Lectures 22 – 23. Central limit theorem.

The name "*central* limit theorem" suggests that it is something of utmost importance; and indeed this is so.

One would imagine that the central limit theorem is some very important theorem of probability theory, having to do with limits. This is not exactly so.

In fact, "the central limit theorem", just like "the law of large numbers", is not the name of some theorem, but rather that of a whole *group* of theorems that have similar formulation.

Theorems of this group are formulated in the following way:

Let $\xi_1, \xi_2, ..., \xi_n, ...$ *be a sequence of random variables with finite expectations and variances*; *let*

$$\zeta_n = \frac{\xi_1 + ... + \xi_n - E(\xi_1 + ... + \xi_n)}{\sqrt{\mathrm{Var}(\xi_1 + ... + \xi_n)}}. \tag{22–23.1}$$

Then, under such and such conditions, *the distribution* μ_{ζ_n} *of the random variable* ζ_n (*of the* normalized sum *of the first* n *random variables*) *converges*, in such and such sense, *to the standard normal distribution.*

(Formula (22 – 23.1) is a particular case of formula (21.6) with $A_n = E(\xi_1 + ... + \xi_n)$, $B_n = \sqrt{\mathrm{Var}(\xi_1 + ... + \xi_n)}$.)

Different theorems of this group differ from one another by the *conditions* imposed on the random variables ξ_k, and by *in what sense* we understand the convergence of distributions.

In these lectures we introduced only one type of convergence of distributions: weak convergence, so we will be considering theorems of this group in which we state weak convergence; but there are theorems belonging to the group of the central limit theorem in which, e. g., convergence of the *densities* is stated – such theorems could end like this:

... then there exists a continuous version $p_{\zeta_n}(x)$ *of the density of the normalized sum* (remember, there are different *versions* of each density, that differ from one another only on sets of zero Lebesgue measure), *and*

$$p_{\zeta_n}(x) \to \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{as } n \to \infty, \quad \text{uniformly in } x \in (-\infty, \infty). \tag{22–23.2}$$

Most of the theorems in the central limit theorem group require that the random variables $\xi_1, \xi_2, ..., \xi_n, ...$ should be *independent88*; but there are some theorems in this group that work for dependent random variables. Perhaps we'll come to such theorems in this course.

In the language of "applied mathematics", what is said around formula (22 – 23.1) can be formulated like this: under such and such conditions, if the number of summands n is very large, the distribution of the normalized sum ζ_n is *approximately* the normal distribution with parameters $(0, 1)$. Remembering what happens with a normal distribution when we take a linear function of a random variable, we can also reformulate it as follows: under such and such conditions, if the number of summands n is very large, the distribution of the sum $\xi_1 + ... + \xi_n$ is approximately the normal distribution with parameters $\big(E(\xi_1 + ... + \xi_n),\ \mathrm{Var}(\xi_1 + ... + \xi_n)\big)$.

Let us formulate and prove our first theorem of this group.

Theorem 22 – 23.1. *Let $\xi_1, \xi_2, ..., \xi_n, ...$ be a sequence of independent and identically distributed random variables with finite expectation $m = E\xi_k$ and variance $\sigma^2 = \mathrm{Var}(\xi_k)$; let*

$$\zeta_n = \frac{\xi_1 + ... + \xi_n - nm}{\sigma\sqrt{n}} \qquad (22-23.3)$$

(the expectation of the sum $\xi_1 + ... + \xi_n$ is clearly equal to nm, its variance to $n\sigma^2$). *Then the distribution μ_{ζ_n} of the normalized sum ζ_n converges weakly, as $n \to \infty$, to the standard normal distribution.*

