STATISTICAL ANALYSIS OF ONE-DIMENSIONAL DISTRIBUTIONS

By

ROBERT SCHMIDT

The present research is to be considered as a contribution to a range of science in which the pioneer work has been done by K. PEARSON. The method for analysing statistical distributions to be developed here differs in principle—as far as the author can see—from the known ones. The mathematical resources are all well known and so simple that their deduction *ab ovo* could be carried through on a few pages; hence this investigation is intelligible to anyone who remembers his mathematical knowledge acquired at school.

The main resource consists of the process of orthogonalization, fundamental in the theory of integral equations. The central idea characterizing the following is, not to deal with a frequency function itself, nor with its integral function, but with the *inverse* of the integral function. The general scope will be given in No. 3.

The author is indebted to his wife and to Mr. J. L. K. GIFFORD, M.A., of Queensland University for kind help in revising the English text.

1. DESIGNATIONS AND GENERAL ASSUMPTIONS

A curve $y=\varphi(x)$, $(-\infty<x<+\infty)$ shall be called a *"frequency curve"*, the function $\varphi(x)$ a *"frequency function"*, if $\varphi(x)$ satisfies the following conditions:

1. $\varphi(x) \geqq 0 \qquad (-\infty<x<+\infty)$
2. *The moments* $\mu_K = \int_{-\infty}^{+\infty} x^K \varphi(x)\, dx$ *exist for* $K=0,1,\dots$ [1]
3. $\mu_0 = 1$.

For our purposes it is convenient—though not necessary—to

---

[1] In this paper we shall not have to make use of the second condition (except in the special case $K=0$); in further notes, too, the condition will never be applied to its full extent.

add a fourth condition which it is simplest to formulate by using the function
$$\phi(x) = \int_{-\infty}^{x} \varphi(t)\,dt.$$

This function is constantly increasing in $-\infty < x < +\infty$, and we have
$$\lim_{x\to-\infty} \phi(x) = 0, \quad \lim_{x\to+\infty} \phi(x) = 1$$
The fourth condition is to guarantee that $\phi(x)$ assumes every value from $0<y<1$ just once, so that $\phi(x)$ possesses a unique inverse function in the ordinary sense.

4. a) $\phi(x)$ is continuous

b) At every x where $0 < \phi(x) < 1$, $\phi(x)$ is increasing (strictly speaking), that is: From $x' < x < x''$ it always follows that $\phi(x') < \phi(x) < \phi(x'')$.

When the conditions 1 - 4 are fulfilled, let us denote $\phi(x)$ as the *"frequency integral"* of the frequency function $\varphi(x)$.

Then there exists one and only one function $\psi(y)$ in $0<y<1$, satisfying $\psi[\phi(x)] \equiv x \qquad (0 < \phi(x) < 1)$
and $\psi(y)$ is called the *inverse function* of $\phi(x)$. This function $\psi(y)$ is continuous and constantly increasing (strictly speaking), and therefore possesses a unique inverse, namely $\phi(x)$:
$$\phi[\psi(y)] \equiv y \qquad (0<y<1).$$

We give here some special examples of frequency curves.

I. The "Step Curve".
$$y = \varphi(x) = \begin{cases} 1 & \text{in } 0 \leqq x \leqq 1 \\ 0 & \text{otherwise.} \end{cases}$$

The moments are $\mu_k = \frac{1}{k+1} \qquad (k = 0, 1, 2, \cdots).$

The frequency integral is
$$\phi(x) = \begin{cases} 0 & \text{in } -\infty < x < 0 \\ x & \text{in } 0 \leqq x < 1 \\ 1 & \text{in } 1 \leqq x < +\infty \end{cases}$$
The inverse is
$$\psi(y) = y \qquad (0<y<1)$$

II. The Normal Law of Error.

$$y = \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

with the moments

$$\mu_K = \begin{cases} \dfrac{(2n)!}{2^n \cdot n!} & \text{for } K = 2n. \\ 0 & \text{for } K = 2n+1 \quad (n = 0, 1, \cdots) \end{cases}$$

and the frequency integral

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

There are a number of tables of the numerical values of this function. Of course these tables can be used to compute the values of $\psi(y)$. Considering the fact that, for our purposes, the values of $\psi(y)$ will often be required for simple rational arguments only, it seems useful to have tables which are converse to those just quoted, that is to say, the tabulated entry of which is $x = \psi(y)$ and the argument $y = \phi(x)$. Such tables have been calculated by KELLEY and WOOD (Statistical Method, New York 1924; Appendix C).

III. The Laplace Curve.

$$y = \varphi(x) = \frac{1}{2} e^{-|x|}$$

$$\mu_K = \begin{cases} (2n)! = K! & \text{for } K = 2n \\ 0 & \text{for } K = 2n+1 \quad (n = 0, 1, \cdots) \end{cases}$$

$$\phi(x) = \begin{cases} \frac{1}{2} e^{x} & \text{in } -\infty < x < 0 \\ 1 - \frac{1}{2} e^{-x} & \text{in } 0 \leq x < +\infty \end{cases}$$

$$\psi(y) = \begin{cases} \log y + \log 2 & \text{in } 0 < y < 1/2 \\ -\log(1-y) - \log 2 & \text{in } 1/2 \leq y < 1 \end{cases}$$

IV. The "Tine Curve".

$$y = \varphi(x) = \begin{cases} 0 & \text{in} \quad -\infty < x < -1 \\ 1+x & \text{in} \quad -1 \leqq x < 0 \\ 1-x & \text{in} \quad 0 \leqq x < +1 \\ 0 & \text{in} \quad +1 \leqq x < +\infty \end{cases}$$

$$\phi(x) = \begin{cases} 0 & \text{in} \quad -\infty < x < -1 \\ \frac{1}{2}(1+x)^2 & \text{in} \quad -1 \leqq x < 0 \\ 1-\frac{1}{2}(1-x)^2 & \text{in} \quad 0 \leqq x < +1 \\ 1 & \text{in} \quad +1 \leqq x < +\infty \end{cases}$$

$$\psi(y) = \begin{cases} -1+\sqrt{2y} & \text{in} \quad 0 < y < 1/2 \\ 1-\sqrt{2-2y} & \text{in} \quad 1/2 < y < 1 \end{cases}$$

2. Ekke's "Best Values"

A. Ekke, in his Kiel dissertation (to appear), deals with the following question among others: Suppose a frequency function $\varphi(x)$ and a natural number n given. Which one among all systems of n values $x_1, \cdots, x_n$ might be considered the "best"?—To give an answer to this question, Ekke divides the total x-axis into n parts $I_1, \cdots, I_n$ with the separating points $\alpha_1, \cdots, \alpha_{n-1}$ in such a manner that

$$\int_{-\infty}^{\alpha_1} \varphi(x)\, dx = \int_{\alpha_1}^{\alpha_2} \varphi(x)\, dx = \cdots = \int_{\alpha_{n-1}}^{+\infty} \varphi(x)\, dx = \frac{1}{n}.$$

Evidently this is possible in one and only one manner, and we have

$$\alpha_1 = \psi\left(\frac{1}{n}\right), \quad \alpha_2 = \psi\left(\frac{2}{n}\right), \cdots, \alpha_{n-1} = \psi\left(\frac{n-1}{n}\right)$$

Each of the parts $I_1, \cdots, I_n$ should contain exactly one value of the system. Furthermore it seems reasonable to fix every point x_ν within its interval I_ν by the conditions

$$\int_{-\infty}^{x_1} \varphi(x)\, dx = \int_{x_1}^{\alpha_1} \varphi(x)\, dx, \cdots\cdots, \int_{\alpha_{n-1}}^{x_n} \varphi(x) dx = \int_{x_n}^{\infty} \varphi(x)\, dx.$$

This also can be done in one and only one way. Let us designate these *"best values"* by $\xi_1, \cdots, \xi_n$. We have

$$(1) \quad \xi_1 = \psi\left(\frac{1}{2n}\right), \quad \xi_2 = \psi\left(\frac{3}{2n}\right), \cdots\cdots, \xi_n = \psi\left(\frac{2n-1}{n}\right).$$

Concerning the best values, EKKE proves two theorems which accentuate the rationality of the definition. If $x_1 \leq \cdots \leq x_n$ are values arranged according to magnitude, and

$$S(x; x_1, \cdots, x_n) = \begin{cases} 0 & \text{in} \quad -\infty < x < x_1 \\ \frac{\nu}{n} & \text{in} \quad x_\nu \leq x < x_{\nu+1} \quad (\nu = 1, \cdots, n-1) \\ 1 & \text{in} \quad x_n \leq x < +\infty . \end{cases}$$

the following theorem holds:

"There is one and only one system $x_1, \cdots, x_n$ *for which*

$$\int_{-\infty}^{+\infty} \{ \phi(x) - S(x; x_1, \cdots, x_n) \}^2 \, dx$$

assumes a minimum, and this system is $x_1 = \xi_1, \cdots, x_n = \xi_n$ *."*
This theorem also holds if the exponent 2 is replaced by an arbitrary positive number.—Furthermore:

"There is one and only one set $x_1, \cdots, x_n$ *for which the lowest upper boundary of*

$$| \phi(x) - S(x; x_1, \cdots, x_n) |$$

assumes a minimum, and this set again is identical with $\xi_1, \cdots, \xi_n$*."*

For normalizing purposes EKKE considers, together with a given frequency function $\varphi(x)$, the totality of the frequency functions which result by linear transformations of the argument, i.e. which result by translations and dilatations in the direction of the x- axis (or by choosing new origins and new units of measurement). With an arbitrary β , and $\alpha > 0$, we have to form

$$\tilde{\varphi}(x) = \frac{1}{\alpha} \varphi\left(\frac{1}{\alpha}(x-\beta)\right),$$

the first factor $\frac{1}{\alpha}$ being required in order to comply with condition 3. The frequency integral corresponding to $\tilde{\varphi}(x)$ is

$$\tilde{\phi}(x) = \phi\left(\frac{1}{\alpha}(x-\beta)\right),$$

and the inverse

$$\tilde{\psi}(y) = \alpha \, \psi(y) + \beta .$$

Due to this simple relation between $\psi(y)$ and $\tilde{\psi}(y)$, we have evidently, if $\xi_1, \cdots, \xi_n$ designate the best values of $\varphi(x)$,

$$\tilde{\xi}_1 = \alpha \, \xi_1 + \beta, \quad \tilde{\xi}_2 = \alpha \, \xi_2 + \beta, \quad \cdots \quad , \tilde{\xi}_n = \alpha \, \xi_n + \beta .$$

This fact can be used to pick out from the multitude of functions $\tilde{\varphi}(x)$ a distinct specimen, and then to operate with its best values only. It is easy to show in a direct manner that there is exactly one specimen in the multitude which complies with the additional conditions $\mu_1 = 0, \mu_2 = 1$.

3. The Starting Point. General Scope.

But the proof of the fact just mentioned can be given indirectly too by considering the inverses $\tilde{\psi}(y)$, and it is this way which gives the starting point of our further developments. Indeed, if we introduce — for simplicity — Stieltjes integrals, the conditions $\mu_1 = 0, \mu_2 = 1$ mean

$$\int_{-\infty}^{+\infty} x \, d\tilde{\varphi}(x) = 0, \quad \int_{-\infty}^{+\infty} x^2 \, d\tilde{\varphi}(x) = 1,$$

and by the substitution $x = \tilde{\psi}(y)$ we get

$$\int_0^1 \tilde{\psi}(y)\, dy = 0, \quad \int_0^1 \tilde{\psi}^2(y)\, dy = 1$$

or

$$\int_0^1 (\beta + \alpha\, \psi(y))\, dy = 0, \quad \int_0^1 (\beta + \alpha\, \psi(y))^2\, dy = 1.$$

Let us put

$$\psi_0(y) = 1, \quad \psi_1(y) = \psi(y)$$

and

$$\chi_0(y) = \gamma_0\, \psi_0(y)$$

$$\chi_1(y) = \beta_1\, \psi_0(y) + \gamma_1\, \psi_1(y).$$

Then our conditions are equivalent to the following demand: Find coefficients $\gamma_0;\ \beta_1, \gamma_1$ $(\gamma_0 > 0, \gamma_1 > 0)$ in such a manner that

$$\int_0^1 \chi_0^2(y)\, dy = 1, \quad \int_0^1 \chi_0(y)\chi_1(y)\, dy = 0, \quad \int_0^1 \chi_1^2(y)\, dy = 1.$$

We add: The functions $\psi_0(y)$ and $\psi_1(y)$ are linearly independent, i.e. $\lambda_0\, \psi_0(y) + \lambda_1\, \psi_1(y) \equiv 0$ cannot hold except for $\lambda_0 = \lambda_1 = 0$.

