Lecture 16. Characteristic functions, continued.

Now the time has come to speak about multidimensional characteristic functions: about characteristic functions of *random vectors*.

Let $\boldsymbol{\xi} = (\xi_1, ..., \xi_n)$ be an n-dimensional random vector. The characteristic function of this random vector (or the characteristic function of its n-dimensional distribution; or the *joint* characteristic function of the random variables $\xi_1, ..., \xi_n$) is the function of an n-dimensional argument $\boldsymbol{t}$ defined by

$$f(\boldsymbol{t}) = f_{\boldsymbol{\xi}}(\boldsymbol{t}) = f_\mu(\boldsymbol{t}) = Ee^{i\,\boldsymbol{t}\cdot\boldsymbol{\xi}} = \int_{\mathbb{R}^n} e^{i\,\boldsymbol{t}\cdot\boldsymbol{x}}\,\mu(d\boldsymbol{x}), \tag{16.1}$$

where $\cdot$ denotes the dot product; or, in coordinates:

$$\begin{aligned} f(t_1, ..., t_n) &= f_{\xi_1, ..., \xi_n}(t_1, ..., t_n) = f_\mu(t_1, ..., t_n) \\ &= Ee^{i\,(t_1\xi_1+...+t_n\xi_n)} = \underset{\mathbb{R}^n}{\int \cdots \int} e^{i\,(t_1x_1+...+t_nx_n)}\,\mu(dx_1 ... dx_n). \end{aligned} \tag{16.2}$$

The following uniqueness theorem is proved in exactly the same way as Theorem 15.1, only one has to use *multidimensional* Fourier series, and multidimensional intervals (parallelepipeds, or, in the two-dimensional case, rectangles with sides parallel to the axes):

Theorem 16.1. *Let* μ *and* ν *be two distributions* (probability measures) *in* $\mathbb{R}^n$. *If their characteristic functions coincide*: $f_\mu(\boldsymbol{t}) = f_\nu(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^n$, *then these distributions coincide.*

We had Theorems 9.6 – 9.8′ expressing independence in terms of probability mass functions, densities, distribution functions. In principle, since there is a one-to-one correspondence between distributions and their characteristic functions, every property of distributions can be expressed in terms of characteristic functions. Independence of n random variables *is* some property of their joint distribution: it means that the joint distribution is the direct product of the one-dimensional distributions; so we should be able to formulate a criterion for independence in terms of characteristic functions.

Theorem 16.2. *Random variables* $\xi_1, ..., \xi_n$ *are independent if and only if their joint characteristic function is the product of their individual one-dimensional characteristic functions, taken each of the argument with the corresponding number*:

$$f_{\xi_1, ..., \xi_n}(t_1, ..., t_n) = f_{\xi_1}(t_1) \cdot ... \cdot f_{\xi_n}(t_n); \tag{16.3}$$

or, if the joint characteristic function factorizes into functions taken each of one one-dimensional argument:

$$f_{\xi_1, ..., \xi_n}(t_1, ..., t_n) = g_1(t_1) \cdot ... \cdot g_n(t_n). \tag{16.4}$$

Proof. That (16.3) (and with it also (16.4)) follows from independence we obtain from the complex version of Theorem 9.1 (formula (14.10)):

$$f_{\xi_1,\,\dots,\,\xi_n}(t_1,\,\dots,\,t_n) = Ee^{i\,(t_1\xi_1+\dots+t_n\xi_n)} = E[e^{it_1\xi_1}\cdot\dots\cdot e^{it_n\xi_n}] = Ee^{it_1\xi_1}\cdot\dots\cdot Ee^{it_n\xi_n}. \tag{16.5}$$

Now, suppose (16.3) holds for all $t_1,\,\dots,\,t_n$. We have to prove that the joint distribution $\mu_{\xi_1,\,\dots,\,\xi_n}$ is the same as the direct product $\mu_{\xi_1}\times\dots\times\mu_{\xi_n}$.