Proof. Better than having to do with the characteristic function $f_{\xi_k}(t)$, let us introduce the random variables $\xi_k^\circ = \xi_k - m$. These random variables are too independent and identically distributed, $E\xi_k^\circ = E\xi_k - m = 0$, $\mathrm{Var}(\xi_k^\circ) = \sigma^2$, $f'_{\xi_k^\circ}(0) = 0$, $f''_{\xi_k^\circ}(0) = -\sigma^2$,

$$\zeta_n = \frac{\xi_1^\circ + ... + \xi_n^\circ}{\sigma\sqrt{n}}, \qquad (22-23.4)$$

and

$$f_{\zeta_n}(t) = \big[f_{\xi_k^\circ}\big(\frac{t}{\sigma\sqrt{n}}\big)\big]^n = \exp\big\{n \cdot \ln_1\big(f_{\xi_k^\circ}\big(\frac{t}{\sigma\sqrt{n}}\big)\big)\big\} \qquad (22-23.5)$$

(for sufficiently large n – such that $\big|f_{\xi_k^\circ}\big(\frac{t}{\sigma\sqrt{n}}\big) - 1\big| < 1$).

In our proof of Khinchin's Theorem, we used the Taylor (Maclaurin) approximation of the characteristic function with terms up to the first power; here we are going to use one with terms up to the *second* power:

$$f_{\xi_k^\circ}(t) \ = f_{\xi_k^\circ}(0) + f'_{\xi_k^\circ}(0) \cdot t + \tfrac{1}{2}\, f''_{\xi_k^\circ}(0) \cdot t^2 + o(t^2) \ = 1 - \frac{\sigma^2}{2} \cdot t^2 + o(t^2) \quad (t \to 0). \quad (22-23.6)$$

We have, for every fixed $t \in (-\infty, \infty)$:

$$f_{\xi_k^\circ}\big(\frac{t}{\sigma\sqrt{n}}\big) = 1 - \frac{\sigma^2}{2} \cdot \big(\frac{t}{\sigma\sqrt{n}}\big)^2 + o\big(\big(\frac{t}{\sigma\sqrt{n}}\big)^2\big) = 1 - \frac{t^2}{2n} + o(1/n) \quad (n \to \infty); \quad (22-23.7)$$

$$\ln_1\big(f_{\xi_k^\circ}\big(\frac{t}{\sigma\sqrt{n}}\big)\big) = -\frac{t^2}{2n} + o(1/n) + o\big(-\frac{t^2}{2n} + o(1/n)\big) = -\frac{t^2}{2n} + o(1/n) \quad (n \to \infty); \qquad (22-23.8)$$

$$\lim_{n\to\infty} n \cdot \ln_1\big(f_{\xi_k^\circ}\big(\frac{t}{\sigma\sqrt{n}}\big)\big) = -\frac{t^2}{2}, \qquad (22-23.9)$$

$$\lim_{n\to\infty} f_{\zeta_n}(t) = e^{-t^2/2}. \qquad (22-23.10)$$

This is the characteristic function of the standard normal distribution; so our theorem is proved.

Let us go to the multidimensional case.

The multidimensional analog of the expectation $E\xi$ of a random variable is the vector of expectations $E\boldsymbol{\xi}$, where $\boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ ... \\ \xi_r \end{pmatrix}$ (I am reserving the letter n for numbering the random vectors in a sequence, so I denote the dimension with r); and that of the variance $\mathrm{Var}(\xi)$ is the *covariance matrix*

$$E\big((\boldsymbol{\xi}^T - E\boldsymbol{\xi}^T)\cdot(\boldsymbol{\xi} - E\boldsymbol{\xi})\big) = \begin{pmatrix} \mathrm{Var}(\xi_1) & \mathrm{Cov}(\xi_1, \xi_2) & ... & \mathrm{Cov}(\xi_1, \xi_r) \\ \mathrm{Cov}(\xi_2, \xi_1) & \mathrm{Var}(\xi_2) & ... & \mathrm{Cov}(\xi_2, \xi_r) \\ .\ .\ .\ . & .\ .\ .\ . & ... & .\ .\ .\ . \\ \mathrm{Cov}(\xi_r, \xi_1) & \mathrm{Cov}(\xi_r, \xi_2) & ... & \mathrm{Var}(\xi_r) \end{pmatrix}. \quad (22-23.11)$$