Now it is obvious that our demand represents a special case of the general problem as follows: *Given a set of linearly independent continuous functions* $\psi_0(y), \psi_1(y), \cdots, \psi_k(y)$ $(0 < y < 1)$. *The scheme* *of coefficients*

$$\begin{array}{cccc} \beta_{00} & 0 & 0 \cdots\cdots & 0 \\ \beta_{10} & \beta_{11} & 0 \cdots\cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ \beta_{K0} & \beta_{K1} & \beta_{K2} \cdots\cdots & \beta_{KK} \end{array}$$

satisfying the additional conditions $\beta_{00} > 0,\ \beta_{11} > 0, \cdots, \beta_{KK} > 0$, *shall be chosen so that the functions*

$$\begin{aligned} Z_0(y) &= \beta_{00}\ \psi_0(y) \\ Z_1(y) &= \beta_{10}\ \psi_0(y) + \beta_{11}\ \psi_1(y) \\ &\cdots\cdots\cdots \\ Z_K(y) &= \beta_{K0}\ \psi_0(y) + \beta_{K1}\ \psi_1(y) + \cdots + \beta_{KK}\ \psi_K(y) \end{aligned}$$

form a normalized orthogonal system, i.e.

$$\int_0^1 Z_p(y) \cdot Z_q(y)\, dy = \begin{cases} 1 & \text{for } p = q \\ 0 & \text{for } p \neq q \end{cases}$$

It is well known that there is one and only one suitable scheme, and it is furnished by the so-called *process of orthogonalization.* Furthermore it is well known that the process of orthogonalization is intimately connected with another problem: *An arbitrary continuous function* $F(y)$ *given, to determine the coefficients* $c_0, \cdots, c_K$

so that $$\int_0^1 \left\{ F(y) - \left[c_0 \psi_0(y) + \cdots\cdots + c_K\ \psi_K(y) \right] \right\}^2 dy$$

assumes a minimum.

Concerning frequency functions, we are led — by pursuing this line—to a general theory of curve types; an account of the results to be obtained will be given in a future article.

Concerning our analysis of statistical data, we do not intend to use from a given frequency function more than its best values. More precisely: *we intend to replace the frequency function by its* *best values.* Our modus procedendi now results by analogy: we have to deal with systems of values (vectors)

$$\begin{array}{l}(u_{11}, u_{12}, \cdots, u_{1n}) \\ (u_{21}, u_{22}, \cdots, u_{2n}) \\ \text{- - - - - - - - -} \\ (u_{k1}, u_{k2}, \cdots, u_{kn})\end{array}$$

which are linearly independent (see No. 5). We have to employ the process of orthogonalization, which gives a normalized orthogonal system (see No. 6)

$$\begin{array}{l}(w_{11}, w_{12}, \cdots, w_{1n}) \\ (w_{21}, w_{22}, \cdots, w_{2n}) \\ \text{- - - - - - - - -} \\ (w_{k1}, w_{k2}, \cdots, w_{kn})\end{array}$$

and we have to direct our attention to the sums of the form

$$\frac{1}{n} \sum_{\nu=1}^{n} \left\{ u_\nu - (c_1 u_{1\nu} + \cdots + c_k u_{k\nu}) \right\}^2,$$

or better

$$\frac{1}{n} \sum_{\nu=1}^{n} \left\{ u_\nu - (b_1 w_{1\nu} + \cdots + b_k w_{k\nu}) \right\}^2.$$

Finally we have to introduce the special set of vectors

$$\begin{array}{l}(1, 1, \cdots, 1) \\ (\xi_1, \xi_2, \cdots, \xi_n) \\ \text{- - - - - - - - -} \\ (\xi_1^{k-1}, \xi_2^{k-1}, \cdots, \xi_n^{k-1})\end{array}$$

where $\xi_1, \cdots, \xi_k$ designate the best values of a frequency function.

We are now in the position to characterize the direction of our research in general words: *A statistical analysis of distributions as an application of the theory of orthogonal systems, based upon the best values of a given frequency function.*

4. Vectors

For our purpose it is convenient to make use of the notations and simplest operations of vector analysis. If $u_1, u_2, \cdots, u_n$ are a set of numbers, we take the symbol $(u_1, \cdots, u_n)$ as an individual, call it a *vector*, and designate it by a gothic letter:

$$\mathfrak{u} = (u_1, \cdots, u_n).$$

Equality of two vectors $\mathfrak{u} = (u_1, \cdots, u_n)$ and $\mathfrak{v} = (v_1, \cdots, v_n)$ is defined by

$$u_1 = v_1, \quad u_2 = v_2, \quad \cdots, \quad u_n = v_n,$$

and is written $\mathfrak{u} = \mathfrak{v}$. The *products of a number* c *with a vector* $\mathfrak{u}$ are defined by

$$c\mathfrak{u} = (cu_1, \cdots, cu_n)$$

$$\mathfrak{u}c = (u_1 c, \cdots, u_n c);$$

the *sum of two vectors* $\mathfrak{u}$ and $\mathfrak{v}$ by

$$\mathfrak{u} + \mathfrak{v} = (u_1 + v_1, \cdots, u_n + v_n).$$

Evidently we have

$$c\mathfrak{u} = \mathfrak{u}c$$

and

$$(\mathfrak{u} + \mathfrak{v}) + \mathfrak{w} = \mathfrak{u} + (\mathfrak{v} + \mathfrak{w}).$$

Hence we may omit the brackets, and the sum of three or more vectors has a definite sense. More general, the meaning of the expression

$$c_1 \mathfrak{u}_1 + \cdots + c_k \mathfrak{u}_k$$

is clear. The *product of two vectors* is (somewhat differently from the customary way) defined as a NUMBER, viz.

$$\mathfrak{u}\mathfrak{v} = \frac{1}{n}(u_1 v_1 + \cdots + u_n v_n),$$

and we have

$$\mathfrak{u}\mathfrak{v} = \mathfrak{v}\mathfrak{u}$$

$$(\mathfrak{u} + \mathfrak{v})\mathfrak{w} = \mathfrak{u}\mathfrak{w} + \mathfrak{v}\mathfrak{w}.$$

But in general the vectors $(\mathfrak{u}\mathfrak{v})\mathfrak{w}$ and $\mathfrak{u}(\mathfrak{v}\cdot\mathfrak{w})$ are entirely different.

Let us put $\qquad \mathfrak{o} = (0, 0, \cdots, 0).$

Every vector $\mathfrak{u}$ satisfies $\mathfrak{u}^2 = \mathfrak{u}\mathfrak{u} \geqslant 0$, and $\mathfrak{u} = \mathfrak{o}$ is the only vector for which $\mathfrak{u}^2 = \theta$ holds. — Whenever the square root of the square of a vector,

$$\sqrt{\mathfrak{u}^2} = \sqrt{\frac{u_1^2 + \cdots + u_n^2}{n}},$$

is met with, we always mean the positive value.

5. Linear Independence

A set of vectors $\breve{\mathfrak{u}}_1, \cdots, \breve{\mathfrak{u}}_K$ is said to be *linearly independent* if the equation

$$\lambda_1 \breve{\mathfrak{u}}_1 + \lambda_2 \breve{\mathfrak{u}}_2 + \cdots + \lambda_K \breve{\mathfrak{u}}_K = \mathfrak{o}$$

does not hold except for $\lambda_1 = \lambda_2 = \cdots = \lambda_K = 0$. Otherwise the vectors are said to be *linearly dependent*. If the vectors

$$\breve{\mathfrak{u}}_1,\ \breve{\mathfrak{u}}_2, \cdots, \breve{\mathfrak{u}}_K$$

are linearly independent, all the more the same is true for every partial system. Especially:

$$\breve{\mathfrak{u}}_1 \neq \mathfrak{o},\quad \breve{\mathfrak{u}}_2 \neq \mathfrak{o},\ \cdots,\ \breve{\mathfrak{u}}_K \neq \mathfrak{o}.$$

Theorem 1. *"Let $\breve{\mathfrak{u}}_1, \cdots, \breve{\mathfrak{u}}_K$ be linearly independent; form the vectors*

$$(2)\qquad \begin{cases} \breve{\mathfrak{u}}_1^* = a_{11} \breve{\mathfrak{u}}_1 \\ \breve{\mathfrak{u}}_2^* = a_{21} \breve{\mathfrak{u}}_1 + a_{22} \breve{\mathfrak{u}}_2 \\ ------ \\ \breve{\mathfrak{u}}_K^* = a_{K1} \breve{\mathfrak{u}}_1 + a_{K2} \breve{\mathfrak{u}}_2 + \cdots + a_{KK} \breve{\mathfrak{u}}_K \end{cases}$$

and suppose

$$a_{11} \neq 0,\quad a_{22} \neq 0, \cdots,\ a_{KK} \neq 0.$$

Then the vectors $\breve{\mathfrak{u}}_1^*, \cdots, \breve{\mathfrak{u}}_K^*$ are also linearly independent."

In fact, if there were a relation of the form

$$\lambda_1 \breve{\mathfrak{u}}_1^* + \cdots + \lambda_K \breve{\mathfrak{u}}_K^* = \mathfrak{o}$$

and the factors $\lambda_1, \cdots, \lambda_K$ were not all equal to zero, then there would be a *last* factor differing from zero, say λ_L, and we should have

$$\lambda_1 \breve{\mathfrak{u}}_1^* + \cdots + \lambda_L \breve{\mathfrak{u}}_L^* = \mathfrak{o} \qquad (\lambda_L \neq 0)$$

if we were to replace $\breve{\mathfrak{u}}_1^*, \cdots, \breve{\mathfrak{u}}_L^*$ by the expressions (2), we should get a relation of the form

$$\mu_1 \breve{\mathfrak{u}}_1 + \cdots + \mu_{L-1} \breve{\mathfrak{u}}_{L-1} + a_{LL} \lambda_L \breve{\mathfrak{u}}_L = \mathfrak{o},$$

which is impossible on account of $a_{LL} \neq 0$, $\lambda_L \neq 0$ and the presupposed linear independency of $\breve{\mathfrak{u}}_1, \cdots, \breve{\mathfrak{u}}_K$.

In order to prove some further theorems it is convenient— but not necessary — to make use of the following fundamental theorem concerning systems of homogeneous linear equations.

"A necessary and sufficient condition that the system of equations

$$a_{11}x_1 + \cdots + a_{1n}x_n = 0$$
$$\text{------------}$$
$$a_{n1}x_1 + \cdots + a_{nn}x_n = 0$$

should have no other solution than $x_1 = x_2 = \cdots = x_n = 0$ *is*

$$\begin{vmatrix} a_{11} \cdots a_{1n} \\ a_{n1} \cdots a_{nn} \end{vmatrix} \neq 0 \text{."}$$

From this statement at once follows:

THEOREM 2. *"A necessary and sufficient condition that the vectors*

$$\breve{u}_1 = (u_{11}, u_{12}, \cdots, u_{1n})$$
$$\text{------}$$
$$\breve{u}_n = (u_{n1}, u_{n2}, \cdots, u_{nn})$$

should be linearly independent is

$$\begin{vmatrix} u_{11} \cdots u_{1n} \\ u_{n1} \cdots u_{nn} \end{vmatrix} \neq 0 \text{."}$$

In fact, linear dependence of $\breve{u}_1, \cdots, \breve{u}_n$ is equivalent to the existence of values $\lambda_1, \cdots, \lambda_n$, not all equal to zero, satisfying

$$\lambda_1 u_{11} + \lambda_2 u_{21} + \cdots + \lambda_n u_{n1} = 0$$
$$\text{----}$$
$$\lambda_1 u_{1n} + \lambda_2 u_{2n} + \cdots + \lambda_n u_{nn} = 0,$$

and the determinant of these equations is equal to the determinant above.

THEOREM 3. *"If* $\breve{u}_1, \cdots, \breve{u}_k$ *are linearly independent, the number* k *of the vectors cannot exceed the number* n *of the components:* $k \leqq n$*."*

We prove this theorem by showing:

"If $n+1$ vectors

$$\breve{u}_1 = (u_{11}, \cdots, u_{1n})$$

$$- - - -$$

$$\breve{u}_{n+1} = (u_{n+1,1}, \cdots, u_{n+1,n})$$

are given, they are linearly dependent."

For obviously, the determinant of the equations

$$\lambda_1 u_{11} + \cdots + \lambda_{n+1} u_{n+1,1} = 0$$

$$- - - - - -$$

$$\lambda_1 u_{1n} + \cdots + \lambda_{n+1} u_{n+1,n} = 0$$

$$\lambda_1\, 0 + \cdots + \lambda_{n+1}\, 0 = 0$$

vanishes, hence the system possesses a solution $\lambda_1, \cdots, \lambda_{n+1}$ different from $0, \cdots, 0$, and with such values $\lambda_1, \cdots, \lambda_{n+1}$ the first n equations mean

$$\lambda_1 \breve{u}_1 + \cdots + \lambda_{n+1} \breve{u}_{n+1} = \breve{o}.$$

6. Normalized Orthogonal Systems of Vectors

If $\breve{u}^2 = 1$, the vector $\breve{u}$ is said to be *normalized.* Every vector $\breve{u} \neq \breve{o}$ can be normalized by multiplying it by $\frac{1}{\sqrt{\breve{u}^2}}$.

If $\breve{u}\breve{v} = 0$, the pair of vectors $\breve{u}$ and $\breve{v}$ is said to be *orthogonal.* The vectors $\breve{u}_1, \cdots, \breve{u}_k$ are said to form an *orthogonal system,* if every pair of them is orthogonal.

