Lecture 21. Limit theorems via characteristic functions.

Now we can solve some problems of ours easier using characteristic functions.

Solving Problem $\boxed{\mathbf{23}}$ (the solution was given, in fact, in some of the previous lectures): ξ_n and η_n are independent, so their joint characteristic function

$$f_{\xi_n,\,\eta_n}(t_1,\,t_2)=f_{\xi_n}(t_1)\cdot f_{\eta_n}(t_2). \tag{21.1}$$

Since the random variables $e^{i(t_1\xi_n+t_2\eta_n)}$, $e^{it_1\xi_n}$, $e^{it_2\eta_n}$ are all dominated by the same integrable random variable, namely by the constant 1, we have by the dominated-convergence theorem:

$$\begin{aligned}\lim_{n\to\infty} f_{\xi_n,\,\eta_n}(t_1,\,t_2)&=\lim_{n\to\infty} Ee^{i(t_1\xi_n+t_2\eta_n)}=E\lim_{n\to\infty} e^{i(t_1\xi_n+t_2\eta_n)}\\&=Ee^{i(t_1\xi+t_2\eta)}=f_{\xi,\,\eta}(t_1,\,t_2),\end{aligned} \tag{21.2}$$

and similarly, $\lim_{n\to\infty} f_{\xi_n}(t_1)=f_\xi(t_1)$, $\lim_{n\to\infty} f_{\eta_n}(t_2)=f_\eta(t_2)$. From this we get

$$f_{\xi,\,\eta}(t_1,\,t_2)=f_\xi(t_1)\cdot f_\eta(t_2), \tag{21.3}$$

which implies independence (Theorem 16.2).

Now the example we had with μ_n being the binomial distribution with parameters $(n,\,p_n)$. We have:

$$f_{\mu_n}(t)=\sum_{k=0}^{n}\binom{n}{k}p_n^k(1-p_n)^{n-k}\cdot e^{itk}=\sum_{k=0}^{n}\binom{n}{k}(p_n\cdot e^{it})^k(1-p_n)^{n-k}=[p_n\cdot e^{it}+(1-p_n)]^n, \tag{21.4}$$

$$\lim_{n\to\infty} f_{\mu_n}(t)=\lim_{n\to\infty}[1+p_n(e^{it}-1)]^n=\exp\{\lim_{n\to\infty} n\cdot p_n(e^{it}-1)\}=\exp\{a(e^{it}-1)\}. \tag{21.5}$$

We recognize this as the characteristic function of the Poisson distribution with parameter a; by Theorem 20.4, the sequence of distributions μ_n converges weakly to this distribution.

Also other examples given in Lectures 17 and 18 can be treated more easily using characteristic functions.

Using characteristic functions we can obtain limit theorems about distributions of sums of a growing number of random variables; i. e., about (weak) limits of distributions of random variables

$$\zeta_n=\frac{\xi_1+\ldots+\xi_n-A_n}{B_n}, \tag{21.6}$$

where A_n, B_n are some sequences of constants.

Why should we subtract a constant A_n and divide by something, why not just find the limit of the distribution of $\eta_n=\xi_1+\ldots+\xi_n$? Let us consider the case of $\xi_1,\,\ldots,\,\xi_n,\,\ldots$ being independent and identically distributed. If $m = E\xi_k > 0$, the law of large numbers tells us that $(\xi_1+\ldots+\xi_n)/n \to_P m$, and the distribution of $\xi_1+\ldots+\xi_n$ "goes to $+\infty$" (if $m < 0$, it goes to $-\infty$). Subtracting nm from $\xi_1+\ldots+\xi_n$ we'll have a sequence of distributions that don't go to $+\infty$, but some part of it goes to $+\infty$, and another part to $-\infty$. So we see that (apart from some exceptional cases) we have to *normalize* the sum, subtracting something and dividing by something (usually going to ∞ as $n \to \infty$).

