P.L. Chebyshev

The Theory of Probability

Translated by Oscar Sheynin
Lectures delivered in 1879 – 1880 as taken down by A.M. Liapunov
Berlin, 2004
© Oscar Sheynin   www.sheynin.de
П.Л. Чебышев   Теория Вероятностей   Лекции, читанные в 1879 – 1880гг.
По записи А.М. Ляпунова   Изданы академиком А.Н. Крыловым   Москва 1936 Ленинград

Contents

Introduction by the Translator

1. Pafnuty Lvovich Chebyshev (Tchébichef) (1821 – 1894) was one of the two most eminent Russian mathematicians of the 19th century (the second, or, rather, the first one was Lobachevsky). In the theory of probability, he proved the law of large numbers in a general form and thoroughly studied the conditions for the central limit theorem (a term introduced by Polya in 1920) providing, in 1887, the necessary framework for its definitive investigations carried out by his former students, Markov and Liapunov. Chebyshev also dealt on both of these most important topics in his lectures on probability theory that he delivered at Petersburg University from 1860 to 1882. The authoritative Russian sources about his work are [1 – 3] and I myself published a paper [4] discussing his lectures on probability; its conclusions would be an useful supplement to this *Introduction* and to my notes in the main text below[1].

In 1936, Aleksei Nikolaevich Krylov, a naval architect and an applied mathematician, published Chebyshev's lecture notes as written down by Liapunov in 1879 – 1880 (see Krylov's *Foreword* which I am now presenting in translation below). In 1999, my translation of the lecture notes appeared as a microfiche edition published by Verlag Dr. Hänsel-Hohenhausen in their series *Deutsche Hochschulschriften*, No. 2665, Egelsbach, but the copyright to ordinary publication is mine.

2. This translation contains Chebyshev's own footnotes as well as notes by Krylov and me (my notes are in curly brackets), all of which are collected at the end of the appropriate subsections, and an Index of names compiled by me.

The readers of the Russian edition undoubtedly noticed many dozens of misprints in mathematical formulas and an almost total absence of periods after displayed formulas and (effectively) new sentences beginning then with a lower-case first letter of their first words. The only possible explanation of this sad state of affairs seems to be that Krylov, in spite of his testimony provided in the *Foreword*, had not rewritten the original manuscript himself. The type-setter (an apprentice?) had contributed to the wrecking of the formulas; and hardly anyone read the proofs. I have corrected the misprints without special notice but I did not check all the formulas; and when I write something like *Chebyshev had not* …, the fault can well lie elsewhere rather than with him or Liapunov.

3. I had not improved on Chebyshev's style of *oral* presentation and I attempted to preserve his mathematical terminology and notation [*value of integral*; *exactly contrary events*; *mathematical expectation* (dropping however the adjective in §3.3); *equations in finite differences*; $\lim P_{x=0}$] or used it either more often than Chebyshev ($\exp[f(x)]$) or throughout rather than exceptionally (i instead of $\sqrt{-1}$). Then, I myself introduced the notation C_m^n and $n!$. On the other hand, disregarding a single exception, I have not retained Chebyshev's notation (§3.3)

$$\prod_{L}^{M} = \sum_{m=L}^{M} p(m)$$

for the sums of probabilities. Chebyshev transformed such sums and arrived at appropriate integrals with limits t_0 and t_1 (say) being functions of L and M respectively. Just the same, I have not preserved his similar use of letter S instead of Π in §§3.3.9 – 3.3.10.

The numeration of the formulas was chaotic; furthermore, it did not enable the author to refer to his previous results and he rewrote many formulas time and time again[2]. I ordered his numeration assigning numbers to those formulas which were marked by asterisks or Greek or Roman letters, and the numbers are now running consecutively through each chapter with no separate systems of them appearing anywhere anymore. In addition, I numbered many more formulas needed for references by using a parallel system of Roman numerals. Note that the formulas which Chebyshev included in his main system are now printed in bold type.

In conclusion, I note that in a few cases Chebyshev had not shown the necessary intermediate steps (§§3.1.5 and 3.3.19).

далее следует дополнение

Notes

1. I have subsequently described Chebyshev's work in probability [5, Chapt. 15]. In particular (pp. 205 and 206), I noted that Tikhomandritsky, in 1898, stated that in 1887 Chebyshev had remarked that it was *necessary to transform the entire theory of probability*. I also described an episode proving that Chebyshev had considered the Riemann geometry and the complex-variable analysis as *trendy disciplines*.
2. In this respect Markov followed his teacher.

References

1. Bernstein, S.N. (1945), Chebyshev's work in the theory of probability. *Собрание сочинений* (Coll. Works), vol. 4. N.p., 1964, pp. 409 – 433. Transl. in DHS 2656, 1999, pp. 67 – 96.
2. --- (1947), Chebyshev and his influence on the development of mathematics. *Uchenye Zapiski Mosk. Gos. Univ.*, No. 91, pp. 35 – 45. Transl. in *Math. Scientist*, vol. 26, 2001, pp. 63 – 73.
3. Youshkevich, A.P. (1971), Chebyshev. *Dict. Scient. Biogr.*, vol 3, pp. 222 – 232.
4. Sheynin, O. (1994), Chebyshev's lectures on the theory of probability. *Arch. Hist. Ex. Sci.*, vol. 46, pp. 321 – 340.
5. --- (2004), *History of the Theory of Probability to the 20th Century*. Berlin.

Chapter 1. Definite Inegrals

1.1. Preliminary Remarks and the Integrals of the First Group

1.1.0. We shall call *definite* only such integrals whose limits are constant magnitudes. Thus, we shall not consider definite integrals of the type

$$\int_1^x \frac{dx}{x} = \ln x.$$

Many definite integrals can be deduced from indefinite integrals but we shall only treat such of them which cannot be determined in an indefinite form. For example, the integral

$$\int \exp(-x^2)\,dx$$

cannot be determined because it represents a transcendental function unknown to us. At the same time, we can find its value if we add the limits 0 and ∞; the integral will then equal $\sqrt{\pi}/2$. This happens because we do not have such a function that shows how the quantity of the area $OAMB$ [1] changes depending on the change of x. This, however, does not preclude the possibility of determining the quantity of all the area restricted by the curve

$$y = \exp(-x^2). \tag{i}$$

In such investigations it is impossibile to apply a direct approach and we must perforce choose an indirect way. For this reason the methods used are extremely diverse and often numerous. Definite integrals are separated into several groups and special tricks exist for each of these. In addition, various scientists determine one and the same integral by different methods. Our method is this: Issuing from known double integrals and changing the order of integration, we shall determine the definite integrals. We do not aim at deriving the value of a certain definite integral; on the contrary, we shall rather determine various definite integrals from a given double integral.

Thus, the change in the order of integration will be the foundation of all our conclusions. Such a change is known to be only possible if the integral might be considered as the limit of a sum; this, in turn, only holds if the integrand remains finite within the limits of integration. For example, the integral

$$\int_{-1}^1 \frac{dx}{x^2} = = -2$$

cannot therefore be regarded as the limit of a sum because the integrand is infinite at $x = 0$. This is even obvious also because the integrand takes positive values for any x lying within the limits of integation so that the equality above is impossible.

The same remark may be made concerning the integral

$$\int_{-1}^1 \frac{dx}{x} = -\ln(-1).$$

Its value can be represented in a somewhat different way. Since

$$e^{\theta i} = \cos\theta + i\sin\theta,$$

it follows that

$$\theta i = \ln(\cos\theta + i\sin\theta).$$

Setting here $\theta = \pi(2n + 1)$, we have

$$\pi(2n + 1)\, i = \ln(-1), \quad -\int_{-1}^{1} \frac{dx}{x} = \pi i(2n + 1).$$

This integral thus has an infinite set of different values, all of them imaginary. This can be explained by noting that, while integrating, we could have led x through imaginary values; many paths of integration are here possible which indeed explains the indefiniteness of the integral.

Note 1. {I do not reproduce the appended figure (which Chebyshev had not mentioned in his text). The equation of the curve there represented was not specified and it was drawn wrongly: at point $B(0; 1)$ the curve (i) was not perpendicular to the axis Oy.}

1.1.1. We begin with the integral

$$\int_{y=\alpha}^{y=\beta} \int_{x=0}^{x=\infty} e^{-xy}\, dx\, dy.$$

For the integrand to remain finite it is neceesary that $\alpha > 0$ and $\beta > 0$. These conditions will indeed restrict our investigation. We have

$$\int e^{-xy}\, dx = -e^{-xy}/y,$$

therefore

$$\int_{0}^{\infty} e^{-xy}\, dx = 1/y.$$

Then

$$\int_{\alpha}^{\beta} \frac{dy}{y} = \ln(\beta/\alpha) = \int_{\alpha}^{\beta} \int_{0}^{\infty} e^{-xy}\, dx\, dy = \int_{0}^{\infty} dx \int_{\alpha}^{\beta} e^{-xy}\, dy.$$

However,

$$\int_{\alpha}^{\beta} e^{-xy}\, dy = \frac{e^{-\beta x}}{-x} - \frac{e^{-\alpha x}}{-x} = \frac{e^{-\alpha x} - e^{-\beta x}}{x}$$

so that

$$\int_{0}^{\infty} \frac{e^{-\alpha x} - e^{-\beta x}}{x}\, dx = \ln(\beta/\alpha). \qquad\qquad (1)$$

This integral is almost the most important one. We shall indicate one of its applications. Supposing that $\beta = n$ and $\alpha = 1$, we have

$$\int_{0}^{\infty} \frac{e^{-\alpha x} - e^{-\beta x}}{x}\, dx = \ln n.$$

Assuming that n takes differenct values from 1 to $(n - 1)$ and adding up the expressions obtained, we get

$$\ln[(n-1)!] = \int_0^\infty \left(\frac{(n-1)e^{-x}}{x} - \frac{e^{-x} + e^{-2x} + \ldots + e^{-(n-1)x}}{x} \right) dx$$

or

$$\ln[(n-1)!] = \int_0^\infty \left((n-1)e^{-x} - \frac{e^{-nx} - e^{-x}}{e^{-x} - 1} \right) \frac{dx}{x}.$$

This integral thus allows us to express the logarithm of the product of natural numbers which is sometimes useful.

Integral (1) is usually written down in a somewhat different form. Suppose that

$$x = \ln(1/z), \text{ then } dx = -dz/z,$$

$$\int_0^\infty \frac{e^{-\alpha x} - e^{-\beta x}}{x} dx = -\int_1^0 \frac{z^\alpha - z^\beta}{-\ln z} \frac{dz}{z} = \ln \frac{\beta}{\alpha}$$

so that

$$\int_0^1 \frac{z^{\alpha-1} - z^{\beta-1}}{\ln z} dz = \ln(\alpha/\beta). \tag{2}$$

We arrived at this conclusion under the assumption that α and β were positive magnitudes which of course presupposes that they were real numbers. Assuming however that they are imaginary magnitudes, and issuing from the integral (1), we can secure some notion about the value of the integral

$$\int_0^\infty \frac{\sin cx}{x} dx.$$

We say *some notion* because our assumption leads to a non-rigorous conclusion, that, like all suchlike inferences, fails to provide, as we shall see now, the desirable results.

Suppose that $\alpha = -ci$, $\beta = ci$ where c is some real magnitude. Then

$$\int_0^\infty \frac{e^{icx} - e^{-icx}}{x} dx = \ln(-1)$$

and, since

$$[(e^{icx}) - (e^{-icx})/2i] = \sin cx,$$

it follows that

$$\int_0^\infty \frac{\sin cx}{x} 2i \, dx = \ln(-1) = \pi i$$

so that

$$\int_0^\infty \frac{\sin cx}{x} dx = \pi/2. \tag{ii}$$

We have thus determined the value of this integral. However, $\ln(-1)$ is an indefinite magnitude and $\pi(2n + 1)i$ should have been taken instead of πi, and already this circumstance indicates that the result is doubtful because the integral ought to have a single and quite definite value. In addition, the expression obtained does

not depend on c so that the determined value holds only for positive c's whereas, according to the assumption made when deriving this integral, we have only suuposed that c was a real magnitude. Strictly speaking, we have not thus obtained the result sought.

1.1.2. Consider now the integral

$$\int_0^\beta dy \int_0^\infty e^{-xy} \sin cx\, dx.$$

Now

$$\int e^{-xy} \sin cx\, dx = -\frac{e^{-xy}\sin cx}{y} - \int \frac{e^{-xy}}{-y} c\cos cx\, dx =$$

$$-\frac{e^{-xy}\sin cx}{y} + (c/y)\int e^{-xy}\cos cx\, dx.$$

Therefore

$$\int_0^\infty e^{-xy}\sin cx\, dx = (c/y)\int_0^\infty e^{-xy}\cos cx\, dx$$

but

$$\int e^{-xy}\cos cx\, dx = \frac{e^{-xy}\cos cx}{-y} - \int \frac{e^{-xy}}{-y}(-c)\sin cx\, dx =$$

$$(1/y) - (c/y)\int_0^\infty e^{-xy}\sin cx\, dx.$$

Thus

$$\int e^{-xy}\sin cx\, dx = (c/y^2) - (c^2/y^2)\int_0^\infty e^{-xy}\sin cx\, dx$$

and it follows that

$$\int_0^\infty e^{-xy}\sin cx\, dx = \frac{c}{c^2+y^2}, \quad \int_0^\infty e^{-xy}\cos cx\, dx = \frac{y}{c^2+y^2}. \qquad (3;4)$$

These integrals are very important because they enable us to express the algebraic fractions in (3; 4) through definiote integrals which very ofdten simplifies the solution of many problems. Note that the integrals (3; 4) can also be determined in the indefinite form.

Supposing that y in these integrals gradually decreases tending to zero, we have, in the limit,

$$\int_0^\infty \sin cx\, dx = 1/c, \quad \int_0^\infty \cos cx\, dx = 0.$$

It is easy to see, however, that the obtained integrals possess no direct sense because neither $\sin cx$ nor $\cos cx$ approaches any definite limit as x increases to infinity. Taken by themselves, these integrals are therefore indefinite magnitudes. Nevertheless, when considered as the limits of integrals (3; 4), they have quite a definite value as found by us.

$$\int_0^\infty \frac{e^{nui} + e^{-nui}}{2} \cdot \frac{du}{p^2 + u^2} = \frac{\pi}{2p} e^{-np}.$$

Substituting positive numbers n_1, n_2, ... instead of n, we get in the right side factors $\exp(-n_1 p)$, $\exp(-n_2 p)$, ... and, multiplying the obtained integrals by A_1, A_2, ... respectively and adding together the results we shall find

$$(1/2) \int_0^\infty [A_1\exp(-n_1\,ui) + A_2\exp(-n_2\,ui) + \ldots\ A_1\exp(n_1\,ui) +$$

$$A_2\exp(n_2\,ui) + \ldots] \frac{du}{p^2 + u^2} =$$

$$\frac{\pi}{2p}[A_1\exp(-n_1\,ui) + A_2\exp(-n_2\,ui) + \ldots].$$

Suppose now that

$$f(x) = A_1\exp(-n_1 x) + A_2\exp(-n_2 x) + \ldots \qquad\qquad\text{(ix)}$$

and the last integral will become

$$\int_0^\infty \frac{f(ui) + f(-ui)}{2} \cdot \frac{du}{p^2 + u^2} = \frac{\pi}{2p}f(p). \qquad\qquad \textbf{(18)}$$

We have thus derived the required formula. The non-rigor of this derivation consists in that (18) holds only for such functions which may be expanded into the series (ix) where A_1, A_2, ... are some constant coefficients, whereas, having no criteria for distinguishing between functions that may, and may not be expanded into such a series, we consider (18) as though valid for any function.

Assuming

$$f(x) = A_1\exp(-n_1 x/a) + A_2\exp(-n_2 x/a) + \ldots$$

we shall find

$$\int_0^\infty \frac{f(aui) + f(-aui)}{2} \cdot \frac{du}{p^2 + u^2} = \frac{\pi}{2p}f(ap).$$

Differentiating this equality with respect to a, we shall obtain

$$\int_0^\infty \frac{f'(aui) - f'(-aui)}{2} \cdot \frac{ui}{p^2 + u^2} du = \frac{\pi}{2}f'(a p)$$

and therefore

$$\int_0^\infty \frac{f'(aui) - f'(-aui)}{2i} \cdot \frac{u\,du}{p^2 + u^2} = -\frac{\pi}{2}f'(a p).$$

Denoting $f'(x) = \varphi(x)$ we get

$$\int_0^\infty \frac{\varphi(aui) - \varphi(-aui)}{2i} \cdot \frac{udu}{p^2 + u^2} = -\frac{\pi}{2}\varphi(ap). \qquad (19)$$

If now $f(x) = \psi'(x)$, the integral (18) will provide

$$\int_0^\infty \frac{\psi'(aui) + \psi'(-aui)}{2} \frac{du}{p^2 + u^2} = \frac{\pi}{2p}\psi'(ap).$$

Integrating this equality with respect to a between the limits 0 and α we shall find

$$\int_0^\infty \int_0^\alpha \frac{\psi'(aui) + \psi'(-aui)}{2} \frac{du}{p^2 + u^2} = \frac{\pi}{2} \cdot \frac{\psi(\alpha p) - \psi(0)}{p^2}.$$

However,

$$\int_0^\alpha \psi'(aui)\, da = \frac{\psi(\alpha ui) - \psi(0)}{ui}, \quad \int_0^\alpha \psi'(-aui)\, da = \frac{\psi(0) - \psi(-\alpha ui)}{ui}$$

and we obtain

$$\int_0^\infty \frac{\psi(\alpha ui) - \psi(-\alpha ui)}{2ui} \cdot \frac{du}{p^2 + u^2} = \frac{\pi}{2} \cdot \frac{\psi(\alpha p) - \psi(0)}{p^2}$$

or

$$\int_0^\infty \frac{\psi(\alpha ui) - \psi(-\alpha ui)}{2i} \cdot \frac{du}{u(p^2 + u^2)} = \frac{\pi}{2} \cdot \frac{\psi(\alpha p) - \psi(0)}{p^2}. \qquad (20)$$

Integrals (18) – (20) are to be found in a contribution by Abel and one of them is in Bertrand's writing, but the general formula that interests us was first given by Cauchy. We conclude here the study of the integrals of the second group.

1.3. Integrals of the Third Group

1.3.1. The integrands of the integrals that we shall now study contain functions which could at first sight seem to be algebraic whereas in essence they are special transcendental functions. These integrals are of the type

$$\int_0^\infty \frac{x^\lambda}{1 + x^2}\, dx \qquad (\text{x})$$

where λ is any number. We shall therefore begin by saying a few words about algebraic functions in general. We call algebraic only such functions that are the roots of the equation

$$A_0 y^n + A_1 y^{n-1} + A_2 y^{n-2} + \dots + A_{n-1} y + A_n = 0$$

where n is an positive integral number and the coefficients $A_0, A_1, \dots, A_n$ are some integral functions of x. Functions not fitting in with this definition no longer represent algebraic functions so that x^λ that cannot satisfy our equation at all values of λ is a transcendental function, but it becomes algebraic as soon as we assume that λ is a commensurable number.[1]

We shall now indeed consider the integral (x) at any λ. But we shall note first of all that this integral will have a finite value only for values of λ within 1 and -1. Indeed, for $\lambda > 1$ the degree of the product $x \cdot x^\lambda = x^{\lambda+1}$ will be higher than 2 and the integral, in virtue of a known theorem, will be infinite. And for $\lambda < -1$, setting $x = 1/z$, $\lambda = -\mu$ and $\mu > 1$, we have

$$\int_0^\infty \frac{x^\lambda}{1+x^2}\,dx = \int_0^\infty \frac{z^\mu}{1+z^2}\,dz.$$

This means that the integral will {again} be infinite. For $\lambda = \pm 1$ we obtain

$$\int_0^\infty \frac{x}{1+x^2}\,dx = (1/2)\,[\ln(1+x^2)]\Big|_0^\infty = \infty;$$

$$\int_0^\infty \frac{x^{-1}}{1+x^2}\,dx = \int_\infty^0 \frac{(1/z)^{-1}}{1+(1/z^2)}\,[-\frac{dz}{z^2}] = \int_0^\infty \frac{z}{1+z^2}\,dz = \infty.$$

If now $\lambda = \alpha + i\beta$ it will be necessary that α be contained between 1 and -1. Thus, we shall consider the integral (x) assuming that $1 > \lambda > -1$.

Note 1. {Chebyshev's term.}

1.3.2. We have

$$\int_0^\infty \frac{x^\lambda}{1+x^2}\,dx = \int_0^1 \frac{x^\lambda}{1+x^2}\,dx + \int_1^\infty \frac{x^\lambda}{1+x^2}\,dx,$$

but

$$\int_0^1 \frac{x^\lambda}{1+x^2}\,dx = \int_0^1 x^\lambda\,(1+x^2)^{-1}\,dx =$$

$$\int_0^1 x^\lambda\,(1-x^2+x^4-x^6+\ldots)\,dx = \int_0^1 x^\lambda\,dx - \int_0^1 x^{\lambda+2}\,dx +$$

$$\int_0^1 x^{\lambda+4}\,dx - \int_0^1 x^{\lambda+6}\,dx + \ldots$$

In general, however,

$$\int_0^1 x^{\lambda+m}\,dx = [\frac{x^{\lambda+m+1}}{\lambda+m+1}]\Big|_0^1 = \frac{1}{\lambda+m+1}$$

because, under our conditions regarding λ, $(\lambda + m + 1)$ is always positive. If $\lambda = \alpha + \beta i$ then

$$\lim (x^{\lambda+m+1})_{x=0} = \lim (x^{\alpha+m+1}\,x^{\beta i})_{x=0}.$$

Since β can be negative it could seem at first sight that $x^{\beta i}$ can be infinite, but it is not difficult to prove that this factor cannot exceed some boundary. Indeed,

$$x^{\beta i} = e^{\beta i \ln x} = \cos(\beta \ln x) + i \sin(\beta \ln x)$$

and therefore in any case

$$\lim (x^{\lambda+m+1})_{x=0} = 0$$

if only α is contained within the boundaries indicated above. Thus,

$$\int_0^1 \frac{x^\lambda}{1+x^2}\,dx = \frac{1}{\lambda+1} - \frac{1}{\lambda+3} + \frac{1}{\lambda+5} - \ldots$$

The second integral is

$$\int_1^\infty \frac{x^\lambda}{1+x^2}\,dx = \int_1^0 \frac{z^{-\lambda}}{1+(1/z^2)}\cdot[-\frac{dz}{z^2}] = \int_0^1 \frac{z^{-\lambda}}{1+z^2}\,dz.$$

But, because of the above, the last integral is equal to

$$\frac{1}{-\lambda+1} - \frac{1}{-\lambda+3} + \frac{1}{-\lambda+5} - \dots$$

And so

$$\int_0^\infty \frac{x^\lambda}{1+x^2}\,dx = (\frac{1}{1+\lambda} + \frac{1}{1-\lambda}) - (\frac{1}{3+\lambda} + \frac{1}{3-\lambda}) + \dots =$$
$$\frac{2\cdot1}{1^2-\lambda^2} - \frac{2\cdot3}{3^2-\lambda^2} + \frac{2\cdot5}{5^2-\lambda^2} - \dots \qquad (21)$$

We have thus expressed our integral as a series whose sum we still ought to determine.
 The similarity of this expansion with the decomposition of rational fractions into partial fractions at once arrests our attention. Indeed, we can present (21) as

$$\int_0^\infty \frac{x^\lambda}{1+x^2}\,dx = \frac{1}{\lambda+1} - \frac{1}{\lambda-1} + \frac{1}{\lambda+3} - \frac{1}{\lambda-3} + \dots$$

However, any rational fraction $f(x)/F(x)$ where $F(x)$ has no multiple roots can be represented as

$$\frac{f(x)}{F(x)} = \frac{f(x_1)}{F'(x_1)}\cdot\frac{1}{x-x_1} + \frac{f(x_2)}{F'(x_2)}\cdot\frac{1}{x-x_2} + \frac{f(x_3)}{F'(x_3)}\cdot\frac{1}{x-x_3} + \dots$$

so that in our case the roots of $F(x)$ are

$$1,\ 3,\ 5,\ 7,\ \dots,\ -1,\ -3,\ -5,\ -7,\ \dots$$

and $f(x)/F(x) = 1$ at any x equal to one of the roots of $F(x)$. And we shall now determine the function $F(x)$ satisfying these conditions.

1.3.3. Let us consider the function

$$F(x) = \cos(n\arccos x)$$

where n is any integer. It is easy to show that $F(x)$ is an integral function of the n-th degree. Indeed, substituting $\arccos x = \varphi$ and noting that

$$\cos n\varphi = [e^{n\varphi i} + e^{-n\varphi i}]/2$$

we shall find that

$$F(\cos\varphi) = \cos n\varphi = [e^{n\varphi i} + e^{-n\varphi i}]/2.$$

However, $e^{\pm n\varphi i} = \cos n\varphi \pm i\sin n\varphi$ so that

$$F(\cos\varphi) = \{[\cos n\varphi + i\sin n\varphi] + [\cos n\varphi - i\sin n\varphi]\}/2 =$$
$$\{[\cos\varphi + i\sin\varphi]^n + [\cos\varphi - i\sin\varphi]^n\}/2.$$

Thus

$$F(x) = \cos(n\arccos x) = (1/2)[(x + \sqrt{x^2-1})^n + (x - \sqrt{x^2-1})^n]. \quad (22)$$

The terms that include odd degrees of the root will depend on it; however, it is not difficult to see that they will finally cancel out so that we shall obtain for $F(x)$ an integral rational function. Since

$$\sqrt{x^2 - 1} = x - (1/2)\cdot(1/x) - (1/2)\cdot(1/2)\cdot(1/2)\cdot(1/x^3) - \dots$$

we have

$$x - \sqrt{x^2 - 1} = (1/2)\cdot(1/x) + (1/2)\cdot(1/2)\cdot(1/2)\cdot(1/x^3) + \dots$$

and the term $[x - \sqrt{x^2 - 1}]^n$ in (22) will contain only negative powers of x; all such terms will finally cancel out with the terms containing negative powers of $[x - \sqrt{x^2 - 1}]^n$ and only the integral part of this expression will be left. Denoting {here and in the sequel} the entire part {of a function} by E, we can express this result as

$$F(x) = E\{[[x - \sqrt{x^2 - 1}]^n / 2]\}.$$

Contributions concerning the study of such functions are mainly due to Chebyshev so that the expression (22) is also called *Chebyshev polynomial.* At present, such investigations are included in many writings on integral calculus,[1] and, for example, in England they can be found in many courses in integral calculus under the heading *Chebyshev's works*. Zolotarev's studies concern similar issues; more precisely, issues relating not to circular but to elliptic functions. Accordingly, they are much more complicated, but also less important than the works of Chebyshev.

Suppose now that $n = 2m$ where m is an integer. Then

$$F(x) = \cos(2m \arccos x) = A_0 x^{2m} + A_1 x^{2m-1} + + A_2 x^{2m-2} + \dots$$

We shall try to expand the function $1/F(x)$ into partial fractions. To attain this goal we shall find the roots of the equation

$$F(x) = 0 \text{ or } \cos(2m \arccos x) = 0$$

and prove that all of them are different. It is not difficult to see that this equation is satisfied if

$$\arccos x = \frac{(2k+1)\pi}{4m}$$

where k is any integer. Assuming that it takes different values from 0 to $(2m - 1)$, we obtain the following roots of this equation:

$k = 0$, $x_1 = \cos(\pi/4m)$; $k = 1$, $x_2 = \cos(3\pi/4m)$;
$k = 2$, $x_3 = \cos(5\pi/4m)$; ...; $k = \mu$, $x_{\mu+1} = \cos[(2\mu + 1)\pi/4m]$;
$k = 2m - \mu - 1$, $x_{2m-\mu} = \cos[(4m - 2\mu - 1)\pi/4m]$; ...
$k = 2m - 1$, $x_{2m} = \cos[(4m - 1)\pi/4m]$.

Since the equation $F(x) = 0$ {above} is of degree $2m$, it cannot have any other roots. It is seen therefore that all of its roots are different and moreover real. For this reason our expansion will be of the type

$$\frac{1}{F(x)} = \frac{1}{F'(x_1)} \cdot \frac{1}{x - x_1} - \frac{1}{F'(x_2)} \cdot \frac{1}{x - x_2} + \dots = \sum \frac{1}{F'(x_\mu)} \cdot \frac{1}{x - x_\mu}.$$

However,

$$F'(x) = \frac{2m \sin(2m \arccos x)}{\sqrt{1 - x^2}},$$

$$F'(x_\mu) = 2m \frac{\sin\{2m \arccos \cos[(2\mu - 1)\pi / 4m]\}}{\sin[(2\mu - 1)\pi / 4m]} = 2m \frac{\sin[(2\mu - 1)\pi / 2]}{\sin[(2\mu - 1)\pi / 4m]}.$$

But $\sin[\pi(2\mu - 1)/2] = (-1)^{\mu-1}$ so that

$$F'(x_\mu) = (-1)^{\mu-1} \frac{2m}{\sin[(2\mu-1)\pi/4m]}.$$

Thus,

$$\frac{1}{\cos(2m\arccos x)} = \sum \frac{\sin[(2\mu-1)\pi/4m]}{2m(-1)^{\mu-1}} \cdot \frac{1}{x - \cos[(2\mu-1)\pi/4m]} =$$

$$(1/2m) \sum (-1)^{\mu-1} \frac{\sin[(2\mu-1)\pi/4m]}{x - \cos[(2\mu-1)\pi/4m]}$$

where the sums should extent consecutively over $\mu = 1, 2, \ldots, 2m$. We now have the following remarkable formula

$$\frac{2m}{\cos(2m\arccos x)} = \sum (-1)^{\mu-1} \frac{\sin[(2\mu-1)\pi/4m]}{x - \cos[(2\mu-1)\pi/4m]}$$

from which, assuming that $\arccos x = \varphi$ and $m = 1$, we can derive the known formula

$$\cos 2\varphi = 2\cos^2\varphi - 1.$$

For $x = \cos\varphi$ we have

$$\frac{2m}{\cos 2m\varphi} = \sum (-1)^{\mu-1} \frac{\sin[(2\mu-1)\pi/4m]}{\cos\varphi - \cos[(2\mu-1)\pi/4m]}. \qquad (23)$$

We shall compare now two terms of this formula, those where $\mu = k$ and $\mu = (2m + 1 - k)$. Introducing, in general,

$$\Phi(\mu) = (-1)^{\mu-1} \frac{\sin[(2\mu-1)\pi/4m]}{\cos\varphi - \cos[(2\mu-1)\pi/4m]}$$

we have

$$\Phi(k) = (-1)^{k-1} \frac{\sin[(2k-1)\pi/4m]}{\cos\varphi - \cos[(2k-1)\pi/4m]},$$

$$\Phi(2m + 1 - k) = (-1)^{2m-k} \frac{\sin[(4m+1-2k)\pi/4m]}{\cos\varphi - \cos[(4m+1-2k)\pi/4m]} =$$

$$(-1)^k \frac{\sin[(2k-1)\pi/4m]}{\cos\varphi + \cos[(2k-1)\pi/4m]}.$$

It is therefore seen that in our expansion the sum of the terms equally distant from the middle is

$$\Phi(k) + \Phi(2m + 1 - k) = (-1)^{k-1} \sin\frac{(2k-1)\pi}{4m} \cdot$$

$$[\frac{1}{\cos\varphi - \cos[(2k-1)\pi/4m]} - \frac{1}{\cos\varphi + \cos[(2k-1)\pi/4m]}] =$$

$$(-1)^{k-1} \frac{\sin[(2k-1)\pi/2m]}{\cos^2\varphi - \cos^2[(2k-1)\pi/4m]} =$$

$$(-1)^{k-1} \frac{\sin[(2k-1)\pi/2m]}{\sin^2[(2k-1)\pi/4m] - \sin^2\varphi} .$$

Formula (23) thus becomes

$$\frac{2m}{\cos 2m\varphi} = \sum (-1)^{k-1} \frac{\sin[(2k-1)\pi/2m]}{\sin^2[(2k-1)\pi/4m] - \sin^2\varphi} \qquad (23')$$

where k should be changed from 1 to m inclusively.