Finally the vectors $\breve{u}_1, \cdots, \breve{u}_k$ are said to form a *normalized orthogonal system,* if they form an orthogonal system and each of them is normalized. Accordingly a normalized orthogonal system is characterized by the conditions

$$(3) \qquad \breve{u}_p \breve{u}_q = \begin{cases} 1 & \text{for } p = q \\ 0 & \text{for } p \neq q. \end{cases}$$

Vectors forming a normalized orthogonal system necessarily are linearly independent. For, from

$$\lambda_1 \breve{u}_1 + \cdots + \lambda_k \breve{u}_k = \breve{o}$$

follows

$$(\lambda_1 \breve{u}_1 + \cdots + \lambda_k \breve{u}_k)\breve{u}_n = \sigma \breve{u}_n = 0 \quad (n = 1, 2, \cdots, k)$$

or

$$\lambda_1 \breve{u}_1 \breve{u}_n + \cdots + \lambda_k \breve{u}_k \breve{u}_n = 0,$$

and from (3):

$$\lambda_n = 0 \qquad (n = 1, 2, \cdots, k).$$

7. The Process of Orthogonalization

Theorem 4. *"If the vectors $\breve{u}_1, \cdots, \breve{u}_k$ are linearly independent, there is one and only one scheme of values*

$$\begin{array}{llll} \beta_{11} & & & (\beta_{11} > 0) \\ \beta_{21} & \beta_{22} & & (\beta_{22} > 0) \\ \cdots\cdots & & & \\ \beta_{k1} & \beta_{k2} & \cdots \; \beta_{kk} & (\beta_{kk} > 0) \end{array}$$

so that the vectors

$$(4) \quad \begin{aligned} w_1 &= \beta_{11} \breve{u}_1 \\ w_2 &= \beta_{21} \breve{u}_1 + \beta_{22} \breve{u}_2 \\ &\cdots\cdots \\ w_k &= \beta_{k1} \breve{u}_1 + \beta_{k2} \breve{u}_2 + \cdots + \beta_{kk} \breve{u}_k \end{aligned}$$

form a normalized orthogonal system."

To prove this theorem, fundamental for our analysis, let us consider

$$(5) \quad \begin{aligned} w_1 &= \breve{u}_1 \\ w_2 &= \gamma_{21} \breve{u}_1 + \breve{u}_2 \\ &\cdots\cdots \\ w_k &= \gamma_{k1} \breve{u}_1 + \gamma_{k2} \breve{u}_2 + \cdots + \breve{u}_k . \end{aligned}$$

From theorem 1 it follows that the vectors $w_1, \cdots, w_k$ are linearly independent. — Let us assume we have already proved that there is one and only one system of values γ so that (5) is an orthogonal system. Then it follows firstly that the coefficients β in (4) can be chosen in *at least one* suitable manner. For we have

$$\lambda_1 = \sqrt{w_1^2} > 0, \cdots, \lambda_k = \sqrt{w_k^2} > 0,$$

and

$$\beta_{11} = \frac{1}{\lambda_1};\ \beta_{21} = \frac{\gamma_{21}}{\lambda_2}, \beta_{22} = \frac{1}{\lambda_2};\cdots;\beta_{k1} = \frac{\gamma_{k1}}{\lambda_k}, \cdots, \beta_{kk} = \frac{1}{\lambda_k}$$

are suitable values. Secondly we can deduce the *uniqueness* of the coefficients β in (4). For suppose β and β^* to be two suitable systems of coefficients; this suggests that we form

$$\gamma_{21}=\frac{\beta_{21}}{\beta_{22}};\ \gamma_{31}=\frac{\beta_{31}}{\beta_{33}},\ \gamma_{32}=\frac{\beta_{32}}{\beta_{33}};\cdots;\ \gamma_{k1}=\frac{\beta_{k1}}{\beta_{kk}},\cdots,\gamma_{k;k-1}=\frac{\beta_{k;k-1}}{\beta_{kk}},$$

associated with

$$\mathfrak{v}_1=\frac{1}{\beta_{11}}\mathfrak{w}_1,\ \cdots,\ \mathfrak{v}_k=\frac{1}{\beta_{kk}}\mathfrak{w}_k,$$

and

$$\gamma^*_{21}=\frac{\beta^*_{21}}{\beta^*_{22}};\ \gamma^*_{31}=\frac{\beta^*_{31}}{\beta^*_{33}},\ \gamma^*_{32}=\frac{\beta^*_{32}}{\beta^*_{33}};\cdots;\ \gamma^*_{k1}=\frac{\beta^*_{k1}}{\beta^*_{kk}},\cdots,\gamma^*_{k:k-1}=\frac{\beta^*_{k:k-1}}{\beta^*_{kk}},$$

associated with

$$\mathfrak{v}^*_1=\frac{1}{\beta^*_{11}}\mathfrak{w}^*_1,\ \cdots,\ \mathfrak{v}^*_k=\frac{1}{\beta^*_{kk}}\mathfrak{w}^*_k.$$

The vectors $\mathfrak{v}_1,\cdots\mathfrak{v}_k$ as well a $\mathfrak{v}^*_1,\cdots,\mathfrak{v}^*_k$ form orthogonal systems of the type (2), hence

$$\mathfrak{v}^*_1=\mathfrak{v}_1,\ \cdots\cdots,\ \mathfrak{v}^*_k=\mathfrak{v}_k$$

and furthermore

$$\mathfrak{w}^*_1=\mathfrak{w}_1,\ \cdots\cdots,\ \mathfrak{w}^*_k=\mathfrak{w}_k.$$

Finally, because of the linear independence of $\mathfrak{u}_1,\cdots,\mathfrak{u}_k$:

$$\beta^*_{11}=\beta_{11};\ \beta^*_{21}=\beta_{21},\ \beta^*_{22}=\beta_{22};\cdots\cdots;\beta^*_{k1}=\beta_{k1};\ldots;\beta^*_{kk}=\beta_{kk}.$$

Accordingly we may confine ourselves to proving the existence and uniqueness of suitable coefficients γ in (5).

This proposition is true for $k=1$. Let $k\geq 2$ arbitrarily, and assume the proposition to be proved up to $k-1$. The vectors $\mathfrak{v}_1,\cdots,\mathfrak{v}_{k-1}$ therefore are orthogonal, and we have to show only: *There is one and only one set of values* $\gamma_{k1},\cdots\cdots,\gamma_{k,k-1}$ *so that the conditions*

(6) $\quad \mathfrak{v}_1\mathfrak{v}_k=0,\quad \mathfrak{v}_2\mathfrak{v}_k=0,\cdots,\ \mathfrak{v}_{k-1}\mathfrak{v}_k=0$

are satisfied.

The vectors $\mathfrak{u}_1,\cdots,\mathfrak{u}_{k-1}$ can be represented as linear combinations of $\mathfrak{v}_1,\cdots,\mathfrak{v}_{k-1}$:

$$\begin{aligned}\mathfrak{u}_1&=\mathfrak{v}_1\\ \mathfrak{u}_2&=c_{21}\mathfrak{v}_1+\mathfrak{v}_2\\ \\ \mathfrak{u}_{v-1}&=c_{k-1,1}\mathfrak{v}_1+\cdots+c_{k-1\,k-2}\mathfrak{v}_{k-2}+\mathfrak{v}_{k-1}.\end{aligned}$$

We introduce this into $\mathfrak{w}_k$, and get

$$(7) \quad \mathfrak{w}_k = C_{k1}\,\mathfrak{w}_1 + \cdots + C_{k,k-1}\,\mathfrak{w}_{k-1} + \breve{\mathfrak{u}}_k$$

with

$$(8) \quad \begin{cases} C_{k1} = \gamma_{k1} + c_{21}\,\gamma_{k2} + \cdots + c_{k-1,1}\,\gamma_{k,k-1} \\ C_{k2} = \qquad\qquad\;\; \gamma_{k2} + \cdots + c_{k-1,2}\,\gamma_{k,k-1} \\ \cdots\cdots \\ C_{k,k-1} = \qquad\qquad\qquad\qquad\qquad\quad \gamma_{k,k-1} \end{cases}$$

From the linear independence of $\mathfrak{w}_1,\cdots,\mathfrak{w}_{k-1}$ we have $\mathfrak{w}_1^2 > 0,\cdots,\mathfrak{w}_{k-1}^2 > 0$, and therefore we can deduce from (7):

$$(9) \quad C_{k1} = -\frac{\breve{\mathfrak{u}}_k\,\mathfrak{w}_1}{\sqrt{\mathfrak{w}_1^2}}, \quad \cdots\cdots, \quad C_{k,k-1} = -\frac{\breve{\mathfrak{u}}_k\,\mathfrak{w}_{k-1}}{\sqrt{\mathfrak{w}_{k-1}^2}}.$$

The coefficients $\gamma_{k1},\cdots,\gamma_{k,k-1}$ having to satisfy the equations (8) with the values (9) of $C_{k1},\cdots,C_{k,k-1}$, there can exist *only one* suitable system $\gamma_{k1},\cdots\gamma_{k,k-1}$.

Conversely: if $C_{k1},\cdots,C_{k,k-1}$ are chosen according to (9), and then $\gamma_{k1},\cdots,\gamma_{k,k-1}$ calculated from (8), evidently there results a vector $\mathfrak{w}_k$ satisfying (6).

We add:

THEOREM 5. *"If $\breve{\mathfrak{u}}_1,\cdots,\breve{\mathfrak{u}}_k$ are linearly independent, and $\mathfrak{w}_1,\cdots,\mathfrak{w}_k$ is the corresponding normalized orthogonal system, then the normalized orthogonal system $\mathfrak{w}_1^*,\cdots,\mathfrak{w}_k^*$ corresponding to*

$$\begin{aligned} \breve{\mathfrak{u}}_1^* &= a_{11}\,\breve{\mathfrak{u}}_1 && (a_{11} > 0) \\ \breve{\mathfrak{u}}_2^* &= a_{21}\,\breve{\mathfrak{u}}_1 + a_{22}\,\breve{\mathfrak{u}}_2 && (a_{22} > 0) \\ &\cdots\cdots && \cdots \\ \breve{\mathfrak{u}}_k^* &= a_{k1}\,\breve{\mathfrak{u}}_1 + a_{k2}\,\breve{\mathfrak{u}}_2 + \cdots + a_{kk}\,\breve{\mathfrak{u}}_k && (a_{kk} > 0) \end{aligned}$$

is identical with $\mathfrak{w}_1,\cdots,\mathfrak{w}_k$."

Obviously the vectors $\mathfrak{w}^*$ are of the form

$$\begin{aligned} \mathfrak{w}_1^* &= B_{11}\,\breve{\mathfrak{u}}_1 && (B_{11} > 0) \\ \mathfrak{w}_2^* &= B_{21}\,\breve{\mathfrak{u}}_1 + B_{22}\,\breve{\mathfrak{u}}_2 && (B_{22} > 0) \\ &\cdots\cdots \\ \mathfrak{w}_k^* &= B_{k1}\,\breve{\mathfrak{u}}_1 + B_{k2}\,\breve{\mathfrak{u}}_2 + \cdots + B_{kk}\,\breve{\mathfrak{u}}_k && (B_{kk} > 0) \end{aligned}$$

The proof of theorem 5 now follows as an immediate application of theorem 4.

8. Complete Systems of Normalized Orthogonal Vectors

A system of normalized orthogonal vectors $\mathfrak{v}_1, \cdots, \mathfrak{v}_k$ is said to be *complete* if, corresponding to every arbitrary vector $\breve{\mathfrak{u}}$, there exist coefficients $c_1, \cdots, c_k$ so that

$$\{\breve{\mathfrak{u}} - (c_1 \mathfrak{v}_1 + \cdots + c_k \mathfrak{v}_k)\}^2 = 0 \tag{10}$$

holds. Evidently, (10) is equivalent to

$$\breve{\mathfrak{u}} = c_1 \mathfrak{v}_1 + \cdots + c_k \mathfrak{v}_k .$$

Theorem 6. *"If the vectors $\mathfrak{v}_1, \cdots, \mathfrak{v}_n$ $(k = n)$ form a normalized orthogonal system, then this system is* complete."

Proof. According to theorem 3, the $n+1$ vectors $\mathfrak{v}_1, \cdots, \mathfrak{v}_k, \breve{\mathfrak{u}}$ are linearly dependent, i.e. there is an equality

$$\lambda_1 \mathfrak{v}_1 + \cdots + \lambda_n \mathfrak{v}_n + \lambda \breve{\mathfrak{u}} = \mathfrak{o},$$

and $\lambda_1, \cdots, \lambda_n, \lambda$ are not all equal to zero. The vectors $\mathfrak{v}$, being linearly independent, we have necessarily $\lambda \neq 0$. Hence

$$\breve{\mathfrak{u}} = -\frac{\lambda_1}{\lambda} \mathfrak{v}_1 - \cdots - \frac{\lambda_n}{\lambda} \mathfrak{v}.$$

The condition $k = n$ is also *necessary* for completeness, but we shall not have to make use of it.

9. Approximation in the Mean

Let us consider a normalized orthogonal system $\mathfrak{v}_1, \cdots, \mathfrak{v}_k$, and an arbitrary vector $\breve{\mathfrak{u}}$. We wish to determine the coefficients $b_1, \cdots, b_k$ in such a way that

$$\{\breve{\mathfrak{u}} - \sum_{\kappa=1}^{k} b_\kappa \mathfrak{v}_\kappa\}^2$$

assumes a minimum. If there exists a suitable set of coefficients, we say that the corresponding linear combination $b_1 \mathfrak{v}_1 + \cdots + b_k \mathfrak{v}_k$ gives a *"best approximation in the mean"* to the vector $\breve{\mathfrak{u}}$.