For the law of large numbers the choice of these constants is:

$$A_n = E\xi_1 + \ldots + E\xi_n, \qquad B_n = n. \tag{21.7}$$

The following simple result helps us in studying the distributions of the sums:

Theorem 21.1. *Let ξ_1, ..., ξ_n be independent random variables. Then the characteristic function of their sum*

$$f_{\xi_1+\ldots+\xi_n}(t) = f_{\xi_1}(t) \cdot \ldots \cdot f_{\xi_n}(t), \qquad t \in \mathbb{R}^1. \tag{21.8}$$

Proof: By the complex version of Theorem 9.1,

$$f_{\xi_1+\ldots+\xi_n}(t) = Ee^{it(\xi_1+\ldots+\xi_n)} = E[e^{it\xi_1}\cdot\ldots\cdot e^{it\xi_n}] = Ee^{it\xi_1}\cdot\ldots\cdot Ee^{it\xi_n} = \prod_{k=1}^{n} f_{\xi_k}(t). \tag{21.9}$$

And of course,

$$f_{\zeta_n}(t) = f_{(\xi_1+\ldots+\xi_n-A_n)/B_n}(t) = f_{\xi_1+\ldots+\xi_n}\Big(\frac{t}{B_n}\Big)\cdot e^{-itA_n/B_n} = \prod_{k=1}^{n} f_{\xi_k}\Big(\frac{t}{B_n}\Big)\cdot e^{-itA_n/B_n}. \tag{21.10}$$

Now we return to Khinchin's Theorem (see Lecture 13):

Let ξ_1, ..., ξ_n, ... be a sequence of independent identically distributed random variables with finite expectation $E\xi_i$. Then

$$\frac{\xi_1+\ldots+\xi_n}{n} \to_P E\xi_i \qquad (n\to\infty). \tag{13.29}$$

Proof. Let us denote $E\xi_k = m$.

The characteristic function of the random variable $\zeta_n = \dfrac{\xi_1+\ldots+\xi_n}{n}$ is

$$f_{\zeta_n}(t) = \prod_{k=1}^{n} f_{\xi_k}(t/n) = [f_{\xi_k}(t/n)]^n. \tag{21.11}$$

We know that the random variable ξ_k has a finite expectation; so its characteristic function is differentiable, and its derivative at $t = 0$ is equal to $i\cdot m$. By the definition of the derivative, we have:

$$f_{\xi_k}(t) = f_{\xi_k}(0) + im\cdot t + o(t) \qquad (t\to 0), \tag{21.12}$$

or, deciphering this:

$$f_{\xi_k}(t) = 1 + im \cdot t + t \cdot \alpha(t), \tag{21.13}$$

where

$$\lim_{t\to 0} \alpha(t) = 0. \tag{21.14}$$

We have:

$$f_{\zeta_n}(t) = [f_{\xi_k}(t/n)]^n = \exp\{n \cdot \ln f_{\xi_k}(t/n)\} = ... \tag{21.15}$$

But stop, stop, stop: we know what the logarithmic function is for positive real values of the argument; but here $f_{\xi_k}(t/n)$ is, in general, a complex number – so how do we handle this?

The logarithmic function – the inverse of the exponential function – *can be* defined for complex non-zero values of the argument; but it is a many-valued function: we must specify what *version* of the logarithmic function, what *branch* of it we consider.

The power series

$$\sum_{k=1}^{\infty} \frac{(-1)^{k-1}(z-1)^k}{k} \tag{21.16}$$

converges for all complex z (including those on the real axis) such that $|z-1| < 1$ (such z's form a circle of radius 1 centered at the point $z = 1$); the function represented by this series – let us denote it $\ln_1 z$ – is differentiable in its complex argument. And the series

$$\exp\{\ln_1(z)\} = 1 + \ln_1(z) + \frac{\big(\ln_1(z)\big)^2}{2!} + \frac{\big(\ln_1(z)\big)^3}{3!} + \frac{\big(\ln_1(z)\big)^4}{4!} + ... \tag{21.17}$$

also converges for all z with $|z-1| < 1$. We want to prove that $\ln_1(z)$ is one of the branches of the inverse exponential function – i. e. that the function (21.17) is equal to z for all complex z, $|z-1| < 1$.