Suppose now that φ tends to zero and m increases indefinitely in such a way that the product $2m\,\varphi$, which we shall denote by $\pi\lambda/2$, remains finite. We shall try to find the limit of our sum under these conditions. We have $2m = \pi\lambda/2\varphi$, therefore

$$\frac{\lambda\pi/2\varphi}{\cos(\lambda\pi/2)} = \sum (-1)^{k-1} \frac{\sin[(2k-1)2\varphi\pi/\pi\lambda]}{\sin^2[(2k-1)2\pi\varphi/2\lambda\pi] - \sin^2\varphi} =$$

$$\sum (-1)^{k-1} \frac{\sin[2(2k-1)\varphi/\lambda]}{\sin^2[(2k-1)\varphi/\lambda] - \sin^2\varphi} ,$$

$$\frac{\pi/2}{\cos(\lambda\pi/2)} = \sum (-1)^{k-1} \frac{\sin[(2k-1)2\varphi/\lambda]}{(\lambda/\varphi)\sin^2[(2k-1)\varphi/\lambda] - (\lambda/\varphi)\sin^2\varphi} =$$

$$\sum (-1)^{k-1} \frac{\sin[2(2k-1)\varphi/\lambda]}{\lambda\{\sin[(2k-1)\varphi/\lambda]/\varphi\}^2 - [(\sin\varphi)/\varphi]^2} .$$

And, since in general

$$\lim [\sin(N\varphi)/\varphi]_{\varphi=0} = N,$$

$$[\frac{\pi/2}{\cos(\lambda\pi/2)}]_{\varphi=0} = \sum (-1)^{k-1} \frac{2(2k-1)/\lambda}{\lambda[(2k-1)/\lambda]^2 - 1} =$$

$$\sum (-1)^{k-1} \frac{2(2k-1)}{(2k-1)^2 - \lambda^2} .$$

Thus, we find the following expression

$$\frac{\pi/2}{\cos(\lambda\pi/2)} = \sum (-1)^{k-1} \frac{2(2k-1)}{(2k-1)^2 - \lambda^2} \qquad (24)$$

where k should take all the integer values from 1 to ∞. Evidently, since λ is here determined by the condition

$$\lambda = (4m\varphi/\pi)_{\varphi=0, \; m=\infty},$$

it may take all possible values including incommensurable ones.[2] If λ has value of the type $(\alpha + i\beta)$, which is only possible if φ takes imaginary values (because m is a real magnitude), the obtained formula is also valid in this case, because, when determining the limit of the sum at $\varphi = 0$ and $m = \infty$ by means explicated in the differential calculus (the method based on geometric

considerations that we applied in this case { ? } evidently will not   do), we would have arrived at the same expression.

Note 1. Alekseev, N.N. *Интегральное исчисление* (Integral Calculus).
Note 2. {Cf. Note 1 to §1.3.1.}

1.3.4. We may represent the expression (24) in such a way:

$$\frac{\pi/2}{\cos(\lambda\pi/2)} = \frac{2\cdot1}{1-\lambda^2} - \frac{2\cdot3}{3-\lambda^2} + \frac{2\cdot5}{5-\lambda^2} - \dots$$

Supposing now that λ is contained between the boundaries -1 and 1, and comparing this equality with (21), we shall have

$$\int_0^\infty \frac{x^\lambda}{1+x^2}dx = \frac{\pi/2}{\cos(\lambda\pi/2)}, \quad -1 < \lambda < 1. \tag{25}$$

We have thus found the integral sought. It is sometimes expressed in a   different way. Substituting $x = e^z$ we shall obtain

$$\frac{\pi/2}{\cos(\lambda\pi/2)} = \int_{-\infty}^{\infty} \frac{e^{\lambda z}}{e^{-z}+e^z}dz = \int_{-\infty}^{0} \frac{e^{\lambda z}}{e^{-z}+e^z}dz + \int_{0}^{\infty} \frac{e^{\lambda z}}{e^{-z}+e^z}dz.$$

However,

$$\int_{-\infty}^{0} \frac{e^{\lambda z}}{e^{-z}+e^z}dz = \int_{0}^{-\infty} \frac{-e^{-\lambda z}}{e^{-z}+e^z}dz = \int_{0}^{\infty} \frac{e^{-\lambda z}}{e^{-z}+e^z}dz$$

and therefore

$$\int_0^\infty \frac{e^{\lambda z}+e^{-\lambda z}}{e^z+e^{-z}}dz = \frac{\pi/2}{\cos(\lambda\pi/2)}. \tag{26}$$

Assuming now that $\lambda = \beta i$, we shall get

$$\int_0^\infty \frac{e^{z\beta i}+e^{-z\beta i}}{e^z+e^{-z}}dz = \int_0^\infty \frac{2\cos\beta z}{e^{-z}+e^z}dz = \frac{\pi/2}{\cos(\beta\pi i/2)} = \frac{\pi}{e^{\beta\pi/2}+e^{-\beta\pi/2}}.$$

Thus

$$\int_0^\infty \frac{\cos\beta z}{e^{-z}+e^z}dz = \frac{\pi/2}{e^{-\beta\pi/2}+e^{\beta\pi/2}}. \tag{27}$$

Even more remarkable is another modification of the integral (25) to which   we can arrive by substituting $x^2 = z$, $dx = (1/2)\,z^{-1/2}\,dz$:

$$\int_0^\infty \frac{z^{\lambda/2}}{1+z}\cdot(1/2)\,z^{-1/2}dz = \frac{\pi/2}{\cos(\lambda\pi/2)},$$

$$\int_0^\infty \frac{z^{\lambda/2-1/2}}{1+z}dz = \frac{\pi}{\cos(\lambda\pi/2)}.$$

Supposing here that $\lambda/2 - 1/2 = n - 1$, we shall find

$$\int_0^\infty \frac{z^{n-1}}{1+z}\,dz = \frac{\pi}{\cos[(n-1/2)\pi]} = \frac{\pi}{\cos(\pi/2 - \pi n)} = \frac{\pi}{\sin \pi n}.$$

If now $z = x/(1 - x)$ we shall have finally

$$\int_0^1 x^{n-1}(1 - x)^{-n}\,dx = \frac{\pi}{\sin \pi n}, \quad 0 < n < 1. \tag{28}$$

In such a form this integral is a particular case of the Euler integral of the   first kind

$$\int_0^1 x^{\lambda-1}(1 - x)^{\mu-1}\,dx$$

with $\lambda = n$ and $\mu = -n - 1$. We shall now go on to considering such   integrals.

The Euler Integrals {§§1.3.5 – 1.3.14}

1.3.5. An Euler integral of the second kind is the integral

$$\int_0^\infty x^{\upsilon - 1} e^{-x}\,dx$$

which is usually denoted $\Gamma(\upsilon)$ as a function of υ. And, as stated above, the   integral of the first kind is

$$\int_0^1 x^{\lambda-1}(1 - x)^{\mu-1}\,dx.$$

There is no special notation for this integral; however, some authors denote it   by $(\lambda; \mu)$, others by B $(\lambda; \mu)$. We shall adopt the fist symbol, so that,   replacing υ by n, λ by p and μ by q, we have

$$\int_0^1 x^{p-1}(1 - x)^{q-1}\,dx = (p; q), \tag{29}$$

$$\int_0^\infty x^{n - 1} e^{-x}\,dx = \Gamma(n). \tag{30}$$

In order to consider these integrals as limits of some sums it is necessary that   the parameters n, p, q be positive; otherwise, each integrand becomes   infinite at one of its limits. We shall therefore assume that $n > 0$, $p > 0$, $q > 0$. True, the integrand in (29) becomes infinite at each of the limits, and in   (30), at the lower limit, when the parameters being positive are less than 1. However, it can be proved that in these cases the integrals are finite. Indeed,

$$\int_0^\infty x^{n - 1} e^{-x}\,dx = \int_0^1 x^{n - 1} e^{-x}\,dx + \int_1^\infty x^{n - 1} e^{-x}\,dx.$$

Supposing now that n is a positive proper fraction,[1] we shall find that

$$\int_0^1 x^{n - 1} e^{-x}\,dx < \int_0^1 x^{n - 1}\,dx$$

or that

$$\int_0^1 x^{n - 1} e^{-x}\,dx < [x^n/n]_0^1.$$

This integral is thus finite at any positive n whereas

$$\int_1^\infty x^{n - 1} e^{-x}\,dx < \int_1^\infty e^{-x}\,dx < 1/e.$$

As to the Euler integral of the first kind, it will also be finite at the stated values of the parameters p and q. Indeed, we shall soon prove that it is connected wuth the integral of the second kind by the remarkable equation

$$(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}. \tag{31}$$

However, before proving this relation, we shall indicate some properties of the Integrals of the first and the second kind beginning with the latter.

When integrating by parts, we find that

$$\int_0^\infty x^{n-1}e^{-x}\,dx = [\frac{x^n e^{-x}}{n}]_0^\infty + \int_0^\infty \frac{x^n e^{-x}}{n}\,dx = (1/n)\int_0^\infty x^n e^{-x}\,dx,$$

i.e., that

$$\Gamma(n) = (1/n)\,\Gamma(n+1).$$

Therefore,

$$\Gamma(n+1) = n\,\Gamma(n). \tag{32}$$

In the same way

$$\Gamma(n+2) = (n+1)\,\Gamma(n+1),\ \Gamma(n+3) = (n+2)\,\Gamma(n+2),\ \dots$$
$$\Gamma(n+m) = (n+m-1)\,\Gamma(n+m-1).$$

Multiplying these equalities we obtain

$$\Gamma(n+m) = n\,(n+1)\,(n+2)\,\dots\,(n+m-1)\,\Gamma(n).$$

Assuming that $n = 1$ and noting that

$$\Gamma(1) = \int_0^\infty e^{-x}\,dx = 1$$

we find for any integral m the following equality

$$\Gamma(m+1) = m!. \tag{33}$$

It is seen now that $\Gamma(2) = 1 = \Gamma(1)$ so that between 1 and 2 the function Γ ought to have a minimum (the second derivative of $\Gamma(n)$ is always positive). Consider now the integral

$$(p; q) = \int_0^1 x^{p-1}(1-x)^{q-1}\,dx$$

where p is supposed to be any positive number and q, a positive integer greater than 1. Integrating by parts, we find that

$$(p; q) = [(x^p/p)(1-x)^{q-1}]_0^1 + (1/p)\int_0^1 x^p(q-1)(1-x)^{q-2}\,dx$$

so that

$$(p; q) = [(q-1)/p]\,(p+1;\ q-1).$$

Hence

$$(p+1;\ q-1) = [(q-2)/(p+1)]\,(p+2;\ q-2),$$
$$(p+2;\ q-2) = [(q-3)/(p+2)]\,(p+3;\ q-3),$$
$$(p+k;\ q-k) = [(q-k-1)/(p+k)]\,(p+k+1;\ q-k-1)$$

if only we admit the inequality $q - k > 1$.

Multiplying these equalities we obtain

$$(p;\ q)\ =\ \frac{(q-1)(q-2)...(q-k-1)}{p(p+1)...(p+q)}\ (p\ +\ k\ +\ 1;\ q\ -\ k\ -\ 1).$$

Assuming that $q\ -\ k\ =\ 2$ we get

$$(p;\ q)\ =\ \frac{(q-1)!}{p\,(p+1)...(p+q-2)}\ (p\ -\ q\ -\ 1;\ 1).$$

This equality will be valid for any positive p and positive integer q. However,

$$(p\ +\ q\ -\ 1;\ 1)\ =\ \int_0^1\ x^{\,p+q-2}\,dx\ =\ \frac{1}{p+q-1}.$$

Therefore, we obtain, for such values of p and q,

$$(p;\ q)\ =\ \frac{(q-1)!}{p\,(p+1)(p+2)...(p+q-1)}. \tag{34}$$

If we assume now that p is also an integer, then it follows that

$$(p;\ q)\ =\ \frac{(q-1)!(p-1)!}{(p+q-1)!}\ =\ \frac{\Gamma(q)\Gamma(p)}{\Gamma(p+q)}.$$

And so, the equation (31) is proved for integer values of p and q.

Note 1. {Here, and in many cases in the sequel, Chebyshev as though avoids irrational numbers.}

1.3.6. In order to prove the validity of (31) in the general case, we consider the double integral

$$I\ =\ \int_0^\infty\ \int_0^\infty\ x^{\,p-1}e^{-xy}\,y^{\,p+q-1}e^{-y}\,dx\,dy.$$

Integrating first with respect to x, and then to y, and substituting $xy\ =\ z,$ we shall have

$$\int_0^\infty\ x^{\,p-1}e^{-xy}\,dx\ =\ (1/y^{\,p})\int_0^\infty\ z^{\,p-1}e^{-z}\,dz\ =\ =\ \frac{\Gamma(p)}{y^{\,p}},$$

$$I\ =\ \int_0^\infty\ \frac{y^{\,p+q-1}}{y^{\,p}}\ \Gamma(p)\,e^{-y}\,dy\ =\ \Gamma(p)\,\Gamma(q).$$

Integrating now in the other order and assuming that $y\,(x\ +\ 1)\ =\ u$, we find   that

$$\int_0^\infty\ y^{\,p+q-1}e^{-y(x+1)}\,dy\ =\ \frac{1}{(x+1)^{p+q}}\ \int_0^\infty\ u^{\,p+q-1}e^{-u}\,du\ =$$

$$\frac{1}{(x+1)^{p+q}}\,\Gamma(p\ +\ q),$$

$$\int_0^\infty\ \frac{x^{\,p-1}}{(x+1)^{p+q}}\,\Gamma(p\ +\ q)\,dx\ =\ \Gamma(p\ +\ q)\int_0^\infty\ \frac{x^{\,p-1}}{(x+1)^{p+q}}\,dx.$$

Therefore

$$\int_0^\infty\ \frac{x^{\,p-1}}{(x+1)^{p+q}}\,dx\ =\ \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}. \tag{xi}$$

Supposing that $p + q = 1$ and using formula (28) we may incidentally derive the following remarkable formula

$$\Gamma(p)\,\Gamma(1 - p) = \frac{\pi}{\sin p\pi} \qquad\qquad (35)$$

which is very important in the practical sense for compiling tables of the values of the function Γ. It shows that, within the interval $(0;\ 1)$, it is sufficient to calculate the values of Γ only for the argument not exceeding $1/2$; the other values will be determined by formula (35).

We return now to the integral (xi). Assuming that $x = z/(1 - z)$ so that $1/(1 + x) = 1 - z$ and $dx = dz/(1 - z)^2$, we obtain

$$\int_0^1 \frac{x^{p-1}}{(x+1)^{p+q}}\,dx = \int_0^1 \frac{z^{p-1}}{(1-z)^{p-1}}\,(1 - z)^{p+q}\,\frac{dz}{(1-z)^2} =$$

$$\int_0^1 z^{p-1}(1 - z)^{q-1}dz = (p;\ q),$$

hence (31). This equation shows that $(p;\ q)$ is a symmetric function of p and q which can also be proved directly by setting $z = 1 - y$:

$$(p;\ q) = \int_0^1 z^{p-1}(1 - z)^{q-1}dz = -\int_1^0 (1 - y)^{p-1}y^{q-1}dy =$$

$$\int_0^1 y^{q-1}(1 - y)^{p-1}dy = (q;\ p).$$

Thus, the validity of the equation (31) is proved for all values of p and q for which the double integral at the very beginning of this subsection might be considered as the limit of a sum. This last is true for any positive p and q because, when integrating in the first order, we arrive at the product $\Gamma(p)\,\Gamma(q)$, and, according to the above, each of these two factors is finite for all positive values of p and q.

1.3.7. Assuming that $p = 1/2$ in (35), we have

$$\Gamma(1/2) = \sqrt{\pi}. \qquad\qquad \text{(xii)}$$

But, in accord with (9),

$$\int_0^\infty \exp(-x^2)\,dx = (1/2)\,\Gamma(1/2).$$

The same result can also be obtained directly by setting $x^2 = z$. Then, indeed,

$$\int_0^\infty \exp(-x^2)\,dx = \int_0^\infty e^{-z}\,(1/2)\,z^{-\frac{1}{2}}\,dz = (1/2)\int_0^\infty z^{1/2-1}\,e^{-z}\,dz =$$

$$(1/2)\,\Gamma(1/2).$$

1.3.8. A remarkable equation connecting the values of the Γ function can be derived from equation (31). Supposing that $p = q$ we have here

$$(p;\ p) = \frac{\Gamma(p)\Gamma(p)}{\Gamma(2p)}.$$

Transforming the left side of this equality we obtain

$$(p;\ p) = \int_0^1 x^{p-1}(1-x)^{p-1}dx = \int_0^1 [x(1-x)]^{p-1}dx =$$

$$\int_0^1 \{(1/4) - [x-(1/2)]^2\}^{p-1}dx = (1/4)^{p-1}\int_0^1 [1-(2x-1)^2]^{p-1}dx.$$

Setting $2x - 1 = z$ we arrive here at

$$(p;\ p) = \frac{1}{4^{p-1}}\int_{-1}^1 (1-z^2)^{p-1}(1/2)\,dz = \frac{1}{4^{p-1}}\int_0^1 (1-z^2)^{p-1}\,dz$$

because the integrand is an even function. Assuming now that $z^2 = u$, we find that

$$(p;\ p) = \frac{1}{4^{p-1}}\int_0^1 (1-u)^{p-1}(1/2)\,u^{-\frac{1}{2}}\,du =$$

$$\frac{1}{2^{2p-1}}\int_0^1 u^{\frac{1}{2}-1}(1-u)^{p-1}du = \frac{1}{2^{2p-1}}(1/2;\ p).$$

Thus,

$$\frac{\Gamma(p)\Gamma(p)}{\Gamma(2p)} = \frac{1}{2^{2p-1}}\frac{\Gamma(1/2)\Gamma(p)}{\Gamma(p+1/2)}$$

and, on the strength of (xii), we obtain

$$\Gamma(p)\,\Gamma(p+1/2) = \frac{\sqrt{\pi}}{2^{2p-1}}\,\Gamma(2p). \tag{36}$$

This remarkable formula was first discovered by Legendre.

1.3.9. Let us now go over to integrals expressed by the logarithm of gamma. In §1.1.1 we found that for any positive and integer n

$$\ln[(n-1)!] = \int_0^\infty \frac{(n-1)e^{-x} - \dfrac{e^{-x} - e^{-nx}}{1-e^{-x}}}{x}dx$$

and on the strength of formula (33) this is equal to $\ln\Gamma(n)$.

We shall prove that this formula is valid for any values of n for which $\Gamma(n)$ can be taken. We have, in general, formula (30). Differentiating it with respect to n, we obtain

$$d\Gamma(n)/dn = \int_0^\infty e^{-x}x^{n-1}\ln x\,dx.$$

But, in virtue of formula (1), we have

$$\ln x = \int_0^\infty \frac{e^{-z} - e^{-xz}}{z}dz.$$

Therefore

$$d\Gamma(n)/dn = \int_0^\infty e^{-x}x^{n-1}\int_0^\infty \frac{e^{-z} - e^{-xz}}{z}dz\,dx =$$

$$\int_0^\infty \int_0^\infty e^{-x}x^{n-1}\frac{e^{-z} - e^{-xz}}{z}dz\,dx.$$

Integrating here with respect to x, we arrive at

$$\int_0^\infty e^{-x}x^{n-1}(e^{-z} - e^{-xz})dx = e^{-z}\int_0^\infty x^{n-1}e^{-x}dx - \int_0^\infty e^{-x(1+z)}x^{n-1}dx$$

and, setting $x(1+z) = t$, we obtain

$$\int_0^\infty e^{-x(1+z)}x^{n-1}dx = \frac{1}{(1+z)^n}\int_0^\infty e^{-t}t^{n-2}dt = \frac{\Gamma(n)}{(1+z)^n}$$

so that

$$d\,\Gamma(n)\,/\,dn = \int_{\infty}^{0} [e^{-z} - (1 + z)^{-n}]\,\Gamma(n)\,\frac{dz}{z},$$

hence

$$\frac{d\,\Gamma(n)}{dn}\,\frac{1}{\Gamma(n)} = \frac{d\ln\Gamma(n)}{dn} = \int_{0}^{\infty} [e^{-z} - (1 + z)^{-n}]\,\frac{dz}{z}.$$

It follows that

$$\int_{1}^{n} d\ln\Gamma(n) = \ln\Gamma(n) = \int_{0}^{\infty}\int_{1}^{n} [e^{-z} - (1 + z)^{-n}]\,\frac{dz}{z}\,dn,$$

$$\ln\Gamma(n) = \int_{0}^{\infty} (n - 1)\,e^{-z}\,\frac{dz}{z} - \int_{0}^{\infty} [\frac{(1+z)^{-n}}{-\ln(1+z)}]_{1}^{n}\,\frac{dz}{z}.$$

Thus

$$\ln\Gamma(n) = \int_{0}^{\infty} [(n - 1)\,e^{-z} + \frac{(1+z)^{-n} - (1+z)^{-1}}{\ln(1+z)}]\,\frac{dz}{z}.$$

Noting now that $\Gamma(2) = 1$, we have

$$0 = \int_{0}^{\infty} [e^{-z} + \frac{(1+z)^{-2} - (1+z)^{-1}}{\ln(1+z)}]\,\frac{dz}{z}$$

and, consequently,

$$\int_{0}^{\infty} [(n - 1)\,e^{-z} + (n - 1)\,\frac{(1+z)^{-2} - (1+z)^{-1}}{\ln(1+z)}]\,\frac{dz}{z} = 0.$$

Subtracting this equality from the preceding one we shall have

$$\ln\Gamma(n) = \int_{0}^{\infty} \frac{(1+z)^{-n} - (1+z)^{-1} + (n-1)[(1+z)^{-1} - (1+z)^{-2}]}{\ln(1+z)}\,\frac{dz}{z}$$

and, assuming that $\ln(1 + z) = x$, we find that

$$\ln\Gamma(n) = \int_{0}^{\infty} [e^{-nx} - e^{-x} + (n - 1)e^{-x}(1 - e^{-x})]\,\frac{e^{x}}{e^{x} - 1}\,\frac{dx}{x}.$$

Finally, it follows that

$$\int_{0}^{\infty} [(n - 1)\,e^{-x} - \frac{e^{-x} - e^{-nx}}{1 - e^{-x}}]\,\frac{dx}{x} = \ln\Gamma(n). \qquad (37)$$

1.3.10. Replacing n by $(n + 1)$ in (37) we have

$$\ln\Gamma(n + 1) = \int_{0}^{\infty} [ne^{-x} - \frac{e^{-x} - e^{-(n+1)x}}{1 - e^{-x}}]\,\frac{dx}{x}. \qquad (38)$$

This formula is extremely important in mathematics and, especially, in the theory of probability where it is applied for integer and very large values of n. We shall dwell on it also because remarkable corollaries follow from it.

If n is a very large number, it will be more convenient to represent this formula in such a way that its right side consists of two parts, one of them including all the finite and very large terms and the other one serving as a supplement and including only very small terms and vanishing at $n = \infty$. This is necessary for example when calculating $\ln\Gamma(n + 1)$ for very large values of n. We shall indeed derive this logarithm. The integrand in (38) can be represented as

$$[ne^{-x} - \frac{e^{-x} - e^{-(n+1)x}}{1 - e^{-x}}] \frac{1}{x} = [ne^{-x} - \frac{e^{-x}}{1 - e^{-x}} + \frac{e^{-x}}{1 - e^{-x}} e^{-nx}] \frac{1}{x}.$$

But

$$\frac{e^{-x}}{1 - e^{-x}} = \frac{1 - x + x^2/2! - x^3/3! + \dots}{x - x^2/2! + x^3/3! - \dots} = \frac{1}{x} - \frac{1}{2} + \varphi(x)$$

where $\varphi(x)$ is the sum of all the other terms of the expansion. It is not difficult to see that $\varphi(x)$ includes only positive powers of x because its term containing x in the lowest degree is $x/12$, so that $\varphi(x)/x$ will have a finite limit at $x = 0$. We thus have

$$[ne^{-x} - \frac{e^{-x} - e^{-(n+1)x}}{1 - e^{-x}}] \frac{1}{x} = [ne^{-x} - \frac{e^{-x}}{1 - e^{-x}} + (\frac{1}{x} - \frac{1}{2}) e^{-nx}] \frac{1}{x} +$$

$$\frac{\varphi(x)}{x} e^{-nx}.$$

However,

$$\varphi(x) = \frac{e^{-x}}{1 - e^{-x}} + \frac{1}{2} \quad \frac{1}{x} \quad \frac{2e^{-x} + 1 - e^{-x}}{2(1 - e^{-x})} - \frac{1}{x} = \frac{1}{2}[\frac{1 + e^{-x}}{1 - e^{-x}} - \frac{2}{x}]$$

and therefore

$$\ln \Gamma(n + 1) = \int_0^\infty [ne^{-x} - \frac{e^{-x}}{1 - e^{-x}} + (\frac{1}{x} - \frac{1}{2}) e^{-nx}] \frac{dx}{x} +$$

$$(1/2) \int_0^\infty [\frac{1 + e^{-x}}{1 - e^{-x}} - \frac{2}{x}] e^{-nx} \frac{dx}{x}.$$

Or, supposing that the two integrals are denoted by $F(n)$ and $\Psi(n)$ respectively, we obtain

$$\ln \Gamma(n + 1) = F(n) + \Psi(n).$$

We thus reduced our expression to the desired form: the function $\Psi(n)$ is represented by an integral of a function taking finite and very small values at all values of x contained within and including the limits of integration, – and, moreover, taking very small values at very large values of n. This function has exactly those properties which, according to our assumption, should be possessed by the second part of the expression sought.

It is only left to determine the type of the functions $F(n)$ and $\Psi(n)$. To find the first of them we differentiate this {yet unknown} function with respect to n:

$$F'(n) = \int_0^\infty [e^{-x} - x(\frac{1}{x} - \frac{1}{2}) e^{-nx}] \frac{dx}{x} =$$

$$\int_0^\infty [e^{-x} - e^{-nx} + (x/2) e^{-nx}] \frac{dx}{x}.$$

It follows that

$$F'(n) = \int_0^\infty \frac{e^{-x} - e^{-nx}}{x} dx + (1/2) \int_0^\infty e^{-nx} dx = \ln n + (1/2) \cdot (1/n)$$

and

$$F(n) = \int \ln n \, dn + (1/2) \int \frac{dn}{n} = n \ln n - n + (1/2) \ln n + C \qquad \text{(xiii)}$$

where C is a constant that for the time being we leave indefinite. Thus,

$$F(n) = C + n \ln n - n + (1/2) \ln n.$$

It ought to be noted here that the differentiation applied by us destroyed the constant C in whose determination all the difficulty really consists. We were able to calculate the function $F(n)$ in such an indefinite form without any trouble exactly because of this differentiation.

Let us now investigate the function $\Psi(n)$. Here, we cannot carry the integration to its conclusion and have to express $\Psi(n)$ by a series. We have

$$\frac{1+e^{-x}}{1-e^{-x}} = \frac{e^{x/2}+e^{-x/2}}{e^{x/2}-e^{-x/2}}$$

so that the integrand becomes

$$(1/2)\,[\frac{e^{x/2}+e^{-x/2}}{e^{x/2}-e^{-x/2}} - \frac{2}{x}]\,\frac{1}{x}\,e^{-nx}.$$

The product of the first two factors is here an even function {see below} so that it is possible to say at once that its expansion into powers of x will be of the form

$$A_1 + A_2 x^2 + A_3\,x^4 + \dots$$

We have

$$e^{\pm x/2} = 1 \pm x/2 + (1/2!)\,(x/2)^2 \pm (1/3!)\,(x/2)^3 + \dots,$$
$$e^{x/2} + e^{-x/2} = 2[1 + (1/2!)\,(x/2)^2 + \dots]$$
$$e^{x/2} - e^{-x/2} = 2[(x/2) + (1/3!)\,(x/2)^3 + \dots]$$

and

$$\frac{e^{x/2}+e^{-x/2}}{e^{x/2}-e^{-x/2}} = \frac{1+x^2/8+\dots}{x/2+x^2/48+\dots} = \frac{2}{x} + \frac{x}{6} + \dots$$

Therefore

$$(1/2)\,[\frac{1+e^{-x}}{1-e^{-x}} - \frac{2}{x}] = (1/2)\,[\frac{2}{x} + \frac{x}{6} + \dots - \frac{2}{x}] = \frac{x}{12} + \dots$$

Thus, $A_1 = 1/12$. Taking into account a larger number of terms in the numerator and the denominator and carrying out the division, we shall obtain, in the same way, A_2, A_3, …The numbers that express these coefficients are very closely related to the Bernoulli numbers; at present, very many mathematicians are studying them.

So, we have

$$\Psi(n) = \int_0^{\infty} [A_1 x + A_2\,x^3 + A_3 x^5 + \dots]\,e^{-nx}\,\frac{dx}{x} =$$

$$A_1\int_0^{\infty} e^{-nx}\,dx + A_2\int_0^{\infty} x^2 e^{-nx}\,dx + A_3\int_0^{\infty} x^4 e^{-nx}\,dx + \dots$$

Noting that, in general,

$$\int_0^{\infty} x^{l-1} e^{-nx}\,dx = \int_0^{\infty} \frac{z^{l-1}}{n^{l-1}}e^{-nx}\,dx/n = (1/n^l)\int_0^{\infty} z^{l-1} e^{-z}\,dz = \frac{\Gamma(l)}{n^l}$$

and that, for any integer and positive values of l, $\Gamma(l) = (l-1)!$, we obtain

$$\Psi(n) = (1!/n)\,A_1 + (2!/n^3)\,A_2 + (4!/n^5)\,A_3 + \dots \tag{39}$$

We have thus expressed $\Psi(n)$ by a series depending on the coefficients A_1, A_2, A_3, ... Therefore, to say something general about this equality, it is necessary to know the law of their composition. However, the method of their determination applied above does not provide this law and we shall derive{other} expressions for these coefficients, awkward for calculation but very convenient for our present purpose.