But what would the square root of this matrix (the covariance matrix of the sum $\boldsymbol{\xi}_1 + ... + \boldsymbol{\xi}_n$) be? and what would division by this square root mean? Well, if this square root is a non-singular matrix, we can imagine that we should multiply the vector $\boldsymbol{\xi}_1 + ... + \boldsymbol{\xi}_n$ by the inverse of this matrix (on the left, of course: the product $\boldsymbol{x} \cdot C$ makes no sense if C is a matrix, and $\boldsymbol{x}$ a column vector). But what if the "square root" (whatever it is) is a singular matrix?

In the one-dimensional case we did not have to care about this: if $\mathrm{Var}(\xi_1+...+\xi_n) = 0$, this random variable is almost surely constant. The same is true in the multidimensional case: if the covariance matrix is the zero matrix, all components of the random vector are just constants, and the random vector (almost surely) is a (vector) constant. But in the multidimensional case there are plenty of matrices that are singular, but not zero (matrices of rank k, $0 < k < r$).

So there may be some complications in formulating theorems of the central limit theorem group in the multidimensional case.

Let me formulate one theorem of this group:

Theorem 22 – 23.2. *Let $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$, ..., $\boldsymbol{\xi}_n$, ... be a sequence of independent identically distributed r-dimensional random vectors with $E\boldsymbol{\xi}_k = \boldsymbol{m}$ and covariance matrix B; let*

$$\boldsymbol{\zeta}_n = \frac{1}{\sqrt{n}} \cdot (\boldsymbol{\xi}_1 + ... + \boldsymbol{\xi}_n - n \cdot \boldsymbol{m}). \quad (22-23.12)$$

The distribution of the random vector $\boldsymbol{\zeta}_n$ converges weakly as $n \to \infty$ to the r-dimensional normal distribution with parameters $(\mathbf{0}, B)$ (with zero expectation and the covariance matrix B).

Writing an analog to formula (22 – 23.3), we just did not introduce anything instead of the division by σ; so the limiting distribution turns out not to be the standard normal, but one corresponding to the matrix B.

The **proof** is done the same way, only we use the *r-dimensional* Taylor formula.

Limit theorems of probability theory form a vast field within it; and from Theorems 22 – 23.1, 22 – 23.2 we can move either to theorems with the summands ξ_k that have different distributions; or to limit theorems with *dependent* summands; or to those with non-normal limiting distributions (see Problem $\boxed{\mathbf{45}}$); etc. But before going away from Theorem 22 – 23.2, I would rather show you one application of this theorem that is useful in mathematical statistics (being the reason why we can use the so-called χ^2-test).

Theorem 22 – 23.3. *Let ξ_1, ξ_2, ..., ξ_n, ... be independent identically distributed random variables taking values x^1, ..., x^m with probabilities p_1, ..., p_m ($\sum_{i=1}^m p_i = 1$, $p_i > 0$, $i = 1$, ..., m).*

Let η_{ni}, $1 \le i \le m$, be the number of times we observe the value x^i in the first n experiments:

$$\eta_{ni} = \#\{k\colon\, 1 \le k \le n,\, \xi_k = x^i\}. \tag{22–23.13}$$

Take

$$\chi_n^2 = \sum_{i=1}^m \frac{(\eta_{ni} - np_i)^2}{np_i}. \tag{22–23.14}$$

The distribution of the random variable χ_n^2 converges weakly as $n \to \infty$ to the χ^2-distribution with $m-1$ degrees of freedom (that is, to the distribution of the sum of squares of $m-1$ squares of independent standard normal random variables).

Proof. Unfortunately, the letters ξ_k are already used for our original random variables (that I have no intention of adding); so I am going to use the letter capital Ξ.