The following transformations will at once clear up the situation:

$$\left\{\breve{u} - \sum_{\chi=1}^{k} b_\chi w_\chi\right\}^2 = \breve{u}^2 - 2\sum_{\chi=1}^{k} b_\chi w_\chi \breve{u} + \sum_{\varsigma,\sigma=1}^{k} b_\varsigma b_\sigma w_\varsigma w_\sigma$$
$$= \breve{u}^2 - \sum_{\chi=1}^{k} (w_\chi \breve{u})^2 + \sum_{\chi=1}^{k} (w_\chi \breve{u})^2 - 2\sum_{\chi=1}^{k} b_\chi w_\chi \breve{u} + \sum_{\chi=1}^{k} b_\chi^2$$
$$= \breve{u}^2 - \sum_{\chi=1}^{k} (w_\chi \breve{u})^2 + \sum_{\chi=1}^{k} \left\{b_\chi - w_\chi \breve{u}\right\}^2,$$

and if we designate

$$a_1 = w_1 \breve{u}, \quad a_2 = w_2 \breve{u}, \ldots, \quad a_k = w_k \breve{u}$$

we have the fundamental equation

$$(11) \qquad \left\{\breve{u} - \sum_{\chi=1}^{k} b_\chi w_\chi\right\}^2 = \breve{u}^2 - \sum_{\chi=1}^{k} a_\chi^2 + \sum_{\chi=1}^{k} (b_\chi - a_\chi)^2.$$

On the right hand, the coefficients $b_1, \cdots, b_k$ are not met with but in the last sum, and this sum assumes a minimum for $b_\chi = a_\chi$ only. By that, we have:

THEOREM 7. *"Among all linear combinations of the normalized orthogonal vectors $w_1, \cdots, w_k$ there is one and only one which gives a best approximation in the mean to the vector $\breve{u}$, and the 'best coefficients' are $a_1, a_2, \cdots, a_k$."*

The equation (11) admits some important conclusions concerning the coefficients $a_1, \cdots, a_k$. By putting

$$b_1 = a_1, \ \ldots, \ b_k = a_k$$

we derive

$$(12) \qquad \left\{\breve{u} - \sum_{\chi=1}^{k} a_\chi w_\chi\right\}^2 = \breve{u}^2 - \sum_{\chi=1}^{k} a_\chi^2$$

The left side herein evidently is not negative, hence

$$(13) \qquad a_1^2 + a_2^2 + \cdots + a_k^2 \leq \breve{u}^2$$

Finally, if $w_1, \cdots, w_n$ is a COMPLETE system of normalized orthogonal vectors, the preceding reasonings of course hold for every $k = 1, 2, \cdots, n$. But we can show more than (13), viz.

$$(14) \qquad a_1^2 + a_2^2 + \cdots + a_n^2 = \breve{u}^2$$

Indeed, according to the definition of completeness we have with suitable $b_1, \cdots, b_n$:

$$\left\{ \breve{u} - \sum_{\nu=1}^{n} b_\nu w_\nu \right\}^2 = 0$$

and a fortiori, by theorem 7,

$$\left\{ \breve{u} - \sum_{\nu=1}^{n} a_\nu w_\nu \right\}^2 = 0,$$

which is, regarding (12), equivalent to (14).

10. THE TCHEBYCHEF COEFFICIENTS

Let $$\xi_1 < \xi_2 < \cdots < \xi_n$$

be a set of best values corresponding to a given frequency function $\varphi(x)$ (see No. 2). We form

$$(15) \qquad \begin{aligned} \mathfrak{T}_0 &= (\ 1,\ 1, \cdots,\ 1\ ) \\ \mathfrak{T}_1 &= (\ \xi_1,\ \xi_2, \cdots,\ \xi_n\ ) \\ \mathfrak{T}_2 &= (\ \xi_1^2,\ \xi_2^2, \cdots,\ \xi_n^2\ ) \\ &\cdots\cdots \\ \mathfrak{T}_{n-1} &= (\ \xi_1^{n-1},\ \xi_2^{n-1}, \cdots,\ \xi_n^{n-1}\ ) \end{aligned}$$

The vectors $\mathfrak{T}_0, \cdots, \mathfrak{T}_{n-1}$ are linearly independent. For

$$\lambda_0 \mathfrak{T}_0 + \cdots + \lambda_{n-1} \mathfrak{T}_{n-1} = \mathfrak{o}$$

means

$$\lambda_0 + \lambda_1 \xi_\nu + \lambda_2 \xi_\nu^2 + \cdots + \lambda_{n-1} \xi_\nu^{n-1} = 0 \quad (\nu = 1, \cdots, n),$$

that is, the polynomial

$$P(x) = \lambda_0 + \lambda_1 x + \cdots + \lambda_{n-1} x^{n-1}$$

of degree $\leq (n-1)$ possesses n different zeros $\xi_1, \cdots, \xi_n$. But the number of zeros of a polynomial cannot exceed its degree unless all coefficients vanish. Hence $\lambda_0 = \lambda_1 = \cdots = \lambda_{n-1} = 0$.

Let us designate the (complete) set of normalized orthogonal vectors corresponding to $\mathfrak{T}_0, \cdots, \mathfrak{T}_{n-1}$ by $\mathfrak{z}_0, \mathfrak{z}_1, \cdots, \mathfrak{z}_{n-1}$.

When we have to deal with a set of observations $x_1, \cdots, x_n$, there will not be any practical loss of generality if we assume these values arranged according to magnitude,

$$x_1 \leqq x_2 \leqq \cdots \leqq x_n,$$

and to be ***not all equal***. Then we define the vector $\mathfrak{X}$ by

$$\mathfrak{X} = (x_1, x_2, \cdots, x_n),$$

and we propose to call the coefficients

$$a_0 = \mathfrak{z}_0 \mathfrak{X}, \quad a_1 = \mathfrak{z}_1 \mathfrak{X}, \cdots, a_{n-1} = \mathfrak{z}_{n-1} \mathfrak{X}$$

"Tchebychef Coefficients" of $\mathfrak{X}$.

The central position of the Tchebychef coefficients for analyzing purposes is pointed out by the following theorems 8 and 9.

THEOREM 8. *"The set $\mathfrak{z}_0, \mathfrak{z}_1, \cdots, \mathfrak{z}_{n-1}$ and a fortiori the Tchebychef coefficients $a_0, a_1, \cdots, a_{n-1}$ of the observations $x_1, x_2, \cdots, x_n$ do not depend on the special frequency function $\varphi(x)$, but on the type only to which $\varphi(x)$ belongs."*

To prove this theorem, let us consider, besides $\varphi(x)$, an arbitrary individual of its type,

$$\tilde{\varphi}(x) = \frac{1}{\alpha} \varphi\left(\frac{1}{\alpha}(x - \beta)\right) \qquad (\alpha > 0).$$

The best values corresponding to $\tilde{\varphi}(x)$ are (see No. 2)

$$\tilde{\xi}_1 = \alpha \cdot \xi_1 + \beta, \cdots, \quad \tilde{\xi}_n = \alpha\, \xi_n + \beta,$$

and we deduce, if $\tilde{\mathfrak{X}}_0, \cdots, \tilde{\mathfrak{X}}_{n-1}$ designate the vectors (15) obtained from $\tilde{\xi}_1, \cdots, \tilde{\xi}_n$ instead of $\xi_1, \cdots, \xi_n$,

$$\begin{aligned}
\tilde{\mathfrak{X}}_0 &= \mathfrak{X}_0 \\
\tilde{\mathfrak{X}}_1 &= \beta \mathfrak{X}_0 + \alpha\, \mathfrak{X}_1 \\
\tilde{\mathfrak{X}}_2 &= \beta^2 \mathfrak{X}_0 + 2\beta\alpha\, \mathfrak{X}_1 + \alpha^2 \mathfrak{X}_2 \\
&\cdots\cdots \\
\tilde{\mathfrak{X}}_{n-1} &= \beta^{n-1} \mathfrak{X}_0 + \binom{n-1}{1} \beta^{n-2} \alpha\, \mathfrak{X}_1 + \cdots + \alpha^{n-1} \mathfrak{X}_{n-1}.
\end{aligned}$$

Hence, by an application of theorem 5, the normalized orthogonal vectors $\tilde{\mathfrak{z}}_0, \cdots, \tilde{\mathfrak{z}}_{n-1}$ are identical with $\mathfrak{z}_0, \cdots, \mathfrak{z}_{n-1}$.

If we choose a new unit of measurement and a new origin, that is to say if we perform a transformation

$$x_\nu^* = \alpha\, x_\nu + \beta, \qquad \xi_\nu^* = \alpha\, \xi_\nu + \beta \qquad (\alpha > 0),$$

the vectors $\mathfrak{z}_0, \cdots, \mathfrak{z}_{n-1}$ do not change (by the reasoning just finished). The vector $\mathfrak{X}$ changes into

$$\mathfrak{X}^* = \alpha\, \mathfrak{X} + \beta\, \mathfrak{N} \qquad (\mathfrak{N} = (1, 1, \cdots, 1)),$$

and we have:

THEOREM 9. *"If a new unit of measurement and a new origin are introduced, say*

$$x^* = \alpha x + \beta \qquad (\alpha > 0),$$

then the Tchebychef coefficients change into

$$a_0^* = \alpha a_0 + \beta; \quad a_1^* = \alpha a_1, \quad a_2^* = \alpha a_2, \cdots, a_{n-1}^* = \alpha a_{n-1}."$$

11. MEAN AND DISPERSION. COEFFICIENTS OF SKEWNESS AND KURTOSIS

Preparatory to the definition in this chapter, let us consider a_0 and a_1 especially. To begin with, we have

$$\mathfrak{z}_0 = \mathfrak{E}_0 = (1, 1, \cdots, 1)$$

and therefore

$$a_0 = \frac{1}{n}(x_1 + x_2 + \cdots + x_n).$$

The proof of theorem 4 furnishes a convenient way to compute $\mathfrak{z}_1$. We put

$$\mathfrak{y}_0 = \mathfrak{E}_0 \; ; \quad \mathfrak{y}_1 = \gamma \mathfrak{E}_0 + \mathfrak{E}_1$$

and determine γ so that $\mathfrak{y}_0 \mathfrak{y}_1 = 0$:

$$\gamma = -\frac{\mathfrak{E}_1 \mathfrak{E}_0}{\mathfrak{E}_0^2} = -\frac{1}{n}(\xi_1 + \cdots + \xi_n).$$

With the designations

$$m_1 = \frac{1}{n}(\xi_1 + \cdots + \xi_n) \qquad m_2 = \frac{1}{n}(\xi_1^2 + \cdots + \xi_n^2)$$

we obtain

$$\mathfrak{E}_1 \mathfrak{E}_0 = m_1, \qquad \mathfrak{E}_1^2 = m_2.$$

Hence

$$\mathfrak{y}_1 = -m_1 \mathfrak{E}_0 + \mathfrak{E}_1, \qquad \mathfrak{y}_1^2 = m_2 - m_1^2$$

$$\mathfrak{z}_1 = \frac{1}{\sqrt{\mathfrak{y}_1^2}} \cdot \mathfrak{y}_1 = \frac{1}{\sqrt{m_2 - m_1^2}}(-m_1 \mathfrak{E}_0 + \mathfrak{E}_1)$$

$$= \frac{1}{\sqrt{m_2 - m_1^2}}(\xi_1 - m_1, \; \xi_2 - m_1, \cdots, \; \xi_n - m_1)$$

$$a_1 = \mathfrak{z}_1 \mathfrak{E} = \frac{1}{\sqrt{m_2 - m_1^2}}(-m_1 \mathfrak{E}_0 \mathfrak{E} + \mathfrak{E}_1 \mathfrak{E})$$

$$= \frac{\frac{1}{n}(\xi_1 x_1 + \cdots + \xi_n x_n) - m_1 \cdot \frac{1}{n}(x_1 + \cdots + x_n)}{\sqrt{m_2 - m_1^2}}.$$

Concerning a_1 we have now to deal with a theorem which is of the greatest importance for our purposes.

THEOREM 10. *"The Tchebychef coefficient a_1 is always positive:*

$$a_1 > 0 ."$$

For a proof we can proceed as follows: if we designate the components of $\mathfrak{z}_1$ by $\varsigma_1, \cdots, \varsigma_n$, we have

$$\varsigma_1 < \varsigma_2 < \cdots < \varsigma_n \quad \text{and} \quad \varsigma_1 + \cdots + \varsigma_n = 0 .$$

From this we deduce the existence of a subscript ν so that

$$\varsigma_1 < \cdots < \varsigma_\nu < 0 \leq \varsigma_{\nu+1} < \cdots < \varsigma_n .$$

Let us put

$$z_1 = \varsigma_1 , \quad z_2 = \varsigma_1 + \varsigma_2 , \cdots \quad , z_n = \varsigma_1 + \cdots + \varsigma_n .$$

Then we have

$$z_1 < 0 , \cdots , z_\nu < 0 ; \quad z_{\nu+1} < z_{\nu+2} < \cdots < z_n = 0,$$

which gives

$$z_1 < 0 , \quad z_2 < 0 , \cdots , \; z_{n-1} < 0.$$

On the other hand, the identity

$$\varsigma_1 x_1 + \cdots + \varsigma_n x_n = -z_1(x_2 - x_1) - z_2(x_3 - x_2) - \cdots - z_{n-1}(x_n - x_{n-1})$$

holds. The differences $x_2 - x_1, \cdots, x_n - x_{n-1}$ are all $\geqq 0$, and $x_1, \cdots, x_n$ being subjected to the condition not to be all equal, at least one difference really is positive. Hence

$$a_1 = \frac{1}{n}(\varsigma_1 x_1 + \cdots + \varsigma_n x_n) > 0 .$$

There are no restrictions for the Tchebychef coefficients different from a_1 as far as their signs are concerned.