1.3.11. Setting $x/2 = \theta i$ we obtain

$$\frac{1+e^{-x}}{1-e^{-x}} = \frac{e^{x/2}+e^{-x/2}}{e^{x/2}-e^{-x/2}} = \frac{e^{\theta i}+e^{-\theta i}}{e^{\theta i}-e^{-\theta i}} = \frac{2\cos\theta}{2i\sin\theta} = \frac{\operatorname{ctg}\theta}{i}.$$

But, in general,

$$\sin\theta = \theta\,[1 - (\theta^2/\pi^2)]\,\{1 - [\theta^2/(2\pi)^2]\}\,\{1 - [\theta^2/(3\pi)^2]\}\,\ldots$$

so that

$$\ln\sin\theta = \ln\theta + \ln[1 - (\theta^2/\pi^2)] + \ln\{1 - [\theta^2/(2\pi)^2]\} + \ln\{1 - [\theta^2/(3\pi)^2]\} + \ldots$$

Differentiating this equality we have

$$\operatorname{ctg}\theta = \frac{1}{\theta} - \frac{2\theta}{\pi^2-\theta^2} - \frac{2\theta}{(2\pi)^2-\theta^2} - \frac{2\theta}{(3\pi)^2-\theta^2} - \ldots,$$

$$(1/2)\,[\operatorname{ctg}\theta - (1/\theta)] = \frac{\theta}{\pi^2-\theta^2} - \frac{\theta}{(2\pi)^2-\theta^2} - \frac{\theta}{(3\pi)^2-\theta^2} - \ldots$$

However, since in general

$$[1/(1-\alpha)] = 1 + \alpha + \alpha^2 + \alpha^3 + \ldots + \alpha^l + \frac{\alpha^{l+1}}{1-\alpha}$$

and

$$\frac{1}{(\pi m)^2-\theta^2} = \frac{1}{(\pi m)^2}\frac{1}{1-(\theta/\pi m)^2},$$

it follows that

$$\frac{(\pi m)^2}{(\pi m)^2-\theta^2} = 1 + \frac{\theta^2}{(\pi m)^2} + \frac{\theta^4}{(\pi m)^4} + \ldots + \frac{\theta^{2l}}{(\pi m)^{2l}} +$$
$$\frac{\theta^{2l+2}}{(\pi m)^{2l+2}}\frac{(\pi m)^2}{(\pi m)^2-\theta^2}.$$

Therefore

$$\frac{\theta}{(\pi m)^2-\theta^2} = \frac{\theta}{(\pi m)^2} + \frac{\theta^3}{(\pi m)^4} + \frac{\theta^5}{(\pi m)^6} + \ldots + \frac{\theta^{2l+1}}{(\pi m)^{2l+2}} +$$
$$\frac{\theta^{2l+3}}{(\pi m)^{2l+4}}\frac{1}{(\pi m)^2-\theta^2}.$$

Assuming that m takes here different values beginning with 1 and adding the thus obtained equalities we shall have

$$\frac{\theta}{\pi^2-\theta^2} + \frac{\theta}{(2\pi)^2-\theta^2} + \frac{\theta}{(3\pi)^2-\theta^2} + \ldots = (\theta/\pi^2)\,[1 + (1/2^2) +$$
$$(1/3^2) + \ldots] + (\theta^3/\pi^4)\,[1 + (1/2^4) + (1/3^4) + \ldots] + \ldots +$$
$$(\theta^{2l+1}/\pi^{2l+2})\,[1 + (1/2^{2l+2}) + (1/3^{2l+2}) + \ldots]$$

$$\frac{\theta^{2l+3}}{\pi^{2l+4}} \left\{ \frac{1}{\pi^2 - \theta^2} + \frac{1}{2^{2l+4}[(2\pi)^2 - \theta^2]} + \frac{1}{3^{2l+4}[(3\pi)^2 - \theta^2]} + \ldots \right\} =$$

$(1/2)\, [\text{ctg}\ \theta - (1/\theta)].$

But $\theta = x/2i = -i\, x/2$ so that

$$(1/2)\, \left[\frac{e^{x/2} + e^{-x/2}}{e^{x/2} - e^{-x/2}} - \frac{2}{x} \right] i = -(S_2/\pi^2)\,(-x\, i/2) - (S_4/\pi^4)\,(-x\, i/2)^3 -$$

$$(S_6/\pi^6)\,(-x\, i/2)^5 - \ldots - \frac{1}{\pi^{2l+2}} \left\{ \frac{1}{\pi^2 + x^2/4} + \frac{1}{2^{l+2}[(2\pi)^2 + x^2/4]} + \ldots \right\} \cdot$$

$[(-x\, i/2)^{2l+3}]$

where the notation introduced is obvious. Hence

$$(1/2)\, \left[\frac{e^{x/2} + e^{-x/2}}{e^{x/2} - e^{-x/2}} - \frac{2}{x} \right] = (S_2/2\pi^2)\, x - (S_4/2^3\pi^4)x^3 +$$

$$(S_6/2^5\pi^6)\, x^5 - \ldots + (-1)^l\, \frac{S_{2l+2}}{2^{2l+1}\pi^{2l+2}}\, x^{2l+1} +$$

$$(-1)^{l+1}\, \frac{1}{2^{2l+3}\pi^{2l+2}} \left[\frac{1}{\pi^2 + x^2/4} + \frac{1}{2^{2l+2}[2\pi^2 + x^2/4]} + \ldots \right] x^{2l+3}.$$

It is seen now that

$A_1 = (1/2\pi^2)\, S_2 = (1/2\pi^2)\, [1 + (1/2^2) + (1/3^2) + \ldots],$
$A_2 = -(1/2^3\pi^4)\, S_4 = (1/2^3\pi^4)\, [1 + (1/2^4) + (1/3^4) + \ldots],$
$A_3 = (1/2^5\pi^6)\, S_6 = (1/2^5\pi^6)\, [1 + (1/2^6) + (1/3^6) + \ldots],$

and in general, that

$$A_{l+1} = \frac{(-1)^l}{2^{2l+1}\pi^{2l+2}}\, S_{2l+2} = \frac{(-1)^l}{2^{2l+1}\pi^{2l+2}}\, \left[1 + \frac{1}{2^{2l+2}} + \frac{1}{3^{2l+2}} + \ldots \right].$$

We thus have

$$\Psi(n) = \int_0^\infty \left\{ A_1 x + A_2 x^3 + A_3 x^5 + A_{l+1}\, x^{2l+1} - \frac{(-1)^l}{2^{2l+3}\pi^{2l+2}} \cdot \right.$$

$$\left. \left[\frac{1}{\pi^2 + x^2/4} + \frac{1}{2^{2l+2}[(2\pi)^2 + x^2/4]} + \ldots \right] x^{2\, l+3} \right\} e^{-nx}\, (dx/x)$$

and, taking into account the equality (39),

$$\Psi(n) = (A_1/n) + A_2\, [\Gamma(3)\, /\, n^3]\ A_3\, [\Gamma(5)\, /\, n^5] + \ldots + A_{l+1} \frac{\Gamma(2l+1)}{n^{2l+1}} -$$

$$\frac{(-1)^l}{2^{2l+3}\pi^{2l+2}} \int_0^\infty \left[\frac{1}{\pi^2 + x^2/4} + \ldots \right] e^{-nx}\, x^{2l+2}\, dx.$$

And so, we have expressed $\Psi(n)$ by a series whose terms are alternatively positive and negative and the sign of whose additional term {remainder} is opposite to that of its preceding term. Such series are called *limitative*. They possess a remarkable property: their additional term is always numerically less than the term subsequent to the last one considered. Indeed, let us take the series

$X = u_1 - u_2 + u_3 - \ldots \pm u_n \mp R_n),$
$X = u_1 - R_1;$ or, $u_1 - u_2 + R_2;$ or, $u_1 - u_2 + u_3 - R_3,$ etc.

It is seen that

$R_1 < u_2;\ R_2 < u_3;\ \ldots;\ R_n < u_{n+1}.$

Therefore, such series always enable us to determine the error taking place when calculating their {approximate} sum and this is their advantage over other series. However, they also suffer from a serious shortcoming: they provide no clue for determining when should we stop in order to obtain the most precise value of the sum. Indeed, in such series each term can be {numerically} either greater or less than its preceding term. When breaking off at some term $\pm u_k$, we will note that in one case the addition of one more term, $\mp u_{k+1}$), can lower, and in another case it can heighten the degree of approximation. In general, as it is seen from what was said about these series, we should stop at such a term u_l whose subsequent neighbor $- u_{l+1}$ is least; the sum, calculated in such a way, will be the most precise and differ from the real sum less than by u_{l+1}.

Formula (39) shows that at first the terms of the series expressing $\Psi(n)$ will decrease, but that, beginning with some term, they will increase. Therefore, in order to determine, in this case, the term at which we should break off, we ought to have $u_l > u_{l-1}$ or $u_l / u_{l+1} > 1$ and to determine the least value of l satisfying this inequality. We shall show how to acquire some notion about the least boundary of l by means of this inequality.

We have

$$u_l = A_l\, \Gamma(2l - 1)\, /n^{\,2l-1},\ \ u_{l-1} = A_{l-1}\, \Gamma(2l - 3)\, /n^{\,2l-3},$$

$$\frac{u_l}{u_{l-1}} = \frac{A_l\, \Gamma(2l-1)n^{2l-3}}{A_{l-1}\, \Gamma(2l-3)n^{2l-1}} = (A_l/A_{l-1})\,(2l - 2)\,(2l - 3)\,(1/n^2)$$

where only the numerical values of A_l and A_{l-1} are taken into account. We shall thus have a conjectural inequality

$$(A_l/A_{l-1})\,(2l - 2)\,(2l - 3) > n^2$$

or

$$(1/2^2\pi^2)\ \frac{1+(1/2^{2l})+(1/3^{2l})+\ldots}{1+(1/2^{2l-2})+(1/3^{2l-2})+\ldots}\ (2l - 2)\,(2l - 3) > n^2. \qquad (40)$$

Now we may conclude that inequality

$$(2l - 2)\,(2l - 3) > 2^2\pi^2 n^2$$

also exists so that $(2l)^2 > (2\pi n)^2$ and

$$2l > 2\pi n,\ l > \pi n. \qquad\qquad\qquad\qquad\qquad\qquad (41)$$

Thus we see that l should not be less than πn. However, the method of obtaining this lower boundary does not provide the possibility of seeing whether, for $l > \pi n$, u_l will really be the *least*, having the least value satisfying this inequality. We went over from inequality (40) to (41), but, evidently, we had no right to go in the other direction. It is therefore seen that πn provides only an approximate notion about how great should l be.

We have remarked (§1.3.10) that the coefficients A_1, A_2, … are very closely related to the Bernoulli numbers B_1, B_2, … The dependence between them is expressed by the equation

$$A_l = (-1)^{l-1}\, \frac{B_l}{(2l)!}$$

so that

$$A_1 = B_1/2!,\ \ A_2 = -B_2/4!,\ \ A_3 = B_3/6!,\ \text{etc.}$$

The Bernoulli numbers play a rather important part in mathematics, and, accordingly, very many treatises are expressly devoted to studying their properties.

1.3.12. Let us now go over to the determination of the constant C in the expression (xiii). Taking the logarithms of both sides of the Legendre formula (36)

$$\Gamma(n)\,\Gamma[n + (1/2)] = (\sqrt{\pi} / 2^{2n-1})\,\Gamma(2n)$$

we have

$$\ln\Gamma(n) + \ln\Gamma[n + (1/2)] = \ln\sqrt{\pi} + \ln\Gamma(2n) - (2n - 1)\ln 2.$$

However,

$$\ln\Gamma(n + 1) = C + n\ln n - n + (1/2)\ln n + \Psi(n)$$

and therefore

$$\ln\Gamma(n) = C + (n - 1)\ln(n - 1) - n + 1 + (1/2)\ln(n - 1) + \Psi(n - 1),$$

$$\ln\Gamma[n + (1/2)] = C + [n - (1/2)]\ln[(n - (1/2)] - n + (1/2) +$$
$$(1/2)\ln[n - (1/2)] + \Psi[n - (1/2)],$$

$$\ln\Gamma(2n) = C + (2n - 1)\ln(2n - 1) - 2n + 1 +$$
$$(1/2)\ln(2n - 1) + \Psi(2n - 1).$$

We thus obtain such an equation:

$$2C + [n - (1/2)]\ln(n - 1) + n\ln[n - (1/2)] - 2n + (3/2) + \Psi(n - 1) +$$
$$\Psi(n - 1/2) = \ln\sqrt{\pi} + C + [2n - (1/2)]\ln(2n - 1) + 1 - 2n +$$
$$\Psi(2n - 1) - (2n - 1)\ln 2$$

so that

$$C = \ln\sqrt{\pi} + [2n - (1/2)]\ln(2n - 1) - [n - (1/2)]\ln(n - 1) -$$
$$n\ln[n - (1/2)] - (1/2) - (2n - 1)\ln 2 - \Psi(n - 1)$$
$$\Psi[n - (1/2)] + \Psi(2n - 1).$$

However, in general we have

$$\ln(x - 1) = \ln\{x[1 - (1/x)]\} = \ln x + \ln[1 - (1/x)] = \ln x -$$
$$(1/x) - (1/2)(1/x^2) - \ldots$$

and therefore

$$\ln(2n - 1) = \ln 2n - (1/2n) - (1/8n^2) - \ldots$$
$$\ln(n - 1) = \ln n - (1/n) - (1/2n^2) - \ldots,$$
$$\ln[n - (1/2)] = \ln(2n - 1) - \ln 2 = \ln n - (1/2n) - (1/8n^2) - \ldots$$

And so

$$C = \ln\sqrt{\pi} + [2n - (1/2)][\ln n + \ln 2 - (1/2n) - (1/8n^2)] - [n - (1/2)]\cdot$$
$$[\ln n - (1/n) - (1/2n^2) - \ldots] - n[\ln n - (1/2n) - (1/8n^2)] - (1/2) -$$
$$(2n - 1)\ln 2 + \Psi(2n - 1) - \Psi(n - 1) - \Psi[n - (1/2)]$$

or

$$C = \ln\sqrt{\pi} + (1/2)\ln 2 - \{[2n - (1/2)]/2n\} + \{[n - (1/2)]/n\} -$$

$\{[(2n - (1/2)]/8n^2\} + \{[n - (1/2)/]2n^2\} + (1/8n) + \ldots + \Psi(2n - 1) - \Psi(n - 1) - \Psi[n - (1/2)]$.

Assuming now that in this equation or identity (it should hold for all values of n) $n = \infty$, and noting that $\Psi(\infty) = 0$, we find that

$$C = \ln\sqrt{\pi} + (1/2) \ln 2 = \ln\sqrt{2\pi}.$$

Thus

$$\ln \Gamma(n + 1) = \ln\sqrt{2\pi} + n \ln n + (1/2) \ln n - n + \Psi(n) \qquad \text{(xiv)}$$

so that

$$\Gamma(n + 1) = \sqrt{2\pi}\, n^{n+1/2}\, e^{-n+\Psi(n)}.$$

But $\quad\cdot$

$$e^{\Psi(n)} = e^{(1/12n) + \ldots} = 1 + (1/12n) + \ldots$$

and we thus finally obtain

$$\Gamma(n + 1) = \sqrt{2\pi}\, n^{n+1/2}\, e^{-n} [1 + (1/12n) + \ldots]. \qquad \textbf{(42)}$$

If n is an integer we shall get

$$n! = \sqrt{2\pi}\, n^{n+1/2}\, e^{-n} [1 + (1/12n) + \ldots].$$

This formula that enables us to calculate the approximate value of the product of natural numbers is due to Stirling.[1]

Note 1. {In 1730, De Moivre derived this formula independently from and simultaneously with Stirling; the latter only communicated to him the value of the constant. De Moivre also published a table of $\lg n!$ for $n = 10 (10) 900$ with 14 decimals of which 11 or 12 were correct.}

1.3.13. In concluding the section on the Euler integrals we shall derive the famous Gauss equation connecting {various values of} gamma. We may always suppose that

$$\Gamma(n)\, \Gamma[n + (1/m)]\, \Gamma[n + (2/m)] \ldots \Gamma[n + (m - 1)/m] = F(n; m)\, \Gamma(nm).$$

Let us now try to determine the function $F(n; m)$. We have

$$\Gamma(a)\, \Gamma[a + (1/m)]\, \Gamma[a + (2/m)] \ldots \Gamma[a + (m - 1)/m] / \Gamma(am) = F(a; m).$$

Substituting here $(a + 1)$ instead of a and recalling equality (32) we shall find that

$$\frac{\prod_{i=0}^{m-1}(a + i/m)\Gamma(a + i/m)}{\Gamma(am + m)} = F(a + 1;\ m)$$

or

$$\frac{a\,(a + 1/m)(a + 2/m)\ldots[a + (m-1)/m]}{(ma + m - 1)\,(ma + m - 2)\ldots ma\Gamma(am)}\cdot$$

$$\Gamma(a)\, \Gamma[a + (1/m)] \ldots \Gamma\{[a + (m - 1)/m]\} = F(a + 1;\ m),$$

$$\frac{ma\,(ma+1)\,(ma+2)...(ma+m-1)}{m^m(ma+m-1)\,(ma+m-2)...ma}\,F\,(a;\,m)\;=\;F\,(a\,+\,1;\,m),$$

$$\frac{F(a+1;m)}{F(a;m)}\;=\;m^{-m}\;=\;\frac{m^{-(a+1)m}}{m^{-am}}$$

and consequently

$$\frac{F(a+1;m)}{m^{-(a+1)m}}\;=\;\frac{F(a;m)}{m^{-a\,m}}.$$

Thus, assuming that in general $m^{km}\,F\,(k;\;m)\;=\;\theta\,(k),$ we have

$\theta\,(a)\;=\;\theta\,(a\,+\,1)$ so that $\theta\,(a)\;=\;\theta\,(\infty).$ Therefore

$$\frac{\Gamma(a)\Gamma(a+1/m)\Gamma(a+2/m)...\Gamma[a+(m-1)/m]}{\Gamma(am)}\;=\;m^{-a\,m}\lim\,[\theta\,(x)]_{\,x\,=\infty}$$

where

$$\theta\,(x)\;=\;\frac{\Gamma(x)\Gamma(x+1/m)\Gamma(x+2/m)...\Gamma[x+(m-1)/m]}{\Gamma(mx)m^{-m\,x}}.$$

Therefore

$\ln\,\theta\,(x)\;=\;\ln\,\Gamma(x)\;+\;\ln\,\Gamma\,[x\,+\,(1/m)]\;+\;...\;+\;\ln\,\Gamma\{x\,+\,[(m\,-\,1)/m]\}\;+$
$\qquad\qquad m\,x\,\ln\,m\;-\;\ln\,\Gamma(m\,x).$

But because of (xiv)

$\ln\,\theta\,(x)\;=\;\ln\sqrt{2\pi}\;+\;[x\,-\,(1/2)]\,\ln\,(x\,-\,1)\;-\;x\,+\,1\;+\;\Psi\,(x\,-\,1)\;+$
$\ln\sqrt{2\pi}\;+\;[x\,-\,(1/2)\,+\,(1/m)]\,\ln[x\,-\,1\,+\,(1/m)]\;-\;x\,+\,1\,-\,(1/m)\;+$
$\Psi\,[x\,-\,1\,+\,(1/m)]\;+\;\ln\sqrt{2\pi}\;+\;[x\,-\,(1/2)\,+\,(2/m)]\,\ln[x\,-\,1\,+\,(2/m)]$
$-\,x\,-\,1\,-\,(2/m)\;+\;\Psi\,[x\,-\,1\,+\,(2/m)]\;+\;...\;+\;\ln\sqrt{2\pi}\;+$
$\{x\,-\,(1/2)\,+\,[(m\,-\,1)/m]\}\,\ln\{x\,-\,1\,+\,[(m-1)/m]\}\;-\;x\,+\,1\,-$
$[(m\,-\,1)/m]\;+\;\Psi\,\{x\,-\,1\,+\,[(m\,-\,1)/m)]\}\;-\;\ln\sqrt{2\pi}\;-$
$[mx\,-\,(1/2)]\,\ln\,(mx\,-\,1)\;+\;mx\,-\,1\;+\;mx\,\ln\,m\;-\;\Psi\,(mx\,-\,1).$

Noting that in general

$\ln\,(x\,-\,r)\;=\;\ln\,[x\,(1\,-\,(r/x)]\;=\;\ln\,x\;+\;\ln[1\,-\,(r/x)]\;=\;\ln\,x\;-\;r/x\;-\;...$

we find that

$\ln\,\theta\,(x)\;=\;(m\,-\,1)\,\ln\sqrt{2\pi}\;+\;[x\,-\,(1/2)]\,[\ln\,x\,-\,(1/x)\,+\,...]\;+$
$\Psi\,(x\,-\,1)\;+\;[x\,-\,(1/2)\,+\,(1/m)]\,\{\ln\,x\,-\,[1\,-\,(1/m)]/x\}\;+$
$\Psi\,[x\,-\,1\,+\,(1/m)]\;+\;[x\,-\,(1/2)\,+\,(2/m)]\,\{\ln\,x\,-\,[1\,-\,(2/m)]/x\}\;+$
$\Psi\,[x\,-\,1\,+\,(2/m)]\;+\;...\;+\;\{x\,-\,(1/2)\,+\,[(m\,-\,1)/m]\}\cdot$
$[\ln\,x\,-\,\dfrac{1-(m-1)/m}{x}]\;+\;\Psi\,\{x\,-\,1\,+\,[(m\,-\,1)/m]\}\;+\;m\,x\,\ln\,m\;+$
$(m\,-\,1)\;-\;\{[(1\,+\,2\,+\,...\,+\,(m\,-\,1)]/m\}\;-$
$[m\,x\,-\,(1/2)]\,[\ln\,mx\,-\,(1/mx)\,-\,...]\;-\;\Psi\,(mx\,-\,1)$

and

$$\ln \theta (x) = (m - 1) \ln \sqrt{2\pi} + mx \ln x + mx \ln m - mx \ln mx + m -$$
$$1 + 1 + (1/2) \ln mx - [1 + 2 + \ldots + (m - 1)]/m - (1/2) m \ln x -$$
$$m + [1 + 2 + \ldots + (m - 1)]/m + \frac{1 + 2 + \ldots + (m-1)}{m} \ln x + \ldots +$$
$$\Psi (x - 1) + \Psi [x - 1 + (1/m)] + \ldots + \Psi \{x - 1 + [(m - 1)/m]\} -$$
$$\Psi (mx - 1).$$

Therefore

$$\ln \theta (x) = (m - 1) \ln \sqrt{2\pi} + (1/2) \ln m + \ldots$$

We have retained only two terms because all the rest of them vanish at $x = \infty$.　　　And so

$$\lim [\,\theta (x)\,]_{x=\infty} = (m - 1) \ln \sqrt{2\pi} + \ln \sqrt{m}$$

which means that

$$\lim [\ln \theta (x)]_{x=\infty} = \sqrt{m} \, (2\pi)^{(m-1)/2}; \quad F(a;\ m) = m^{-am+1/2} (2\pi)^{(m-1)/2}$$

and

$$\frac{\Gamma(a)\Gamma(a+1/m)\Gamma(a+2/m)\ldots\Gamma[a+(m-1)/m]}{\Gamma(am)} = m^{-a\,m+1/2} (2\pi)^{(m-1)/2}.$$

Thus we have

$$\Gamma(\lambda) \, \Gamma[\lambda + (1/m)] \ldots \Gamma\{\lambda + [(m - 1)/m]\} = m^{-\lambda m-1/2} (2\pi)^{(m-1)/2} \Gamma(m\lambda). \qquad (\mathbf{43})$$

This equation which is valid for any λ and any integral and positive m is a generalization of equation (36). Assuming here that $\lambda = 1$ and supposing that　　　n is an integral positive number, we obtain

$$\Gamma[1 + (1/n)] \, \Gamma[1 + (2/n)] \ldots \Gamma\{1 + [(n - 1)/n]\} = n^{-n+1/2} (2\pi)^{(n-1)/2} \Gamma(n).$$

It follows that
$$\frac{(n-1)!}{n^{n-1}} \Gamma(1/n) \, \Gamma(2/n) \ldots \Gamma[(n - 1)/n] = n^{-n+1/2} (2\pi)^{(n-1)/2} \Gamma(n)$$
and in virtue of the equality (33)

$$\Gamma(1/n) \, \Gamma(2/n) \ldots \Gamma[(n - 1)/n] = (2\pi)^{(n-1)/2} n^{-1/2}. \qquad\qquad (\mathbf{44})$$

1.3.14. We defined $\Gamma(\lambda)$ as the value of the definite integral {see (30)}

$$\int\limits_0^\infty x^{\lambda-1} e^{-x} dx.$$

Accordingly, we have necessarily considered the gamma function only for such values　　　of λ for which this integral had sense, i.e., for which it was the limit of some sum. We　saw that these values were contained within 0 and ∞. There exists, however, another　more general definition of the gamma {function}. The values of this function are extended onto such cases where the variable also takes negative values. Exactly this　　　last fact gave the occasion for the new definition and in order to go over to it we note　　　that for any integral and positive λ

$$\Gamma(\lambda) = \frac{(n-1)!\Gamma(\lambda)\,n^\lambda}{(n+\lambda+1)!}\Big|_{n=\infty}.$$

This equality might however be expressed in the forms

$$\Gamma(\lambda) = \frac{\Gamma(n)\Gamma(\lambda) \, n^{\lambda}}{\Gamma(n+\lambda)} \bigg|_{n=\infty} = \frac{(n-1)! \, n^{\lambda}}{\lambda \, (\lambda+1)(\lambda+2)\dots(\lambda+n-1)} \bigg|_{n=\infty}. \qquad \textbf{(45)}$$

The last equality is indeed adopted as the definition of gamma and it is also extended onto fractional and negative values of λ.

1.3.15. We are now going over to integrals that have a very close connection with the Euler integrals and are usually derived from these latter by assuming imaginary values for x. However, so as to follow quite a rigorous path, we shall determine them independently from the Euler integrals. We shall consider the integral

$$I = \int_0^\infty \int_0^\infty e^{-z} \cos(x z) \, x^{\mu-1} \, dx \, dz.$$

Integrating at first with respect to x, and then to z, we find, when substituting $x z = y$,

$$\int_0^\infty \cos(x z) \, x^{\mu-1} \, dx = \int_0^\infty \cos y \, (y/z)^{\mu-1} \, dy/z = z^{-\mu} \int_0^\infty y^{\mu-1} \cos y \, dy$$

so that

$$I = \int_0^\infty y^{\mu-1} \cos y \, dy \int_0^\infty z^{-\mu} e^{-z} \, dz = \Gamma(1 - \mu) \int_0^\infty y^{\mu-1} \cos y \, dy$$

where μ is supposed to be positive and less than 1.

Integrating now in the other order, we have, in accord with formula (4) of §1.1.2

$$\int_0^\infty e^{-z} \cos(x z) \, dz = \frac{1}{1+x^2}.$$

Then, on the strength of formula (25) we obtain

$$I = \int_0^\infty \frac{x^{\mu-1}}{1+x^2} \, dx = \frac{\pi}{2\cos[(\mu-1)\pi/2]}$$

and therefore

$$\int_0^\infty y^{\mu-1} \cos y \, dy = \frac{\pi}{2\Gamma(1-\mu) \cos[(\mu-1)\pi/2]}.$$

Assuming that μ is here positive and using formula (35) we have

$$\int_0^\infty y^{\mu-1} \cos y \, dy = \frac{\Gamma(\mu) \sin \mu\pi}{2\cos[(\mu-1)\pi/2]} = \frac{2\Gamma(\mu) \sin \mu\pi/2 \cos \mu\pi/2}{2\cos[(\mu-1)\pi/2]} =$$

$\Gamma(\mu) \cos(\mu\pi/2)$.

Integrating this equality by parts we obtain a similar formula containing a sine:

$$\int_0^\infty y^{\mu-1} \cos y \, dy = [y^{\mu-1} \sin y]_0^\infty - (\mu - 1) \int_0^\infty y^{\mu-2} \sin y \, dy =$$

$$(1 - \mu) \int_0^\infty y^{\mu-2} \sin y \, dy$$

so that

$$\int_0^\infty y^{\mu-2} \sin y \, dy = \frac{\Gamma(\mu)}{1-\mu} \cos(\mu\pi/2).$$

Substituting $\mu - 2 = \lambda - 1$ we arrive at

$$\int_0^\infty y^{\lambda-1} \sin y \, dy = \frac{\Gamma(\lambda+1)}{-\lambda} \cos[(\lambda + 1)\,\pi/2] = \Gamma(\lambda) \sin(\lambda\,\pi/2). \qquad \text{(xv)}$$

We have replaced $[\Gamma(\lambda + 1)/\lambda]$ by $\Gamma(\lambda)$; in this case, however, since μ is contained within the boundaries 0 and 1, λ takes values between -1 and 0 whereas

the substantiation of the formula (32) held only for positive values of n because this restriction follows from the definition of the gamma function that underlies the proof. So as to remove this difficulty we can prove the obtained formula also for negative values of n and we shall now have to use the definition of the gamma function included in formula (45). Noting that

$$\lim [n/(n + \lambda)]_{n=\infty} = 1$$

we may assume that

$$[\frac{(n-1)!\,n^{\lambda}}{\lambda\,(\lambda+1)\,(\lambda+2)\ldots(\lambda+n-1)}]_{n=\infty} = [\frac{(n-1)!\,n^{\lambda+1}}{\lambda\,(\lambda+1)\ldots(\lambda+n-1)\,(\lambda+n)}]_{n=\infty}$$

so that

$$\Gamma(\lambda) = [\frac{(n-1)!\,n^{\lambda+1}}{\lambda\,(\lambda+1)\ldots(\lambda+1+n-1)}]_{n=\infty}, \quad \lambda\,\Gamma(\lambda) = \Gamma(\lambda + 1)$$

and

$$\int_0^{\infty} y^{\lambda-1} \sin y \, dy = \Gamma(\lambda) \sin (\lambda\,\pi/2).$$

And so, formula (xv) is justified. We derived it for negative values of λ but it will also hold for positive λ's because it can be obtained in the same way as the formula containing the cosine. We thus have the formulas

$$\int_0^{\infty} x^{\mu-1} \cos x \, dx = \Gamma(\mu) \cos (\mu\,\pi/2), \quad 0 < \mu < 1 \qquad \textbf{(46a)}$$

$$\int_0^{\infty} x^{\mu-1} \sin x \, dx = \Gamma(\mu) \sin (\mu\,\pi/2), \quad -1 < \mu < 1 \qquad \textbf{(46b)}$$

We are concluding here the section on the Euler integrals.

