

Theorem 18.1. *Let* $\mu_1, \mu_2, ..., \mu_n, ...,$ *be a sequence of probability distributions in the real line*; ν *a probability distribution. Let* $F_1(x), F_2(x), ..., F_n(x), ..., G(x)$ *be the corresponding distribution functions*:

$$F_k(x) = \mu(-\infty, x], \qquad G(x) = \nu(-\infty, x], \qquad x \in \mathbb{R}^1. \qquad (18.10)$$

If

$$\lim_{n\to\infty} F_n(x) = G(x) \qquad (18.11)$$

for all x *belonging to a set* K *that is dense in* $\mathbb{R}^1$, *then* $\mu_n \to_w \nu$ $(n \to \infty)$.

Proof. Let $f(x)$ be a bounded continuous function. We have to prove that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f(x)\ \mu_n(dx) = \int_{-\infty}^{\infty} f(x)\ \nu(dx). \qquad (18.12)$$

This means that for every positive ε there exists a natural n_* such that for every $n \ge n_*$

$$\Big| \int_{-\infty}^{\infty} f(x)\ \mu_n(dx) - \int_{-\infty}^{\infty} f(x)\ \nu(dx) \Big| < \varepsilon. \qquad (18.13)$$

Let $\|f\| = \sup_x |f(x)|$.

We know that $\lim_{x\to\infty} G(x) = 1$, $\lim_{x\to-\infty} G(x) = 0$. This means, in particular, that there exist such negative a and positive b that

$$G(a) < \frac{\varepsilon}{8\|f\|}, \qquad G(b) > 1 - \frac{\varepsilon}{8\|f\|}. \qquad (18.14)$$

Without restriction of generality, we can assume that $a, b \in K$: indeed, if, say, $a \notin K$, we can replace it by an $a' < a$ such that $a' \in K$, and $G(a') \le G(a) = \nu(-\infty, a] < \varepsilon/8\|f\|$ (or $b' > b$, $b' \in K$, $G(b') \ge G(b) > 1 - \varepsilon/8\|f\|$, $\nu(b, \infty) < \varepsilon/8\|f\|$).

Now, on the interval $[a,\, b]$ the continuous function $f(x)$ is *uniformly* continuous; and there exists a positive δ such that from $x,\, y \in [a,\, b]$, $|x - y| < \delta$ it follows that $|f(x) - f(y)| < \dfrac{\varepsilon}{8}$. Let us divide the interval $[a,\, b]$ into N intervals of length less than δ by the points $a = x_0 < x_1 < x_2 < \ldots < x_{k-1} < x_k < \ldots < x_{N-1} < x_N = b$.

Again, *without restriction of generality* we can assume that all $x_k \in K$: otherwise we can move each x_k a little so that it is in K now, but still its distances from its neighbors $x_k - x_{k-1}$, $x_{k+1} - x_k$ are still $< \delta$.

We have:

$$\begin{aligned} \Big| \int_{(-\infty,\,\infty)} f(x)\ \mu_n(dx) - \int_{(-\infty,\,\infty)} f(x)\ \nu(dx) \Big| \\ \le \int_{(-\infty,\,a]} |f(x)|\ \mu_n(dx) + \int_{(-\infty,\,a]} |f(x)|\ \nu(dx) \\ + \Big| \int_{(a,\,b]} f(x)\ \mu_n(dx) - \int_{(a,\,b]} f(x)\ \nu(dx) \Big| \\ + \int_{(b,\,\infty)} |f(x)|\ \mu_n(dx) + \int_{(b,\,\infty)} |f(x)|\ \nu(dx). \end{aligned} \tag{18.16}$$

The first, second, fourth, and fifth summands in the right-hand side are not greater than, respectively, $\|f\| \cdot \mu_n(-\infty,\, a] = \|f\| \cdot F_n(a)$, $\|f\| \cdot \nu(-\infty,\, a] = \|f\| \cdot G(a) < \varepsilon/8$, $\|f\| \cdot \mu_n(b,\, \infty) = \|f\| \cdot [1 - F_n(b)]$, $\|f\| \cdot \nu(b,\, \infty) = \|f\| \cdot [1 - G(b)] < \varepsilon/8$.

To estimate the third summand in the right-hand side of (18.16), let us choose a point y_k in each interval $(x_{k-1},\, x_k]$, $k = 1,\, 2,\, \ldots,\, N$, and define a function $\tilde{f}(x)$ on the interval $(a,\, b]$ by

$$\tilde{f}(x) = f(y_k) \qquad \text{for } x \in (x_{k-1},\, x_k] \tag{18.17}$$

(draw a picture).