The reader, after having verified the truth of the following statement, will now be prepared to accept the definition below.

"If the vector $\mathfrak{b}$ is of the form

$$\mathfrak{b} = b_0 \mathfrak{t}_0 + b_1 \mathfrak{t}_1 + b_2 \mathfrak{t}_2 ,$$

the sign of a_2 coincides with that of b_2 ; if it is of the form

$$\mathfrak{b} = b_0 \mathfrak{t}_0 + b_1 \mathfrak{t}_1 + b_2 \mathfrak{t}_2 + b_3 \mathfrak{t}_3$$

the sign of a_3 coincides with that of b_3 ; and so on."

DEFINITION. *"A type of frequency function being given, the Tchebychef coefficients a_0, a_1, a_2, a_3 of the observations $x_1, x_2, \cdots, x_n$ shall be called:*

$a_0 = M$ = Mean *of the Observations*

$a_1 = \sigma$ = Dispersion *of the Observations*

a_2 = Tchebychef Coefficient of Skewness *of the Observations*

a_3 = Tchebychef Coefficient of Kurtosis *of the Observations."*

We do not believe the Tchebychef coefficients with a higher subscript than 3 to be of any practical interest.

12. Measures of Skewness and Kurtosis

No matter how the mean and the dispersion of a set of observations are defined, the dispersion will always have to depend on the unit of measurement, and the mean furthermore on the origin. But the case is a different one concerning the concepts of skewness and kurtosis. Here it is reasonable to raise the question for measures in the strict sense. It is obvious that such measures will be obtained if the set of observations is—by a convenient choice of a new unit—brought to the dispersion 1; the new Tchebychef coefficients of skewness and kurtosis will be suitable. This leads to the

Definition. *"With the designations of the preceding chapter, the ratios* $\frac{a_2}{a_1}$ *and* $\frac{a_3}{a_1}$ *shall be called:*

$\frac{a_2}{a_1} = S$ = Measure of Skewness *of the Observations*

$\frac{a_3}{a_1} = K$ = Measure of Kurtosis *of the Observations."*

There will be no misunderstanding if we use the words *"Skewness"* and *"Kurtosis"* instead of *"Measure of Skewness"* and *"Measure of Kurtosis"*.—Utilizing theorems 8 and 9, we have at once:

Theorem 11. *"The measures of skewness and kurtosis depend on the type of frequency function and on the observations* ζ *only; they are independent of origin and unit of measurement."*

13. Meaning of Skewness and Kurtosis

To secure an idea of the mechanism of skewness and kurtosis, let us construct some examples which show these phenomena in complete purity. We will use the step function, and we intend to choose the values $x_1, \cdots, x_n$ so that they are affected—apart from the inevitable dispersion—in the first place with skewness only, in the second place with kurtosis only.

We take $n = 10$, and for the convenience of the reader we actually write down the vectors $\mathfrak{z}_0, \cdots, \mathfrak{z}_3$. We observe however that in practice one will never evaluate these vectors, but rather compute the Tchebychef coefficients in the direct manner described in No. 15.

We obtain

ν	$\mathfrak{z}_0$	$\mathfrak{z}_1$	$\mathfrak{z}_2$	$\mathfrak{z}_3$
1	1	− 1.56670	+ 1.65145	− 1.43388
2	1	− 1.21855	+ .55048	+ .47796
3	1	− .87039	− .27524	+ 1.19490
4	1	− .52223	− .82572	+ 1.05834
5	1	− .17408	− 1.10096	+ .40968
6	1	+ .17408	− 1.10096	− .40968
7	1	+ .52223	− .82572	− 1.05834
8	1	+ .87039	− .27524	− 1.19490
9	1	+ 1.21855	+ .55048	− .47796
10	1	+ 1.56670	+ 1.65145	+ 1.43388

We shall have to come back to these vectors in No. 17. For this reason they have been calculated more accurately than is necessary here.

1 a. Positive skewness.

$\mathfrak{x} = \mathfrak{z}_1 + \frac{1}{5}\mathfrak{z}_2 \qquad (a_1 = 1, a_2 = +\frac{1}{5};\ a_\nu = 0$ otherwise$)$.

1 b. Negative skewness.

$\mathfrak{x} = \mathfrak{z}_1 - \frac{1}{5}\mathfrak{z}_2 \qquad (a_1 = 1, a_2 = -\frac{1}{5};\ a_\nu = 0$ otherwise$)$.

2 a. Positive kurtosis.

$\mathfrak{x} = \mathfrak{z}_1 + \frac{1}{10}\mathfrak{z}_3 \qquad (a_1 = 1, a_3 = +\frac{1}{10};\ a_\nu = 0$ otherwise$)$.

2 b. Negative kurtosis.

$\mathfrak{x} = \mathfrak{z}_1 - \frac{1}{10}\mathfrak{z}_3 \qquad (a_1 = 1, a_3 = -\frac{1}{10};\ a_\nu = 0$ otherwise$)$.

The components of the different vectors $\mathfrak{E}$ are put together in the table below:

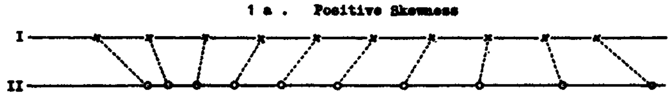
ν	$\mathfrak{z}_1+\frac{1}{5}\mathfrak{z}_2$	$\mathfrak{z}_1-\frac{1}{5}\mathfrak{z}_2$	$\mathfrak{z}_1+\frac{1}{10}\mathfrak{z}_3$	$\mathfrak{z}_1-\frac{1}{10}\mathfrak{z}_3$
1	- 1.236	- 1.897	- 1.710	- 1.423
2	- 1.108	- 1.329	- 1.171	- 1.266
3	- .925	- .815	- .751	- .990
4	- .687	- .357	- .416	- .628
5	- .394	+ .046	- .133	- .215
6	- .046	+ .394	+ .133	+ .215
7	+ .357	+ .687	+ .416	+ .628
8	+ .815	+ .925	+ .751	+ .990
9	+ 1.329	+ 1.108	+ 1.171	+ 1.266
10	+ 1.897	+ 1.236	+ 1.710	+ 1.423

To illustrate the preceding, we compare the vectors $\mathfrak{E}$ with their corresponding *"best systems of best values"*, that is to say with the vectors

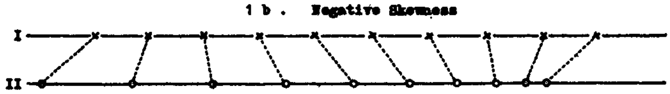
$$\bar{\mathfrak{E}} = a_0 \mathfrak{z}_0 + a_1 \mathfrak{z}_1 ,$$

and carry it through with some figures. We place the components of $\bar{\mathfrak{E}}$ on a horizontal straight line I, the components of $\mathfrak{E}$ on a second straight line II below:

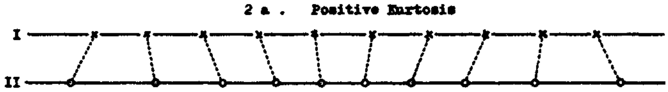
1 a . Positive Skewness



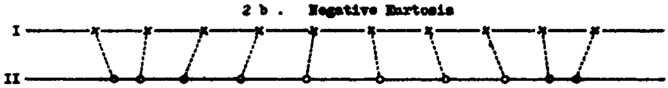
1 b . Negative Skewness



2 a . Positive Kurtosis



2 b . Negative Kurtosis



The reader should settle his mind upon the fact that the general behaviour of observations affected with skewness only or kurtosis only is always the same, no matter which type of frequency function is considered.—The meaning of skewness and kurtosis can be, generally speaking, expressed by:

Positive Skewness = Overconcentration to the Left
Negative Skewness = Overconcentration to the Right
Positive Kurtosis = Overconcentration near the Mean
Negative Kurtosis = Underconcentration near the Mean.

14. Measures of Approximation

Let T be a type of frequency function, $\xi = (x_1, \cdots, x_n)$ a set of observations, and $k \geq 1$ a *"degree of approximation"*, that is the subscript in the sum $a_0 \mathfrak{z}_0 + \cdots + a_k \mathfrak{z}_k$. The expression

$$\{\xi - (a_0 \mathfrak{z}_0 + \cdots + a_k \mathfrak{z}_k)\}^2$$

will give us a clue to the quality of approximation to the vector ξ which is obtained on the basis of the type T and the degree of approximation k. But the expression above of course is not yet fit to be taken as a measure of the quality of approximation. Therefore it will be necessary in the first instance to modify it so that it will become not only independent of the origin, but also independent of the unit of measurement.

Regarding theorem 9, and making reflections customary in situations of this kind, we are almost compulsorily led to the

Definition. *"The values*

$$M_k = \sqrt{\frac{a_1^2 + \cdots + a_k^2}{\xi^2 - a_0^2}} \qquad (k = 1, 2, \cdots, \overline{n-1})$$

shall be called MEASURES OF APPROXIMATION OF THE DEGREES k."

Theorem 12. *"The measures of approximation M_k depend on the type T and on the observations ξ only; they are independent of origin and unit of measurement. Furthermore they satisfy*

$$0 < M_1 \leq M_2 \leq \cdots \leq M_{n-1} = 1."$$

All is clear if we write—utilizing the relation (14) in No. 9—

M_κ in the form

$$M_\kappa^2 = \frac{a_1^2 + a_2^2 + \cdots\cdots + a_\kappa^2}{a_1^2 + a_2^2 + \cdots\cdots + a_{n-1}^2},$$

paying attention to theorems 6, 8 and 9.

If $M_{1\kappa}$ is not much smaller than 1, the approximation of degree κ will be estimated to be good. If T and T^* are two types of frequency function, M_κ and M_κ^* the corresponding measures of approximation, and if $M_\kappa^* \geqslant M_\kappa$, we say: T^* is, *for the degree* κ, *better than* T (equivalence not excluded). If $M_1^* \geqslant M_1, \cdots, M_\kappa^* \geqslant M_\kappa$, we say: T^* is, *up to the degree* κ, *better than* T.

Clearly we may base upon these concepts a method of curve-fitting. A full account will be given in a future note.

15. Computation of the Tchebychef Coefficients

If the vectors $\mathfrak{z}_0, \cdots, \mathfrak{z}_\kappa$ are already known, the finding of $a_0, \cdots, a_\kappa$ is, according to their definition, very simple. But the actual calculation of $\mathfrak{z}_0, \cdots, \mathfrak{z}_\kappa$ is embarrassing, especially if n is large. We already mentioned that this can be and should be avoided, and we recommend the following procedure.

We form, just as in the proof of theorem 4,

$$\begin{aligned} \mathfrak{y}_0 &= \mathfrak{t}_0 \\ \mathfrak{y}_1 &= \gamma_{10}\mathfrak{t}_0 + \mathfrak{t}_1 \\ &\cdots\cdots \\ \mathfrak{y}_\kappa &= \gamma_{\kappa 0}\mathfrak{t}_0 + \cdots + \gamma_{\kappa,\kappa-1}\mathfrak{t}_{\kappa-1} + \mathfrak{t}_\kappa, \end{aligned}$$

and to determine the coefficients γ, we demand that the vectors $\mathfrak{y}$ be orthogonal. Let χ be an arbitrary subscript among $0, 1, \cdots, \kappa$. Then at least it must be true that

$$\mathfrak{y}_\chi \mathfrak{y}_0 = 0, \quad \cdots\cdots, \quad \mathfrak{y}_\chi \mathfrak{y}_{\chi-1} = 0, \tag{16}$$

and a fortiori

$$\mathfrak{y}_\chi (c_0 \mathfrak{y}_0 + \cdots + c_{\chi-1}\mathfrak{y}_{\chi-1}) = 0$$

for arbitrary values $c_0, \cdots, c_{\chi-1}$. But $\mathfrak{t}_0, \cdots, \mathfrak{t}_{\chi-1}$ are linear combinations of $\mathfrak{y}_0, \cdots, \mathfrak{y}_{x-1}$, hence

(17) $$\mathfrak{y}_x \mathfrak{k}_0 = 0, \cdots\cdots, \mathfrak{y}_x \mathfrak{k}_{x-1} = 0.$$

For abbreviation let us designate the moments of the best values $\xi_1, \cdots, \xi_n$ by

$$m_r = \frac{1}{n}\left(\xi_1^r + \xi_2^r + \cdots + \xi_n^r\right) \qquad (r = 0, 1, \cdots).$$

Obviously we have

$$\mathfrak{k}_p \mathfrak{k}_q = m_{p+q} \qquad (p, q = 0, 1, \cdots).$$

and the equation (17) produces

(18) $$\begin{cases} m_0 \gamma_{x0} + m_1 \gamma_{x1} + \cdots + m_{x-1} \gamma_{x, x-1} + m_x = 0 \\ m_1 \gamma_{x0} + m_2 \gamma_{x1} + \cdots + m_x \gamma_{x, x-1} + m_{x+1} = 0 \\ \cdots\cdots\cdots \\ m_{x-1} \gamma_{x0} + m_x \gamma_{x1} + \cdots + m_{2x-2} \gamma_{x, x-1} + m_{2x-1} = 0 \end{cases}$$

Conversely, from (18) follows (16), hence the equations (18) must have exactly one solution $\gamma_{x0}, \cdots, \gamma_{x, x-1}$.