Let us introduce the random variables

$$\Xi_{ki} = \begin{cases} 1, & \xi_k = x^i, \\ 0, & \xi_k \ne x^i, \end{cases} \tag{22–23.15}$$

and the random vectors

$$\mathbf{\Xi}_k = \begin{pmatrix} \Xi_{k1} \\ \Xi_{k2} \\ \dots \\ \Xi_{km} \end{pmatrix}. \tag{22–23.16}$$

Of course the random vectors $\mathbf{\Xi}_1$, $\mathbf{\Xi}_2$, ..., $\mathbf{\Xi}_n$, ..., defined each as the same function of the random variable ξ_k with the same number, are independent and identically distributed (taking the values $\begin{pmatrix} 1 \\ 0 \\ \dots \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \\ \dots \\ 0 \end{pmatrix}$, ..., $\begin{pmatrix} 0 \\ 0 \\ \dots \\ 1 \end{pmatrix}$ with probabilities p_1, p_2, ..., p_m). Let us find the expectation $\boldsymbol{m}$ and the covariance matrix B of these random vectors.

We have:

$$E\Xi_{ki} = p_i \cdot 1 + (1 - p_i) \cdot 0 = p_i; \tag{22–23.17}$$

$$\mathrm{Cov}(\Xi_{ki},\, \Xi_{ki}) = \mathrm{Var}(\Xi_{ki}) = E\Xi_{ki}^2 - (E\Xi_{ki})^2 = E\Xi_{ki} - (E\Xi_{ki})^2 = p_i(1 - p_i), \tag{22–23.18}$$

and for $i \ne j$

$$\mathrm{Cov}(\Xi_{ki},\, \Xi_{kj}) = E(\Xi_{ki} \cdot \Xi_{kj}) - E\Xi_{ki} \cdot E\Xi_{kj} = 0 - p_i p_j = -p_i p_j. \tag{22–23.19}$$

So the covariance matrix is

$$B = \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \dots & -p_1p_m \\ -p_2p_1 & p_2(1-p_2) & \dots & -p_2p_m \\ \dots\dots & \dots\dots & \dots & \dots\dots \\ -p_mp_1 & -p_mp_2 & \dots & p_m(1-p_m) \end{pmatrix}. \qquad (22-23.20)$$

This matrix is singular: its columns are linearly dependent. Indeed, the sum of these columns with coefficients equal to 1 is equal to $\begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \end{pmatrix}$, because $\sum_{j:\, j\neq i} p_j = 1 - p_i$. In fact, the rank of the matrix B is equal to $m-1$. Problems $\boxed{\mathbf{34}}$, $\boxed{\mathbf{35}}$ suggest us that the distribution of a random vector $\boldsymbol{\zeta}$ having $E\boldsymbol{\zeta} = \mathbf{0}$ and the covariance matrix (22 – 23.20) must be concentrated on a linear subspace of some lower dimension. Indeed, it is easy to check that almost surely

$$\zeta_1 + \zeta_2 + \dots + \zeta_m = 0. \qquad (22-23.21)$$

Checking: the expectation of the sum (22 – 23.21) is equal to 0; let us evaluate the variance of this random variable:

$$\begin{aligned} \mathrm{Var}\Bigl(\sum_{i=1}^{m} \zeta_i\Bigr) &= E\Bigl(\sum_{i=1}^{m} \zeta_i\Bigr)^2 = E\Bigl(\sum_{i=1}^{m} \zeta_i \cdot \sum_{j=1}^{m} \zeta_j\Bigr) \\ &= \sum_{i,\, j=1}^{m} E(\zeta_i \cdot \zeta_j) = \sum_{i,\, j=1}^{m} \mathrm{Cov}(\zeta_i,\, \zeta_j) = 0. \end{aligned} \qquad (22-23.22)$$

Haven't I mentioned that if the variance of a random variable is equal to 0, then it is almost surely equal to its expectation? It seems not. Let me do it now:

Theorem 22 – 23.4. *If* $\mathrm{Var}(\xi) = 0$, *then almost surely* $\xi = E\xi$.