1.4. Integrals of the Fourth Group

1.4.1. We are now going over to the integrals whose characteristic feature is that their limits are magnitudes having special significance for such angles as 0, $\pi/2$, π, 2π, etc. We begin by considering the integral

$$\int_{-\pi}^{\pi} e^{m\varphi i}e^{-n\varphi i}d\varphi$$

and we shall prove that its value is either 0 or 2π depending on whether m and n are different or equal to each other. In the first case we obtain by integration

$$\int_{-\pi}^{\pi} e^{(m-n)\varphi i}d\varphi = [\frac{e^{(m-n)\varphi i}}{(m-n)i}]_{-\pi}^{\pi} = \frac{e^{(m-n)\pi i} - e^{-(m-n)\pi i}}{(m-n)i} =$$

$$\frac{2i\sin[(m-n)\,\pi]}{(m-n)\,i} = \frac{2\sin[(m-n)\,\pi]}{m-n}.$$

It is seen however that, when m and n are integers, and, in addition, different, the integral vanishes. If $m = n$, the formula takes an indefinite form $0/0$. When determining its real magnitude [1] in accord with the rules of differential calculus, we shall have 2π. The same result can also be gotten directly: in this case, the integral will be

$$\int_{-\pi}^{\pi} d\varphi = 2\pi.$$

Thus

$$\int_{-\pi}^{\pi} e^{m\varphi i}e^{-n\varphi i}d\varphi = \begin{array}{l} 0, \text{ if } m \neq n \\ 2\pi \text{ if } m = n \end{array} \qquad \textbf{(47)}$$

A large number of mathematicians studied, and are studying, integrals of this kind. Their importance for {mathematical} analysis is based on the fact that, through their property consisting in {the existence of} equalities similar to (47), any function can be expanded into powers of one of the factors of the integrand. Thus, by means of

equality (47) we may expand $F(e^{i\varphi})$ into powers of $e^{i\varphi}$. Note that we may attain the same goal by differential calculus and it would seem that this latter method is preferable because no difficulties can be encountered there, but actually this is not so: differentiation presents no trouble only when we determine derivatives of known orders expressed by numbers, and becomes as difficult as integration is as soon as we desire to calculate the expression for the n-th derivative.

In addition, it is very often important to determine the term at which we should break off, and this problem is reduced to finding out how the general term of the expansion changes with the change of its number. In studies of this kind the terms expressed by integrals, even when these cannot be calculated, provide unquestionable advantage over expressions depending on derivatives of a known order. It is for this reason that the integrals of the considered kind are important for analysis.

By applying the properties of these integrals it is evidently possible to solve converse problems as well: Given an expansion, to determine the value of the integral that expresses its general term. Indeed, let us suppose that

$$F(e^{\varphi i}) = A_o e^{o\varphi i} + A_1 e^{\varphi i} + A_2 e^{2\varphi i} + \ldots + A_n e^{n\varphi i} + \ldots$$

Now, multiplying both sides of this equality by $e^{-in\varphi}$ and integrating within the limits $-\pi$ and π we have, by means of the formula (47),

$$\int_{-\pi}^{\pi} F(e^{\varphi i}) e^{-n\varphi i} d\varphi = A_n \cdot 2\pi$$

so that

$$A_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(e^{\varphi i}) e^{-n\varphi i} d\varphi.$$

We thus obtain a formula that enables us to derive both the general term of the expansion, and the integrals given the expansion. The coefficient of the general term can also be represented as

$$A_n = F^{(n)}(0)/n!.$$

Substitute now $F(x) = f(\lambda x)$ so that $F(e^{i\varphi}) = f(\lambda e^{i\varphi})$. Then

$$F'(x) = dF(x)/dx = df(\lambda x)/dx = \lambda f'(\lambda x),$$
$$F''(x) = d^2 F/dx^2 = \lambda\, df'(\lambda x)/dx = \lambda^2 f''(\lambda x), \ldots, \ F^{(n)}(x) = \lambda^n f^{(n)}(\lambda x)$$

and therefore

$$A_n = \lambda^n f^{(n)}(0)/n!.$$

We thus arrive at the formula

$$\int_{-\pi}^{\pi} f(\lambda e^{\varphi i}) e^{-n\varphi i} d\varphi = (2\pi/n!)\, f^{(n)}(0)\, \lambda^n. \tag{48}$$

We can somewhat generalize this formula by setting $f(z) = \theta(a + z)$ so that $f^{(n)}(z) = \theta^{(n)}(a + z)$. We thus obtain

$$\int_{-\pi}^{\pi} \theta(a + \lambda e^{\varphi i}) e^{-n\varphi i} d\varphi = (2\pi/n!)\, \theta^{(n)}(a)\, \lambda^n. \tag{49}$$

Note 1. {A dated concept.}

1.4.2. Suppose now that $f(\lambda x) = \ln(1 - \lambda x)$ so that

$$f(\lambda x) = -\lambda x - \lambda^2 x^2/2 - \lambda^3 x^3/3 - \ldots,$$
$$f(\lambda e^{\varphi i}) = -\lambda e^{\varphi i} - (\lambda^2/2) e^{2\varphi i} - (\lambda^3/3) e^{3\varphi i} - \ldots$$

or

$$f(\lambda e^{\varphi i}) = -(\lambda/1)(\cos\varphi + i\sin\varphi) - (\lambda^2/2)(\cos 2\varphi + i\sin 2\varphi) - \ldots$$

It is seen now that we ought to assume here that $\lambda < 1$. Only under this condition the series will always converge: the series of its coefficients, $-(\lambda/1)$, $-(\lambda^2/2)$, $-(\lambda^3/3)$, ... will always converge only if $|\lambda| < 1$ because, for such values of λ, the series $(\lambda + \lambda^2 + \lambda^3 + \ldots)$ will represent an infinitely decreasing geometric progression.

And so, supposing that $\lambda < 1$ and applying formula (48), we find that

$$I = \int_{-\pi}^{\pi} \ln (1 - \lambda e^{\varphi i}) \, d\varphi = 0$$

Transforming this integral, we have

$$I = \int_{-\pi}^{0} \ln (1 - \lambda e^{\varphi i}) \, d\varphi + \int_{0}^{\pi} \ln (1 - \lambda e^{\varphi i}) \, d\varphi = I_1 + I_2.$$

Suppose now that $\varphi = -\psi$, then

$$I_1 = - \int_{-\pi}^{0} \ln (1 - \lambda e^{-\psi i}) \, d\psi = \int_{0}^{\pi} \ln (1 - \lambda e^{-\varphi i}) \, d\varphi,$$

$$I = I_2 + \int_{0}^{\pi} \ln (1 - \lambda e^{-\varphi i}) \, d\varphi = \int_{0}^{\pi} \ln [(1 - \lambda e^{\varphi i}) (1 - \lambda e^{-\varphi i})] \, d\varphi$$

and therefore, for $\lambda < 1$,

$$I = \int_{0}^{\pi} \ln [1 - \lambda (e^{\varphi i} + e^{-\varphi i}) + \lambda^2] \, d\varphi = \int_{0}^{\pi} \ln [1 - 2\lambda \cos \varphi + \lambda^2] \, d\varphi = 0.$$

If $\lambda = 1/R$ where $R > 1$, we shall find that

$$\int_{0}^{\pi} \ln [1 - 2R \cos \varphi + R^2] \, d\varphi = \ln R^2 \int_{0}^{\pi} d\varphi = \pi \ln R^2.$$

We do not replace $\ln R^2$ by $2\ln R$ because R can be negative so that the formula   would have provided an indefinite result whereas it should be quite definite. We thus   come to such an integral:

$$\int_{0}^{\pi} \ln [\rho^2 - 2\rho \cos \varphi + 1] \, d\varphi = 0 \text{ if } \rho < 1 \text{ and } = \pi \ln \rho^2 \text{ if } \rho > 1 \qquad \textbf{(50)}$$

which is usually attributed to Poisson.

1.4.3. Issuing from the integral (50) we can derive several remarkable integrals.   Setting $\rho = 1$ we shall find that

$$\int_{0}^{\pi} \ln (2 - 2 \cos \varphi) \, d\varphi = 0.$$

We may admit this as a limiting equality because each of the formulas (50) leads to it   at $\rho = 1$. And so

$$\int_{0}^{\pi} \ln (2\sin \varphi/2)^2 \, d\varphi = 0, \quad \int_{0}^{\pi} \ln (2\sin \varphi/2) \, d\varphi = 0$$

so that

$$\int_{0}^{\pi} \ln \sin \varphi/2 \, d\varphi = - \int_{0}^{\pi} \ln 2 \, d\varphi = - \pi \ln 2.$$

Assuming here that $\varphi = 2\psi$ we get

$$\int_{0}^{\pi/2} \ln \sin \psi \, d\psi = - (\pi/2) \ln 2. \qquad \textbf{(51)}$$

If $\sin \psi = x$ then $d\psi = dx / \sqrt{1-x^2}$ and

$$\int_{0}^{1} \frac{\ln x}{\sqrt{1-x^2}} \, dx = -\frac{\pi \ln 2}{2}. \qquad \textbf{(52)}$$

Integrating the expression (51) by parts, we get

$$\int_{0}^{\pi/2} \ln \sin \psi \, d\psi = [\psi \ln \sin \psi]_{0}^{\pi/2} - \int_{0}^{\pi/2} (\psi/\sin \psi) \cos \psi \, d\psi = - \int_{0}^{\pi/2} \frac{\psi \, d\psi}{\operatorname{tg} \psi}.$$

Since

$$\lim (\psi \ln \sin \psi)_{\psi=0} = 0$$

we have

$$\int_0^{\pi/2} \frac{\psi \, d\psi}{\operatorname{tg} \psi} = \frac{\pi \ln 2}{2}. \qquad\qquad (53)$$

1.4.4. We shall show now the application of the formula (49) to determining the upper boundaries of derivatives, but before that we shall say when this formula may be used without any danger of encountering contradictions. The derivation of this formula was   based on the possibility of expanding the function $F(e^{\varphi i})$ into powers of $e^{\varphi i}$, but it is known that not any function may be expanded into powers of its independent variable; in other words, that the series obtained will not converge always. It is seen now that the convergence of the series

$$A_o + A_1 e^{\varphi i} + A_2 e^{2\varphi i} + \dots$$

which, according to our supposition, expresses the function $F(e^{\varphi i})$, should be formulated   as the condition for the validity of formula (49). Since we replace F by an identical   function $f[\lambda (e^{\varphi i})]$ expanded into a series of the kind

$$A_o + A_1 \lambda e^{\varphi i} + A_2 \lambda^2 e^{2\varphi i} + \dots \qquad\qquad (xvi)$$

the convergence of this series is necessary for the possibility of the existence of the   formula (49). But (xvi) can be represented as a sum of two series arranged in the order   of cosines, and sines, of multiple arcs. It follows that (xvi) will always converge if only   the series of the numerical values of the coefficients A_o, $A_1 \lambda$, $A_2 \lambda^2$, $\dots$ of its terms converges.

We may form an opinion about the convergence of this series only if $\lambda < 1$ because, under this condition, as it is not difficult to see, the series will always converge if only A_k decreases, or at least remains finite with an increasing k. We thus see that the formula   (49) may be adopted only for such functions the coefficients of whose expansion into powers of the variable always remain finite for the values of $\lambda < 1$.

After these remarks we proceed to solve the issue now interesting us. We have, in general,

$$\operatorname{mod}(k_1 + k_2 + \dots) < \operatorname{mod} k_1 + \operatorname{mod} k_2 + \dots$$

Therefore, considering the integral as the limit of a sum, we find that

$$\operatorname{mod} \int \Phi(u) \, du < \int \operatorname{mod} \Phi(u) \, du.$$

From formula (49) and from the one just obtained, since

$$\operatorname{mod} \frac{2\pi f^{(n)}(a) \lambda^n}{n!} = \operatorname{mod} \int_{-\pi}^{\pi} f(a + \lambda e^{\varphi i}) e^{-n\varphi i} \, d\varphi,$$

it follows however that the left side is less than

$$\int_{-\pi}^{\pi} \operatorname{mod} f(a + \lambda e^{\varphi i}) e^{-n\varphi i} \, d\varphi.$$

But, since mod $e^{-n\varphi i} = 1$,

$$\operatorname{mod}[f(a + \lambda e^{\varphi i}) e^{-n\varphi i}] = \operatorname{mod} f(a + \lambda e^{\varphi i}) \operatorname{mod} e^{-n\varphi i} = \operatorname{mod} f(a + \lambda e^{\varphi i}).$$

Let us assume that we have somehow determined the upper boundary R of the   modulus, $\operatorname{mod} f(a + \lambda e^{\varphi i})$:

$$\operatorname{mod} f(a + \lambda e^{\varphi i}) \leq R. \qquad\qquad (xvii)$$

Then

$$\operatorname{mod} \frac{2\pi f^{(n)}(a) \lambda^n}{n!} < \int_{-\pi}^{\pi} R \, d\varphi = 2\pi R,$$

hence

$$\operatorname{mod} f^{(n)}(a) \lambda^n < R \, n!.$$

Supposing now that both the function $f^{(n)}(a)$ and λ^n are real and stipulating that these expressions only denote the appropriate numerical values, we obtain the formula

$$f^{(n)}(a) \; < \; R\,n! \,/\, \lambda^n \qquad\qquad\qquad\qquad\qquad\qquad \textbf{(54)}$$

where R is determined by the condition (xvii). To provide an example of applying formula (54) let us take $F\,(e^{\varphi\,i}) = 1/[k - e^{i\,\varphi}]$ with $\mathrm{mod}\; k > 1$. Then the series

$$F\,(e^{\varphi\,i}) \; = \; (1/k) \; + \; (1/k^2)\,e^{\varphi\,i} \; + \; (1/k^3)\,e^{2\varphi\,i} \; + \; \dots$$

will be convergent. Assuming that $a = 0$ we find that

$$f\,(a \; + \; \lambda\,e^{\varphi\,i}) \; = \; \frac{1}{k - \lambda\,e^{\varphi i}}$$

where, in accord with the remark above, we ought to suppose that $\lambda < 1$.

In order to determine R we note that, in general,

$$\mathrm{mod}\,(A \; + \; Bi) \; = \; \sqrt{(A + Bi)\,(A - Bi)}$$

so that

$$(\mathrm{mod}\;\frac{1}{k - \lambda\,e^{\varphi i}}\,)^2 \; = \; (\frac{1}{k - \lambda\,e^{\varphi i}}\,)^2 \; = \; \frac{1}{k^2 - k\lambda\,(e^{\varphi i} + e^{-\varphi i}) + \lambda^2}\;.$$

It is seen now that the modulus will be maximal at $\varphi = 0$ so that we should assume that

$$R^2 \; = \; \frac{1}{k^2 - 2k\lambda + \lambda^2}$$

and, since $k > \lambda$, $R = 1/(k - \lambda)$. We thus obtained

$$\frac{d^n[1/(k - x)]}{dx^n}\Big|_{x=0} \; < \; \frac{n!}{(k - \lambda)\,\lambda^n}\;.$$

And so, we determined the upper boundary of the n-th derivative of the function $1/(k - x)$ at $x = 0$. Note that it is advantageous to derive the least value of the upper boundary which the studied magnitude cannot exceed; and since the inequality obtained by us takes place for any values of λ, we ought to find such of its values that will minimize the determined upper boundary. This problem reduces to the derivation of the maximal value of the function

$$(k \; - \; \lambda)\,\lambda^n \; = \; k\,\lambda^n \; - \; \lambda^{n+1}.$$

For calculating this value we have the equation

$$n\,k\,\lambda^{n-1} \; = \; -\,(n \; + \; 1)\,\lambda^n \; = \; 0 \;\; \text{and} \;\; \lambda \; = \; n\,k/(n \; + \; 1).$$

If the thus determined value of λ will be less than 1, we might use it, and then

$$k\,\lambda^n \; - \; \lambda^{n+1} \; = \; \frac{n^n k^{n+1}}{(n+1)^n} \; - \; \frac{(n+1)^{n+1} k^{n+1}}{(n+1)^{n+1}} \; = \; \frac{n^n k^{n+1}}{(n+1)^{n+1}}\;.$$

Therefore, if $k < [1 \; + \; (1/n)]$, we shall have

$$\frac{d^n[1/(k - x)]}{dx^n}\Big|_{x=0} \; < \; \frac{n!(n+1)^{n+1}}{n^n k^{n+1}}\;.$$

The Fourier Formulas {§§1.4.5 – 1.4.12}

1.4.5. We are now going over to multiple integrals and to deriving the Fourier formula that was previously considered very important; recently, however, it is ever more losing its significance. This happens because we are unable, while deriving this formula, to provide the conditions determining the functions for which it remains valid.

Let us take the integral

$$P_\infty = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) \cos [y (x - \alpha)] dx \, dy.$$

Integrating with respect to y, we find that

$$\int_{-\infty}^{\infty} \cos [y (x - \alpha)] \, dy = \frac{\sin[(x - \alpha) y]}{x - \alpha} \Big\}_{-\infty}^{\infty}.$$

It is seen now that the value of this integral is indefinite so that instead of the limits $-\infty$ and $+\infty$ we first assume $-A$ and A, and only then, in the final result, we set $A = \infty$. Consequently, we have

$$\int_{-A}^{A} \cos [y (x - \alpha)] \, dy = \frac{2\sin[A(x - \alpha)]}{x - \alpha}, \quad P_A = \int_{-\infty}^{\infty} f(x) \frac{2\sin[A(x - \alpha)]}{x - \alpha} dx.$$

Substituting now

$$A(x - \alpha) = z \tag{xviii}$$

so that $dx = dz/A$, we find that

$$P_A = 2 \int_{-\infty}^{\infty} f[\alpha + (z/A)] \, [(\sin z)/z] \, dz.$$

Assuming here $A = \infty$, we arrive at

$$P_\infty = 2 \int_{-\infty}^{\infty} f(\alpha) \, [(\sin z)/z] \, dz = 2 f(\alpha) \int_{-\infty}^{\infty} [(\sin z)/z] \, dz$$

and, on the strength of formula (5), $P_\infty = 2\pi f(\alpha)$. We thus derive the famous Fourier formula

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) \cos [y (x - \alpha)] dx \, dy = 2\pi f(\alpha). \tag{55}$$

The non-rigor of this derivation consists in that, having obtained the limits $-\infty$ and $+\infty$ for z from formula (xviii) or from $x = \alpha + z/A$, and knowing that these are the limits for x, we {nevertheless} are not always able to formulate the inverse statement: Assuming that $A = \infty$ in the final result for z, we cannot go over to the limits $-\infty$ and $+\infty$ for x. It follows that we would be unable to go back, i.e., to pass to the integral from the expression $2\pi f(\alpha)$, when deriving this formula.

It is seen now that in general the formula (55) will not be valid for any function universally. Let for example $f(x) = \exp(-x^2)$. By formula (55) we would have found

$$\exp(-\alpha^2) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-x^2) \cos [y (x - \alpha)] \, dx \, dy.$$

We shall integrate so as to check this result. We have

$$I = \int_{-\infty}^{\infty} \exp(-x^2) \cos [y (x - \alpha)] dx = \int_{-\infty}^{\infty} \cos (y \alpha) \exp(-x^2) \cos (y x)] \, dx +$$

$$\int_{-\infty}^{\infty} \sin (y \alpha) \exp(-x^2) \sin (y x) \, dx.$$

The second integral vanishes because its integrand is an odd function; the integrand in the first integral is even, so that in accord with formula (13)

$$I = 2 \cos (y \alpha) \int_{0}^{\infty} \exp(-x^2) \cos (y x) \, dx = \sqrt{\pi} \exp(-y^2/4) \cos (y \alpha).$$

Once more in virtue of this formula we have

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-x^2) \cos [y (x - \alpha)] \, dx \, dy = \sqrt{\pi} \int_{-\infty}^{\infty} \exp(-y^2/4) \cos (y \alpha) \, dy = 2\pi \exp(-\alpha^2).$$

We thus arrived at the same result as when applying the Fourier formula.

1.4.6. We shall now modify formula (55). We have

$$\cos \varphi = e^{\varphi i} - i \sin \varphi$$

so that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) \cos [y(x-\alpha)]\, dx\, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)\, e^{iy(x-\alpha)}\, dx\, dy -$$

$$i \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) \sin [y(x-\alpha)]\, dx\, dy.$$

But the second integral contains an odd function with respect to y and therefore vanishes. We have

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) \cos [y(x-\alpha)]\, dx\, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)\, e^{y(x-\alpha)i}\, dx\, dy.$$

Consequently

$$f(\alpha) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)\, e^{yxi}\, e^{-\alpha y i}\, dx\, dy,$$

$$f(\alpha) = \int_{-\infty}^{\infty} \varphi(y)\, e^{-\alpha y i}\, dy \qquad\qquad\qquad (56)$$

where

$$\varphi(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x)\, e^{yxi}\, dx.$$

Equation (56) solves a particular case of determining a function satisfying the equation

$$\int_{A}^{B} \varphi(y)\, F(\alpha;\, y)\, dy = f(\alpha).$$

Such issues were studied among others by Abel. In general, it ought to be noted that their solution leads to very remarkable results. Until now {however} they were completely solved only for the particular case in which $A = -\infty$, $B = \infty$, $F(\alpha;\, y) = e^{-i\alpha y}$.

1.4.7. We shall now modify the equation (56) and apply it, in its new form, in the sequel. Assuming that

$$\alpha = u\, i,\ f(u\, i) = F(u),\ \text{so that}\ f(u) = F(-u\, i),\ \text{we find that}$$

$$f(u\, i) = \int_{-\infty}^{\infty} \varphi(y)\, e^{uy}\, dy = F(u),\quad \varphi(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(-x\, i)\, e^{yxi}\, dx.$$

And so, we use the following equations:

$$F(u) = \int_{-\infty}^{\infty} \varphi(y)\, e^{uy}\, dy,\quad \varphi(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(-x\, i)\, e^{yxi}\, dy. \qquad (57)$$

Issuing from them, we can derive formula (18). Indeed, on the strength of (15) we have

$$\int_{0}^{\infty} \frac{\cos mx}{a^2 + x^2}\, dx = \frac{\pi}{2a}\, e^{-am},\quad \int_{0}^{\infty} \frac{e^{mxi} + e^{-mxi}}{2}\, \frac{dx}{a^2 + x^2} = \frac{\pi}{2a}\, e^{-am}.$$

Setting in the second equality $x = z$, $m = n\, y$, multiplying both its sides by $\varphi(y)\, dy$ and integrating with respect to y from $-\infty$ to $+\infty$, we obtain

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{e^{mxi} + e^{-mxi}}{2}\, \frac{\varphi(y)\, dy\, dx}{a^2 + x^2} = \frac{\pi}{2a} \int_{-\infty}^{\infty} e^{-any}\, \varphi(y)\, dy,$$

$$\int_{0}^{\infty} (1/2)[\int_{-\infty}^{\infty} e^{nyzi}\, \varphi(y)\, dy + \int_{-\infty}^{\infty} e^{-nyzi}\, \varphi(y)\, dy] \frac{dz}{a^2 + z^2} =$$

$$\frac{\pi}{2a} \int_{-\infty}^{\infty} e^{-a\,n\,y}\, \varphi\,(y)\, dy$$

and, because of (57),

$$\int_{0}^{\infty} \frac{F(nzi)+F(-nzi)}{2}\, \frac{dz}{a^2+z^2} \; = \; \frac{\pi}{2a}\, F\,(-\,an).$$

Now we can proceed to the derivation of the celebrated Dirichlet formula concerning multiple integrals. To this end, we note that, in general,

$$\int_{0}^{\infty} \upsilon^{\rho-1} e^{-\alpha\,\upsilon} d\upsilon \;=\; \int_{0}^{\infty} (t/\alpha)^{\rho-1} e^{-t}\,(d\,t/\alpha) \;=\; \Gamma(\rho)\,/\,\rho^{\,\alpha}$$

so that

$$\int_{0}^{\infty} \int_{0}^{\infty} \dots \int_{0}^{\infty} x^{\lambda-1} e^{-\alpha\,x}\, y^{\mu-1} e^{-\alpha\,y}\, z^{\nu-1} e^{-\alpha\,z} \dots dx\, dy\, dz \dots =$$

$$\int_{0}^{\infty} x^{\lambda-1} e^{-\alpha\,x}\, dx \int_{0}^{\infty} y^{\mu-1} e^{-\alpha\,y}\, dy \dots = \frac{\Gamma(\lambda)\Gamma(\mu)\Gamma(\nu)\dots}{\alpha^{\lambda+\mu+\nu+\dots}}.$$

Then, we have

$$\int_{0}^{\infty} s^{\lambda+\mu+\upsilon+\dots-1}\, e^{-\alpha\,s}\, ds \;=\; \frac{\Gamma(\lambda+\mu+\nu+\dots)}{\alpha^{\lambda+\mu+\nu+\dots}}$$

and consequently

$$\int_{0}^{\infty} \int_{0}^{\infty} \dots \int_{0}^{\infty} x^{\lambda-1}\, y^{\mu-1}\, z^{\upsilon-1} \dots e^{-\alpha\,(x+y+z+\dots)} \dots dx\, dy\, dz \dots =$$

$$\frac{\Gamma(\lambda)\Gamma(\mu)\Gamma(\nu)\dots}{\Gamma(\lambda+\mu+\nu+\dots)} \int_{0}^{\infty} s^{\lambda+\mu+\upsilon+\dots-1}\, e^{-\alpha\,s}\, ds.$$

Multiplying both sides of this equality by $\varphi\,(\alpha)\,d\alpha$ and integrating with respect to α within the limits $-\infty$ and $+\infty$, we obtain

$$\int_{0}^{\infty} \int_{0}^{\infty} \dots \int_{0}^{\infty} x^{\lambda-1}\, y^{\mu-1}\, z^{\upsilon-1} \dots [\int_{-\infty}^{\infty} \varphi\,(\alpha)\, e^{-\alpha\,(x+y+z+\dots)}\, d\alpha]\, dx\, dy\, dz \dots =$$

$$\frac{\Gamma(\lambda)\Gamma(\mu)\Gamma(\nu)\dots}{\Gamma(\lambda+\mu+\nu+\dots)} \int_{0}^{\infty} s^{\lambda+\mu+\upsilon+\dots-1}\, ds \int_{-\infty}^{\infty} \varphi\,(\alpha)\, e^{-\alpha\,s} d\alpha.$$

Consequently, on the strength of (57), we get

$$\int_{0}^{\infty} \int_{0}^{\infty} \dots \int_{0}^{\infty} x^{\lambda-1}\, y^{\mu-1}\, z^{\upsilon-1} \dots F\,[-\,(x\,+\,y\,+\,z\,+\,\dots)]\, dx\, dy\, dz \dots =$$

$$\frac{\Gamma(\lambda)\Gamma(\mu)\Gamma(\nu)\dots}{\Gamma(\lambda+\mu+\nu+\dots)} \int_{0}^{\infty} s^{\lambda+\mu+\upsilon+\dots-1}\, F(-\,s)\, ds$$

and, substituting $F\,(-\,t)\;=\;f\,(t)$, we indeed arrive at the Dirichlet formula

$$\int_{0}^{\infty} \int_{0}^{\infty} \dots \int_{0}^{\infty} x^{\lambda-1}\, y^{\mu-1}\, z^{\upsilon-1} \dots f\,(x\,+\,y\,+\,z\,+\,\dots)\, dx\, dy\, dz \dots =$$

$$\frac{\Gamma(\lambda)\Gamma(\mu)\Gamma(\nu)\dots}{\Gamma(\lambda+\mu+\nu+\dots)} \int_{0}^{\infty} s^{\lambda+\mu+\upsilon+\dots-1} f\,(s)\, ds. \qquad\qquad \textbf{(58)}$$

1.4.8. Formula (58) can be somewhat generalized by introducing new parameters whose particular values would have led to it. Suppose that $x\;=\;au,\;y\;=\;bv,\;z\;=\;cw,\;\dots$ and that $a,\;b,\;c,\;\dots$ are positive so that the limits of integration persist. Noting that, consequently, a constant factor $a^{\,\lambda}\,b^{\,\mu}\,c^{\,\upsilon}\dots$will be included in the left side of the equality and dividing both its sides by this factor, we obtain

$$\int_0^\infty \int_0^\infty \ldots \int_0^\infty u^{\lambda-1} v^{\mu-1.} w^{\upsilon-1} \ldots f(au + bv + cw + \ldots)\, du\, dv\, dw \ldots =$$

$$\frac{\Gamma(\lambda)\Gamma(\mu)\ldots}{a^\lambda b^\mu \ldots \Gamma(\lambda + \mu + \ldots)} \int_0^\infty s^{\lambda+\mu+\upsilon+\ldots-1} f(s)\, ds.$$

Setting now $u = x^m$, $v = y^n$, $w = z^p$, ... we arrive at

$$\int_0^\infty \int_0^\infty \ldots \int_0^\infty x^{m\lambda-m} y^{n\mu-n.} z^{p\upsilon-p} \ldots f(ax^m + by^n + cz^p + \ldots) \cdot$$
$$mx^{m-1}\, dx\, ny^{n-1}\, dy\, pz^{p-1}\, dz \ldots =$$
$$m\, n\, p \ldots \int_0^\infty \int_0^\infty \ldots \int_0^\infty x^{m\lambda-1} y^{n\mu-1.} z^{p\upsilon-1} \ldots f(ax^m + by^n + cz^p + \ldots)\, dx\, dy\, dz =$$

$$\frac{\Gamma(\lambda)\Gamma(\mu)\ldots}{a^\lambda b^\mu \ldots \Gamma(\lambda + \mu + \ldots)} \int_0^\infty s^{\lambda+\mu+\upsilon+\ldots-1} f(s)\, ds.$$

Substituting finally $m\lambda = \alpha$, $n\mu = \beta$, $p\upsilon = \gamma$, ... and dividing both sides of the equality by $m\, n\, p \ldots$, we get

$$\int_0^\infty \int_0^\infty \ldots \int_0^\infty x^{\alpha-1} y^{\beta-1.} z^{\gamma-1} \ldots f(ax^m + by^n + cz^p + \ldots)\, dx\, dy\, dz\ldots =$$

$$\frac{\Gamma(\alpha/m)\Gamma(\beta/n)\Gamma(\gamma/p)\ldots}{mnp\ldots a^{\alpha/m} b^{\beta/n} c^{\gamma/p} \ldots \Gamma(\alpha/m + \beta/n + \gamma/p \ldots)} \int_0^\infty s^{\alpha/m+\beta/n+\gamma/p+\ldots-1} f(s)\, ds. \qquad \textbf{(59)}$$

Here, the parameters m, n, p, ... are of course supposed to be positive; otherwise, 0 and ∞ would not be the limits of the new integral.

Formula (58) can be derived as a particular case of (59) when substituting in the latter $a = b = c \ldots = 1$, $m = n = p = 1$, $\alpha = \lambda$, $\beta = \mu$, $\gamma = \upsilon$, ...