Concerning the normalizing factor in $\mathfrak{z}_x = \frac{1}{\lambda_x} \mathfrak{y}_x$, we have

$$\begin{aligned} \lambda_x^2 = \mathfrak{y}_x^2 &= (\gamma_{x0} \mathfrak{k}_0 + \gamma_{x1} \mathfrak{k}_1 + \cdots + \gamma_{x, x-1} \mathfrak{k}_{x-1} + \mathfrak{k}_x)^2 \\ &= (m_0 \gamma_{x0} + \cdots + m_{x-1} \gamma_{x, x-1} + m_x) \gamma_{x0} \\ &+ (m_1 \gamma_{x0} + \cdots + m_x \gamma_{x, x-1} + m_{x+1}) \gamma_{x1} \\ &\cdots\cdots \\ &+ (m_x \gamma_{x0} + \cdots + m_{2x-1} \gamma_{x, x-1} + m_{2x}) \end{aligned},$$

and from (18):

$$\lambda_x^2 = m_x \gamma_{x0} + \cdots + m_{2x-1} \gamma_{x, x-1} + m_{2x}.$$

With the abbreviations

$$X_0 = \mathfrak{k}_0 \mathfrak{k}, \cdots\cdots, X_k = \mathfrak{k}_k \mathfrak{k}$$

we have

$$\begin{aligned} a_0 &= \frac{1}{\lambda_0} X_0 \\ a_1 &= \frac{1}{\lambda_1} (\gamma_{10} X_0 + X_1) \\ &\cdots\cdots \\ a_k &= \frac{1}{\lambda_k} (\gamma_{k0} X_0 + \gamma_{k1} X_1 + \cdots + \gamma_{k, k-1} X_{k-1} + X_k). \end{aligned}$$

For the calculation of $a_0, a_1, \cdots$ we recommend operating according to the following recipe, in the demonstration of which we confine ourselves to the most important case $\kappa = 3$. The modus procedendi for other values of the degree κ will be clear.

1. *Compute*

$$\xi_1 = \psi\left(\tfrac{1}{2n}\right), \quad \xi_2 = \psi\left(\tfrac{3}{2n}\right), \cdots, \quad \xi_n = \psi\left(\tfrac{2n-1}{2n}\right).$$

In the interest of the accuracy of the results it is advisable to take care that the equations

$$\tfrac{1}{n}\left(\xi_1 + \cdots + \xi_n\right) = 0, \qquad \tfrac{1}{n}\left(\xi_1^2 + \cdots + \xi_n^2\right) = 1$$

are precisely or approximately satisfied. This will be the case if $\mu_1 = 0$, $\mu_2 = 1$ hold; otherwise introduce

$$\xi_1^* = \alpha\,\xi_1 + \beta, \cdots, \quad \xi_n^* = \alpha\,\xi_n + \beta$$

instead of $\xi_1, \cdots, \xi_n$, with convenient constants $\alpha > 0$, β.

2. *Compute*

$$m_0, m_1, \cdots, m_6; \quad \chi_0, \cdots, \chi_3; \quad \varepsilon^2$$

Again it is useful to take care that the equations

$$\tfrac{1}{n}\left(x_1 + \cdots + x_n\right) = 0; \quad \tfrac{1}{n}\left(x_1^2 + \cdots + x_n^2\right) = 1$$

are precisely or approximately satisfied. This will be secured if $x_1, \cdots, x_n$ are distributed over nearly the same interval as $\xi_1, \cdots, \xi_n$.

3. *Form the scheme*

$$\begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ a_{10} & a_{11} & a_{12} & a_{13} \\ a_{20} & a_{21} & a_{22} & a_{23} \\ a_{30} & a_{31} & a_{32} & a_{33} \end{pmatrix} \doteq \begin{pmatrix} m_0 & m_1 & m_2 & m_3 \\ m_1 & m_2 & m_3 & m_4 \\ m_2 & m_3 & m_4 & m_5 \\ m_3 & m_4 & m_5 & m_6 \end{pmatrix}$$

4a. *To every element of the second, third and fourth row in this scheme add the corresponding element of the first row multiplied by* $-\frac{a_{10}}{a_{00}}, -\frac{a_{20}}{a_{00}}$, *and* $-\frac{a_{30}}{a_{00}}$ *respectively, so that there results a scheme*

$$\begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ 0 & a'_{11} & a'_{12} & a'_{13} \\ 0 & a'_{21} & a'_{22} & a'_{23} \\ 0 & a'_{31} & a'_{32} & a'_{33} \end{pmatrix}$$

4b. *To every element of the second and third row in the scheme*

$$\begin{pmatrix} a'_{11} & a'_{12} & a'_{13} \\ a'_{21} & a'_{22} & a'_{23} \\ a'_{31} & a'_{32} & a'_{33} \end{pmatrix}$$

add the corresponding element of the first row multiplied by $-\dfrac{a'_{21}}{a'_{11}}$ *and* $-\dfrac{a'_{31}}{a'_{11}}$ *respectively, so that there results a scheme*

$$\begin{pmatrix} a'_{11} & a'_{12} & a'_{13} \\ 0 & a''_{22} & a''_{23} \\ 0 & a''_{32} & a''_{33} \end{pmatrix}$$

4c. *To every element of the second row of the scheme*

$$\begin{pmatrix} a''_{22} & a''_{23} \\ a''_{32} & a''_{33} \end{pmatrix}$$

add the corresponding element of the first row multiplied by $-\dfrac{a''_{32}}{a''_{22}}$, *so that there results a scheme*

$$\begin{pmatrix} a''_{22} & a''_{23} \\ 0 & a'''_{33} \end{pmatrix}$$

5. *Extract*

$$\lambda_0 = \sqrt{a_{00}}\,, \qquad \lambda_1 = \sqrt{a'_{11}}\,, \quad \lambda_2 = \sqrt{a''_{22}}\,, \quad \lambda_3 = \sqrt{a'''_{33}}\,.$$

6. *Multiply the elements of the first, second and third row in the scheme*

$$\begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ 0 & a'_{11} & a'_{12} & a'_{13} \\ 0 & 0 & a''_{22} & a''_{23} \end{pmatrix}.$$

by $\dfrac{1}{a_{00}}$, $\dfrac{1}{a'_{11}}$, *and* $\dfrac{1}{a''_{22}}$ *respectively, so that there results a scheme*

$$B = \begin{pmatrix} 1 & b_{01} & b_{02} & b_{03} \\ 0 & 1 & b_{12} & b_{13} \\ 0 & 0 & 1 & b_{23} \end{pmatrix},$$

and extract

$$\gamma_{10} = -\,b_{01}$$

7. *To every element of the first row in the scheme* B *add the corresponding element of the second row multiplied by* $-b_{01}$, *so*

that there results a scheme

$$B' = \begin{pmatrix} 1 & 0 & b'_{02} & b'_{03} \\ 0 & 1 & b_{12} & b_{13} \\ 0 & 0 & 1 & b_{23} \end{pmatrix},$$

and extract

$$\gamma_{20} = -b'_{02}, \qquad \gamma_{21} = -b_{12}.$$

8. *To every element of the first and second row in the scheme* B' *add the corresponding element of the third row multiplied by* $-b'_{02}$ *and* $-b_{12}$ *respectively, so that there results a scheme*

$$B'' = \begin{pmatrix} 1 & 0 & 0 & b''_{03} \\ 0 & 1 & 0 & b'_{13} \\ 0 & 0 & 1 & b_{23} \end{pmatrix},$$

and extract

$$\gamma_{30} = -b''_{03}, \quad \gamma_{31} = -b'_{13}, \quad \gamma_{32} = -b_{23}.$$

9. *Form*

$$\begin{aligned} \alpha_0 &= \chi_0 \\ \alpha_1 &= \gamma_{10}\chi_0 + \chi_1 \\ \alpha_2 &= \gamma_{20}\chi_0 + \gamma_{21}\chi_1 + \chi_2 \\ \alpha_3 &= \gamma_{30}\chi_0 + \gamma_{31}\chi_1 + \gamma_{32}\chi_2 + \chi_3, \end{aligned}$$

$$M = a_0 = \frac{\alpha_0}{\lambda_0}, \quad \sigma = a_1 = \frac{\alpha_1}{\lambda_1}, \quad a_2 = \frac{\alpha_2}{\lambda_2}, \quad a_3 = \frac{\alpha_3}{\lambda_3}$$

$$S = \frac{a_2}{a_1}, \quad K = \frac{a_3}{a_1}$$

$$M_1 = \sqrt{\frac{a_1^2}{t^2 - a_0^2}}, \quad M_2 = \sqrt{\frac{a_1^2 + a_2^2}{t^2 - a_0^2}}, \quad M_3 = \sqrt{\frac{a_1^2 + a_2^2 + a_3^2}{t^2 - a_0^2}}$$

16. Controls of Computation

It is easy to point out controls for the process of evaluation of a_0, a_1, a_2, a_3, which do not require any considerable extra work, and yet indicate every occurring miscalculation with almost absolute safety. Such general controls, of course, can not bear upon the ascertainment of $\xi_1, \cdots, \xi_n$.

A. *Control of* m_0, m_1, m_2, m_3;

$$m_0 + 3(m_1 + m_2) + m_3 = \frac{1}{n}\sum_{\nu=1}^{n}(1+\xi_\nu)^3.$$

B. *Control of* m_4, m_5, m_6;

$$m_3 + 3(m_4 + m_5) + m_6 = \frac{1}{n}\sum_{\nu=1}^{n}\xi_\nu^3(1+\xi_\nu)^3.$$

C. *Control of* X_0, X_1, X_2, X_3;

$$X_0 + 3(X_1 + X_2) + X_3 = \frac{1}{n}\sum_{\nu=1}^{n}(1+\xi_\nu)^3 \cdot x_\nu .$$

D. *Control of* $\mathcal{E}^2$;

$$1 + 2X_0 + \mathcal{E}^2 = \frac{1}{n}\sum_{\nu=1}^{n}(1+x_\nu)^2.$$

E. *Control of* γ_{10}; γ_{20}, γ_{21}; $\gamma_{30}, \gamma_{31}, \gamma_{32}$;

The operations indicated under 3 - 8 in No. 15 are essentially nothing else than the solution of three systems of linear equations for one, two and three unknowns respectively, contracted into one uniform process of reckoning. Hence we can make use of the method of control by sums. We have to add the sums

$$S_0 = a_{00} + \cdots + a_{03}$$
$$\cdots\cdots$$
$$S_3 = a_{30} + \cdots + a_{33}$$

as elements of a fifth column:

$$\begin{pmatrix} a_{00} & \cdots\cdots & a_{03} & S_0 \\ \cdots & \cdots & \cdots & \cdots \\ a_{30} & & a_{33} & S_3 \end{pmatrix}$$

and to transform this expanded scheme in the way described in No. 15. Then everywhere the sum of the first four elements of each row must equal its fifth element. If this is true for the scheme B'' especially, it is practically impossible that $\gamma_{10}, \cdots, \gamma_{32}$ should have been wrongly computed.

F. *Control of* $\alpha_0, \alpha_1, \alpha_2, \alpha_3$;

The computation of $\alpha_0, \cdots, \alpha_3$ should be performed by starting from the scheme

$$\begin{pmatrix} 1 & \gamma_{10} & \gamma_{20} & \gamma_{30} & S_0 \\ 0 & 1 & \gamma_{21} & \gamma_{31} & S_1 \\ 0 & 0 & 1 & \gamma_{32} & S_2 \\ 0 & 0 & 0 & 1 & S_3 \end{pmatrix}$$

with the meaning

$$\begin{aligned} 1 + \gamma_{10} + \gamma_{20} + \gamma_{30} &= S_0 \\ 1 + \gamma_{21} + \gamma_{31} &= S_1 \\ 1 + \gamma_{32} &= S_2 \\ 1 &= S_3 . \end{aligned}$$

Multiply the elements of the first, second, third and fourth row by χ_0, χ_1, χ_2 *and* χ_3 *respectively, and form the sums of the elements of each column.* The sums of the first four columns are $\Upsilon_0, \Upsilon_1, \Upsilon_2, \Upsilon_3$, and we have the *control* $\Upsilon_0 + \Upsilon_1 + \Upsilon_2 + \Upsilon_3 = R$, where R designates the sum of the fifth column

17. EXAMPLES

I. Let the observations $(n = 30)$

$\zeta = (-1.5, -1.0, -.7, -.5, -.3, 0, +.3, +6, +1.0, +1.8)$ be given, and let us first assume the normal type. The normal law of error being symmetric, we have $m_1 = m_3 = m_5 = 0$, and in this case we are able to write down the Tchebychef coefficients required:

$$a_0 = \chi_0, \quad a_1 = \frac{\chi_1}{\sqrt{m_2}}, \quad a_2 = \frac{-m_2 \chi_0 + \chi_2}{\sqrt{m_4 - m_2^2}}, \quad a_3 = \frac{-m_4 \chi_1 + m_2 \chi_3}{\sqrt{m_2(m_2 m_6 - m_4^2)}} .$$

Nevertheless we will proceed according to No. 15. But we will confine ourselves to give the resulting data of the different steps of computation only. A full reproduction of the complete process of reckoning is to be found in No. 19, dealing with a somewhat more general situation.