Proof. By Chebyshev's inequality, for every positive ε

$$P\{|\xi - E\xi| \geq \varepsilon\} \leq \frac{\mathrm{Var}(\xi)}{\varepsilon^2} = 0. \qquad (22-23.23)$$

Taking $\varepsilon = 1/n$, and using the obvious set equality:

$$\{\omega\colon\ \xi(\omega) \neq E\xi\} = \bigcup_{n=1}^{\infty} \{\omega\colon\ |\xi(\omega) - E\xi| \geq 1/n\}, \qquad (22-23.24)$$

we get:

$$P\{\xi \neq E\xi\} = \lim_{n\to\infty} P\{|\xi - E\xi| \geq 1/n\} = 0. \qquad (22-23.25)$$

Now let us return to our random vectors $\boldsymbol{\Xi}_k$ and the random variable χ_n^2. We have:

$$\boldsymbol{\zeta}_n = \frac{1}{\sqrt{n}}\,(\boldsymbol{\Xi}_1 + ... + \boldsymbol{\Xi}_n - n\cdot\boldsymbol{m}) = \begin{pmatrix} \dfrac{\eta_{n1} - np_1}{\sqrt{n}} \\ \dfrac{\eta_{n2} - np_2}{\sqrt{n}} \\ \cdots\cdot \\ \dfrac{\eta_{nm} - np_m}{\sqrt{n}} \end{pmatrix}; \qquad (22-23.26)$$

and

$$\chi_n^2 = h(\boldsymbol{\zeta}_n), \qquad (22-23.27)$$

where

$$h(\boldsymbol{z}) = \sum_{i=1}^m \frac{z_i^2}{p_i}. \qquad (22-23.28)$$

By Theorem 22 – 23.2, the distribution of $\boldsymbol{\zeta}_n$ converges weakly as $n \to \infty$ to the normal distribution with parameters $(\mathbf{0},\, B)$.

Does it mean that the (one-dimensional) distribution of the random variable $h(\boldsymbol{\zeta}_n)$ also converges weakly to the distribution of the random variable $h(\boldsymbol{\zeta}_\infty)$, where $\boldsymbol{\zeta}_\infty$ is some random vector having the normal distribution with parameters $(\mathbf{0},\, B)$?

I am including the proof of Theorem 22 – 23.3 in this lecture note even though it was given in the next lecture.

Theorem 22 – 23.5 (I am calling this a theorem rather than a lemma because the statement can be interesting in its own right). *Let $\boldsymbol{\eta}_1,\, \boldsymbol{\eta}_2,\, ...,\, \boldsymbol{\eta}_n,\, ...$ be a sequence of r-dimensional random vectors; and let $\boldsymbol{\eta}_\infty$ be some other random vector. Let $h(\boldsymbol{x})$ be a continuous function $\mathbb{R}^r \mapsto \mathbb{R}^l$.*

If the distribution of $\boldsymbol{\eta}_n$ converges weakly to that of $\boldsymbol{\eta}_\infty$ as $n \to \infty$: $\mu_{\boldsymbol{\eta}_n} \to_w \mu_{\boldsymbol{\eta}_\infty}$, then the distribution of the l-dimensional random vector $h(\boldsymbol{\eta}_n)$ converges weakly as $n \to \infty$ to the distribution of $h(\boldsymbol{\eta}_\infty)$.

Proof. Convergence $\mu_{\boldsymbol{\eta}_n} \to_w \mu_{\boldsymbol{\eta}_\infty}$ means that for every bounded continuous function $f(\boldsymbol{x})$ on $\mathbb{R}^r$

$$\lim_{n\to\infty} Ef(\boldsymbol{\eta}_n) = Ef(\boldsymbol{\eta}_\infty) \qquad (22-23.29)$$

(see formula (17.29)). To prove that $\mu_{h(\boldsymbol{\eta}_n)} \to_w \mu_{h(\boldsymbol{\eta}_\infty)}$, we have to check that

$$\lim_{n\to\infty} Eg\big(h(\boldsymbol{\eta}_n)\big) = Eg\big(h(\boldsymbol{\eta}_\infty)\big) \qquad (22-23.30)$$

for every bounded continuous function g on $\mathbb{R}^l$.