1.4.9. Suppose that $f(t)$ is a function satisfying the conditions

$$f(t) = 1, \ t \le L \text{ and } = 0, \ t > L.$$

Then, assuming {also} the condition

$$ax^m + by^n + cz^p + \ldots \le L \qquad\qquad\qquad \text{(xix)}$$

where L is a given positive magnitude and applying formula (59), we reduce the determination of the integral

$$\int_0^\infty \int_0^\infty \ldots \int_0^\infty x^{\alpha-1} y^{\beta-1.} z^{\gamma-1} \ldots dx\, dy\, dz\ldots \qquad\qquad\qquad \text{(xx)}$$

to the calculation of

$$\frac{\Gamma(\alpha/m)\Gamma(\beta/n)\Gamma(\gamma/p)\ldots}{mnp\ldots a^{\alpha/m} b^{\beta/n} c^{\gamma/p} \ldots \Gamma(\alpha/m + \beta/n + \gamma/p \ldots)} \int_0^L s^{\alpha/m+\beta/n+\gamma/p+\ldots-1}\, ds$$

since in this case

$$\int_0^\infty s^{\alpha/m+\beta/n+\gamma/p+\ldots-1} f(s)\, ds = \int_0^L s^{\alpha/m+\beta/n+\gamma/p+\ldots-1} 1\, ds + \int_L^\infty s^{\alpha/m+\beta/n+\gamma/p+\ldots-1} 0\, ds.$$

But the first of these integrals is

$$\frac{L^{\alpha/m+\beta/n+\gamma/p+\ldots}}{\alpha/m + \beta/n + \gamma/p + \ldots}$$

and, on the strength of formula (32), we find that

Integral (xx) equals $\dfrac{\Gamma(\alpha/m)\Gamma(\beta/n)\Gamma(\gamma/p)...L^{\,\alpha/m+\beta/n+\gamma/p+...}}{mnp...a^{\alpha/m}b^{\beta/n}c^{\gamma/p}...\Gamma(\alpha/m+\beta/n+\gamma/p+...+1)}$. **(60)**

This formula thus solves the problem of determining the integral (xx) extended over all the positive values of the variables x, y, z, … connected by condition (xix).

1.4.10. Problems about the determination of areas and volumes as well as those touching on the attraction of bodies of a known form are easily solved by formula (60). For example, we shall calculate the area of an ellipse; the condition connecting the variables x and y will therefore be
$ax^2 + by^2 \le 1$ where $a = 1/A^2$, $b = 1/B^2$ with A and B being the semi-axes of the ellipse. In this case, $m = n = 2$, $\alpha = \beta = 1$ so that formula (60) provides

$$\int \int dx\, dy = \frac{\Gamma(1/2)\Gamma(1/2)L}{4\Gamma(2)\sqrt{ab}}$$

where the integral is extended over all the positive values of the variables x and y obeying the condition $ax^2 + by^2 \le L$. In our case, $L = 1$; and $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(2) = 1$ so that
$$\int \int dx\, dy = \pi/4 \sqrt{ab} = (\pi/4)\, AB.$$

As expected, we thus obtained the magnitude of a quarter of the area sought: the variables x and y are positive only for the quarter of the ellipse determined by the equation

$x^2/A^2 + y^2/B^2 = 1.$

Let us also consider a triple integral that, for $\alpha = \beta = \gamma = 1$ and under the condition of the type

$ax^{m} + by^{n} + cz^{p} \le L,$

represents some volume. Assume also that $m = n = p = 2$. Then formula (60) provides

$$\int \int \int dx\, dy\, dz = \frac{\Gamma(1/2)\Gamma(1/2)\Gamma(1/2)L^{3/2}}{8\Gamma(5/2)\sqrt{abc}}.$$

However,

$\Gamma(5/2) = (3/2)\,\Gamma(3/2) = (3/2)\,(1/2)\,\Gamma(1/2) = 3\sqrt{\pi}/4$

and

$$\int \int \int dx\, dy\, dz = (\pi/6)\frac{L^{3/2}}{\sqrt{abc}}.$$

Suppose now that the equation transformed to the normal form

$ax^2 + by^2 + cz^2 = L$

belongs to an ellipsoid. Consequently, $L = 1$, $a = 1/A^2$, $b = 1/B^2$, $c = 1/C^2$ where A, B and C are the ellipsoid's semi-axes and

$$\int \int \int dx\, dy\, dz = (\pi/6)\, ABC$$

will be the expression for 1/8 of its volume.

1.4.11. We deduced the Fourier formula in the form (55). Now, we shall impart a more general form to it by choosing some magnitudes L and M ($L < M$) as the limits of integrating with respect to x. In order to accomplish this we might have acted in the following way. Since $f(x)$ is an absolutely arbitrary function (restricted however by conditions indicated when deriving formula (55)) that can even be discontinuous, we may assume that

$f(x) = 0,\ -\infty < x < L;\ f(x) = \varphi(x),\ L < x < M;\ f(x) = 0,$

$M < x < +\infty.$

Consequently, we have

$$f(\alpha) = (1/2\pi)\{0 + \int_{-\infty}^{\infty} \int_{L}^{M} \varphi(x) \cos[y(x - \alpha)]\, dx\, dy + 0\} =$$

$$(1/2\pi) \int_{-\infty}^{\infty} \int_{L}^{M} \varphi(x) \cos[y(x - \alpha)]\, dx\, dy.$$

Here, for α included within $-\infty$ an L or between M and $+\infty$, the function $f(\alpha) = 0$ and the integral vanishes. If, however, α is within the boundaries L and M the integral equals $2\pi\, \varphi(\alpha)$.

We shall now derive the same result by applying considerations similar to those used when deducing formula (55). To this end let us study the integral

$$\int_{y=-A}^{y=A} \int_{x=L}^{x=M} f(x) \cos[y(x - \alpha)]\, dx\, dy.$$

When integrating with respect to y, we find that

$$\int_{-A}^{A} f(x) \cos[y(x - \alpha)]\, dy = 2f(x)\, \frac{\sin[A(x - \alpha)]}{x - \alpha}.$$

Supposing now that $A(x - \alpha) = z$ and integrating with respect to z (whose limits will be $A(L - \alpha)$ and $A(M - \alpha)$) we get

$$\int_{-A}^{A} \int_{L}^{M} f(x) \cos[y(x - \alpha)]\, dx\, dy = 2 \int_{A(L-\alpha)}^{A(M-\alpha)} f(\alpha + z/A)\,(\sin z / z)\, dz$$

so that

$$\int_{-\infty}^{\infty} \int_{L}^{M} f(x) \cos[y(x - \alpha)]\, dx\, dy = 2f(\alpha) \int_{A(L-\alpha)}^{A(M-\alpha)} f(\alpha + z/A)\,(\sin z / z)\, dz_{\;A=\infty.}$$

In order to calculate this integral we ought to consider separately two cases. If α is included between the boundaries L and M, then, when A increases to $+\infty$, the lower limit of the integral will tend to $-\infty$, and the upper limit, to $+\infty$, so that the double integral will be equal to $2\pi f(\alpha)$. Otherwise, both limits of integration will approach $+\infty$ (if $L > \alpha$) or $-\infty$ (if $M < \alpha$) and

$$\int_{-\infty}^{\infty} \int_{L}^{M} f(x) \cos[y(x - \alpha)]\, dx\, dy = 2f(\alpha) \int_{+\infty}^{+\infty} (\sin z / z)\, dz = 0$$

or

$$\int_{-\infty}^{\infty} \int_{L}^{M} f(x) \cos[y(x - \alpha)]\, dx\, dy = 2f(\alpha) \int_{-\infty}^{-\infty} (\sin z / z)\, dz = 0.$$

We thus arrive at the result that can be expressed in the following way

$$(1/2\pi) \int_{-\infty}^{\infty} \int_{L}^{M} f(x) \cos[y(x - \alpha)]\, dx\, dy = 0 \text{ if } \alpha < L$$

$$\text{or if} > M;\; = f(\alpha) \text{ otherwise.} \qquad \textbf{(61)}$$

This is identical with what was obtained above in a different way.

1.4.12. In §1.4.11, while considering the case in which α was not included between the boundaries L and M, we arrived at the integrals

$$\int_{+\infty}^{+\infty} (\sin z/z)\, dz, \quad \int_{-\infty}^{-\infty} (\sin z/z)\, dz \qquad\qquad\qquad \text{(xxi)}$$

and, since in both cases the upper and the lower limits of integration were equal to each other, we concluded that the integrals vanished. Evidently, however, we do not always have the right to make such inferences. An integral taken between infinite limits of the same sign is, in general, an indefinite magnitude representing the area included between two ordinates infinitely distant from the origin. This area is obviously not always equal to zero and it can even be infinitely large.

For this reason we believe that it is not amiss to say now a few words about when such integrals really have indefinite values and when they ought to be considered equal to zero. We shall therefore examine, in general, the integral

$$\int_S^T f(x)\,dx \qquad\qquad (xxii)$$

and assume that S and T tend to infinity. It is not difficult to see that, if the integral

$$\int_0^\infty f(x)\,dx \qquad\qquad (xxiii)$$

has a finite and definite value C, then (xxii) will always be equal to 0 when S and T increase to infinity. Indeed, we have

$$\lim \int_0^T f(x)\,dx\,_{T=\infty} = C,\ \lim \int_0^S f(x)\,dx\,_{S=\infty} = C$$

so that

$$\int_{+\infty}^{+\infty} f(x)\,dx = \lim \int_S^T f(x)\,dx\,_{T=\infty,\ S=\infty} =$$

$$\lim [\int_0^T f(x)\,dx - \int_0^S f(x)\,dx\,]\,_{T=\infty,\ S=\infty} = C - C = 0. \quad (xxiv) \qquad\qquad \text{If,}$$

however, the value of the integral (xxiii) is infinitely large or indefinite, then the integral (xxiv), being represented by the difference of either two infinities or of two indefinite expressions, will also have an indefinite value.

Thus, we see that a necessary condition for the integral (xxiv) to vanish is the definiteness and finiteness of (xxiii). In §1.4.11 the latter is {indeed} finite and definite because

$$\int_0^\infty [(\sin z)\,/\,z]\,dz = \pi/2$$

and

$$\int_0^{-\infty} [(\sin z)\,/\,z]\,dz = \int_0^\infty \{[(\sin(-z)]\,/(-z)\}\,d(-z) = -\int_0^\infty [(\sin z)\,/\,z]\,dz = -\pi/2$$

so that we have the right to assume that the integrals (xxi) are both zero.

We are here concluding our course {in definite integrals}.

Supplement

In §1.3.14 the gamma function was defined as

$$\Gamma(\lambda) = \frac{(n-1)!\,n^\lambda}{\lambda(\lambda+1)...(\lambda+n-1)}\,_{n=\infty}. \qquad\qquad (\mathbf{45})$$

The correctness of this definition is obvious only for integer and positive values of λ. We shall now justify it for any positive values of λ. Consider the function

$$F(x;\,n) = \frac{(n-1)!\,n^x}{x(x+1)...(x+n-1)} \qquad\qquad (\mathbf{62})$$

where x is arbitrary and n is an integral and positive number. We shall prove that $F(x;\,n)$ has a limit at $n = \infty$ for any *non-integer* x. Obviously,

$$F(0;\,n) = \infty;\ \lim F(\varepsilon;\,n)\,_{\varepsilon=0} = +\infty,\ \lim F(-\varepsilon;\,n)\,_{\varepsilon=0} = -\infty,$$

and $F(-x;\,n) = \pm\infty$ for any integer and positive x not greater than $(n-1)$. To prove our proposition we shall apply the formula

$$\frac{\beta(\beta+1)(\beta+2)...(\beta+n-1)}{\alpha(\alpha+1)(\alpha+2)...(\alpha+n-1)} = \theta\,[(\alpha/(\alpha+n)]^{\alpha-\beta} \qquad\qquad (\mathbf{63})$$

where $\beta > 0$, $\alpha > 0$, $\alpha > \beta$, $1 > \theta > 0$. This formula can be derived in the following way.

We have in general

$$\varphi(a + h) = \varphi(a) + h\,\varphi'(a + \theta h)$$

so that, if $\varphi(x) = x^{1+\mu}$,

$$(1 + h)^{1+\mu} = 1 + h(1 + \mu)(1 + \theta h)^{\mu}.$$

If $h > 0$ and $\mu > 0$, then $(1 + \theta h)^{\mu} < (1 + h)^{\mu}$ and

$$(1 + h)^{1+\mu} < 1 + h(1 + \mu)(1 + h)^{\mu}$$

or

$$(1 - \mu h)(1 + h)^{\mu} < 1 \text{ if } \mu h < 1.$$

Dividing by $(1 - \mu h)$ we have

$$(1 + h)^{\mu} < 1/(1 - \mu h), \quad 0 < \mu < 1/h.$$

Let now $h = 1/(\alpha + m)$ and $\mu = \alpha - \beta$ where $\alpha > \beta > 0$. Then

$$\left(\frac{\alpha + m + 1}{\alpha + m}\right)^{\alpha - \beta} < \frac{\alpha + m}{\beta + m}$$

or

$$\frac{\beta + m}{\alpha + m} < \left(\frac{\alpha + m}{\alpha + m + 1}\right)^{\alpha - \beta}.$$

Supposing that, consecutively, $m = 0, 1, 2, \ldots, (n - 1)$ and multiplying the obtained inequalities, we have

$$0 < \frac{\beta(\beta + 1)\ldots(\beta + n - 1)}{\alpha(\alpha + 1)\ldots(\alpha + n - 1)} < \left(\frac{\alpha}{\alpha + n}\right)^{\alpha - \beta}$$

{QED}.

We return now to the function $F(x; n)$. For $x > 0$ formulas (62) and (63) provide

$$F(x; n) = \frac{(n-1)!\,n^x}{x(x+1)\ldots(x+n-1)} = \theta\left\{\left(\frac{n}{x+n}\right)^n\left(\frac{(n+1)^x}{x}\right)\right\}$$

And so, if $x > 0$, $F(x; n)$ is always less than $[(x + 1)^x/x]$ and, since

$$\frac{F(x; n+1)}{F(x; n)} = \frac{n}{n+x}\left(\frac{n+1}{n}\right)^x$$

with this ratio approaching its limit equal to 1, then, for $x > 0$, $F(x; n)$ remains always positive and tends to a finite limit.

Suppose now that $x = -y$ and let $k - 1 < y < k$. If $k - y = \lambda = k + x$ then $1 > \lambda > 0$ so that

$$F(x; n) = \frac{(n-1)!\,n^{\lambda-k}}{x(x+1)\ldots(x+k-1)(x+k)\ldots(x+n-1)} =$$

$$\frac{1}{x(x+1)\ldots(x+k)}\,u$$

where

$$u = \frac{(n-1)!\,n^{\lambda-k}}{(\lambda+1)(\lambda+2)...(\lambda+n-k-1)} =$$

$$\frac{(n-1)!\,(\lambda+n-k)(\lambda+n-k+1)...(\lambda+n-1)}{(\lambda+1)(\lambda+2)...(\lambda+n-1)\,n^k}\,n^\lambda.$$

But

$$\frac{\lambda+n-k}{n}\,\frac{\lambda+n-k+1}{n}\,...\,\frac{\lambda+n-1}{n} = [1 - (k - \lambda)/n]$$

$$[1 - (k - \lambda - 1)/n] \,...\, [1 - (1 - \lambda)/n] < [1 - (1 - \lambda)/n]^k$$

and

$$\frac{(n-1)!}{(\lambda+1)(\lambda+2)...(\lambda+n-1)} = \theta\left(\frac{\lambda+1}{\lambda+n}\right)^\lambda.$$

Therefore

$$u = \theta\,(\lambda + 1)^\lambda\,\frac{n}{\lambda+n}\,[1 - (1 - \lambda)/n]^k.$$

It is seen now that

$$u < (1 + \lambda)^\lambda = (1 + x + k)^{x+k}$$

so that $F(x;\,n)$ cannot be infinitely large; and because

$$\lim \frac{F(x;n+1)}{F(x;n)} = 1$$

it has a finite limit equal to

$$\frac{\theta\,(1+k+x)^{k+x}}{x\,(x+1)...(x+k)}.$$

Suppose that $x > 0$. We shall prove that

$$F(x;\,\infty) = \Gamma(x). \tag{xxv}$$

Taking the logarithm of both sides of (62) we have

$$\ln F(x;\,n) = \ln(n - 1)! + x \ln n - [\ln x + \ln(x + 1) + \,...\, + \ln(x + n - 1)]$$

and consequently

$$\frac{F'_x(x;n)}{F(x;n)} = \ln n - [\frac{1}{x} + \frac{1}{x+1} + \,...\, + \frac{1}{x+n-1}].$$

But

$$\ln n = \int_0^\infty \frac{e^{-z} - e^{-nz}}{z}\,dz, \quad \frac{1}{x-k} = \int_0^\infty e^{-(x+k)z}\,dz$$

and therefore

$$\frac{\partial \ln F(x;n)}{\partial x} = \int_0^\infty \{\frac{e^{-z} - e^{-nz}}{z} - e^{-xz}[1 + e^{-z} + e^{-2z} + \,...\, + e^{-(n-1)z}]\}\,dz =$$

$$\int\limits_{0}^{\infty} \, [\frac{e^{-z} - e^{-nz}}{z} \; - \; e^{-xz} \, \frac{e^{-nz} - 1}{e^{-z} - 1}] \, dz.$$

Formula (62) shows that $F(1; n) = 1$ so that, integrating the equation above with respect to x within the limits 1 and x, we obtain

$$\ln F(x; n) = \int\limits_{0}^{\infty} \, [(x - 1)(e^{-z} - e^{-nz}) - (e^{-xz} - e^{-z}) \, \frac{1 - e^{-nz}}{e^{-z} - 1}] \, \frac{dz}{z}.$$

At $n = \infty$ this formula becomes

$$\ln F(x; \infty) = \int\limits_{0}^{\infty} \, [(x - 1) e^{-z} - \frac{e^{-xz} - e^{-z}}{e^{-z} - 1} \, \frac{dz}{z}] \; = \; \ln \Gamma(x),$$

hence (xxv).

<div align="center">

Collection of Formulas Occurring in This Course

</div>

{Chebyshev adduced a list of 50 integrals which I am omitting here.}

<div align="center">

Chapter 2. The Theory of Finite Differences

2.1. Direct Calculus of {Finite} Differences

</div>

2.1.1. Let us take some function $f(x)$ and assume that the independent variable x gets equal finite increments Δx (as in the differential calculus, x is here supposed to vary uniformly). We denote the corresponding values of the function in the following way:

$$u_{\mathrm{o}} = f(x); \; u_1 = f(x + \Delta x); \; u_2 = f(x + 2\Delta x); \; \ldots; \; u_n = f(x + n\Delta x).$$

Its increments will correspondingly be

$$\Delta u_{\mathrm{o}} = u_1 - u_{\mathrm{o}}; \; \Delta u_1 = u_2 - u_1; \; \Delta u_2 = u_3 - u_2; \; \ldots; \Delta u_n = u_{n+1} - u_n.$$

We thus obtain the series of functions

$$\Delta u_{\mathrm{o}}; \; \Delta u_1; \; \Delta u_2; \; \ldots; \; \Delta u_n \qquad\qquad\qquad\qquad (\mathrm{i})$$

from the series $u_{\mathrm{o}}; \; u_1; \; u_2; \; \ldots; \; u_n; \; u_{n+1}.$

Considering (i) as the series of initial functions and calculating the differences between its adjacent terms (always subtracting the preceding term from the subsequent term) we get a series of new functions

$$\Delta^2 u_{\mathrm{o}}; \; \Delta^2 u_1; \; \Delta^2 u_2; \; \ldots; \; \Delta^2 u_n$$

where, in general,

$$\Delta^2 u_n = \Delta(\Delta u_n) = \Delta u_{n+1} - \Delta u_n.$$

Reasoning further on in the same way, we shall each time obtain a new series of functions. The general form of these series is

$$\Delta^{\lambda} u_{\mathrm{o}}; \; \Delta^{\lambda} u_1; \; \Delta^{\lambda} u_2; \; \ldots; \; \Delta^{\lambda} u_n; \quad \Delta^{\lambda+1} u_{\mathrm{o}}; \; \Delta^{\lambda+1} u_1; \; \Delta^{\lambda+1} u_2; \; \ldots; \; \Delta^{\lambda+1} u_n$$

where

$$\Delta^{\lambda+1} u_n = \Delta(\Delta^{\lambda} u_n) = \Delta^{\lambda} u_{n+1} - \Delta^{\lambda} u_n.$$

It is not difficult to see now that

$$u_{n+1} = \mathrm{u}_n + \Delta \mathrm{u}_n; \; \Delta u_{n+1} = \Delta u_n + \Delta^2 u_n; \; \ldots; \; \Delta^{\lambda} u_{n+1} = \Delta^{\lambda} u_n + \Delta^{\lambda+1} \mathrm{u}_n.$$

We thus arrive at the following practical rule for calculating the differences of any order: When compiling a table, each of whose vertical columns includes all the functions corresponding to one and the same increment Δx of the independent

variable, and determining some function located at the intersection of a column and a line, it is necessary to choose the function of the same column situated directly above it and to add to it the directly following function of the horizontal line.

2.1.2. Let us now derive the general formulas that enable us to express the difference of any order in terms of the initial functions and, conversely, to express the initial functions through the differences of various orders.

We have $\Delta u_n = u_{n+1} - u_n$ so that, replacing n by $(n + 1)$, $\Delta u_{n+1} = u_{n+2} - u_{n+1}$. Therefore

$$\Delta^2 u_n = \Delta u_{n+1} - \Delta u_n = (u_{n+2} - u_{n+1}) - (u_{n+1} - u_n) = u_{n+2} - 2u_{n+1} + u_n,$$
$$\Delta^2 u_{n+1} = u_{n+3} - 2u_{n+2} + u_{n+1}$$

so that

$$\Delta^3 u_n = \Delta^2 u_{n+1} - \Delta^2 u_n = (u_{n+3} - 2u_{n+2} + u_{n+1}) - (u_{n+2} - 2u_{n+1} + u_n) =$$
$$u_{n+3} - 3u_{n+2} + 3u_{n+1} - u_n.$$

By analogy we may conclude that, in general,

$$\Delta^\lambda u_n = u_{n+\lambda} - \lambda u_{n+\lambda-1} + [\lambda (\lambda - 1) / 2!] u_{n+\lambda-2} -$$
$$[\lambda (\lambda - 1) (\lambda - 2) / 3!] u_{n+\lambda-3} + \ldots +$$
$$(-1)^\mu C_\lambda^\mu u_{n+\lambda-\mu}. \tag{1}$$

In order to justify this formula we shall prove that it holds for $(\lambda + 1)$ if it is valid for λ. Replacing n by $(n + 1)$ in (1) we have

$$\Delta^\lambda u_{n+1} = u_{n+1+\lambda} - \lambda u_{n+\lambda} + [\lambda (\lambda - 1) / 2!] u_{n+\lambda-1} -$$
$$[\lambda (\lambda - 1) (\lambda - 2) / 3!] u_{n+\lambda-2} + \ldots + (-1)^{\mu+1} C_\lambda^{\mu+1} u_{n+\lambda-\mu+1}$$

Therefore

$$\Delta^{\lambda+1} u_n = \Delta^\lambda u_{n+1} - \Delta^\lambda u_n = u_{n+1+\lambda} - (\lambda + 1) u_{n+\lambda} + [(\lambda + 1) \lambda / 2!] u_{n+\lambda-1} -$$
$$[(\lambda + 1) \lambda (\lambda - 1) / 3!] u_{n+\lambda-2} + \ldots$$

The general term will be

$$(-1)^{\mu+1} C_\lambda^\mu [\frac{\lambda - \mu}{\mu + 1} - 1] u_{n+\lambda-\mu} = (-1)^{\mu+1} C_{\lambda+1}^{\mu+1} u_{n+\lambda-\mu}.$$

Thus, formula (1) is proved.

Let us now solve the inverse problem, that is, derive a formula enabling us to express the initial function through the differences of various orders corresponding to one and the same increment of the variable. We have the formulas {see §2.1.1}

$$\Delta u_n = u_{n+1} - u_n, \ \Delta^2 u_n = \Delta u_{n+1} - \Delta u_n, \ \Delta^3 u_n = \Delta^2 u_{n+1} - \Delta^2 u_n, \ \ldots$$

so that

$$u_{n+1} = u_n + \Delta u_n,$$
$$u_{n+2} = u_{n+1} + \Delta u_{n+1} = (u_n + \Delta u_n) + (\Delta u_n + \Delta^2 u_n) = u_n + 2\Delta u_n + \Delta^2 u_n,$$
$$u_{n+3} = u_{n+1} + 2\Delta u_{n+1} + \Delta^2 u_{n+1} = u_n + 3\Delta u_n + 3\Delta^2 u_n + \Delta^3 u_n.$$

By analogy we suppose that

$$u_{n+\lambda} = u_n + \lambda \Delta u_n + [\lambda (\lambda - 1)/2!] \Delta^2 u_n + \ldots + C_\lambda^\mu \Delta^\mu u_n. \tag{ii}$$

To substantiate this formula we shall consider $\Delta u_{n+\lambda}$ as an initial function. In accord with (ii) we have

$$\Delta u_{n+\lambda} = \Delta u_n + (\lambda/1!)\Delta^2 u_n + [\lambda (\lambda - 1)/2!] \Delta^3 u_n + \ldots + C_\lambda^\mu \Delta^{\mu+1} u_n.$$

Adding this formula to (ii) we obtain

$$u_{n+\lambda} + \Delta u_{n+\lambda} = u_n + \Delta u_n + \lambda \Delta u_n + [\lambda(\lambda-1)/2!]\,\Delta^2 u_n + (\lambda/1!)\,\Delta^2 u_n +$$
$$\ldots + [C_\lambda^{\mu+1} + C_\lambda^{\mu}]\,\Delta^{\mu+1} u_n + \ldots,$$

$$u_{n+\lambda+1} = u_n + (\lambda+1)\,\Delta u_n + [(\lambda+1)\lambda/2!]\,\Delta^2 u_n + \ldots + C_{\lambda+1}^{\mu+1}\,\Delta^{\mu+1} u_n + \ldots$$

We thus proved our formula. Assuming now that $n = 0$ we arrive at the Newton interpolation formula

$$u_\lambda = u_o + (\lambda/1!)\,\Delta u_o + [\lambda(\lambda-1)/2!]\,\Delta^2 u_o + C_\lambda^3 \Delta^3 u_o + \ldots \qquad \textbf{(2)}$$

2.1.3. Suppose now that

$$u_o = f(a), \quad \Delta x = h, \quad u_\lambda = f(a + \lambda h).$$

Formula (2) will then become

$$f(a + \lambda h) = f(a) + (\lambda/1!)\Delta f(a) + [\lambda(\lambda-1)/2!]\,\Delta^2 f(a) + \ldots$$

Substituting $\lambda h = x$ we obtain for each x being a multiple of h

$$f(a + x) = f(a) + (x/1!)[\Delta f(a)/h] + [x(x-h)/2!][\Delta^2 f(a)/h^2] + \ldots \qquad \text{(iii)}$$

It is not difficult now to derive the Taylor series by assuming that h tends to zero so that, in the limit,

$$f(a + x) = f(a) + (x/1!)[df(a)/dx] + (x^2/2!)[d^2 f(a)/dx^2] + \ldots \qquad \text{(iv)}$$

where, in general, $d^k f(a)/dx^k$ denotes the value of $d^k f(x)/dx^k$ at $x = a$.
  Suppose that

$$F(x) = f(a) + (x/1!)[\Delta f(a)/h] + [x(x-h)/2!][\Delta^2 f(a)/h^2] + \ldots +$$
$$\frac{x(x-h)\ldots(x-kh+h)}{k!}\,\frac{\Delta^k f(a)}{h^k},$$

$$\Phi(x) = f(a) + (x/1!)[df(a)/dx] + (x^2/2!)[d^2 f(a)/dx^2] + \ldots +$$
$$(x^k/k!)[d^k f(a)/dx^k]$$

so that in general

$$f(a + x) - \Phi(x) = A' x^{k+1} + B' x^{k+2} + C' x^{k+3} + \ldots$$

Assume now that $y = f(a + x)$ is the equation of the curve MN and consider also curves MN' and $PABCDQ$ with equations $y = \Phi(x)$ and

$$y = F(x) \qquad\qquad\qquad\qquad (v)$$

(Figures 1 and 2). The curve MN' has a common point with the curve MN at $x = 0$ where the curves have an osculation with contact of the k-th order so that in its vicinity they generally very little differ from each other. And, if the expansion of the function $f(a + x)$ into powers of x is a convergent series, then, with a sufficiently large k, the curve MN may be replaced by the curve MN' which is what we are doing every time when we calculate the values of a function by expanding it into a Taylor series and neglecting the terms of the series beginning with some $(k + 1)$-th order.
  When interpolating, we replace a given function in exactly the same way by another one, – by a simpler one, – and the geometric sense of this substitution is that we replace a given curve MN by a "parabolic" curve PQ having equation (v). It is not difficult to see that this curve has $(k + 1)$ points in common with the curve MN, and, namely, points A, B, C, D, … with abscissas 0, $0a = h$, $0b = 2h$, …, $(k - 1)h$, kh.
  If a sufficiently small magnitude is chosen as h, the curve PQ will in general very little differ from the curve MN, and it is this fact that underlies the calculation of the values of the function $f(a + x)$ corresponding to the intermediate

values of x between 0 and $(k-1)h$. The calculation is replaced by the determination of the values of the function $F(x)$ corresponding to the same values of x.

It is seen now that interpolation has much in common with the calculation of the values of a function by expanding {it} into a series in powers of the variable. The difference consists in that in one case the given curve is replaced by another one that intersects it a known number of times at points whose abscissas increase in an arithmetic progression, whereas in the second instance the given curve is substituted by a curve having with it an osculation with contact of a certain order, but, in general, withdrawing from it afterwards.

2.1.4. In §2.1.2 we derived the Newton interpolation formula. We shall now derive another formula, and, to this end, we shall solve the following problem: To find the simplest polynomial, that, at various given values of the variable x, takes values identical to the known values of an {otherwise} unknown function $f(x)$.

We shall suppose that the given values of x are

$$x_1, \; x_2, \; \ldots, \; x_{n-1}, \; x_n \qquad\qquad\qquad\qquad\qquad \text{(vi)}$$

and that $f(x_1)$, $f(x_2)$, …, $f(x_{n-1})$, $f(x_n)$ are the known values of the function $f(x)$ whose form might be unknown. Since the sought polynomial ought to take n different values at n values of the variable, it should in general have at least n coefficients so that its form will be

$$\Phi(x) = A_n + A_{n-1}x + \ldots + A_1 x^{n-1}.$$

To determine these n coefficients it would suffice to substitute consecutively, in this equation, the values (vi) instead of x and to replace $\Phi(x_1)$, $\Phi(x_2)$, … by $f(x_1)$, $f(x_2)$, … This would have provided n equations necessary, and, in general, sufficient for determining the coefficients sought. However, we can also write out the unknown polynomial at once. It is not at all difficult to see that the polynomial

$$\begin{aligned}
\Phi(x) = {} & \frac{(x-x_2)(x-x_3)\ldots(x-x_n)}{(x_1-x_2)(x_1-x_3)\ldots(x_1-x_n)} f(x_1) + \\[2mm]
& \frac{(x-x_1)(x-x_3)\ldots(x-x_n)}{(x_2-x_1)(x_2-x_3)\ldots(x_2-x_n)} f(x_2) + \ldots + \\[2mm]
& \frac{(x-x_1)(x-x_2)\ldots(x-x_{n-1})}{(x_n-x_1)(x_n-x_2)\ldots(x_n-x_{n-1})} f(x_n)
\end{aligned} \qquad\qquad \textbf{(3)}$$

satisfies the formulated conditions and it only remains to show that it really is the simplest polynomial of the $(n-1)$-th degree {from among those} satisfying them.