In the KELLEY-WOOD tables we find

$$\xi_1 = -1.644854 \quad \xi_2 = -1.036433 \quad \xi_3 = -.764490 \quad \xi_4 = -.385320$$

$$\xi_5 = -.125661 \quad \xi_6 = .125661 \quad \xi_7 = .385320 \quad \xi_8 = .764490$$

$$\xi_9 = 1.036433 \quad \xi_{10} = 1.644854 .$$

We obtain $m_0 = 1 \qquad m_1 = 0 \qquad m_2 = +.87979 \qquad m_3 = 0$

$$m_4 = +1.74062 \qquad m_5 = 0 \qquad m_6 = +4.22829$$

$$\chi_0 = -.03000 \qquad \chi_1 = +.87237 \qquad \chi_2 = +.07317 \qquad \chi_3 = +1.73577$$

$$\mathcal{T}^2 = .87700;$$

$$\begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ 0 & a_{11} & a_{12} & a_{13} \\ 0 & 0 & a_{22} & a_{23} \\ 0 & 0 & 0 & a_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & +.87779 & 0 \\ 0 & +.87979 & 0 & +1.74062 \\ 0 & 0 & +.96659 & 0 \\ 0 & 0 & 0 & +.78456 \end{pmatrix}$$

$$\lambda_0 = 1 \qquad \lambda_1 = .93797 \qquad \lambda_2 = .98315 \qquad \lambda_3 = .88575$$

$$B = B' = \begin{pmatrix} 1 & 0 & +.87979 & 0 \\ 0 & 1 & 0 & +1.97845 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\Upsilon_{1c} = 0; \qquad \Upsilon_{20} = -.87979 \qquad \Upsilon_{21} = 0;$$

$$B'' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ c & 1 & 0 & +1.97845 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\Upsilon_{3c} = 0 \qquad \Upsilon_{31} = -1.97845 \qquad \Upsilon_{32} = 0;$$

$$M = a_0 = -.03000 \qquad \sigma = a_1 = +.93006 \qquad a_2 = +.10127 \qquad a_3 = +.01110$$

$$S = +.10889 \qquad K = +.01193$$

$$M_1 = .99382 \qquad M_2 = .99953 \qquad M_3 = .99959$$

For comparison we give the value which is furnished by the traditional concept of dispersion:

$$\sqrt{\frac{1}{n}\sum(x_\nu - M)^2} = \sqrt{\mathcal{T}^2 - a_0^2} = .93600$$

II. Let the same observations as above be given, but now let us assume the step type. We can make use of the vectors in No. 13, which give at once

$$M = a_0 = -.03000 \qquad \sigma = a_1 = +.92087 \qquad a_2 = +.10184 \qquad a_3 = +.12529$$

$$S = +.11059 \qquad K = +.13606$$

$$M_1 = .98383 \qquad M_2 = .98989 \qquad M_3 = .99941$$

We note that for our observations $\mathcal{T}$ the normal type is, up to the degree 3, better than the step type.

18. Analysis of Frequency Groups

In economic statistics, observations very often do not appear in the form dealt with in the preceding chapters. Instead, they usually are gathered into groups, so that there is given a set of values $x_1 < x_2 < \cdots < x_n$ and a set of corresponding positive values $N_1, \cdots, N_n$, not necessarily integers. If N means the sum of $N_1 + \cdots + N_n$, the ratios

$$f_1 = \frac{N_1}{N}, \qquad f_2 = \frac{N_2}{N}, \quad \cdots\cdots, \quad f_n = \frac{N_n}{N}$$

are called the "*frequencies*" of the "*observations*" $x_1, \cdots, x_n$. The frequencies satisfy

$$f_1 > 0, \quad f_2 > 0, \cdots, f_n > 0 \quad \text{and} \quad f_1 + \cdots + f_n = 1.$$

We shall now have to extend our developments to make them applicable in situations as stated above. To anyone who is familiar with integrals and sums in the sense of Stieltjes, it is clear that no special difficulty can arise.

Again we have to start from a frequency function $\varphi(x)$, and to agree which values $\xi_1, \cdots, \xi_n$ should be designated as "best values". Reflections similar to those of No. 2 make it reasonable to choose

$$\xi_1 = \psi(\tfrac{1}{2}f_1), \qquad \xi_2 = \psi(f_1 + \tfrac{1}{2}f_2), \quad \xi_3 = \psi(f_1 + f_2 + \tfrac{1}{2}f_3);$$

$$, \cdots\cdots, \; \xi_n = \psi(f_1 + f_2 + \cdots + f_{n-1} + \tfrac{1}{2}f_n).$$

Apart from the best values, we only have to modify the definition of the product of vectors (No. 4). We define

$$\breve{u}\, v = u_1 v_1 f_1 + u_2 v_2 f_2 + \cdots + u_n v_n f_n.$$

If these modifications are kept in mind, all the definitions, theorems, proofs and remarks of Nos. 4 - 16 remain unaltered. Of course, the abbreviations m_λ and χ_κ (No. 15) must now be read

$$m_\lambda = \xi_1^\lambda f_1 + \xi_2^\lambda f_2 + \cdots + \xi_n^\lambda f_n \qquad (\lambda = 0, 1, 2, \cdots)$$

$$\chi_\nu = \breve{\zeta}_\nu \zeta = \xi_1^\nu x_1 f_1 + \cdots\cdots + \xi_n^\nu x_n f_n \qquad (\nu = 0, 1, \cdots k),$$

and the controls A - D (No. 16):

$$\text{A.} \qquad m_0 + 3(m_1 + m_2) + m_3 = \sum_{\nu=1}^{n} (1 + \xi_\nu)^3 f_\nu$$

B. $$m_3 + 3(m_4 + m_5) + m_6 = \sum_{\nu=1}^{n} (1+\xi_\nu)^3 \xi_\nu^3 f_\nu$$

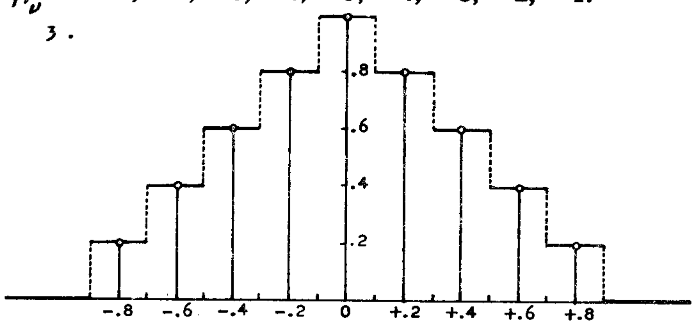
C. $$\chi_0 + 3(\chi_1 + \chi_2) + \chi_3 = \sum_{\nu=1}^{n} (1+\xi_\nu)^3 x_\nu f_\nu$$

D. $$1 + 2\chi_0 + \mathcal{E}^2 = \sum_{\nu=1}^{n} (1+x_\nu)^2 f_\nu .$$

We are now in the position to illustrate the mechanism of skewness and kurtosis still more impressively than in No. 13. For this purpose we start from the frequency curve represented in Fig. 3; we choose $n = 9$ and

N_ν : 1, 2, 3, 4, 5, 4, 3, 2, 1.

3.



Then the best values become equidistant, and they are given by the abscissae of the points marked by small circles:

$-.8, \ -6, \ -.4, \ -.2, \ 0, \ +.2, \ +.4, \ +.6, \ +.8.$

The ordinates η_ν of these points are proportional to N_ν, namely:

$$\eta_\nu = \frac{1}{0.2} f_\nu = 5\frac{N_\nu}{N} .$$

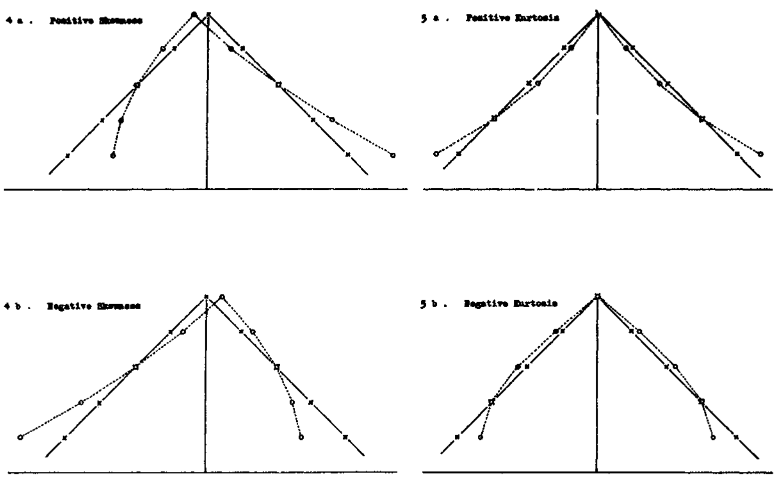
The table below gives the corresponding vectors $\mathfrak{z}_0, \cdots, \mathfrak{z}_3$, and also the vectors $\mathfrak{z}_1 \pm \frac{1}{4}\mathfrak{z}_2$ and $\mathfrak{z}_1 \pm \frac{1}{8}\mathfrak{z}_3$ as examples of distributions which show skewness or kurtosis in all purity.

ν	$\mathfrak{z}_0$	$\mathfrak{z}_1$	$\mathfrak{z}_2$	$\mathfrak{z}_3$	$\mathfrak{z}_1+\frac{1}{4}\mathfrak{z}_2$	$\mathfrak{z}_1-\frac{1}{4}\mathfrak{z}_2$	$\mathfrak{z}_1+\frac{1}{8}\mathfrak{z}_3$	$\mathfrak{z}_1-\frac{1}{8}\mathfrak{z}_3$
1	1	−2.0	+2.582	−2.561	−1.355	−2.645	−2.320	−1.680
2	1	−1.5	+1.076	+.116	−1.231	−1.769	−1.485	−1.515
3	1	−1.0	.000	+1.048	−1.000	−1.000	−.869	−1.131
4	1	−.5	−.645	+.815	−.661	−.339	−.398	−.602
5	1	.0	−.861	.000	−.215	+.215	.000	.000
6	1	+.5	−.645	−.815	+.339	+.661	+.398	+.602
7	1	+1.0	.000	−1.048	+1.000	+1.000	+.869	+1.131
8	1	+1.5	+1.076	−.116	+1.769	+1.231	+1.485	+1.515
9	1	+2.0	+2.582	+2.561	+2.645	+1.355	+2.320	+1.680

As in No. 13, let us illustrate the relations between the vector $\mathfrak{z}_1$ and the vectors $\mathfrak{z}_1 \pm \frac{1}{4}\mathfrak{z}_2$ and $\mathfrak{z}_1 \pm \frac{1}{8}\mathfrak{z}_3$ by means of some figures. This time however, we shall not only consider the components of the vectors, but also operate with the values f_ν. We do that by associating every vector $(u_1, \cdots, u_n)$ with the system of points

$$(u_1, f_1), \quad (u_2, f_2), \quad \cdots\cdots, (u_n, f_n).$$

Thus, in the figures 4a - 5b, the vector $\mathfrak{z}_1$ is every time associated with the system of points marked by crosses, whereas the system of points marked by circles successively correspond to the vectors $\mathfrak{z}_1 \pm \frac{1}{4}\mathfrak{z}_2$ and $\mathfrak{z}_1 \pm \frac{1}{8}\mathfrak{z}_3$.



The statements in No. 13 concerning the meaning of skewness as overconcentration to the left or to the right, and of kurtosis as overconcentration or underconcentration near the mean should be recognized.

Until now, the values N_ν were supposed to be really positive, but there is no difficulty in allowing some of them to equal zero. Then, it is true, the formulation of some intermediary theorems must be changed. Yet, the existence and the main properties of the Tchebychef coefficients remain untouched, and *their values are independent of those* x_ν *for which the corresponding* f_ν *are equal to zero.* To know this is sometimes useful in order to get a scheme of computation of the highest possible uniformity.

19. Example

To conclude, we reproduce the reckoning of an example, frequently discussed, concerning observations of the right ascension of the pole star (see: A. L. Bowley, Elements of Statistics, 4th ed., p. 255). The given data are

x_ν^*:	-7	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6
N_ν:	1	6	12	21	36	61	73	82	72	63	38	16	5	1,

and the normal type shall be assumed.

Because the function

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

satisfies $\mu_1 = 0, \mu_2 = 1$, it will be suitable to start from the best values of this specimen. These best values $\xi_1, \cdots, \xi_{14}$ stretch from -3 to $+3$ approximately. In order to have the values x_ν, with which we intend to work, in coextension with $\xi_1, \cdots, \xi_{14}$, we choose

$$x_\nu = \frac{1}{2} x_\nu^* \quad \text{i.e.} \quad x_\nu^* = 2 x_\nu \qquad (\nu = 1, \cdots, 14)$$

Between the means M, M^* and the dispersions σ, σ^* of the observations x_ν, x_ν^* there exist the connections (theorem 9)

$$M^* = 2M, \quad \sigma^* = 2\sigma,$$

whereas the measures of skewness and kurtosis as well as the measures of approximation do not change at the transition from x_ν to x_ν^* (theorems 11 and 12).

Computation of $m_0, \cdots, m_6$; $\chi_0, \cdots, \chi_3$; $\mathcal{E}^2$.