But the function $f(\boldsymbol{x}) = g\big(h(\boldsymbol{x})\big)$ is bounded (by the same constant as the function g) and continuous, so (22 – 23.30) follows from (22 – 23.29).

Note that we need this theorem now only for $h\colon \mathbb{R}^r \mapsto \mathbb{R}^1$; but the theorem clearly holds also for random variables and functions of them taking values in arbitrary metric (or topological) spaces.

So by Theorem 22 – 23.5 we have that the weak limit of the distributions of the random variables χ_n^2 exists, and it is the same as the distribution of the random variable

$$h(\boldsymbol{\zeta}_\infty) = \sum_{i=1}^{m} \frac{\zeta_{\infty\, i}^2}{p_i}, \tag{22 – 23.31}$$

where the m-dimensional random vector $\boldsymbol{\zeta}_\infty = \begin{pmatrix} \zeta_{\infty\, 1} \\ \zeta_{\infty\, 2} \\ \dots \\ \zeta_{\infty\, m} \end{pmatrix}$ has the m-dimensional normal distribution with zero expectation and the covariance matrix (22 – 23.20).

According to (22 – 23.21), we can express everything through $m-1$ random variables only. With probability 1 we have:

$$\zeta_{\infty\, m} = -\,\zeta_{\infty\, 1} - \dots - \zeta_{\infty,\, m-1}, \tag{22 – 23.32}$$

and we can rewrite the expression (22 – 23.31) as

$$h(\boldsymbol{\zeta}_\infty) = \sum_{i=1}^{m-1} \frac{\zeta_{\infty i}^2}{p_i} + \frac{(\sum_{i=1}^{m-1} \zeta_{\infty i})^2}{p_m} = \sum_{i,\, j=1}^{m-1} q_{ij} \cdot \zeta_{\infty i} \zeta_{\infty j}, \tag{22 – 23.33}$$

the $(m-1) \times (m-1)$-matrix $Q = (q_{ij})_{i,\, j=1}^{m-1}$ of this quadratic form being given by

$$Q = \begin{pmatrix} \frac{1}{p_1} + \frac{1}{p_m} & \frac{1}{p_m} & \dots & \frac{1}{p_m} \\ \frac{1}{p_m} & \frac{1}{p_2} + \frac{1}{p_m} & \dots & \frac{1}{p_m} \\ \dots\dots & \dots\dots & \dots & \dots\dots \\ \frac{1}{p_m} & \frac{1}{p_m} & \dots & \frac{1}{p_{m-1}} + \frac{1}{p_m} \end{pmatrix}. \tag{22 – 23.34}$$

The $(m-1)$-dimensional random vector $\begin{pmatrix} \zeta_{\infty\, 1} \\ \zeta_{\infty\, 2} \\ \dots \\ \zeta_{\infty,\, m-1} \end{pmatrix}$ has the normal distribution with zero expectation and the covariance matrix

$$B_{(m-1)\times(m-1)} = \begin{pmatrix} p_1(1-p_1) & -p_1 p_2 & \dots & -p_1 p_{m-1} \\ -p_2 p_1 & p_2(1-p_2) & \dots & -p_2 p_{m-1} \\ \dots\dots & \dots\dots & \dots & \dots\dots \\ -p_{m-1} p_1 & -p_{m-1} p_2 & \dots & p_{m-1}(1-p_{m-1}) \end{pmatrix}, \tag{22 – 23.35}$$

obtained form the matrix (22 – 23.20) by deleting the last row and the last column.

As I mentioned before, the $m \times m$-matrix B given by (22 – 23.20) has the rank $m-1$, so we have a good chance of having the matrix $B_{(m-1)\times(m-1)}$ being nonsingular, and the joint distribution of $\zeta_{\infty\,1}$, $\zeta_{\infty\,2}$, ..., $\zeta_{\infty,\,m-1}$ (the $(m-1)$-dimensional distribution of the corresponding random vector) being a continuous one.