I thought at first that my students were exposed to some of complex analysis; but it turned out not to be so. So I'll try and give an independent proof.

Let us put the expression (21.16) for the function $\ln_1(z)$ in formula (21.17). The powers of the power series (21.16) can also be expressed as power series: we just multiply this power series by itself several times, and collect the terms with the same power of $(z-1)$:

$$\begin{array}{lcl} \ln_1(z) & = & (z-1) - \frac{(z-1)^2}{2} + \frac{(z-1)^3}{3} - \frac{(z-1)^4}{4} + ..., \\ \big(\ln_1(z)\big)^2 & = & (z-1)^2 - (z-1)^3 + (\frac{2}{3} + \frac{1}{4})(z-1)^4 - ..., \\ \big(\ln_1(z)\big)^3 & = & (z-1)^3 - \frac{3}{2}(z-1)^4 + ..., \\ \big(\ln_1(z)\big)^4 & = & (z-1)^4 - ..., \\ \cdots\cdots & & \end{array} \tag{21.18}$$

(of course I couldn't write infinitely many powers, infinitely many terms in each: so I stopped at the fourth power of $(z-1)$).

After putting these into (21.17) we get:

$$\exp\{\ln_1(z)\} = 1 + c_1 \cdot (z-1) + c_2 \cdot (z-1)^2 + c_3 \cdot (z-1)^3 + c_4 \cdot (z-1)^4 + \ldots, \tag{21.19}$$

where c_1, c_2, c_3, c_4, ... are some coefficients (it takes some arithmetic to evaluate these coefficients). So we divide the first equality in (21.18) by $1! = 1$, the second one by $2! = 2$, the third by $3! = 6$, the fourth by $4! = 24$, etc., and add them; we get:

$$\begin{aligned} \exp\{\ln_1(z)\} = 1 + (z-1) - \tfrac{1}{2}\,(z-1)^2 + \tfrac{1}{3}\,(z-1)^3 - \tfrac{1}{4}\,(z-1)^4 + \ldots \\ + \tfrac{1}{2}\,(z-1)^2 - \tfrac{1}{2}(z-1)^3 + \tfrac{11}{24}\,(z-1)^4 - \ldots \\ + \tfrac{1}{6}\,(z-1)^3 - \tfrac{1}{4}(z-1)^4 + \ldots \\ + \tfrac{1}{24}\,(z-1)^4 - \ldots . \end{aligned} \tag{21.20}$$

So we have:

$$c_1 = 1, \quad c_2 = -\tfrac{1}{2} + \tfrac{1}{2} = 0, \quad c_3 = \tfrac{1}{3} - \tfrac{1}{2} + \tfrac{1}{6} = 0, \quad c_4 = -\tfrac{1}{4} + \tfrac{11}{24} - \tfrac{1}{4} + \tfrac{1}{24} = 0. \tag{21.21}$$

If all coefficients c_k, $k > 1$, are equal to 0, we have $\exp\{\ln_1(z)\} = z$ for all z in the circle $|z-1| < 1$. But we don't know, perhaps c_5 or $c_{287} \neq 0$; and I am not so talented in arithmetic to prove this arithmetically. Fortunately, we know that for *real* z, $|z-1| < 1$ (i. e., for $z \in (0,\, 2)$) the natural logarithm *is* the inverse of the exponential function, and that this natural logarithm *is* represented, for real $z \in (0,\, 2)$, by the power series (21.16); so that for these z the sum of the power series (21.19) is equal to z. The coefficients of a power series (of a real variable) is determined uniquely by the values of its sum in the interval of convergence (if this interval does not reduce to one point): e. g., we can take the derivatives of the sum at the midpoint of this interval (in our case, at $z = 1$). The series (21.19) expressing the function $\exp\{\ln_1(z)\}$ for complex values of z is the same as that for real values of z (in fact, the expression (21.19) was written for *all* complex z, $|z-1| < 1$, including the real ones in the interval $(0,\, 2)$); zo we have $c_1 = 1$, $c_k = 0$ for $k > 1$, and $\exp\{\ln_1(z)\} = z$ for all complex z, $|z-1| < 1$.