The characteristic function of the first of these two measures is equal to $f_{\xi_1,\,\dots,\,\xi_n}(t_1,\,\dots,\,t_n)$; that of the direct product

$$f_{\mu_{\xi_1}\times\dots\times\mu_{\xi_n}}(t_1,\,\dots,\,t_n) = \underset{\mathbb{R}^n}{\int\cdots\int} e^{i\,(t_1x_1+\dots+t_nx_n)}\,\mu_{\xi_1}(dx_1)\dots\mu_{\xi_n}(dx_n). \tag{16.6}$$

By Fubini's Theorem, this is equal to the iterated integral

$$\begin{gathered}\int_{-\infty}^{\infty}\Big[\dots\Big[\int_{-\infty}^{\infty} e^{i\,(t_1x_1+\dots+t_nx_n)}\,\mu_{\xi_n}(dx_n)\Big]\dots\Big]\,\mu_{\xi_1}(dx_1)\\ = \int_{-\infty}^{\infty} e^{it_1x_1}\,\mu_{\xi_1}(dx_1)\cdot\dots\cdot\int_{-\infty}^{\infty} e^{it_nx_n}\,\mu_{\xi_n}(dx_n) = f_{\xi_1}(t_1)\cdot\dots\cdot f_{\xi_n}(t_n).\end{gathered} \tag{16.7}$$

By (16.3) we see that the characteristic functions of the measures $\mu_{\xi_1,\,\dots,\,\xi_n}$ and $\mu_{\xi_1}\times$ $\times\dots\times\mu_{\xi_n}$ coincide for all $\boldsymbol{t}\in\mathbb{R}^n$, so these measures coincide: the random variables are independent.

Finally, if (16.4) is satisfied, but we don't know whether $g_k(t_k) = f_{\xi_k}(t_k)$: It is clear that

$$f_{\xi_k}(t_k) = Ee^{it_k\xi_k} = f_{\xi_1,\,\dots,\,\xi_n}(0,\,\dots,\,0,\,t_k,\,0,\,\dots,\,0) \tag{16.8}$$

with t_k in the k-th place, and all other arguments equal to 0. If (16.4) holds, we have:

$$f_{\xi_k}(t_k) = \prod_{j\neq k} g_j(0)\cdot g_k(t_k), \tag{16.9}$$

$$\prod_{k=1}^{n} f_{\xi_k}(t_k) = g_1(0)^{n-1}\cdot\dots\cdot g_n(0)^{n-1}\cdot\prod_{k=1}^{n} g_k(t_k). \tag{16.10}$$

Since

$$f_{\xi_1,\,\dots,\,\xi_n}(0,\,\dots,\,0) = \prod_{k=1}^{n} f_{\xi_k}(0) = 1, \tag{16.11}$$

we have

$$g_1(0)^{n-1}\cdot\dots\cdot g_n(0)^{n-1} = 1, \tag{16.12}$$

and (16.3) is satisfied.

Let us go to properties of characteristic functions that are more elementary than the uniqueness theorem (Theorems 15.1, 16.1).

Theorem 16.3. *Let ξ be a random variable, c and d real numbers; $\eta = c\xi + d$. Then*

$$f_\eta(t) = f_\xi(ct) \cdot e^{idt}. \tag{16.13}$$

Proof:

$$Ee^{it(c\xi+d)} = Ee^{i(ct)\xi} \cdot e^{idt}. \tag{16.14}$$

As an example, let us consider normal distributions. It turns out that normal distributions are easiest handled by means of their characteristic functions.

Let ξ be a random variable having the standard normal distribution, i. e., the normal distribution with parameters (0, 1). Its characteristic function is $f_\xi(t) = e^{-t^2/2}$ (see formula (14.34)). Let us take $\eta = c\xi + d$, where c and d are some real constants. From the elementary probability course we know that $c\xi + d$ has again a normal distribution with parameters... what are the parameters? The parameters of the normal distribution are the expectation (the mean) and the variance; they are easily found for the random variable η: $E\eta = E(c\xi + d) = c \cdot E\xi + d = c \cdot 0 + d = d$, $\mathrm{Var}(\eta) = E(\eta - d)^2 = E(c^2\xi^2) = c^2 \cdot E\xi^2 = c^2 \cdot \mathrm{Var}(\xi) = c^2 \cdot 1 = c^2$: the random variable $\eta = c\xi + d$ has a normal distribution with parameters $(d,\, c^2)$. (I don't know whether it was proved in your elementary probability course, and if yes, how it was proved; one possible way is to find the distribution function $F_\eta(y) = P\{c\xi + d \le y\}$, and after that, differentiating, find the density $p_\eta(y)$ – of course, this is not in the case of $c = 0$, in which case the random variable $\eta = 0 \cdot \xi + d$ is just a constant, and as such has no probability density.)