ν	ξ_ν	f_ν	$\xi_\nu f_\nu$	$\xi_\nu^2 f_\nu$	$\xi_\nu^3 f_\nu$
1	-3.08242	.00205 339	-.00632 941	.01950 990	- .06013 77
2	-2.39928 8	.01232 033	-.02956 002	.07092 300	- .17016 47
3	-1.93176 9	.02464 066	-.04760 006	.09195 232	- .17763 06
4	-1.54996 6	.04312 11	-.06683 62	.10359 38	- .16056 69
5	-1.17951 6	.07392 20	-.08719 22	.10284 46	- .12130 67
6	- .77663 9	.12525 7	-.09727 9	.07555 1	- .05867 6
7	- .36846 6	.14989 7	-.05523 2	.02035 1	- .00749 9
8	+ .03861 3	.16837 8	+.00650 2	.00025 1	+ .00001 0
9	+ .44963 0	.14784 4	+.06647 5	.02988 9	+ .01343 9
10	+ .88571 7	.12936 34	+.11457 94	.10148 49	+ .08988 69
11	+1.37743 5	.07802 87	+.10747 95	.14804 60	+ .20392 37
12	+1.89953 0	.03285 421	+.06240 756	.11854 503	+ .22517 98
13	+2.44778 6	.01026 694	+.02513 127	.06151 597	+ .15057 79
14	+3.08242	.00205 339	+.00632 941	.01950 990	+ .06013 77
		$+1.00000 = m_0$	$-.00113 = m_1$	$+.96397 = m_2$	$-.01283 = m_3$

ν	$\xi_\nu^4 f_\nu$	$\xi_\nu^5 f_\nu$	$\xi_\nu^6 f_\nu$
1	.18536 96	- .57138 7	1.76125 5
2	.40827 31	- .97956 7	2.35026 3
3	.34314 13	- .66287 0	1.28051 2
4	.24887 3	- .38574 5	.59789 2
5	.14308 3	- .16876 9	.19906 6
6	.04557 0	- .03539 1	.02748 6
7	.00276 3	- .00101 8	.00037 5
8	.00000 0	+ .00000 0	.00000 0
9	.00604 3	+ .00271 7	.00122 2
10	.07961 4	+ .07051 5	.06245 6
11	.28089 2	+ .38691 0	.53294 3
12	.42773 58	+ .81249 7	1.54336 2
13	.36858 25	+ .90221 1	2.20841 9
14	.18536 96	+ .57138 7	1.76125 5
	$+ 2.72531 = m_4$	$- .05851 = m_5$	$+12.32651 = m_6$

ν	X_ν	$X_\nu f_\nu$	$\xi_\nu X_\nu f_\nu$	$\xi_\nu^2 X_\nu f_\nu$	$\xi_\nu^3 X_\nu f_\nu$	$X_\nu^2 f_\nu$
1	-3.5	- .00718 687	.02215 295	- .06828 47	.21048 2	.02515 4
2	-3.0	- .03696 099	.08868 006	- .21276 90	.51049 4	.11088 3
3	-2.5	- .06160 164	.11900 014	- .22988 08	.44407 7	.15400 4
4	-2.0	- .08624 23	.13367 26	- .20718 80	.32113 4	.17248 5
5	-1.5	- .11088 30	.13078 83	- .15426 69	.18196 0	.16632 4
6	-1.0	- .12525 67	.09727 9	- .07555 1	.05867 6	.12525 7
7	- .5	- .07494 9	.02761 6	- .01017 6	.00374 9	.03747 4
8	.0	.00000 0	.00000 0	.00000 0	.00000 0	.00000 0
9	+ .5	+ .07392 2	.03323 8	+ .01494 5	.00672 0	.03696 1
10	+1.0	+ .12936 34	.11457 94	+ .10148 49	.08988 7	.12936 3
11	+1.5	+ .11704 31	.16121 93	+ .22206 91	.30588 6	.17556 5
12	+2.0	+ .06570 842	.12481 512	+ .23709 01	.45036 0	.13141 7
13	+2.5	+ .02566 735	.06282 818	+ .15378 99	.37644 5	.06416 8
14	+3.0	+ .00616 016	.01898 820	+ .05852 96	.18041 3	.01848 0
		- .08522 = X_0	+1.13486 = X_1	- .17021 = X_2	+ 3.14028 = X_3	1.34754 = $\mathcal{E}^2$

Controls A - D.

ν	$1+\xi_\nu$	$(1+\xi_\nu)^3$	$(1+\xi_\nu)^3 f_\nu$	$(1+\xi_\nu)^3 \xi_\nu^3 f_\nu$	$(1+\xi_\nu)^3 X_\nu f_\nu$	$(1+X_\nu)^2 f_\nu$
1	-2.08242	- 9.0304	- .01854 3	+ .54306 5	+ .06490 0	.01283 4
2	-1.39928 8	- 2.73982	- .03375 5	+ .46622 0	+ .10126 6	.04928 1
3	- .93176 9	- .80896	- .01993 3	+ .14369 5	+ .04983 3	.05544 1
4	- .54996 6	- .16634	- .00717 3	+ .02670 9	+ .01434 6	.04312 1
5	- .17951 6	- .00579	- .00042 8	+ .00070 2	+ .00064 2	.01848 1
6	+ .22336 1	+ .01114	+ .00139 6	- .00065 4	- .00139 5	.00000 0
7	+ .63153 4	+ .25188	+ .03775 6	- .00188 9	- .01887 8	.03747 4
8	+1.03861 3	+ 1.12037	+ .18864 5	+ .00001 1	+ .00000 0	.16837 8
9	+1.44963 0	+ 3.04629	+ .45037 6	+ .04093 9	+ .22518 8	.33264 9
10	+1.88571 7	+ 6 70548	+ .86744 3	+ .60273 4	+ .86744 4	.51745 4
11	+2.37743 5	+13.43773	+ 1.04852 9	+ 2.74027 2	+1.57279 4	.48767 9
12	+2.89953 0	+24.37714	+ .80089 2	+ 5.48924 0	+1.60178 3	.29568 8
13	+3.44778 6	+40.98462	+ .42078 7	+ 6.17137 8	+1.05196 7	.12577 0
14	+4.08242	+68.0382	+ .13970 9	+ 4.09166 2	+ .41912 6	.03285 4
			+ 3.87570	+ 20.31408	+ 5.94902	2.17710

$$m_0 + 3(m_1+m_2)+m_3 = 3.87569$$
$$m_3+3(m_4+m_5)+m_6 = 20.31408$$
$$X_0+3(X_1+X_2)+X_3 = 5.94901$$
$$1+2X_0+\mathcal{E}^2 = 2.17710$$

Computation of $a_{00}, \cdots, a_{03}$; $a'_{11}, \cdots, a'''_{33}$ (twice underlined) and of $\lambda_0, \cdots ; \lambda_3$, with control by sums.

1	- .00113	+ .96397	- .01283	+1.95001
- .00113	+ .96397	- .01283	+2.72531	+3.67532
	- .00000	+ .00109	- .00001	+ .00220
	+ .96397	- .01174	+2.72530	+3.67752
+ .96397	- .01283	+2.72531	- .05851	+3.61794
	+ .00109	- .92924	+ .01237	- 1.87975
	- .01174	+1.79607	- .04614	+1.73819
- .01283	+2.72531	- .05851	+12.32651	+14.98048
	- .00001	+ .01237	- .00016	+ .02502
	+2.72530	- .04614	+12.32635	+15.00550

+ .96397	- .01174	+2.72530	+ 3.67752
- .01174	+1.79607	- .04614	+ 1.73819
	- .00014	+ .03319	+ .04479
	+1.79593	- .01295	+ 1.78298
+2.72530	- .04614	+12.32635	+15.00550
	+ .03319	- 7.70486	- 10.39694
	- .01295	+ 4.62149	+ 4.60856

+ 1.79593	- .01295	+ 1.78298
- .01295	+ 4.62149	+ 4.60856
	- .00009	+ .01286
	+ 4.62140	+ 4.62142

$\lambda_0 = 1$

$\lambda_1 = .98182$

$\lambda_2 = 1.34012$

$\lambda_3 = 2.14974$

Computation of $\gamma_{10}, \cdots, \gamma_{32}$, with control by sums.

$\gamma_{10} = +.00113$

1	− .00113	+ .96397	− .01283	+ 1.95001
	+ .00113	− .00001	+ .00319	+ .00431
1	0	+ .96396	− .00964	+ 1.95432
	1	− .01218	+ 2.82716	+ 2.81497
		1	− .00721	+ .99279

$\gamma_{20} = -.96396$
$\gamma_{21} = +.01218$

1	0	+ .96396	− .00964	+ 1.95432
		− .96396	+ .00695	− .95701
1	0	0	− .00269	+ .99731
	1	− .01218	+ 2.82716	+ 3.81497
		+ .01218	− .00009	+ .01209
	1	0	+ 2.82707	+ 3.82706
		1	− .00721	+ .99279

$\gamma_{30} = + .00269$
$\gamma_{31} = - 2.82707$
$\gamma_{32} = + .00721$

Computation of $\alpha_0, \cdots, \alpha_3$, with control by sums.

1	+ .00113	− .96396	+ .00269	+ .03986
	1	+ .01218	− 2.82707	− 1.81489
		1	+ .00721	+ 1.00721
			1	+ 1.00000
− .08522	− .00010	+ .08214	− .00023	− .00340
	+ 1.13486	+ .01382	− 3.20833	− 2.05965
		− .17021	− .00123	− .17144
			+ 3.14028	+ 3.14028
− .08522	+ 1.13476	− .07425	− .06951	+ .90579
$= \alpha_0$	$= \alpha_1$	$= \alpha_2$	$= \alpha_3$	

$\alpha_0 + \cdots + \alpha_3 = + .90578$

Finishing computations.

$M = a_0 = .08522$	$\sigma = a_1 = 1.15577$	$a_2 = -.05541$	$a_3 = -.03234$
$M^* = .17044$	$\sigma^* = 2.31154$	$S = -.04794$	$K = -.02799$

$\mathfrak{E}^2 = 1.34754$	$a_1^2 = 1.33580$	$M_1^2 = .99666$	$M_1 = .99833$
$a_0^2 = .00726$	$a_2^2 = .00307$		
$\mathfrak{E}^2 - a_0^2 = 1.34028$	$a_1^2 + a_2^2 = 1.33887$	$M_2^2 = .99895$	$M_2 = .99947$
	$a_3^2 = .00105$		
	$a_1^2 + a_2^2 + a_3^2 = 1.33992$	$M_3^2 = .99973$	$M_3 = .99987$

So long as we pay regard to the Tchebychef coefficients a, $\cdots a_3$ only, the purport of our results is that *the observations are somewhat overconcentrated to the right, and somewhat underconcentrated near the mean.* The sum of the squares of the Tchebychef coefficients with higher subscripts than 3 is

$$a_4^2 + \cdots\cdots + a_{13}^2 = \mathfrak{E}^2 - (a_0^2 + \cdots + a_3^2) = .00036;$$

it is small compared with $a_2^2 = .00307$ and $a_3^2 = .00105$. The vectors $\mathfrak{z}_0, \cdots\cdots, \mathfrak{z}_{13}$ being normalized, we are sure that *the influence of $a_4, \cdots, a_{13}$ cannot essentially disturb our statements.*

Finally we give an illustration by computing and drawing the "best curve" of the normal type, corresponding to the observations x_ν. With it we mean that curve $y = \frac{1}{\alpha}\varphi\left(\frac{x-\beta}{\alpha}\right)$, the best values of which are the components of the vector $a_0\mathfrak{z}_0 + a_1\mathfrak{z}_1$. The values α, β (see No. 2) have to satisfy

$$\beta\mathfrak{E}_0 + \alpha\mathfrak{E}_1 = a_0\mathfrak{z}_0 + a_1\mathfrak{z}_1;$$

substituting

$$\mathfrak{z}_0 = \frac{1}{\lambda_0}\mathfrak{E}_0, \qquad \mathfrak{z}_1 = \frac{1}{\lambda_1}(\gamma_{10}\mathfrak{E}_0 + \mathfrak{E}_1)$$

we get

$$\left\{\beta - \left(\frac{a_0}{\lambda_0} + \gamma_{10}\frac{a_1}{\lambda_1}\right)\right\}\mathfrak{E}_0 + \left\{\alpha - \frac{a_1}{\lambda_1}\right\}\mathfrak{E}_1 = \mathfrak{o},$$

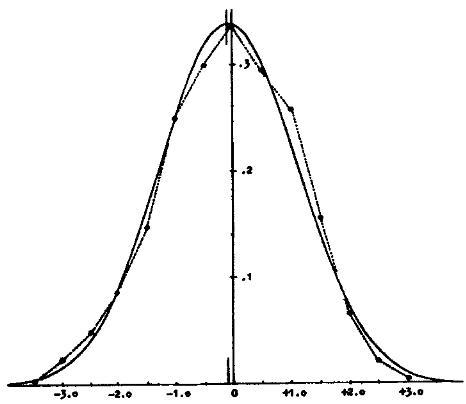
hence

$$\sigma = \frac{a_1}{\lambda_1} \qquad\qquad = +1.17717$$

$$\beta = \frac{a_0}{\lambda_0} + \gamma_{10}\,\frac{a_1}{\lambda_1} \qquad = -\ .08389.$$

With these values σ and β, the curve in Fig. 6 represents the function

$$y = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\beta)^2}{2\sigma^2}}$$



The abscissae of the points marked by circles are the observations x_ν, their ordinates are equal to the corresponding f_ν divided by the length 0.5 of the group intervals.

University of Kiel, Germany.