I *mentioned* this, but did not prove it. But luckily, the inverse matrix of the matrix $B_{(m-1)\times(m-1)}$ has, in fact, already been written: it is the matrix Q given by (22 – 23.34). Indeed, let us check that $Q \cdot B_{(m-1)\times(m-1)}$ is the identity matrix. The $(i,\, i)$-th entry of this product $(1 \le i \le m-1)$ is equal to

$$\Big(\frac{1}{p_i} + \frac{1}{p_m}\Big) \cdot p_i(1-p_i) + \sum_{k:\,1\le k\le m-1,\,k\ne i} \frac{1}{p_m} \cdot (-\,p_i p_k) = (1-p_i) + \frac{p_i}{p_m} \cdot \Big(1 - \sum_{k=1}^{m-1} p_k\Big) = 1; \tag{22 – 23.36}$$

its $(i,\, j)$-th entry $(i \ne j)$ is

$$\begin{aligned}\Big(\frac{1}{p_i} + \frac{1}{p_m}\Big) \cdot (-\,p_i p_j) + \frac{1}{p_m} \cdot p_j(1-p_j) + \sum_{k:\,1\le k\le m-1,\,k\ne i,\,j} \frac{1}{p_m} \cdot (-\,p_j p_k) \\ = -\,p_j + \frac{p_j}{p_m} \cdot \Big(1 - \sum_{k=1}^{m-1} p_k\Big) = 0.\end{aligned} \tag{22 – 23.37}$$

Now let us formulate a general result which proves Theorem 22 – 23.3:

Theorem 22 – 23.6. *Let an r-dimensional random vector $\boldsymbol{\zeta}$ have the normal distribution with zero expectation and a nonsingular covariance matrix B; let $Q = (q_{ij})_{i,\,j=1}^{r} = B^{-1}$. Then the random variable*

$$\eta = \sum_{i,\,j=1}^{r} q_{ij}\zeta_i\zeta_j = \boldsymbol{\zeta}^T Q\, \boldsymbol{\zeta} \tag{22 – 23.38}$$

has the χ^2-distribution with r degrees of freedom.

Proof. Remember how we found that the formula (16.21) defined a characteristic function: we took independent standard normal random variables ξ_i forming a random vector $\boldsymbol{\xi}$ having the normal distribution with parameters $(\mathbf{0},\, I)$, where I was the identity matrix; we represented the matrix B as $B = C \cdot C^T$; and we took the random vector $C\boldsymbol{\xi} + \boldsymbol{a}$. Let us do the same now.

Let $B = C \cdot C^T$ (where C is an $(r \times r)$-matrix); then

$$Q = B^{-1} = (C \cdot C^T)^{-1} = (C^T)^{-1} \cdot C^{-1}. \tag{22 – 23.39}$$

The random vector $\boldsymbol{\zeta} = C\boldsymbol{\xi}$ has the normal distribution with parameters $(\mathbf{0},\, B)$; and we have:

$$\begin{aligned}\eta = \boldsymbol{\zeta}^T Q\, \boldsymbol{\zeta} = (C\boldsymbol{\xi})^T Q\, C\boldsymbol{\xi} = \boldsymbol{\xi}^T C^T [(C^T)^{-1} \cdot C^{-1}]\, C\boldsymbol{\xi} \\ = \boldsymbol{\xi}^T [C^T (C^T)^{-1} C^{-1} C]\boldsymbol{\xi} = \boldsymbol{\xi}^T I \boldsymbol{\xi} = \sum_{i=1}^{r} \xi_i^2.\end{aligned} \tag{22 – 23.40}$$

This is exactly what random variable has the χ^2-distribution by definition.

Theorem 22 – 23.6, and Theorem 22 – 23.3 with it, are proved.