By the choice of our $\delta > 0$, we have for all $x \in (a,\, b]$:

$$|f(x) - \tilde{f}(x)| < \frac{\varepsilon}{8}, \tag{18.18}$$

and from this,

$$\begin{gathered} \Big| \int_{(a,\,b]} f(x)\ \mu_n(dx) - \int_{(a,\,b]} \tilde{f}(x)\ \mu_n(dx) \Big| < \frac{\varepsilon}{8}, \\ \Big| \int_{(a,\,b]} f(x)\ \nu(dx) - \int_{(a,\,b]} \tilde{f}(x)\ \nu(dx) \Big| < \frac{\varepsilon}{8}. \end{gathered} \tag{18.19}$$

Now, the third summand in the right-hand side of (18.16) is not greater than

$$\begin{aligned} \Big| \int_{(a,\,b]} f(x)\ \mu_n(dx) - \int_{(a,\,b]} \tilde{f}(x)\ \mu_n(dx) \Big| + \Big| \int_{(a,\,b]} f(x)\ \nu(dx) - \int_{(a,\,b]} \tilde{f}(x)\ \nu(dx) \Big| \\ + \Big| \int_{(a,\,b]} \tilde{f}(x)\ \mu_n(dx) - \int_{(a,\,b]} \tilde{f}(x)\ \nu(dx) \Big| \\ < \frac{2\varepsilon}{8} + \sum_{k=1}^{N} \Big| \int_{(x_{k-1},\,x_k]} \tilde{f}(x)\ \mu_n(dx) - \int_{(x_{k-1},\,x_k]} \tilde{f}(x)\ \nu(dx) \Big|. \end{aligned} \tag{18.20}$$

The k-th summand here is equal to

$$|f(y_k)| \cdot \big|\mu_n(x_{k-1}, x_k] - \nu(x_{k-1}, x_k]\big| = |f(y_k)| \cdot |F_n(x_k) - F_n(x_{k-1}) - G(x_k) + G(x_{k-1})|. \tag{18.21}$$

Now, for every k, $0 \le k \le N$, we can find a natural n_k so that for $n \ge n_k$ we have:

$$|F_n(x_k) - G(x_k)| < \frac{\varepsilon}{8(N+1)\|f\|}. \tag{18.22}$$

If we take $n_* = \max(n_0, n_1, n_2, ..., n_N)$, for $n \ge n_*$ *all* inequalities (18.22) hold.

In particular, for $k = 0$, from which we get for $n \ge n_*$:

$$F_n(a) = F_n(x_0) < G(a) + \frac{\varepsilon}{8(N+1)\|f\|} < \frac{\varepsilon}{8\|f\|} + \frac{\varepsilon}{8(N+1)\|f\|}; \tag{18.23}$$

similarly,

$$1 - F_n(b) < 1 - G(b) + \frac{\varepsilon}{8(N+1)\|f\|} < \frac{\varepsilon}{8\|f\|} + \frac{\varepsilon}{8(N+1)\|f\|}. \tag{18.24}$$

So the first and the fourth summand in the right-hand side of (18.16) are less than $\frac{\varepsilon}{8} + \frac{\varepsilon}{8(N+1)}$.

By inequality (18.22) for k and for $k-1$, the expression (18.21) is less than $\frac{2\varepsilon}{8(N+1)}$. Putting together formulas (18.16), (18.20) and the estimates obtained, we get

$$\Big|\int_{(-\infty,\,\infty)} f(x)\ \mu_n(dx) - \int_{(-\infty,\,\infty)} f(x)\ \nu(dx)\Big| \\ \le \frac{2\varepsilon}{8} + \frac{2\varepsilon}{8(N+1)} + \frac{2\varepsilon}{8} + \frac{2\varepsilon}{8} + N \cdot \frac{2\varepsilon}{8(N+1)} = \varepsilon, \tag{18.25}$$

which was to be proved.

Let me formulate another theorem, sort of inverse to Theorem 18.1.

Theorem 18.2. *Let* $\mu_1, \mu_2, ..., \mu_n, ...$ *be a sequence of distributions on* $\mathbb{R}^1$, ν *a distribution on it*; *let* $F_1(x), F_2(x), ..., F_n(x), ..., G(x)$ *be the corresponding distribution functions*: $F_n(x) = \mu_n(-\infty, x]$, $G(x) = \nu(-\infty, x]$.

If $\mu_n \to_w \nu$ *as* $n \to \infty$, *then* $F_n(x) \to G(x)$ $(n \to \infty)$ *for all* x *that are continuity points of the function* G.

I'll give the proof of this theorem in the next lecture; but in the meantime, let us take Theorems 18.1 and 18.2 together:

Theorem 18.3. *A sequence of distributions* μ_n *converges weakly to a probability distribution* ν *if and only if the sequence of the corresponding distribution functions* $F_n(x) = \mu_n(-\infty, x]$ *converges to the distribution function* $G(x) = \nu(-\infty, x]$ *at all points* x *at which the limiting function* G *is continuous*; *and also if and only if there* *exists a set* K *that is dense in* $(-\infty, \infty)$ *such that the convergence* $F_n(x) \to G(x)$ *takes place for all* $x \in K$.

The only thing that remains to be **proved** here is that automatically the set of points at which the distribution function $G(x)$ is continuous is dense in $(-\infty, \infty)$.

But this follows from the fact that discontinuities of a monotone function cannot be anything other than jumps, and the number of its jumps is at most countable; the complement of a countable set in the real line is always dense.