Another version of complex logarithm is given by

$$\mathrm{Ln}(z) = 2\pi i + \ln_1(z), \qquad |z-1| < 1; \tag{21.22}$$

another one, by

$$\mathrm{LN}(z) = \pi i - \sum_{k=1}^{\infty} \frac{(z+1)^k}{k}, \qquad |z+1| < 1; \tag{21.23}$$

etc.

For our function $\ln_1$, which is differentiable at $z = 1$ with respect to the complex argument z, we have:

$$\ln_1(z) = (z-1) + o(z-1) \qquad (z \to 1), \tag{21.24}$$

which means that

$$\ln_1(z) = (z-1) + (z-1)\cdot\beta(z), \qquad \beta(z)\to 0 \quad (z\to 1). \tag{21.25}$$

Is our $\ln_1(z)$ defined for $z = f_{\xi_k}(t/n)$ (we want to apply it to formula (21.15))?

For every fixed t the limit $\lim_{n\to\infty} f_{\xi_k}(t/n) = f_{\xi_k}(0) = 1$, so for sufficiently large n we have

$$|f_{\xi_k}(t/n) - 1| < 1: \tag{21.26}$$

yes. So we can, for sufficiently large n, rewrite (21.15) as follows:

$$f_{\zeta_n}(t) = \exp\bigl\{n\cdot\ln_1\bigl(f_{\xi_k}(t/n)\bigr)\bigr\}. \tag{21.27}$$

By (21.13), the quantity under the $\ln_1$ sign is

$$f_{\xi_k}(t/n) = 1 + im\cdot\frac{t}{n} + \frac{t}{n}\cdot\alpha(t/n); \tag{21.28}$$

for every fixed $t\in\mathbb{R}^1$ we have $t/n\to 0$ $(n\to\infty)$, so by (21.14) we can write:

$$\alpha(t/n)\to 0 \qquad (n\to\infty). \tag{21.29}$$

By (21.25) we have (again for every fixed $t\in\mathbb{R}^1$):

$$\begin{aligned}
\ln_1\bigl(f_{\xi_k}(t/n)\bigr) &= [f_{\xi_k}(t/n)-1] + [f_{\xi_k}(t/n)-1]\cdot\beta\bigl(f_{\xi_k}(t/n)\bigr)\\
&= \Bigl[im\cdot\frac{t}{n} + \frac{t}{n}\cdot\alpha(t/n)\Bigr]\cdot\Bigl[1+\beta\bigl(f_{\xi_k}(t/n)\bigr)\Bigr]\\
&= im\cdot\frac{t}{n} + im\cdot\frac{t}{n}\cdot\beta\bigl(f_{\xi_k}(t/n)\bigr) + im\cdot\frac{t}{n}\cdot\alpha(t/n)\cdot\Bigl[1+\beta\bigl(f_{\xi_k}(t/n)\bigr)\Bigr],
\end{aligned} \tag{21.30}$$

where $\alpha(t/n)\bigr)$, $\beta\bigl(f_{\xi_k}(t/n)\bigr)\to 0$ as $n\to\infty$. So by (21.27) we have:

$$\lim_{n\to\infty} f_{\zeta_n}(t) = \exp\bigl\{\lim_{n\to\infty} n\cdot\ln_1\bigl(f_{\xi_k}(t/n)\bigr)\bigr\} = e^{imt}. \tag{21.31}$$

The limiting function is, of course, continuous at $t=0$, by Theorem 20.4 it must be the characteristic function of some probability distribution, and μ_n converges weakly to this distribution.