Indeed, when assuming that $\psi(x)$ is the simplest polynomial of the same degree satisfying the demanded conditions, the difference $\Phi(x) - \psi(x)$ will represent a polynomial of a degree not higher than $(n-1)$.[1] It vanishes at n values of x which is only possible if $\psi(x)$ is identically equal to $\Phi(x)$. Thus, $\Phi(x)$ is the polynomial sought.

Equation (3) is called the Lagrange formula of interpolation. For example, let $x_1 = 0$, $x_2 = 1$, $x_3 = 2$ and $f(x_1) = 0$, $f(x_2) = 1$, $f(x_3) = 8$. In accord with the Lagrange formula we obtain

$$\begin{aligned}
\Phi(x) = {} & \frac{(x-1)(x-2)}{(0-1)(0-2)} \cdot 0 + \frac{(x-0)(x-2)}{(1-0)(1-2)} \cdot 1 + \frac{(x-0)(x-1)}{(2-0)(2-1)} \cdot 8 = \\
& 3x^2 - 2x.
\end{aligned}$$

In order to solve the same problem by the Newton formula, we note that in this case $u_0 = 1$, $u_1 = 1$, $u_2 = 8$; $\Delta u_0 = 1$, $\Delta u_1 = 7$, $\Delta u_2 = 6$,

$$\begin{aligned}
\Phi(x) = {} & u_0 + (x/1!)\,\Delta u_0, + [x(x-1)/2!]\,\Delta^2 u_0 = 0 + 1x + \\
& 6[x(x-2)/2!] = 3x^2 - 2x.
\end{aligned}$$

Note 1. Because this difference can only be of a degree higher than $(n-1)$ if $\psi(x)$ is {also} of a degree higher than $(n-1)$, but then $\psi(x)$ would not have been the simplest polynomial.

2.1.5. There also exists a method of interpolation based on replacing the unknown function $f(x)$ by a linear function $(A + Bx)$ most suitable to it. Suppose that we know the values of this function, $f(x_1)$, $f(x_2)$, …, $f(x_n)$, corresponding to the given values of the variable x. We take the function

$$u = [f(x_1) - A - Bx_1]^2 + [f(x_2) - A - Bx_2]^2 + \ldots +$$

$$[f(x_n) - A - Bx_n]^2$$

of two variables, A and B, and determine these latter according to the rules of the differential calculus by issuing from the condition that this function is minimal. The values of A and B thus obtained will indeed make the function $(A + Bx)$ the most suitable to $f(x)$, at least within the boundaries of the minimal and the maximal values of x_1, x_2, ..., x_n.

2.1.6. We are now going over to the determination of the finite differences of the simplest functions. This section corresponds to that on the differentials of the simplest functions in the differential calculus.

a) One of these functions there was x^m, but in the theory of finite differences it will not be simplest because its difference is represented by a series. So as to find out what function will here correspond to x^m, we compare the Newton formula (iii) with the Taylor series (iv). It is seen that the function

$$x(x - h)(x - 2h) \ldots [x - (m - 1)h]$$

corresponds to x^m. We shall now show that its difference has indeed a very simple form similar to the differential of x^m. We have

$$\Delta\{x(x - h)(x - 2h) \ldots [x - (m - 1)h]\} = (x + h)x(x - h) \ldots \\ [x - (m - 2)h] - x(x - h) \ldots [x - (m - 1)h] = \\ x(x - h) \ldots [x - (m - 2h)](x + h - x + mh - h).$$

And so

$$\Delta\{x(x - h)(x - 2h) \ldots [x - (m - 1)h]\} = \\ m\,h\,x(x - h)(x - 2h) \ldots [x - (m - 2h)]. \tag{4}$$

This formula is similar to the formula $dx^m = mdx \cdot x^{m-1}$ of the differential calculus. From (4) we find that

$$\Delta^2\{x(x - h) \ldots [x - (m - 1)h]\} = \\ m(m - 1)h^2 x(x - h) \ldots [x - (m - 3h)], \\ \Delta^3\{x(x - h) \ldots [x - (m - 2)h]\} = \\ m(m - 1)(m - 2)h^3 x(x - h) \ldots [x - (m - 4h)], \ldots,$$

$$\Delta^{m-1}\{x(x - h) \ldots [x - (m - 1)h]\} = m!\,h^{m-1}x, \\ \Delta^m\{x(x - h) \ldots [x - (m - 1)h]\} = m!\,h^m.$$

The next differences corresponding to a constant will identically be equal to zero. It is not difficult to derive on this basis the Newton formula. To this end let us assume the following expansion:

$$f(a + x) = A_0 + A_1 x + A_2 x(x - h) + A_3 x(x - h)(x - 2h) + \ldots$$

Supposing that here $x = 0$ we have $A_0 = f(a)$ and, in addition,

$$\Delta f(a + x) = A_1 h + 2A_2 hx + 3A_3 h x(x - h) + \ldots, \\ \Delta^2 f(a + x) = 1 \cdot 2 A_2 h^2 + 2 \cdot 3 A_3 h^2 x + \ldots, \quad \Delta^3 f(a + x) = 3!\,A_3 h^3 + \ldots$$

Assuming that in these equalities $x = 0$ we obtain

$$A_1 = \Delta f(a)/h, \quad A_2 = \Delta^2 f(a)/(2!\,h^2), \quad A_3 = \Delta^3 f(a)/(3!\,h^3), \ldots$$

hence the Newton formula (iii).

In the theory of finite differences, the function

$$\frac{1}{x(x + h)(x + 2h) \ldots (x + nh)},$$

whose difference is

$$\frac{1}{(x + h)(x + 2h) \ldots (x + nh)} - \frac{1}{x(x + h)(x + 2h) \ldots [x + (n - 1)h]} =$$

$$\frac{x - x - nh}{x\,(x+h)\,(x+2h)...(x+nh)} = -\,nh\ \frac{1}{x\,(x+h)...(x+nh)}, \tag{5}$$

corresponds to $1/x^n$. This formula is similar to the formula $d\,(1/x^n) = -\,n\,dx/x^{n+1}$.

Let us determine now the difference of a **fraction** in terms of the differences of its numerator and denominator:

$$\Delta\,(u_n/v_n) = u_{n+1}/v_{n+1} - u_n/v_n = \frac{(u_n + \Delta u_n)v_n - (v_n + \Delta v_n)u_n}{v_n v_{n+1}} =$$

$$\frac{v_n \Delta u_n - u_n \Delta v_n}{v_n v_{n+1}}. \tag{6}$$

This formula which is similar to $d\,(u/v) = (v\,du - u\,dv)/v^2$ is far less important in the practical sense than the latter.

b) The function a^x. We have

$$\Delta\,a^x = a^{x+h} - a^x = a^x\,(a^h - 1) = a^x h\,(a^h - 1)/h. \tag{7}$$

In order to go over from (7) to the {corresponding} formula of the differential calculus it only suffices to assume here that h tends to zero; noting also that

$$\lim\,[(a^h - 1)/h]_{\,h=0} = \ln a$$

we indeed have $d\,a^x = a^x \ln a\,dx$.

From (7) we obtain

$$\Delta^2\,a^x = (a^h - 1)\,\Delta\,a^x = a^x\,(a^h - 1)^2$$

and in general

$$\Delta^m\,a^x = a^x\,(a^h - 1)^m. \tag{8}$$

c) The function $\sin x$:

$$\Delta \sin x = \sin\,(x + h) - \sin x = 2\cos\,[x + (h/2)]\,\sin\,(h/2). \tag{9}$$

Noting that

$$\Delta \sin x = \cos\,[x + (h/2)]\,\frac{\sin h/2}{h/2}\,h$$

and assuming that h tends to zero, we have $d \sin x = \cos x\,dx$.

d) Let us now consider the function $\cos x$:

$$\Delta \cos x = \cos\,(x + h) - \cos x = -\,2\sin\,[(x + h)/2]\,\sin\,(h/2).$$

Thus,

$$\Delta \cos x = -\,\sin\,[(x + h)/2]\,2\,\sin\,(h/2). \tag{10}$$

On the basis of formulas (9) and (10) we find that

$$\Delta^2 \sin x = \Delta \cos\,[(x + h)/2]\,2\,\sin\,(h/2) = -\,\sin\,(x + h)\,[2\sin\,(h/2)]^2,$$
$$\Delta^3 \sin x = -\,\Delta\sin\,(x + h)\,[2\sin\,(h/2)]^2 = -\,\cos\,[x + (3h/2)]\,[2\sin\,(h/2)]^3, \ ...$$

and in general

$$\Delta^n \sin x = \sin\{x + [n(\pi + h)/2]\}\,[2\sin\,(h/2)]^n, \tag{11}$$
$$\Delta^n \cos x = -\,\cos\{x + [n(\pi + h)/2]\}\,[2\sin\,(h/2)]^n. \tag{12}$$

In conclusion, we shall derive the difference of **the product of two functions**:

$$\Delta(u_n\,v_n) = u_{n+1}\,v_{n+1} - u_n\,v_n = (u_n + \Delta u_n)\,v_{n+1} - u_n\,v_n =$$

$$u_n \, \Delta \, v_n \; + \; v_{n+1} \, \Delta \, u_n. \hspace{4cm} \textbf{(13)}$$

Like formula (6), this one is not really important.

2.1.7. We go over to consider a new section, the derivation of the dependences {connections} between finite differences and differentials. Supposing that $u = f(x)$, we have

$$\Delta u = f(x + h) - f(x) = (h/1!) f'(x) + (h^2/2!) f''(x) + \ldots =$$
$$(h/1!) \, du/dx \; + \; (h^2/2!) \, d^2u/dx^2 \; + \ldots$$

Thus,

$$\Delta u = (h/1!) \, du/dx \; + \; (h^2/2!) \, d^2u/dx^2 \; + \ldots, \hspace{2cm} \textbf{(14)}$$
$$\Delta^2 u = (h/1!) \, \Delta \, (du/dx) \; + \; (h^2/2!) \, \Delta \, (d^2u/dx^2) + \ldots$$

Replacing u consecutively by du/dx, d^2u/dx^2 etc in formula (14) we obtain

$$\Delta \, (du/dx) = (h/1!) \, d^2u/dx^2 \; + \; (h^2/2!) \, d^3u/dx^3 \; + \ldots,$$
$$\Delta \, (d^2u/dx^2) = (h/1!) \, d^3u/dx^3 \; + \; (h^2/2!) \, d^4u/dx^4 \; + \ldots,$$
$$\Delta \, (d^3u/dx^3) = (h/1!) \, d^4u/dx^4 \; + \ldots$$

so that

$$\Delta^2 u = (h/1!) \, [(h/1!) \, d^2u/dx^2 \; + \; (h^2/2!) \, d^3u/dx^3 \; + \ldots] +$$
$$(h^2/2!) \, [(h/1!) \, d^3u/dx^3 \; + \; (h^2/2!) \, d^4u/dx^4 \; + \ldots] +$$
$$(h^3/3!) \, [(h/1!) \, d^4u/dx^4 \; + \ldots] + \ldots =$$
$$h^2 \, d^2u/dx^2 \; + \; h^3 \, d^3u/dx^3 \; + \; (7/12) \, h^4 \, d^4u/dx^4 \; + \ldots =$$
$$A \, d^2u/dx^2 \; + \; B \, d^3u/dx^3 \; + \; C \, d^4u/dx^4 \; + \ldots$$

It is seen now that in general

$$\Delta^\lambda u = A_1 du/dx \; + \; A_2 \, d^2u/dx^2 \; + \ldots + A_{\lambda-1} \, d^{\lambda-1}u/dx^{\lambda-1} + A_\lambda \, d^\lambda u/dx^\lambda \; + \ldots$$

where A_1, A_2, … are magnitudes *not dependent* on the form of the function $f(x) = u$. Using this fact, we can easily determine these coefficients. Since $\Delta^\lambda u$ is here represented by a linear and homogeneous function of u and its derivatives, it only suffices to choose $f(x)$ in such a way that both its finite differences and derivatives were of the simplest form; and we saw that a^x was such a function. And so, suppose that $u = a^x$, then

$$\Delta^\lambda u = a^x (a^h - 1)^\lambda, \quad d^m u/dx^m = a^x (\ln a)^m.$$

Consequently, we find that

$$a^x (a^h - 1)^\lambda = A_1 a^x \ln a \; + \; A_2 a^x (\ln a)^2 \; + \; A_3 a^x (\ln a)^3 \; + \ldots$$

or

$$(a^h - 1)^\lambda = A_1 \ln a \; + \; A_2 (\ln a)^2 \; + \; A_3 (\ln a)^3 \; + \ldots$$

Assuming here $\ln a = s$ and noting that

$$e^{hs} = 1 \; + \; (hs/1!) \; + \; (h^2 s^2/2!) \; + \ldots$$

we obtain the following equality

$$[(hs/1!) \; + \; (h^2 s^2/2!) \; + \ldots]^\lambda = A_1 s \; + \; A_2 s^2 \; + \ldots + A_\lambda s^\lambda \; + \ldots$$

from which we shall indeed determine the coefficients A_1, A_2, …by making use of its being an identity. Thus, it is not difficult to see that

$$A_1 = A_2 = \ldots A_{\lambda-1} = 0$$

so that

$$\Delta^\lambda u = A_\lambda \, d^\lambda u/dx^\lambda + A_{\lambda+1} \, d^{\lambda+1} u/dx^{\lambda+1} + \dots$$

and

$$[(hs/1!) + (h^2 s^2/2!) + \dots]^\lambda = A_\lambda \, s^\lambda + A_{\lambda+1} \, s^{\lambda+1} + \dots \tag{15}$$

It is not difficult to conclude now that if u is an integral function of a power not higher than $(\lambda - 1)$, all of its derivatives beginning with those of the λ-th order, and all of its differences beginning with those of the same order, are identically equal to zero.

We are now going over to the solution of the inverse problem: To express the derivatives of any order through the differences. Suppose that we found that

$$\Delta u = C_1 \, du/dx + C_2 \, d^2 u/dx^2 + \dots,$$
$$\Delta^2 u = D_2 \, d^2 u/dx^2 + D_3 \, d^3 u/d^3 x + \ ,$$
$$\Delta^3 u = E_3 \, d^3 u/dx^3 + E_4 \, d^4 u/d^4 x + \ , \dots$$

then

$$du/dx = (1/C_1) \, \Delta u - (C_2/C_1) \, (d^2 u/dx^2) - (C_3/C_1) \, (d^3 u/dx^3) - \dots,$$
$$d^2 u/dx^2 = (1/D_2) \, \Delta^2 u - (D_3/D_2) \, (d^3 u/dx^3) - (D_4/D_2) \, (d^4 u/dx^4) - \dots,$$
$$d^3 u/dx^3 = (1/E_3) \, \Delta^3 u - (E_4/E_3) \, (d^4 u/dx^4) - \dots$$

It is seen now that in general

$$d^\mu u/dx^\mu = N_\mu \Delta^\mu u + N_{\mu+1} \Delta^{\mu+1} u + \dots \tag{16}$$

In order to determine the coefficients of this series we assume that $u = a^x$ so that

$$(\ln a)^\mu = \ = N_\mu (a^h - 1)^\mu + N_{\mu+1} (a^h - 1)^{\mu+1} + \dots$$

Substituting here $(a^h - 1) = s$ and noting that

$$\ln (1 + s) = (s/1) - (s^2/2) + (s^3/3) - \dots$$

we ought to have such an identity:

$$(\Delta/h^\mu) \, [(s/1) - (s^2/2) + (s^3/3) - \dots]^\mu =$$
$$N_\mu s^\mu + N_{\mu+1} s^{\mu+1} + \dots$$

from which we shall indeed determine the coefficients N. Thus we have (16) and [1]

$$\left(\frac{\ln(1+s)}{h} \right)^\mu = N_\mu s^\mu + N_{\mu+1} s^{\mu+1} + N_{\mu+2} s^{\mu+2} + \dots \tag{17}$$

Supposing that $\mu = 1$ we obtain

$$(s/1h) - (s^2/2h) + (s^3/3h) - \dots = N_1 s + N_2 s^2 + N_3 s^3 + \dots$$

so that

$$N_1 = 1/h, \ N_2 = - [1/(2h)], \ N_3 = 1/(3h), \dots$$

and

$$du/dx = \Delta u \, / \, h - \Delta^2 u \, / \, 2h + \Delta^3 u / 3h - \dots \tag{18}$$

Formulas (16) and (17) can be represented symbolically and they will {then} be easier to memorize. Considering u in (16) as a factor and Δ as some magnitude, we may write them in the following way:

$$d^\mu u/dx^\mu = (N_\mu \Delta^\mu + N_{\mu+1} \Delta^{\mu+1} + N_{\mu+2} \Delta^{\mu+2} + \dots) \, u,$$

$$\left(\frac{\ln(1+\Delta)}{h}\right)^{\mu} = N_{\mu}\Delta^{\mu} + N_{\mu+1}\Delta^{\mu+1} + N_{\mu+2}\Delta^{\mu+2} + \dots$$

so that

$$d^{\mu}u/dx^{\mu} = \left(\frac{\ln(1+\Delta)}{h}\right)^{\mu}u. \tag{19}$$

Another symbolic formula replacing (15) can be deduced symbolically from (19). To this end we shall consider there u as a factor, μ as a power and d/dx as a magnitude, so that we obtain

$$d/dx = \ln(1+\Delta)/h, \tag{vii}$$

hence $\Delta = (e^{h\,d/dx} - 1)$ and

$$\Delta^{\lambda}u = (e^{h\,d/dx} - 1)^{\lambda}u. \tag{20}$$

In order to provide one more example of the "symbolic" method of deriving symbolic formulas, we shall obtain, issuing from (20), a formula showing the dependence between the difference of any order corresponding to a given increment of the independent variable, and the differences corresponding to another increment. Let

$$\Delta u = f(x+h) - f(x) \text{ and } \Delta_1 u = f(x+H) - f(x).$$

From (19) we had, symbolically, (vii). In the same way, from

$$\Delta_1^{\lambda}u = (e^{H\,d/dx} - 1)^{\lambda}u$$

we obtain $d/dx = \ln(1+\Delta_1)/H$ so that

$$(1+\Delta)^{1/h} = (1+\Delta_1)^{1/H},$$

hence $\Delta_1 = (1+\Delta)^{H/h} - 1$ and

$$\Delta_1^{\lambda}u = [(1+\Delta)^{H/h} - 1]^{\lambda}u. \tag{21}$$

Thus, for example, setting $\lambda = 2$, $h = 1$, $H = 2$, we shall find that

$$\Delta_1^{2}u = [(1+\Delta)^{2} - 1]^{2}u = \Delta^{4}u + 4\Delta^{3}u + 4\Delta^{2}u. \tag{22}$$

Suppose that we have the following table (Table 1) for the increment $h = 1$. Then, for the increment $H = 2$ we obtain Table 2, and from the equation (22) it follows that

$$\Delta_1^{2}u = 0 + 4.6 + 4.12 = 72,$$

a result coinciding with that provided by Table 2.

Table 1

u	Δu	$\Delta^2 u$	$\Delta^3 u$	$\Delta^4 u$
1	7	12	6	0
8	19	18	6	
27	37	24		
64	61			
125				

Table 2

u	$\Delta_1 u$	$\Delta_1^2 u$	$\Delta_1^3 u$	$\Delta_1^4 u$
1	26	72	48	0
27	98	120	48	
125	218	168		
343	386			
729				

Formula (21) transforms into (19) if H becomes infinitesimal:

$$\left(\frac{\Delta_1^{\lambda}u}{H^{\lambda}}\right)_{H=0} = \left(\frac{(1+\Delta)^{H/h}-1}{H}\right)^{\lambda}u = \left(\left[\frac{(1/h)\ln(1+\Delta)(1+\Delta)^{H/h}}{1}\right]^{\lambda}\right)_{H=0} =$$

$$\left(\frac{\ln(1+\Delta)}{h}\right)^{\lambda}u,$$

but

$$\lim \left(\frac{\Delta_1^{\lambda} u}{H^{\lambda}} \right)_{H=0} = \lim [(\Delta^{\lambda} u / (\Delta x)^{\lambda}] = d^{\lambda} u/dx^{\lambda}$$

so that we find

$$d^{\lambda} u/dx^{\lambda} = \left(\frac{\ln(1+\Delta)}{h} \right)^{\lambda} u = \left(\frac{\ln(1+\Delta_1)}{H} \right)^{\lambda} u.$$

Note 1. {The left side of the identity above is apparently written wrongly. Moreover, Chebyshev wrote out here the equality (16) for the second time without indicating that it was already provided somewhat above, and it is this equality that is really needed.}

2.2. The Inverse Calculus of Finite Differences

2.2.1. We are going over to a section of the theory of finite differences similar to the integral calculus in the doctrine of infinitesimals; indeed, we shall now determine the function given its differences.

Suppose that $\Delta u_x = v_x$ where the subscripts show the value of the variable to which the value of the function is corresponding; or, in a simpler way, $\Delta u = v$. We shall show that all the functions satisfying this equation can differ only by a constant for those intervals at whose ends {for those values of x for which} the values of the function are taken. Indeed; suppose that some function w {also} satisfies this equation, then $\Delta w = v$ and $\Delta u - \Delta w = \Delta (u - w) = 0$. But the difference of a function can equal zero only when it takes one and the same value for all the values of the independent variable following each other after equal intervals which we assume as the constant increment $h = \Delta x$ of the variable. In other words, we may then regard the function as a constant; this follows from the fact that the values of a function corresponding to the intermediate values of the variable are not at all considered in the theory of finite differences.

We thus have $u - w = C$ where C should not be considered as an "absolutely" constant magnitude; because of the said just above, it can depend on such a function that takes one and the same value at the values of x differing one from another by a constant $h = \Delta x$. Thus, if $h = 1$, C can equal $\sin(2\pi x)$, $\cos(2\pi x)$, etc., because, in general, $\sin(2\pi x_o) = \sin[2\pi(x_o + n)]$ where x_o is the initial value of the variable and n is an integer. However, we do not need intermediate values and we shall consider C as a constant.

It is not difficult to show now that the function

$$w_x = v_m + v_{m+1} + v_{m+2} + \dots + v_{x-1}$$

satisfies our equation. Indeed,

$$\Delta w_x = v_m + v_{m+1} + v_{m+2} + \dots + v_{x-1} + v_x - (v_m + v_{m+1} \dots + v_{x-1}) = v_x.$$

We shall denote the sum

$$v_m + v_{m+1} + v_{m+2} + \dots + v_{x-1} = \sum_m^x v.$$

Therefore, if $\Delta u = v$, $u = \sum_m^x v + C.$

If $x = m$, noticing that $\sum_m^m v = 0$, we find that $u_m = C$ and

$$\sum_m^x v = u_x - u_m.$$

It is not difficult to see that

$$\sum_m^x (v \pm w) = \sum_m^x v \pm \sum_m^x w$$

and

$$\sum_m^x Av = A \sum_m^x v$$

where A is a constant magnitude.

2.2.2. We shall now try to determine the sums of some simplest functions assuming that $h = 1$.
a) First, let

$$u = x(x - 1)(x - 2) \dots (x - l + 1),$$

then

$$\Delta u = l\, x(x - 1)(x - 2) \dots (x - l + 2)$$

and

$$\Delta (u/l) = \frac{\Delta\, x(x-1)(x-2)\dots(x-l+1)}{l} = x(x - 1)(x - 2) \dots (x - l + 2).$$

Denote now $l - 1 = n$ so that

$$\frac{\Delta\, x(x-1)(x-2)\dots(x-n)}{n+1} = x(x - 1)(x - 2) \dots (x - n + 1)$$

and

$$\sum x(x - 1) \dots (x - n + 1) = \frac{x(x-1)(x-2)\dots(x-n)}{n+1} + C$$

because in general $\sum \Delta u = u + C$.

Consequently, we also have

$$\sum_{m}^{x} x(x - 1)\dots(x - n + 1) = \frac{x(x-1)\dots(x-n) - m(m-1)\dots(m-n)}{n+1}. \qquad (23)$$

Assuming here $m = 0$ we obtain

$$\sum_{0}^{x} x(x - 1)\dots(x - n + 1) = \frac{x(x-1)(x-2)\dots(x-n)}{n+1}. \qquad (24)$$

These formulas are similar to the following formulas of the integral calculus:

$$\int x^n\, dx = \frac{x^{n+1}}{n+1} + C, \quad \int_0^x x^n\, dx = \frac{x^{n+1}}{n+1}.$$

b) The sums of the type

$$\sum x^m \qquad\qquad\qquad \text{(viii)}$$

are expressed by very involved formulas and are therefore usually determined by reducing the calculation to the computation of

$$\sum x^{m-1}.$$

Before going on to these calculations we shall show now another method of computation similar to the approximate integration, but instead of the Taylor series applied in the integral calculus we shall, however, use the Newton interpolation formula, cf. (2),

$$u = u_0 + (x/1!)\, \Delta u + [x(x - 1)/2!]\, \Delta^2 u + \\ [x(x - 1)(x - 2)/3!]\, \Delta^3 u + \dots$$

As an illustration, we shall thus find the sum

$$\sum_{0}^{x} x^3.$$

We shall have Table 3 so that $u_0 = 0$, $\Delta u_0 = 1$, $\Delta^2 u = 6$, $\Delta^3 u = 6$, $\Delta^4 u = \dots = 0$ and

$$x^3 = x + 3x(x - 1) + x(x - 1)(x - 2).$$

Table 3

x	u	Δu	$\Delta^2 u$	$\Delta^3 u$	$\Delta^4 u$
0	0	1	6	6	0
1	1	7	12	6	
2	8	19	18		
3	27	37			
4	64				

Consequently,

$$\sum_0^x x^3 = \sum_0^x x + 3\sum_0^x x(x-1) + \sum_0^x x(x-1)(x-2).$$

But

$$\sum_0^x x = \frac{x(x-1)}{2}, \quad \sum_0^x x(x-1) = \frac{x(x-1)(x-2)}{3},$$

$$\sum_0^x x(x-1)(x-2) = \frac{x(x-1)(x-2)(x-3)}{4}$$

and therefore

$$\sum_0^x x^3 = \frac{x(x-1)}{2} + 3\frac{x(x-1)(x-2)}{3} + \frac{x(x-1)(x-2)(x-3)}{4} =$$

$$\frac{x(x-1)}{2} \cdot \frac{x(x-1)}{2}.$$

And so

$$\sum_0^x x^3 = \left(\frac{x(x-1)}{2}\right)^2 = \left(\sum_0^x x\right)^2. \tag{25}$$

This remarkable formula shows that

$$1^3 + 2^3 + \ldots + N^3 + = (1 + 2 + \ldots + N)^2.$$

We are now going on to the abovementioned method of **calculating the sums** of the type of (viii). We have

$$\Delta x^m = (x+1)^m - x^m = (m/1!)x^{m-1} + [m(m-1)/2!]x^{m-2} + \ldots + mx + 1$$

so that

$$x^m = m\sum_0^x x^{m-1} + [m(m-1)/2!]\sum_0^x x^{m-2} + \ldots + m\sum_0^x x + x$$

because $\Delta x = (x+1) - x = 1$ and

$$\sum_0^x 1 = x.$$

Therefore

$$m\sum_0^x x^{m-1} = x^m - \{[m(m-1)/2!]\sum_0^x x^{m-2} + $$

$$C_m^3 \sum_0^x x^{m-3} + \ldots + m\sum_0^x x + x\}$$

and

$$\sum_0^x x^{m-1} = (x^m/m) - [(m-1)/2!]\sum_0^x x^{m-2} - $$

$$[(m-1)(m-2)/3!]\sum_0^x x^{m-3} - \ldots - \sum_0^x x - (1/m)x.$$

Setting here $m = n + 1$ we shall indeed arrive at the formula sought:

$$\sum_0^x x^n = [x^{n+1}/(n+1)] - (n/2!) \sum_0^x x^{n-1}$$

$$- [n(n-1)/3!] \sum_0^x x^{n-2} - \ldots - \sum_0^x x - [1/(n+1)x. \tag{26}$$

Assuming that $n = 1$ we obtain

$$\sum_0^x x = (x^2/2) - (1/2) \sum_0^x 1 = [(x^2 - x)/2] = [x(x-1)/2].$$

If $n = 2$ we shall find that

$$\sum_0^x x^2 = (x^3/3) - \sum_0^x x - (x/3) = (x^3/3) - [x(x-1)/2] - (x/3) =$$

$$[x(2x^2 - 3x + 1)/6]$$

and if $n = 3$

$$\sum_0^x x^3 = (x^4/4) - (3/2) \sum_0^x x^2 - \sum_0^x x - (x/4) =$$

$$(x^4/4) - [x(2x^2 - 3x + 1)/4] - x(x-1)/2 - (x/4) = \left(\frac{x(x-1)}{2}\right)^2 \text{ etc.}$$

Formula (26) thus enables us to calculate consecutively the sums of the type (viii).
c) Suppose now that

$$u = \frac{1}{x(x+1)(x+2)\ldots(x+l-1)}.$$

According to formula (5) we have then

$$\Delta u = \frac{l}{x(x+1)(x+2)\ldots(x+l)}$$

or

$$\Delta\left[-\frac{1}{l\,x(x+1)(x+2)\ldots(x+l-1)}\right] = \frac{1}{x(x+1)(x+2)\ldots(x+l)}.$$

Assuming that $l = n - 1$ we obtain

$$\frac{1}{x(x+1)(x+2)\ldots(x+n-1)} = \Delta\left[-\frac{1}{(n-1)\,x(x+1)(x+2)\ldots(x+n-2)}\right].$$

Consequently

$$\sum \frac{1}{x(x+1)(x+2)\ldots(x+n-1)} =$$

$$-\frac{1}{(n-1)\,x(x+1)(x+2)\ldots(x+n-2)} + C,$$

hence

$$\sum_m^x \frac{1}{x(x+1)(x+2)\ldots(x+n-1)} = -\frac{1}{(n-1)\,x(x+1)(x+2)\ldots(x+n-2)} +$$

$$\frac{1}{n-1}\cdot\frac{1}{m(m+1)\ldots(m+n-2)}, \tag{27}$$

$$\sum_m^\infty \frac{1}{x(x+1)(x+2)\ldots(x+n-1)} = \frac{1}{n-1}\cdot\frac{1}{m(m+1)\ldots(m+n-2)}. \tag{28}$$

These formulas are similar to the formulas of the integral calculus

$$\int \frac{dx}{x^n} = \frac{-1}{n+1} \cdot \frac{1}{x^{n-1}} + C, \quad \int_m^\infty \frac{dx}{x^n} = \frac{1}{n-1} \cdot \frac{1}{m^{n-1}}.$$

d) Suppose now that $u = a^x$. Then

$$\Delta u = a^x (a - 1), \quad \Delta [u/(a - 1)] = a^x$$

so that

$$\sum a^x = \frac{a^x}{a-1} + C$$

and therefore

$$\sum_m^x a^x = \frac{a^x - a^m}{a-1}. \tag{29}$$

This formula is similar to the formula of the integral calculus

$$\int_m^x a^x \, dx = \frac{a^x - a^m}{\ln a}.$$

e) We shall also determine the sums of the trigonometric functions $\sin x$ and $\cos x$. We have

$$\Delta \cos x = - 2 \sin (h/2) \sin [x + (h/2)]$$

so that

$$\Delta \frac{\cos x}{- 2\sin(h/2)} = \sin [x + (h/2)].$$

Supposing that $x + (h/2) = z$ we obtain

$$\Delta\left\{- \frac{\cos[z - (h/2)]}{2\sin(h/2)}\right\} = \sin z$$

and, setting $z = a + by$, we arrive at

$$\Delta \frac{- \cos [a + by - (h/2)]}{2 \sin (h/2)} = \sin (a + by).$$

Assuming now that $\Delta y = 1$ we have

$$h = \Delta x = \Delta z = \Delta (x + by) = b \Delta y = b$$

and therefore

$$\Delta \frac{- \cos[a + by - (b/2)]}{2\sin(b/2)} = \sin (a + by).$$

Hence

$$\sum \sin (a + by) = \frac{- \cos[a + by - (b/2)]}{2\sin(b/2)} + C \tag{30}$$

or

$$\sum 2\sin(b/2)\sin(a + by) = -\cos[a + by - (b/2)] + C_1.$$

When substituting here $by = x$, assuming that b decreases to zero and y increases to infinity in such a way that x remains finite, we transform this formula into the known formula

$$\int \sin(a + x)\,dx = -\cos(a + x) + C_1$$

of the integral calculus because

$$2\sin(b/2)_{b=0} = b_{b=0} = [b(y + 1) - by]_{b=0} = (b\,\Delta y)_{b=0} = dx.$$

The formula for $\cos x$ can be obtained from (30) if we set $a = (\pi/2) + f$. This transforms (30) into

$$\sum \cos(f + by) = \frac{\sin[f + by - (b/2)]}{2\sin(b/2)} + C. \qquad (31)$$

2.2.3. In most cases, summation, like integration, cannot be accomplished precisely and we therefore need methods enabling us to sum approximately. We are now indeed going over to describing these methods. We have the formula (14), that, as we saw, is included in the general symbolic formula, cf. (20),

$$\Delta^n u = (e^{h\,d/dx} - 1)^n u.$$

Setting $h = 1$ we obtain

$$\Delta u = du/dx + (1/2)\,d^2u/dx^2 + [1/(2\cdot3)]\,d^3u/dx^3 + \ldots$$

so that

$$u = \sum du/dx + (1/2)\sum d^2u/dx^2 + [1/(2\cdot3)]\sum d^3u/dx^3 + \ldots + C'$$

where C' is a general arbitrary constant.