Now let us find the characteristic function of the random variable $\eta = c\xi + d$: by formula (16.13)

$$f_\eta(t) = e^{idt} \cdot f_\xi(ct) = e^{idt} \cdot e^{-(ct)^2/2} = e^{idt - c^2t^2/2}. \tag{16.15}$$

So: the characteristic function of the normal distribution with parameters $(a,\, b)$ (a being an arbitrary real number, and b positive) is

$$f(t) = e^{iat - bt^2/2} \tag{16.16}$$

(we took $d = a$, and $c = \sqrt{b}$ or $-\sqrt{b}$).

Now let us go to several dimensions.

Before now, it wasn't important to us whether we write our vectors as row vectors or as column ones. Let us set a standard now: we are going to write them as *column vectors.*

Theorem 16.4. *Let $\boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ \dots \\ \xi_n \end{pmatrix}$ be an n-dimensional random vector, C an $(m \times n)$-matrix, $\boldsymbol{d} = \begin{pmatrix} d_1 \\ \dots \\ d_m \end{pmatrix}$ an m-dimensional* (column) *vector* (C and $\boldsymbol{d}$ non-random). *Let $\boldsymbol{\eta} = C\boldsymbol{\xi} + \boldsymbol{d}$. Then*

$$f_{\boldsymbol{\eta}}(\boldsymbol{t}) = f_{\boldsymbol{\eta}}(t_1, \dots, t_m) = f_{\boldsymbol{\xi}}(C^T\boldsymbol{t}) \cdot e^{i\boldsymbol{d}\cdot\boldsymbol{t}}, \qquad \boldsymbol{t} \in \mathbb{R}^m \tag{16.17}$$

(^{T} here denotes the transposed matrix).

This is a multidimensional analog of Theorem 16.3. The **proof** is essentially the same:

$$Ee^{i\boldsymbol{t}\cdot(C\boldsymbol{\xi}+\boldsymbol{d})} = Ee^{i\boldsymbol{t}^T(C\boldsymbol{\xi})} \cdot e^{i\boldsymbol{t}^T\boldsymbol{d}} = e^{i\boldsymbol{t}^T\boldsymbol{d}} Ee^{i(\boldsymbol{t}^T C)\boldsymbol{\xi}} = e^{\boldsymbol{d}\cdot\boldsymbol{t}} Ee^{i(C^T\boldsymbol{t})\cdot\boldsymbol{\xi}} = e^{\boldsymbol{d}\cdot\boldsymbol{t}} f_{\boldsymbol{\xi}}(C^T\boldsymbol{t}). \tag{16.18}$$

Example: Let $\xi_1, \ldots, \xi_n$ be independent standard normal random variables; their joint characteristic function (the characteristic function of the random vector $\boldsymbol{\xi}$) is

$$f_{\boldsymbol{\xi}}(\boldsymbol{t}) = f_{\xi_1,\ldots,\xi_n}(t_1, \ldots, t_n) = \prod_{k=1}^{n} f_{\xi_k}(t_k) = e^{-(t_1^2+\ldots+t_n^2)/2} = e^{-|\boldsymbol{t}|^2/2}. \tag{16.19}$$

Let C be an $(n \times n)$-matrix, $\boldsymbol{a}$ an n-dimensional vector. Then the random vector $\boldsymbol{\eta} = C\boldsymbol{\xi} + \boldsymbol{a}$ has the characteristic function equal to

$$\begin{aligned} f_{\boldsymbol{\eta}}(\boldsymbol{t}) &= f_{\boldsymbol{\xi}}(C^T\boldsymbol{t}) \cdot e^{i\boldsymbol{a}\cdot\boldsymbol{t}} = \exp\{-|\boldsymbol{t}C|^2/2 + i\boldsymbol{a}\cdot\boldsymbol{t}\} \\ &= \exp\{-(\boldsymbol{t}^T C)\cdot(C^T\boldsymbol{t})/2 + i\boldsymbol{a}\cdot\boldsymbol{t}\} = \exp\{-\boldsymbol{t}^T(C\cdot C^T)\boldsymbol{t}/2 + i\boldsymbol{t}\boldsymbol{a}\}. \end{aligned} \tag{16.20}$$

In terms of $t_1, \ldots, t_n$ we can rewrite this formula as

$$f_{\eta_1,\ldots,\eta_n}(t_1, \ldots, t_n) = \exp\Bigl\{-\tfrac{1}{2}\sum_{k,\,l=1}^{n} b_{kl}t_kt_l - i\sum_{k=1}^{n} a_kt_k\Bigr\}, \tag{16.21}$$

where b_{kl} are the entries of the matrix $B = C \cdot C^T$:

$$b_{kl} = \sum_{j=1}^{n} c_{kj}c_{lj}, \tag{16.22}$$

c_{kj} being the entries of the matrix C. The first sum in (16.21) is the quadratic form with the matrix B.