Moreover, we *recognize* the function e^{imt} as the characteristic function of the distribution δ_m concentrated at one point m: the characteristic function of the random variable identically equal to the constant m (or equal to m almost surely – with probability 1). So by Theorem 20.4 we have:

$$(w)\lim_{n\to\infty}\mu_{\zeta_n} = \delta_m: \tag{21.32}$$

the distribution of the arithmetic mean ζ_n converges weakly to that of the constant m.

Have we proved our theorem, which was not about convergence of *distributions*, but about convergence (in probability) of *random variables*?

I have warned you one million times that if distributions of two random variables are the same, this *doesn't* mean that it is the same random variable (or that both are *almost* the same random variable, coinciding with each other almost surely); and convergence of distributions *does not* imply convergence of the random variables: not almost sure convergence, and not even convergence in probability.

So why did we pursue this convergence of distributions thing?

It turns out that convergence of distributions does not imply convergence of the random variables – *except in one case*: that of convergence of distributions to a probability distribution of a constant. The following simple result holds:

Theorem 21.2. *Let* m *be a real number*; ξ_1, ξ_2, ..., ξ_n, ... *a sequence of random variables. If the sequence of distributions* μ_{ξ_n} *of these random variables converges weakly to the distribution* δ_m *concentrated at* m, *then*

$$\xi_n \to_P m \qquad (n \to \infty). \tag{21.33}$$

Proof. We want to prove that for every positive ε

$$P\{|\xi_n - m| < \varepsilon\} = P\{m - \varepsilon < \xi_n < m + \varepsilon\} \to 1 \qquad (n \to \infty). \tag{21.34}$$

The distribution function $G(x)$ corresponding to the distribution δ_m concentrated at m is given by

$$G(x) = \begin{cases} 0, & x < m, \\ 1, & x \geq m. \end{cases} \tag{21.35}$$

By Theorem 18.2, the sequence of distribution functions $F_{\xi_n}(x)$ converges to $G(x)$ for every x being a continuity point of G; that is, for all $x \neq m$. In particular, this will be at $x = m - \varepsilon$ and at $x = m + \varepsilon/2$ (why not $m + \varepsilon$? because when we are handling distribution functions, we consider intervals *with* their right ends, and in (21.34) we have an interval *without* its right end):

$$\lim_{n \to \infty} F_{\xi_n}(m - \varepsilon) = 0, \qquad \lim_{n \to \infty} F_{\xi_n}(m + \varepsilon/2) = 1. \tag{21.36}$$

So we have:

$$\begin{aligned} P\{m - \varepsilon < \xi_n < m + \varepsilon\} &\geq P\{m - \varepsilon < \xi_n \leq m + \varepsilon/2\} \\ &= F_{\xi_n}(m + \varepsilon/2) - F_{\xi_n}(m - \varepsilon) \to 1 - 0 = 1. \end{aligned} \tag{21.37}$$

This proves Khinchin's Theorem.

Note that Khinchin's original proof was much more complicated: the theory of weak convergence in terms of characteristic functions was developed only later.

Now we have come closer to answering the following interesting question: for independent identically distributed random variables ξ_1, ξ_2, ..., ξ_n, ... having *no expectation*, is it possible that the arithmetic mean $\zeta_n = \dfrac{\xi_1 + ... + \xi_n}{n}$ converges in probability to some number m (*not* to their expectation: remember, it does not exist)?

If we have an example (see Problem $\boxed{\mathbf{33^*}}$) of a distribution without expectation, but with the corresponding characteristic function $f(t)$ differentiable at 0, the same reasoning as in the above proof leads us to $\zeta_n \to_P m$ $(= i^{-1} \cdot f'(0))$.