Suppose now that $du/dx = v$. Accordingly, our series will become

$$\int v\,dx = \sum v + (1/2)\sum du/dx + [1/(2\cdot3)]\sum d^2u/dx^2 + \ldots + C',$$

$$\sum v = \int v\,dx - (1/2)\sum du/dx - [1/(2\cdot3)]\sum d^2u/dx^2 + \ldots + C'. \qquad (32)$$

Substituting here dv/dx instead of v we shall find that

$$\sum du/dx = v - (1/2)\sum d^2u/dx^2 - [1/(2\cdot3)]\sum d^3u/dx^3 - \ldots$$

In the same way

$$\sum d^2v/dx^2 = dv/dx - (1/2)\sum d^3v/dx^3 - [1/(2\cdot3)]\sum d^4v/dx^4 - \ldots,$$

$$\sum d^3v/dx^3 = d^2v/dx^2 - (1/2)\sum d^4v/dx^4 - [1/(2\cdot3)]\sum d^5v/dx^5 - \ldots$$

etc.

Inserting the values of these sums into the expression (32) we shall obtain, in general,

$$\sum v = \int v\,dx + A_o v + A_1\,dv/dx + A_2\,d^2v/dx^2 + \ldots + C.$$

In order to determine the coeffficients which, as is not difficult to see, do not depend on the type of the function v, we set $v = a^x$. Therefore, if $-\infty$ and x are taken as the limits of summation and integration.

$$[a^x/(a - 1)] = [a^x/\ln a] + A_o\,a^x + A_1\,a^x\ln a + A_2\,a^x(\ln a)^2 + \ldots$$

It follows, when setting $x = 0$, that

$[1/(a - 1)] = (1/\ln a) + A_1 \ln a + A_2 (\ln a)^2 + \ldots$

And, if $\ln a = \alpha$,

$[1/(e^\alpha - 1)] = (1/\alpha) + A_1 \alpha + A_2 \alpha^2 + \ldots + A_k \alpha^k + \ldots$

This identity indeed enables us to determine the coefficients A_1, A_2, ..., A_k, ... Let us calculate some of them. We have

$e^\alpha - 1 = \alpha + (\alpha^2/2!) + (\alpha^3/3!) + \ldots$

Dividing 1 by this series we obtain

$[1/(e^\alpha - 1)] = (1/\alpha) - (1/2) + (1/12)\alpha + 0\,\alpha^2 + \ldots$

and thus $A_o = -(1/2)$, $A_1 = (1/12)$, $A_2 = 0$, ... It is easy to show that $A_4 = A_6 = \ldots = A_{2n} = 0$. Indeed,

$[1/(e^\alpha - 1)] = (1/\alpha) - (1/2) + A_1 \alpha + A_2 \alpha^2 + \ldots$

therefore

$[1/(e^\alpha - 1)] + (1/2) = (1/\alpha) + A_1 \alpha + A_2 \alpha^2 + \ldots$

but

$[1/(e^\alpha - 1)] + (1/2) = (1/2)\dfrac{e^\alpha + 1}{e^\alpha - 1} = (1/2)\dfrac{e^{\alpha/2} + e^{-\alpha/2}}{e^{\alpha/2} - e^{-\alpha/2}}$

so that

$(1/2)\dfrac{e^{\alpha/2} + e^{-\alpha/2}}{e^{\alpha/2} - e^{-\alpha/2}} = (1/\alpha) + A_1 \alpha + A_2 \alpha^2 + \ldots$

Now, the left side of this equality is an odd function. Consequently, the right side should also be odd and

$A_2 \alpha^2 + A_4 \alpha^4 + \ldots = 0.$

Since this is an identity, $A_2 = A_4 = \ldots = A_{2n} = 0$. Thus we have

$$\sum v = \int v\,dx - (1/2)v + A_1\,dv/dx + A_3\,d^3v/dx^3 + \ldots + C \qquad\qquad (33)$$

where the coefficients are determined by the equality

$(1/2)\dfrac{e^{\alpha/2} + e^{-\alpha/2}}{e^{\alpha/2} - e^{-\alpha/2}} = (1/\alpha) + A_1 \alpha + A_3 \alpha^3 + \ldots$

or by

$[1/(e^\alpha - 1)] + (1/2) = (1/\alpha) + A_1 \alpha + A_3 \alpha^3 + \ldots$

Suppose for example that $v = x$, then

$$\sum x = \int x\,dx - (1/2)x + (1/12) + C = [x^2 - x)/2] + (1/12) + C$$

and

$$\sum_1^n x = 1 + 2 + 3 + \ldots + (n - 1) = [n(n - 1)/2].$$

As a second example, let $v = x^3$, then

$$\sum x^3 = \int x^3\,dx - (1/2)x^3 + (1/4)x^2 + 6A_3 + C =$$

$$[(x^4 - 2x^3 + x^2)/4] + 6A_3 + C$$

and

$$\sum_{1}^{n} x^3 = [n^4 - 2n^3 + n^2)/4] = [n^2(n-1)^2/4] = [n(n-1)/2]^2 = \left(\sum_{1}^{n} x\right)^2.$$

2.2.4. Suppose now that $A_1 = B_1/2!$, $A_3 = -B_2/4!$ and that in general

$$A_{2\lambda+1} = (-1)^\lambda B_{\lambda+1}/[2(\lambda+1)]!.$$

The magnitudes B_1, B_2, ... are called the Bernoulli numbers. We shall now derive a formula for determining them and it will also enable us to make some general conclusions about them. These numbers should obviously satisfy the equality

$$(1/2)\frac{e^{\alpha/2} + e^{-\alpha/2}}{e^{\alpha/2} - e^{-\alpha/2}} = (1/\alpha) + (B_1/2!)\,\alpha - (B_2/4!)\,\alpha^3 + ... +$$

$$(-1)^\lambda \frac{B_{\lambda+1}}{[2(\lambda+1)]!}\,\alpha^{2\lambda+1}.$$

However, we have in general

$$\sin\psi = \psi[1 - (\psi^2/\pi^2)][1 - (\psi^2/2^2\pi^2)][1 - (\psi^2/3^2\pi^2)]\ ...$$

and therefore

$$\ln\sin\psi - \ln\psi = \ln[1 - (\psi^2/\pi^2)] + \ln[1 - (\psi^2/2^2\pi^2)] + ...$$

Differentiating this equality, we obtain after simplification

$$\operatorname{ctg}\psi = \frac{1}{\psi} - \frac{2\psi}{\pi^2 - \psi^2} - \frac{2\psi}{2^2\pi^2\psi^2} - \frac{2\psi}{3^2\pi^2\psi^2} - ...$$

Denote for the sake of brevity $\sqrt{-1} = i$, then

$$\operatorname{ctg}\psi = (\cos\psi\,/\sin\psi) = i\frac{e^{i\psi} + e^{-i\psi}}{e^{i\psi} - e^{-i\psi}}$$

so that, substituting $2\psi = \alpha\,i$, we arrive at

$$i\,\frac{e^{\alpha/2} + e^{-\alpha/2}}{e^{\alpha/2} - e^{-\alpha/2}} = (2/\alpha\,i) -$$

$$\alpha i\left(\frac{1}{\pi^2 + \alpha^2/4} + \frac{1}{2^2\pi^2 + \alpha^2/4} + \frac{1}{3^2\pi^2 + \alpha^2/4} + ...\right)\ ,$$

$$(1/2)\frac{e^{\alpha/2} + e^{-\alpha/2}}{e^{\alpha/2} - e^{-\alpha/2}} = (1/\alpha) +$$

$$\frac{\alpha}{2}\left(\frac{1/\pi^2}{1 + (\alpha/2)^2(1/\pi^2)} + \frac{1/2^2\pi^2}{1 + (\alpha/2)^2(1/2^2\pi^2)} + \frac{1/3^2\pi^2}{1 + (\alpha/2)^2(1/3^2\pi^2)} + ...\right).$$

However, in general,

$$[1/(1 + x)] = 1 - x + x^2 - ... + (-1)^n x^n + ...$$

and consequently

$$\frac{1/\pi^2}{1+(\alpha/2)^2/\pi^2} = (1/\pi^2)\left(1-\frac{\alpha^2}{4\pi^2}+\frac{\alpha^4}{2^4\pi^4}-...+(-1)^n\frac{\alpha^{2n}}{2^{2n}\pi^{2n}}+...\right),$$

$$\frac{1/2^2\pi^2}{1+(\alpha/2)^2/2^2\pi^2} =$$

$$(1/2^2\pi^2)\left(1-\frac{\alpha^2}{2^22^2\pi^2}+\frac{\alpha^4}{2^42^4\pi^4}-...+(-1)^n\frac{\alpha^{2n}}{2^{2n}2^{2n}\pi^{2n}}+...\right),$$

$$\frac{1/3^2\pi^2}{1+(\alpha/2)^2/3^2\pi^2} =$$

$$(1/3^2\pi^2)\left(1-\frac{\alpha^2}{2^23^2\pi^2}+\frac{\alpha^4}{2^43^4\pi^4}-...+(-1)^n\frac{\alpha^{2n}}{2^{2n}3^{2n}\pi^{2n}}+...\right).$$

We thus obtain

$$(1/2)\frac{e^{\alpha/2}+e^{-\alpha/2}}{e^{\alpha/2}-e^{-\alpha/2}} = (1/\alpha) + (\alpha/2\pi^2)[1 + (1/2^2) + (1/3^2) + ...] -$$
$$(\alpha^3/2^3\pi^4)[1 + (1/2^4) + (1/3^4) + ...] +$$
$$(\alpha^5/2^5\pi^6)[1 + (1/2^6) + (1/3^6) + ...] + ...$$
$$(-1)^\lambda\frac{\alpha^{2\lambda+1}}{2^{2\lambda+1}\pi^{2\lambda+2}}[1 + (1/2)^{2\lambda+2} + (1/3^{2\lambda+2}) + ...].$$

Comparing this result with the previous one we get the following series for the Bernoulli numbers:

$$B_1 = (2!/2\pi^2)S_2, \quad B_2 = (4!/2^3\pi^4)S_4, \quad ..., \quad B_{\lambda+1} = \frac{[2(\lambda+1)]!}{2^{\lambda+1}\pi^{2\lambda+2}}S_{2\lambda+2},$$

$$S_{2\lambda+2} = 1 + (1/2^{2\lambda+2}) + (1/3^{2\lambda+2}) + ...$$

Details concerning these series can be found in Ostrogradsky's "Sur les quadratures définies".

2.2.5. We had to do with series in which the difference Δx was assumed to be 1; wishing to introduce any difference h, we would only have to set $x = z/h$ so that, when x changed by 1, z would have changed by h, and we would have obtained

$$\sum v = \int v\, dz/h - (1/2)v + A_1\, dv/d(z/h) + A_2\, d^2v/d(z/h)^2 + ...$$

or

$$\sum v = (1/h)\int v\, dz - (1/2)v + hA_1\, dv/dz + h^2A_2\, d^2v/dz^2 + ... \qquad (34)$$

Therefore

$$\sum v = \sum vh - h[(1/2)v - hA_1\, dv/dz - h^2A_2\, d^2v/dz^2 - ...].$$

Supposing that h is here tending to zero, we get

$$\lim[h\sum v] = \lim \sum vh = \int v\, dz.$$

Thus, the integral is the limit of the product of the sum by the increment of the independent variable.

2.2.6. Noticing that $A_1 = (1/12)$, we have

$$\sum u = \int u\, dx - (1/2)u + (1/12)\, du/dx + C.$$

Assuming that $u = \ln x$ and assigning 1 and x as the limits for x, we obtain

$$\sum_1^x \ln x = C + \int \ln x\, dx - (1/2) \ln x + (1/12)\, d \ln x/dx + \ldots$$

or

$$\sum_1^x \ln x = C + x \ln x - x - (1/2) \ln x + (1/12)(1/x) + \ldots$$

where C already has an absolutely definite value which we shall indeed try to determine. We may express this equality in the following form

$$\ln 1 + \ln 2 + \ln 3 + \ldots + \ln (x - 1) =$$
$$C + x \ln x - x - (1/2) \ln x + (1/12)(1/x) + \ldots,$$

$$\ln x! = C + x \ln x - x - (1/2) \ln x + (1/12)(1/x). \tag{ix}$$

However,

$$(\pi/2) = \frac{2 \cdot 2 \cdot 4 \cdot 4 \cdot 6 \cdot 6}{1 \cdot 3 \cdot 3 \cdot 5 \cdot 5 \cdot 7} \ldots \frac{2n}{2n-1} \frac{2n}{2n+1} \ldots$$

or

$$(\pi/2) = \lim \left(\frac{2^2 \cdot 4^2 \cdot 6^2 \ldots (2n)^2}{1^2 \cdot 3^2 \cdot 5^2 \ldots (2n-1)} \right)_{n=\infty} = \lim \left(\frac{2^4 \cdot 4^4 \cdot 6^4 \ldots (2n)^4}{1^2 \cdot 2^2 \cdot 3^2 \ldots (2n)^2 (2n+1)} \right)_{n=\infty} =$$

$$\lim \left(\frac{2^4 \cdot 1^4 \cdot 2^4 \cdot 2^4 \cdot 2^4 \cdot 3^4 \ldots \cdot 2^4 \cdot n^4}{1 \cdot 2^2 \cdot 3^2 \cdot 4^2 \ldots (2n)^2 (2n+1)} \right)_{n=\infty} = \lim \left(\frac{2^{4n} (n!)^4}{[(2n)!]^2 (2n+1)} \right)_{n=\infty}.$$

Consequently,

$$\ln (\pi/2) = \lim \{4n \ln 2 + 4 \ln (n!) - 2 \ln [(2n!)] - \ln (2n + 1)\}_{n=\infty},$$

but, cf. (ix),

$$\ln (n!) = C + n \ln n - n + (1/2) \ln n + (1/12)(1/n) + \ldots,$$
$$\ln [(2n)!] = C + 2n \ln n - 2n + (1/2) \ln (2n) + (1/12)(1/2n) + \ldots$$

with next terms of the order higher than $(1/n)$ in both cases. We thus have

$$\ln (\pi/2) = \lim [4n \ln 2 + 4n + 4C + 4n \ln n - 4n + 2 \ln n + (1/3n) +$$
$$\ldots - \ln (2n + 1) - 2C - 4n \ln n - \ln (2n) - (1/12n)]_{n=\infty} =$$
$$\lim [2C + \ln n - \ln 2 - \ln (2n + 1) + (1/4n)]_{n=\infty} =$$

$$\lim \left\{ 2C + \ln \frac{n}{2(2n+1)} + (1/4n) \right\}_{n=\infty}.$$

Finally,

$$\ln (\pi/2) = 2C + \ln (1/4), \quad C = \ln \sqrt{2\pi}$$

and we get

$$\ln (x!) = \ln \sqrt{2\pi} + x \ln x - x + (1/2) \ln x + (1/12x) + \ldots$$

or

$$x! = \sqrt{2\pi}\, x^{x+1/2} e^{-x} e^{(1/12x)+\ldots}$$

But $e^{(1/12x)} = 1 + (1/12x) + \ldots$ so that, as already derived in §1.3.12,

$$x! = \sqrt{2\pi}\, x^{x+1/2} e^{-x} [1 + (1/12x) + \ldots]. \tag{35}$$

2.2.7. We shall now provide another proof of (35). Consider the function

$$\frac{x!}{x^{x+(1/2)}e^{-x}} = \varphi(x)$$

where x is supposed to be integer and positive. We have

$$\frac{(x-1)!}{(x-1)^{x-(1/2)}e^{-x+1}} = \varphi(x-1)$$

so that

$$\frac{\varphi(x)}{\varphi(x-1)} = x\frac{(x-1)^{x-(1/2)}}{x^{x+(1/2)}}e = [1-(1/x)]^{x-(1/2)}e$$

and consequently

$$\ln\varphi(x) - \ln\varphi(x-1) = 1 + [x-(1/2)]\ln[1-(1/x)].$$

Expanding $\ln[1-(1/x)]$ we obtain

$$\ln\varphi(x) - \ln\varphi(x-1) = 1 + [x-(1/2)]\{-(1/x)-(1/2x^2)-(1/3x^3)-\ldots-$$
$$[1/(l+1)x^{l+1}]-\ldots\} = 1-1-(1/2x)-(1/3x^2)-\ldots-(1/l\,x^{l-1})-$$
$$[1/(l+1)x^l]-\ldots+(1/2x)+(1/2)(1/2x^2)+(1/6x^3)+\ldots+(1/2l\,x^l)+$$
$$[1/2(l+1)x^{l+1}]+\ldots = -(1/12)(1/x^2)-(1/12x^3)-\ldots-$$
$$\frac{l-1}{2l(l+1)}\frac{1}{x^l}-\ldots \qquad\qquad\qquad (x)$$

But

$$\frac{l-1}{2l(l+1)} < \frac{l}{2l(l+1)};\quad \frac{l-1}{2l(l+1)} < \frac{1}{2(l+1)}.$$

If $l=5$ it follows that the left sides of these inequalities are less than $(1/12)$. For $l=4$ it is $3/40$ and, since it decreases with an increasing l, $(1/12)$ is the maximal coefficient in the right side of (x) so that

$$(1/12x^2) + (1/12x^3) + \ldots + \frac{l-1}{2l(l+1)}\frac{1}{x^l} <$$
$$(1/12x^2) + (1/x^3) + \ldots + (1/x^l) + \ldots]$$

which leads to the equality

$$(1/12)(1/x^2) + (1/12x^3) + \ldots + \frac{l-1}{2l(l+1)}\frac{1}{x^l} = (\theta/12)[1/x(x-1)]$$

where θ is some proper fraction. And so,

$$\ln\varphi(x) - \ln\varphi(x-1) = -(\theta/12)[1/x(x-1)],$$

hence

$$\ln\varphi(x) - \ln\varphi(x-1) < 0, \qquad\qquad\qquad (36)$$
$$\ln\varphi(x) - \ln\varphi(x-1) > -(1/12)[1/x(x-1)]. \qquad\qquad (37)$$

Inequality (36) leads to $\ln\varphi(x) < \ln\varphi(x-N)$ where N is any integer and inequality (37) can be written as

$$\ln\varphi(x) - (1/12x) > \ln\varphi(x-1) - [1/12(x-1)]$$

so that

$$\ln \varphi(x) - (1/12x) > \ln \varphi(x - N) - [1/12(x - N)].$$

We set now $x - N = z$ and obtain

$$\ln \varphi(N + z) < \ln \varphi(z),$$
$$\ln \varphi(N + z) > \ln \varphi(z) - (1/12z) + [1/12(N + z)].$$

And, assuming that

$$\lim [\ln \varphi(N + z)]_{N=\infty} = C,$$

we arrive at

$$\ln \varphi(z) > C > \ln \varphi(z) - (1/12z).$$

Consequently,

$$C = \ln \varphi(z) - (\theta/12z), \quad \ln \varphi(z) = C + (\theta/12z)$$

where θ is a proper fraction. However,

$$\ln \varphi(x) = \ln(x!) - [(x + (1/2)] \ln x + x$$

so that

$$\ln(x!) = [(x + (1/2)] \ln x - x + C + (\theta/12x),$$

which is what we ought to have derived.

  We are thus concluding the issue on summing and are going over to the integration of equations in finite differences.

2.3. Integration of Equations in Finite Differences

2.3.1. We are now going over to that section of the theory of finite differences that is similar to integrating differential equations. The problems of this theory are much more difficult than the similar problems of the theory of infinitesimals, and the integration of equations in finite differences is also incomparably more troublesome than the solution of differential equations. Even the reduction of this integration to summation which is similar to expressing the solution of the latter equations in quadratures, – even this would have been of comparatively little use because we are only able to calculate the sums of a very small number of functions, by far less than the number of those which we can integrate.

  It is therefore clear that the integrating factor that transforms the left side of an equation into a total differential, or, in our case, into a difference, after which the problem is reduced to summation, can have no significance in the theory of finite differences. Even in the theory of differential equations the determination of this factor leads to integrating partial equations, i.e., to the integration presenting more difficulties so that it cannot be important as a method of integrating [1] whereas in the theory of finite differences we would have encountered incomparably more difficulties when determining it. We begin by integrating linear equations of the first order, then we shall consider linear equations of any order both with and without the last term {the right side}. The methods of integrating these equations will be quite similar to those of solving differential equations except that now we shall not encounter integrating factors.

  In their form, the equations in finite differences will not be quite similar to differential equations because now we will show the dependence not between the differences of the function and the independent variable, but between the initial and the changed values of the function and the independent variable so that these equations will be of the following type:

$$f(y_{x+n}; y_{x+n-1}; y_{x+n-2}; \ldots; y_{x+1}; y_x; x) = 0$$

where y_x is the initial value of the function. Since

$$y_{x+1} = y_x + \Delta y_x, \quad y_{x+2} = y_{x+1} + \Delta y_{x+1} = y_x + 2\Delta y_x + \Delta^2 y_x, \text{ etc,}$$

the form of this equation is easily made similar to that of differential equations.

Note 1. Concerning the integrating factor, Lagrange expressed himself in such a way: "It is good for various theorems about it, but not as a method of integration". At present, his idea is being ever more confirmed.

2.3.2. Let us consider a linear equation of the first order of the type

$$P\,y_{x+1} + Q\,y_x = V$$

where P, Q and V are functions of x only. Suppose that $y_x = u_x\,v_x$, then

$$y_{x+1} = (u_x + \Delta\,u_x)\,v_{x+1}$$

so that the form of the equation becomes

$$P\,(u_x\,v_{x+1} + \Delta\,u_x\,v_{x+1}) + Q\,v_x\,u_x = V$$

or

$$u_x\,(P\,v_{x+1} + Q\,v_x) + P\,\Delta\,u_x\,v_{x+1} = V.$$

Since the function v_x is arbitrary, we assume that

$$P\,v_{x+1} + Q\,v_x = 0,$$

therefore

$$P\,\Delta\,u_x\,v_{x+1} = V.$$

Let

$$v_x = \exp(w_x),\ \ v_{x+1} = \exp(w_x)\cdot\exp(\Delta\,w_x),$$

then

$$P\exp(w_x)\exp(\Delta\,w_x) + Q\exp(w_x) = 0$$

and consequently

$$P\exp(\Delta\,w_x) + Q = 0,$$
$$\Delta\,w_x = \ln(-Q/P),\ \ w_x = \sum^{x}\ln(-Q/P).$$

We shall not assign any lower limit because it ought to remain arbitrary. And so,

$$v_x = \exp\sum^{x}\ln(-Q/P)$$

and

$$\Delta\,u_x = \frac{V}{P}\cdot\frac{1}{v_{x+1}} = \frac{V/P}{\sum^{x+1}\ln(-Q/P)} = \frac{V}{P}\exp[-\sum^{x+1}\ln(-Q/P)].$$

Therefore,

$$u_x = \sum^{x}\frac{V}{P}\exp[-\sum^{x+1}\ln(-Q/P)] + C,$$

$$y = \exp\sum^{x}\ln(-Q/P)\,[C + \sum^{x}\frac{V}{P}\exp[-\sum^{x+1}\ln(-Q/P)].$$

For example, let us integrate the equation

$$x\, y_{x+1} - (x + 1)\, y_x = 1.$$

We substitute $y_x = u_x\, v_x$, then

$$y_{x+1} = (u_x + \Delta u_x)\, \Delta v_{x+1} = u_x\, v_{x+1} + u_x\, v_{x+1}$$

and

$$u_x[x\, v_{x+1} - (x + 1)\, v_x] + x\, \Delta u_x\, v_{x+1} = 1.$$

Suppose that

$$x\, v_{x+1} - (x + 1)\, v_x = 0,$$

then

$$\Delta u_x = (1/x)\, v_{x+1}^{-1}.$$

Let $v_x = \exp(w_x)$, then

$$x\, \exp(\Delta w_x) - (x + 1) = 0, \quad \Delta w_x = \ln[(x + 1)/x] = \Delta \ln x,$$

$$w_x = \sum \ln[(x + 1)/x] = \ln x, \quad \exp(w_x) = x,$$

$$\Delta u_x = \frac{1}{x(x+1)}, \quad u_x = \sum^{x} \frac{1}{x(x+1)} = -(1/x) + C.$$

Thus,

$$y_x = [C - (1/x)]\, x = Cx - 1.$$

2.3.3. We go over now to linear equations of the higher orders and we begin by considering those of the second order

$$y_{x+2} + P\, y_{x+1} + Q\, y_x = V \tag{38}$$

where $P,\ Q,\ V$ are some functions of x. We shall show how to find its integral when knowing the integral of the same equation without its right side

$$y_{x+2} + P\, y_{x+1} + Q\, y_x = 0. \tag{xi}$$

Suppose that functions v_x and u_x satisfy this; we shall show that the function

$$C\, u_x + C'\, v_x \tag{xii}$$

will also satisfy it. Indeed, since u_x satisfies the equation, we have

$$u_{x+2} + P\, u_{x+1} + Q\, u_x = 0. \tag{39}$$

Multiplying (39) by a constant C we obtain

$$C\, u_{x+2} + C P\, u_{x+1} + C Q\, u_x = 0$$

so that $C\, u_x$ {also} satisfies it. In the same way we find that

$$C'\, v_{x+2} + C' P\, v_{x+1} + C' Q\, v_x = 0.$$

Adding these two equalities we obtain

$$C\, u_{x+2} + C'\, v_{x+2} + P(C\, u_{x+1} + C'\, v_{x+1}) + Q(C\, u_x + C'\, v_x) = 0,$$

which means that (xii) is an integral of equation (39) containing two arbitrary constants. So as to determine now the integral of the initial complete {non-homogeneous} equation we shall apply the Lagrange method of varying the arbitrary

constants. Going over to this complete equation, we ought to assume that C and C' are some functions of x so that its general integral will have the form

$$y_x = C_x u_x + C'_x v_x.$$

Now we shall indeed determine C_x and C'_x. We have

$$y_{x+1} = C_{x+1} u_{x+1} + C'_{x+1} v_{x+1} = (C_x + \Delta C_x) u_{x+1} + (C'_x + \Delta C'_x) v_{x+1}.$$

Suppose that

$$\Delta C_x u_{x+1} + \Delta C'_x v_{x+1} = 0$$

so that

$$y_{x+1} = C_x u_{x+1} + C'_x v_{x+1} \tag{40}$$

and

$$y_{x+2} = C_{x+2} u_{x+2} + C'_{x+2} v_{x+2} = (C_x + \Delta C_x) u_{x+2} + (C'_x + \Delta C'_x) v_{x+2}.$$

Substituting the values of y_{x+2}, y_{x+1} and y_x into equation (38) we have

$$C_x [u_{x+2} + P u_{x+1} + Q u_x] + C'_x [v_{x+2} + P v_{x+1} + Q v_x] +$$
$$\Delta C_x u_{x+2} + \Delta C'_x v_{x+2} = V.$$

But

$$u_{x+2} + P u_{x+1} + Q u_x = v_{x+2} + P v_{x+1} + Q v_x = 0$$

because u_x and v_x satisfy equation (xi). We thus determine that

$$\Delta C_x u_{x+2} + \Delta C'_x v_{x+2} = V. \tag{41}$$

Adding to this equation (40) we have

$$\Delta C_x u_{x+1} + \Delta C'_x v_{x+1} = 0$$

and, solving now {this together with (41)} with respect to ΔC_x and $\Delta C'_x$, we obtain

$$\Delta C_x = \varphi_0(x), \ \Delta C'_x = \varphi_1(x)$$

so that
$$C_x = \sum \varphi_0(x) + C_o, \ C'_x = \sum \varphi_1(x) + C_1$$
and

$$y_x = C_o u_x + C_1 v_x + u_x \sum \varphi_0(x) + v_x \sum \varphi_1(x).$$

We have thus reduced the integration of equation (38) to the integration of (xi) for whose solution mathematics in its present state has no methods. To illustrate, let us take up equation

$$y_{x+2} - 5y_{x+1} + 6y_x = x. \tag{xiii}$$

The equation

$$y_{x+2} - 5y_{x+1} + 6y_x = 0$$

has solutions 2^x and 3^x and we assume that

$$y_x = C_x 2^x + C'_x 3^x.$$

Consequently,

$$y_{x+1} = C_x 2^{x+1} + C'_x 3^{x+1} + \Delta C_x 2^{x+1} + \Delta C'_x 3^{x+1}.$$

Assuming that

$$\Delta C_x 2^{x+1} + \Delta C'_x 3^{x+1} = 0 \qquad\qquad (42)$$

we have

$$y_{x+1} = C_x 2^{x+1} + C'_x 3^{x+1}$$

and

$$y_{x+2} = C_x 2^{x+2} + C'_x 3^{x+2} + \Delta C_x 2^{x+2} + \Delta C'_x 3^{x+2}$$

and also

$$\Delta C_x 2^{x+2} + \Delta C'_x 3^{x+2} = x.$$

Solving this equation together with (42), we obtain

$$\Delta C_x = -x/2^{x+1}, \quad \Delta C'_x = x/3^{x+1}$$

so that

$$C_x = -(1/2)\sum x\,(1/2)^x + C_0, \quad C'_x = (1/3)\sum x\,(1/3)^x + C_1.$$

In order to obtain the values of these sums we shall derive the formula for "summing by parts". Choose some functions S_x and K_x. Then

$$\Delta(S_x K_x) = (S_x + \Delta S_x) K_{x+1} - S_x K_x = S_x(K_x + \Delta K_x) + K_{x+1}\Delta S_x - S_x K_x,$$
$$\Delta(S_x K_x) = S_x \Delta K_x + K_{x+1}\Delta S_x$$

and consequently

$$S_x K_x = \sum S_x \Delta K_x + \sum K_{x+1}\Delta S_x.$$

Set now $\Delta K_x = T_x$, then

$$K_x = \sum^{x} T_x, \quad K_{x+1} = \sum^{x+1} T_x,$$

$$\sum S_x T_x = S_x \sum T_x - \sum \Delta S_x \sum^{x+1} T_x. \qquad\qquad (43)$$

When applying this formula that is similar to the formula

$$\int S_x T_x\, dx = S \int T_x\, dx - \int [\frac{dS_x}{dx} \int T_x\, dx]\, dx$$

of the integral calculus, we have

$$\sum x\,(1/2)^x = x \sum (1/2)^x - \sum \Delta x \sum^{x+1} (1/2)^x,$$

$$\sum x \, (1/3)^x = x \sum (1/3)^x - \sum \Delta x \overset{x+1}{\sum} (1/3)^x.$$

However, in general

$$\sum a^x = a^x / (a - 1)$$

and

$$\sum (1/2)^x = -2 \, (1/2)^x; \quad \sum (1/3)^x = -(3/2) \, (1/3)^x;$$

$$\overset{x+1}{\sum} (1/2)^x = -2 \, (1/2)^x; \quad \overset{x+1}{\sum} (1/3)^x = -(1/2) \, (1/3)^x;$$

$$-\sum 2 \, (1/2)^{x+1} = 2 \, (1/2)^x; \quad -(1/2) \sum (1/3)^x = (3/4) \, (1/3)^x$$

so that

$$\sum x \, (1/2)^x = -2x \, (1/2)^x - 2 \, (1/2)^x = -2 \, (1/2)^x (x + 1),$$

$$\sum x \, (1/3)^x = -(3/2) \, x \, (1/3)^x - (3/4) \, (1/3)^x = -3 \, (1/3)^x [(x/2) + (1/4)].$$

Thus,

$$C_x = (1/2)^x (x + 1) + C_o, \quad C'_x = -(1/3)^x [(x/2) + (1/4)] + C_1$$

and

$$y = C_o 2^x + C_1 3^x + (x + 1) (2^x / 2^x) - (1/3)^x [(x/2) + (1/4)] 3^x =$$
$$(1/2) \, x + (3/4) + C_o 2^x + C_1 3^x.$$

This is indeed the general integral of the equation (xiii).