Let $\boldsymbol{a}$ be an n-dimensional vector, and B an $(n \times n)$-matrix. By definition, the n-dimensional normal distribution with parameters $(\boldsymbol{a}, B)$ is one whose characteristic function is given by formula (16.21).

Our example tells us that such a thing as the multidimensional normal distribution *exists*. But can we assign its parameters $(\boldsymbol{a}, B)$ arbitrarily? It turns out that no: the matrix B has to be nonnegative definite:

$$\boldsymbol{t}^T B\boldsymbol{t} = \sum_{k,\,l=1}^{n} b_{kl}t_kt_l \ge 0 \qquad \text{for all } \boldsymbol{t} = \begin{pmatrix} t_1 \\ \ldots \\ t_n \end{pmatrix} \in \mathbb{R}^n. \tag{16.23}$$

Why? If $\sum_{k,\,l=1}^{n} b_{kl}t_kt_l < 0$ for some $t_1, \ldots, t_n$, we have

$$|f(t_1, \ldots, t_n)| = \exp\Bigl\{-\tfrac{1}{2}\sum_{k,\,l} b_{kl}t_kt_l\Bigr\} > 1, \tag{16.24}$$

which is impossible.

Also, we know that the matrix of a quadratic form $\sum_{k,\,l} b_{kl}t_kt_l$ can be chosen symmetric: if it is not, we replace (b_{kl}) with $(\tilde{b}_{kl})$, $\tilde{b}_{kl} = \dfrac{b_{kl} + b_{lk}}{2}$ with the same value of the quadratic form: $\sum_{k,\,l} \tilde{b}_{kl}t_kt_l = \sum_{k,\,l} b_{kl}t_kt_l$.

Let $\boldsymbol{a} = \begin{pmatrix} a_1 \\ ... \\ a_n \end{pmatrix}$ *be an arbitrary column vector* $\in \mathbb{R}^n$; B *a symmetric nonnegative-definite* $(n \times n)$*-matrix. Then there exists an* n*-dimensional distribution with characteristic function*

$$f(\boldsymbol{t}) = \exp\bigl\{-\tfrac{1}{2}\boldsymbol{t}^T B\boldsymbol{t} + i\boldsymbol{a}^T\boldsymbol{t}\bigr\}. \tag{16.25}$$

Proof: Every symmetric matrix can be reduced to the diagonal form:

$$B = A \cdot \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \cdot A^T \tag{16.26}$$

with zeros in the matrix $\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$ outside the diagonal. For a nonnegative-definite matrix B all λ_k are nonegative. Let us take

$$C = A \cdot \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix}; \tag{16.27}$$

we have:

$$\begin{aligned} C \cdot C^T &= A \cdot \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix} \cdot \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix}^T \cdot A^T \\ &= A \cdot \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix} \cdot \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix} \cdot A^T \\ &= A \cdot \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \cdot A^T = B. \end{aligned} \tag{16.28}$$

So if we take independent standard normal random variables ξ_1, ..., ξ_n, form a column vector $\boldsymbol{\xi}$, and take $\boldsymbol{\eta} = C\boldsymbol{\xi} + \boldsymbol{a}$, the n-dimensional random vector $\boldsymbol{\eta}$ will have the characteristic function (16.25).

Our new definition of the n-dimensional normal distribution changes something in the one-dimensional case: before now, we considered only normal distributions with *positive* second parameter b; and now we can consider the normal distribution with parameters $(a,\ 0)$ (the (1×1)-matrix 0 is certainly nonnegative definite): this is the distribution with characteristic function $f(t) = e^{iat}$. It is easy to see that this is the distribution δ_a concentrated at the point a: $\delta_a(C) = 1$ if $C \ni a$, and $= 0$ for $C \not\ni a$.