2.3.4. Let us apply now the same method to equations of the third order which we shall consider in the form of

$$y_{x+3} + P \, y_{x+2} + Q \, y_{x+1} + R \, y_x = V. \tag{44}$$

Suppose that three functions, u_x, v_x, w_x, satisfy the equation

$$y_{x+3} + P \, y_{x+2} + Q \, y_{x+1} + R \, y_x = 0 \tag{45}$$

and that its general integral can be formed by them. A necessary condition for this is that the determinant

$$\begin{vmatrix} u_{x+1} & v_{x+1} & w_{x+1} \\ u_{x+2} & v_{x+2} & w_{x+2} \\ u_{x+3} & v_{x+3} & w_{x+3} \end{vmatrix} \tag{46}$$

does not vanish. Otherwise these three functions will be connected by such a dependence that equation (45) (after their substitution there and its solution with respect to 1, P and Q)[1] would have provided for P and Q either indefinite or infinite values. The three functions should be linearly indendent one from another.
    In this case the function

$$y_x = Cu_x + C'v_x + C''w_x$$

will satisfy equation (45). For a function of the same type to satisfy equation (44) we ought to replace here C, C' and C'' by C_x, C'_x and C''_x and consider them {the new magnitudes} as functions of x.

And so, let the function

$$y_x = C_x u_x + C'_x v_x + C''_x w_x \qquad \text{(xiv)}$$

satisfy equation (44). Then

$$y_{x+1} = C_x u_{x+1} + C'_x v_{x+1} + C''_x w_{x+1} + \Delta C_x u_{x+1} + \Delta C'_x v_{x+1} + \Delta C''_x w_{x+1}.$$

Let

$$\Delta C_x u_{x+1} + \Delta C'_x v_{x+1} + \Delta C''_x w_{x+1} = 0 \qquad (47)$$

so that

$$y_{x+1} = C_x u_{x+1} + C'_x v_{x+1} + C''_x w_{x+1}, \qquad \text{(xv)}$$
$$y_{x+2} = C_x u_{x+2} + C'_x v_{x+2} + C''_x w_{x+2} + \Delta C_x u_{x+2} + \Delta C'_x v_{x+2} + \Delta C''_x w_{x+2}.$$

We suppose now that

$$\Delta C_x u_{x+2} + \Delta C'_x v_{x+2} + \Delta C''_x w_{x+2} = 0, \qquad (48)$$

then

$$y_{x+2} = C_x u_{x+2} + C'_x v_{x+2} + C''_x w_{x+2} \qquad \text{(xvi)}$$

and

$$y_{x+3} = C_x u_{x+3} + C'_x v_{x+3} + C''_x w_{x+3} + \Delta C_x u_{x+3} + \Delta C'_x v_{x+3} + \Delta C''_x w_{x+3}. \qquad \text{(xvii)}$$

Substituting now the expressions (xvii), (xvi), (xv) and (xiv) into equation (44) and noticing that

$$u_{x+3} + Pu_{x+2} + Qu_{x+1} + Ru_x = 0, \quad v_{x+3} + Pv_{x+2} + Qv_{x+1} + Rv_x = 0,$$
$$w_{x+3} + Pw_{x+2} + Qw_{x+1} + Rw_x = 0$$

because u_x, v_x and w_x are solutions of the equation (45), we get

$$\Delta C_x u_{x+3} + \Delta C'_x v_{x+3} + \Delta C''_x w_{x+3} = V. \qquad (49)$$

Solving now equations (47) – (49) with respect to ΔC_x, $\Delta C'_x$ and $\Delta C''_x$, we obtain finite and definite values because the determinant (46) is not equal to zero. And so, we shall have

$$\Delta C_x = \varphi_0(x), \ \Delta C'_x = \varphi_1(x), \ \Delta C''_x = \varphi_2(x)$$

and therefore

$$C_x = \sum \varphi_0(x) + A_0, \ C'_x = \sum \varphi_1(x) + A_1, \ C''_x = \sum \varphi_2(x) + A_2.$$

The general integral of equation (44) will thus be

$$y_x = A_0 u_x + A_1 v_x + A_2 w_x + u_x \sum \varphi_0(x) + v_x \sum \varphi_1(x) + w_x \sum \varphi_2(x). \qquad \text{(xviii)}$$

We have reduced the integration of the equation (44) to the solution of a simpler equation (45). Note that the expression (xviii) shows that the general integral consists of two parts: of the general integral of equation (45) and some function of x. It follows that, when substituting the expression (xviii) in equation (44), this function will in itself satisfy it because the first part of (xviii) turns the left side of (44) into zero.

Thus, in order to determine the general integral of the equation (44) it suffices to find some solution satisfying it, and the sum of that function and the general integral of the equation (45) will indeed be the general integral of the equation (44).

Note 1. {Chebyshev's own expression.}

2.3.5. Let us extend now the derivations of the previous sections onto equations of any order. We take the equation

$$y_{x+n} + Py_{x+n-1} + \ldots + Ry_{x+2} + Sy_{x+1} + Ty_x = V \qquad (50)$$

where $P, \ldots, R, S, T, V$ are functions of only one {variable} x, and suppose that n functions $u_x, v_x, w_x, \ldots, \omega_x$ satisfy the equation

$$y_{x+n} + Py_{x+n-1} + \ldots + Ry_{x+2} + Sy_{x+1} + Ty_x = 0 \qquad (51)$$

so that its general integral is

$$y_x = C'u_x + C''v_x + C^{(3)}w_x + \ldots + C^{(n)}\omega_x.$$

Assume also that the general integral of the {non-homogeneous} equation (50) with the last term included is

$$y_x = C'_x u_x + C''_x v_x + C^{(3)}_x w_x + \ldots + C^{(n)}_x \omega_x$$

and set

$$\begin{cases} \Delta C'_x u_{x+1} + \Delta C''_x v_{x+1} + \ldots + \Delta C^{(n)}_x \omega_{x+1} = 0 \\ \Delta C'_x u_{x+2} + \Delta C''_x v_{x+2} + \ldots + \Delta C^{(n)}_x \omega_{x+2} = 0, \ldots, \\ \Delta C'_x u_{x+n-1} + \Delta C''_x v_{x+n-1} + \ldots + \Delta C^{(n)}_x \omega_{x+n-1} = 0 \end{cases} \qquad (52)$$

Then

$$\begin{aligned} y_{x+1} &= C'_x u_{x+1} + C''_x v_{x+1} + \ldots + C^{(n)}_x \omega_{x+1}, \\ y_{x+2} &= C'_x u_{x+2} + C''_x v_{x+2} + \ldots + C^{(n)}_x \omega_{x+2}, \ldots, \\ y_{x+n-1} &= C'_x u_{x+n-1} + C''_x v_{x+n-1} + \ldots + C^{(n)}_x \omega_{x+n-1} \end{aligned}$$

and

$$\begin{aligned} y_{x+n} = \ &C'_x u_{x+n} + C''_x v_{x+n} + \ldots + C^{(n)}_x \omega_{x+n} + \\ &\Delta C'_x u_{x+n} + \Delta C''_x v_{x+n} + \ldots + \Delta C^{(n)}_x \omega_{x+n}. \end{aligned}$$

Substituting these expressions instead of $y_x, y_{x+1}, \ldots, y_{x+n}$ into equation (50) and noting that the functions $u_x, v_x, \ldots$ satisfy the equation (51), we obtain

$$\Delta C'_x u_{x+n} + \Delta C''_x v_{x+n} + \ldots + \Delta C^{(n)}_x \omega_{x+n} = V.$$

When solving this equation together with (52) with respect to $\Delta C'_x, \ldots, \Delta C^{(n)}_x$, we have

$$\begin{aligned} &\Delta C'_x = \varphi_1(x), \ \Delta C''_x = \varphi_2(x), \ldots, \Delta C^{(n)}_x = \varphi_n(x); \\ &C'_x = \sum \varphi_1(x) + A_1, \ldots, C^{(n)}_x = \sum \varphi_n(x) + A_n, \end{aligned}$$

and

$$\begin{aligned} y_x = \ &A_1 u_x + A_2 v_x + \ldots + A_n \omega_x + \\ &u_x \sum \varphi_1(x) + v_x \sum \varphi_2(x) + \ldots + \omega_x \sum \varphi_n(x) \end{aligned}$$

will be the general integral of the equation (50).

Далее с. 102 до 110 со сканнера и после этого:

$$y_0, \ y_1, \ y_2, \ldots, y_x, \ y_{x+1}, \ y_{x+2}, \ldots, y_{n-1}, y_n$$

arranged in the same order and satisfying the inequality

y_x not $> R_{n-x}$ with $y_0 = 2$ and $y_1 = 3$. [1]

It is easy to see that this condition will be satisfied if, in general,

$$y_{x+2} = y_{x+1} + y_x. \qquad \text{(xxiv)}$$

Indeed, assuming that

$$y_{x+1} \leq R_{n-x-1} \text{ and } y_x \leq R_{n-x}$$

we have

$$y_{x+2} \leq R_{n-x-1} + R_{n-x}.$$

But

$$R_{n-x-2} = R_{n-x-1} \, q_{n-x} + R_{n-x}, \; R_{n-x-2} > R_{n-x-1} + R_{n-x}$$

so that

$$y_{x+2} \leq R_{n-x-2}.$$

However, $y_0 \leq R_n$, $y_1 \leq R_{n-1}$.and, consequently, in general $y_x \leq R_{n-x}$. Nevertheless, when integrating equation (xxiv), we obtain

$$y_x = C_1 [(1 + \sqrt{5})/2]^x + C_2 [(1 - \sqrt{5})/2]^x$$

with the arbitrary constants C_1 and C_2 determined by the conditions $y_0 = 2$ and $y_1 = 3$, so that

$$C_1 + C_2 = 2, \; C_1[(1 + \sqrt{5})/2] + C_2 [(1 - \sqrt{5})/2] = 3.$$

We thus obtain

$$C_1 = (\sqrt{5} + 2)/\sqrt{5}, \; C_2 = (\sqrt{5} - 2)/\sqrt{5},$$
$$y_x = [(\sqrt{5} + 2)/\sqrt{5}] [(1 + \sqrt{5})/2]^x + [(\sqrt{5} - 2)/\sqrt{5}] [(1 - \sqrt{5})/2]^x \leq R_{n-x}.$$

Assuming here $n = x$ and noting that $R_0 = A$, we get

$$[(\sqrt{5} + 2)/\sqrt{5}] [(1 + \sqrt{5})/2]^n + [(\sqrt{5} - 2)/\sqrt{5}] [(1 - \sqrt{5})/2]^n \leq A.$$

But

$$[(\sqrt{5} - 1)/2] \leq 0.7 \text{ and } [(1 - \sqrt{5})/2]^n \leq 0.7,$$
$$[(\sqrt{5} - 2)/\sqrt{5}] [(1 - \sqrt{5})/2]^n \leq 0.7 [(\sqrt{5} - 2)/\sqrt{5}] < 0.08.$$

Therefore

$$[(\sqrt{5} + 2)/\sqrt{5}] [(1 + \sqrt{5})/2]^n - 0.08 < A,$$
$$[(\sqrt{5} + 2)/\sqrt{5}] [(1 + \sqrt{5})/2]^{n+1} < (A + 0.08) [(1 + \sqrt{5})/2]$$

and, since $A \geq 2$, $A + 0.08 = A(1 + 0.08/A) < 1.04A$, we have

$$[(1 + \sqrt{5})/2]^{n+1} < 1.04A\,(\sqrt{5}/2)\,\frac{1+\sqrt{5}}{2+\sqrt{5}} = 0.52A\,\sqrt{5}\,\frac{1+\sqrt{5}}{2+\sqrt{5}}.$$

Taking logarithms we get [2]

$$(n + 1) < \frac{\lg A + \lg 0.52 + \lg(1+2.236) + \lg 2.236 - \lg(2+2.236)}{\lg(1+2.236) - \lg 2},$$

$$(n + 1) < \frac{1}{\lg[1+2.236)/2]}\,[\lg A - \lg \frac{2+2.236}{0.52\,(1+2.236)\,2.236}].$$

But

$$\frac{2+2.236}{0.52\,(1+2.236)\,2.236} > \frac{4}{0.52\,(2.236+5)} = \frac{1}{0.13\,(7+\alpha)}$$

where α is a proper fraction; [3] then, since $\sqrt{5} < 2.24$, $\alpha < 0.24$ or $\alpha < 1/4$. Thus,

$$\frac{2+2.236}{0.52\,(1+2.236)\,2.236} > \frac{1}{0.13\,(7+1/4)} > 1$$

and consequently

$$n + 1 > \frac{\lg A}{\lg[(1+2.236)/2]}.$$

But $\sqrt{5} > 2.236$ and $\lg[(1 + \sqrt{5})/2] > \lg 1.618 = 0.2090$,

$\lg[(1 + \sqrt{5})/2] > 0.2 = 1/5$ and consequently $(n + 1) < 5\lg A$.

Denote the number of digits in number A by N, then $\lg A < N$ and $\qquad$ $(n + 1) < 5N$. We thus infer that, when calculating the greatest common divisor of numbers A and B with $A < B$, we shall have to perform a number of divisions in any case less than five times the number of digits in the lesser number A.

We conclude here the theory of finite differences.

Note 1. {Below, I do not anymore follow Chebyshev in writing a not $> b$ (say); instead, I adopt the simpler notation $a \leq b$.}

Note 2. {In several lines below I write 2.236 instead of $\sqrt{5}$ as given by Chebyshev.}

Note 3. {Strictly speaking, α is irrational. Here and below Chebyshev considered common logarithms but did not change his notation.}

Chapter 3. The Theory of probability

3.1. The Laws of Probability

3.1.1. The theory of probability aims at determining the chances for the occurrence of some event. The word *event* means, in general, everything whose probability is being determined. In mathematics, the word *probability* thus serves to denote some magnitude subject to measurement.

Probability evidently depends only on two magnitudes: on the number of cases favorable to the event and on the number of all the equally possible cases. Therefore, when denoting the probability of some event by E, the first number by m and the second one by n, we have

$$E = F(m; n) = F[n(m/n); n] = \varphi(m/n; n)$$

where φ is such a function that increases with m and decreases with an increasing n. But it is not difficult to agree that probability should not change when the numbers of all the possible cases and of those favoring the event increase in the same ratio. In the theory of probability, this property is being assumed as an *axiom*. Consequently, the function φ should not change if we replace m and n by λm and λn where λ is an arbitrary factor. It follows that

$$\varphi(m/n; n) = \varphi[(\lambda m / \lambda n); \lambda n].$$

This equality shows that the function φ does not depend on n. Thus,

$$E = f(m/n).$$

That is, the probability is a function of the ratio of the number of favorable cases to that of all the equally possible cases, and this function, if the mentioned ratio be assumed as an independent variable, should be an increasing function. As to its form, this is unknown to us so that when defining probability we may arbitrarily take any increasing function of the ratio m/n. In mathematics, this very ratio, which is the simplest function, is indeed *assumed* as an expression of probability.

{To repeat,} in mathematics, the ratio m/n, which we shall now denote by p, is indeed usually assumed as the *definition* of probability in the sense of a magnitude subject to measurement. If $p = 0$, the probability turns into certainty that the event will not occur; in the same way, if $p = 1$, probability turns to certainty that the event will occur. As everywhere in mathematics, we consider these two extreme cases, which go beyond the province of probability, as the limiting cases of the general, and it is in this sense that we regard probabilities equal to zero or unity.

Above, we defined the word *event* as one of the terms occurring in the theory of probability; now we add that we shall call events *incompatible* if they cannot take place at one and the same chance {trial}. Thus, the throwing of a card of clubs and of an ace from a deck of cards are compatible events, but the drawing of a single card being a club and an ace of hearts provides an example of incompatible events. Let us consider now the main properties of probabilities expressing them as theorems.

3.1.2. Theorem 1. *The probability that* {any} *one of the two incompatible events will occur is equal to the sum of their probabilities.* Suppose that E and E_1 are the two incompatible events and let $p = m/n$ be the probability of the first one, and $p_1 = m_1/n_1$, the probability of the second one. Since the events E and E_1 are incompatible, a case favorable for the first will not be favorable for the second one, and vice versa. The number of cases favorable for the event $(E + E_1)$ is therefore $(m + m_1)$, whereas the number of all the equally possible cases remains, as it was, equal to n. The probability that one of these events will occur is therefore

$$P = (m + m_1) / n = p + p_1. \qquad\qquad (i)$$

It is easy to extend this theorem onto any number of incompatible events. Indeed, let E, E_1, E_2, ... be incompatible events with probabilities $p = m/n$, $p_1 = m_1/n$, $p_2 = m_2/n$, ... The probability that one of the two events, E and E_1, occurs, or the probability of the event $(E$ or $E_1)$, is $(m + m_1)/n$. It follows that the probability of the event $(E, E_1$ or $E_2)$ is $[(m + m_1) + m_2]/n$. Continuing to reason in this manner, we shall find that, in general, the probability P of the event $(E, E_1, E_2, ...,$ or $E_k)$ is

$$P = (m + m_1 + ... + m_k) / n = p + p_1 + p_2 + ... + p_k.$$

Note that if, in formula (i), $P = 1$, the events E and E_1 whose probabilities are p and p_1 are called *contrary*. For such events the probability of one of them thus complements to unity the probability of the other one.

The probability that some event occurs or not is always 1, so that knowing the probability p that the event will take place, we shall find the probability p_1 that it will not occur in accord with the formula

$$p_1 = 1 - p.$$

The Theorem above can also be formulated thus: *If one and the same event has several different incompatible forms, then its probability is the sum of the probabilities of its forms. If the probabilities of all the forms are equal one to another, the probability of the event is proportional both to the probability of each separate form and to the number of the forms.*

3.1.3. Theorem 2. *If the probability of event E is p and the probability for the event F to occur after event E happened is q, then the product pq is the probability of the compatibility of these events (of event E and then of event F to take place).* Suppose that

$$C_1, \ C_2,\ldots, \ C_m, \ \ldots, \ C_\mu, \ C_{\mu+1}, \ \ldots, \ C_{n-1}, \ C_n$$

represent equally possible cases and μ of them, taken in that order in which they are written above, are cases favorable for the event F. Then,

$$p \ = \ \mu/n \ \text{ and } \ q \ = \ m/\mu$$

because, when determining q, we ought to take μ as the number of all the possible cases. It follows that

$$p\,q \ = \ m/n,$$

but m is the number of cases favorable for both the first and the second event and n is the number of all the equally possible cases for each of these events so that m/n is the probability that they both occur at the same time {one after another}.

If r is the probability of event G occurring after event F took place, qr will be the probability that the events F and G occur at the same time {one after another}, and pqr, the probability of the joint occurrence of the three events, E, F and G. Reasoning in this manner, we shall extend our theorem onto any number of events.

In the particular case in which the probability of event E_1 to occur after event E took place is q_1; the probability of event E_2 to occur after event E_1 took place is q_2; etc, and, finally, in which the probability of event E_k to occur after event E_{k-1} took place is q_k; and if $q_1 = q_2 = \ldots = q_k = p$ where p is the probability of the event E, then the probability of the compatibility of these $(k + 1)$ events will be $p\,q_1\,q_2\,q_3 \ldots q_k = p^{k+1}$.

3.1.4. To illustrate, let us solve the following problem: Suppose that we have a vessel containing white and black balls and that we draw a ball, return it to the vessel, draw a ball again, etc, and repeat this operation l times. It is inquired, what is the probability to extract a black ball during these *trials* {exactly once}.

We understand the word *trial* as such a concurrence of circumstances under which the event can take place. Let the probability of drawing a black ball at a trial be p, then $(1 - p)$ is the probability that this event will not occur at one trial. The probability that this event will not take place at the second trial is also $(1 - p)$ because the chances of its occurrence remain as they were previously.

And so, the probability that the event will not happen in two trials is $(1 - p)^2$; in three trials, it equals $(1 - p)^3$, …, and in l trials it is $(1 - p)^l$. Therefore, the probability that this event occurs, i.e., that the black ball will be extracted, is

$$P \ = \ 1 - (1 - p)^l.$$

If p is a very small fraction, we may break off at the first term of the expansion

$$\ln (1 - p)^l \ = \ l \ln (1 - p) \ = \ l\,[-p \ - \ (p^2/2) \ - (p^3/3) \ - \ldots]$$

so that we will have

$$(1 - p)^l \ = \ e^{-l\,p}, \ P \ = \ 1 - e^{-l\,p}.$$

3.1.5. As a second example, we shall try to solve the following problem: Determine the probability that a randomly chosen fraction can be reduced. Let A/B be this fraction and P, the probability that it cannot be reduced. It is easy to see that this probability is *composed* from probabilities p_2, p_3, p_5, ..., p_m, where m is any prime number, that A/B cannot be reduced by 2, by 3, ..., by m. Therefore,

$$P = p_2 p_3 p_5 \dots p_{m\dots}$$

Let us determine p_m. We shall calculate the probability that the fraction cannot be reduced by m if we find the probability that the numbers A and B are not divisible by m. Suppose that we divide A by m; then the remainder can only equal 0, 1, 2, ..., $(m - 1)$. It is seen therefore that the probability that A is divisible by m is $1/m$. In the same way the probability that B is divisible by m is $1/m$ so that $1/m^2$ is the probability that these events coincide, i.e., that the fraction can be reduced by m. Therefore $p_m = 1 - (1/m^2)$. Thus

$$P = [1 - (1/2^2)] [1 - (1/3^2)] [1 - (1/5^2)] \dots [1 - (1/m^2)]\dots$$

where m is a prime number. It follows that

$$1/P = \frac{1}{1-(1/2^2)} \frac{1}{1-(1/3^2)} \frac{1}{1-(1/5^2)} \dots = 1 + 1/2^2 + 1/3^2 + 1/4^2 + 1/5^2 + \dots$$

and the sum of the series is $\pi^2/6$.[1]

Neither is it difficult to derive this result directly. The expansion

$$\frac{\sin x}{x} = [1 - (x^2/\pi^2)] [1 - (x^2/2^2\pi^2)] [1 - (x^2/3^2\pi^2)] \dots$$

is known, and we also have

$$\frac{\sin x}{x} = 1 - (x^2/6) + (x^4/120) - \dots$$

so that

$$\ln [1 - (x^2/6) + (x^4/120) - \dots] = \ln [1 - (x^2/\pi^2)] + \ln [1 - (x^2/2^2\pi^2)] + \dots$$

or

$$-\frac{x^2}{6} + \dots = -\frac{x^2}{\pi^2} - \frac{x^2}{2^2\pi^2} - \frac{x^2}{3^2\pi^2} - \dots$$

Therefore, equating the coefficients of the same degrees of x in both sides of this equality, we shall obtain, in part,

$$1+ (1/2^2) + (1/3^2) + (1/4^2) + \dots = \pi^2/6$$

so that $1/P = \pi^2/6$ and $P = 6/\pi^2$ which is about 6/10. If we denote the probability that the fraction can be reduced by Q, we shall have $Q = 1 - 6/\pi^2$.

Denoting now the probability that some fraction is irreducible once it is known that it caanot be reduced by 2, 3 or 5 by P_0, we shall have

$$P = P_0 [1 - 1 (1/2^2)][1 - (1/3^2)] [1 - (1/5^2)]$$

and

$$P_\text{o} = [6/\pi^2]\ \frac{1}{1-(1/2^2)}\cdot\frac{1}{1-(1/3^2)}\cdot\frac{1}{1-(1/5^2)} = \frac{75}{8\pi^2}\,.$$

But $8\pi^2 = 78.97$ and $P_\text{o} = 75/79$, $1 - P_\text{o} = 4/79$,

$$1/19 > 1 - P_\text{o} > 1/20.$$

Thus, if the fraction A/B is known to be irreducible by 2, 3 and 5, the probability that it cannot be reduced by other numbers either is contained between 1/19 and 1/20.[2]

Note 1. {Chebyshev refers here to his §1.3.11 where this well-known series is not even considered. He had not explained the transition from product to this series, but it can be found in Euler's *Introduction to the Analysis of Infinitesimals*, Chapt. 15, §275.}

Note 2. {An obvious mistake: $(1 - P_\text{o})$ is the probability of the contrary event. More important: It should have been specified that A and B with equal probabilities and independently from each other take any value from $\sqrt{N}$ to N where $\sqrt{N}$ is a large natural number and $N \to \infty$. Bernstein severely criticized Chebyshev's solution (also mentioning Markov who had followed his teacher) noting that the application of probability in the number theory is of a peculiar nature. See his paper The present state of the theory of probability (1928; translated in Deutsche Hochschulschriften 2579. Egelsbach, 1998, pp. 109 – 129 (pp. 111 – 112)). Also see Postnikov, A.G., *Вероятностная теория чисел* (Stochastic Theory of Numbers). Moscow, 1974.}

3.1.6. Until now, we studied the laws which enable us to determine *prior* probabilities; now we shall go over to the laws concerning probabilities known as *posterior*. And we note that these laws are far from being distinguished by the rigor possessed by the two expounded by us {above} so that they should rather be regarded as hypotheses than as laws.

Theorem 3. *Knowing that event E occurred, and that it could have taken place together with events F_1, F_2, …, F_i, whose probabilities are P_1, P_2, …, P_i, and which are independent one from another, we shall find, that the probability Q that event E took place together with event F_j, is expressed as*

$$Q_j = P_j\, p_j\, /\, (P_1\, p_1 + P_2\, p_2 + \ldots + P_i\, p_i)$$

where p_k is the probability of the event E after event F_k took place.

Let n be the number of all the equally possible cases for the events F_1, F_2, …, F_i, and m_j, the number of cases favorable for event F_j. Since the events F_1, F_2, …, F_i are independent one from another, the number of cases m_j does not include those favoring the other events. Thus, $P_j = m_j/n$. Suppose now that among these m_j cases favorable for the event F_j there are λ_j cases favoring the event E, then $p_j = \lambda_j/m_j$ and we will have

$$Q_j = \lambda_j\, /\, (\lambda_1 + \lambda_2 + \ldots + \lambda_j + \ldots + \lambda_i).$$

Substituting instead of λ_1, λ_2, … their values we obtain

$$Q_j = p_j\, m_j\, /\, (p_1 m_1 + p_2 m_2 + \ldots + p_i m_i) =$$
$$P_j\, p_j\, /\, (P_1\, p_1 + P_2\, p_2 + \ldots + P_i\, p_i).$$

Suppose for example that we have two groups of cards, A and B, with group A consisting of two small piles, two red cards and one black card in each of them, and with group B being a small pile consisting only of three red cards.

Suppose now that we draw a red card n times in succession. It is inquired, What is the probability that the extracted card {cards} belongs {belong} to group A.[1] We also suppose that, having put our hand in some group, we make all our drawings from this group; and that, after each extraction, the card is returned to the pile from which it was drawn.

In this problem, the event F_1 is the drawing of a card from group A, and the event F_2, its being drawn from B; P_1 is the probability that the card is drawn from group A; P_2, the probability of its being drawn

from *B*. Since *A* consists of two piles and *B*, of one pile, and since we consider it equally possible that any pile is chosen, we will have $P_1 = 2/3$ and $P_2 = 1/3$.

The probability of drawing a red card from group *A* is 2/3 so that the extraction of *n* red cards in succession from this group is $(2/3)^n$ and the probability of drawing a red card from group *B* is 1. It follows that $p_1 = (2/3)^n$ and $p_2 = 1$. The probability that the red card {cards} extracted *n* times belongs {belong} to group *A* is

$$Q = (2/3)^n\,(2/3)\,/\,[(2/3)^n\,(2/3)\,+\,(1/2)\,1] = 2^{n+1}\,/\,[2^{n+1}\,+\,3^n]$$
$$= 1\,/\,[1\,+\,(1/2)\,(3/2)^n].$$

It is seen therefore that with *n* increasing to infinity this probability tends to zero whereas the probability $(1 - Q)$ that the extracted card {cards} belongs {belong} to group *B* approaches unity.

To provide another example let us determine the probability that an examined student who successfully answered *l* questions [2] will successfully answer the other ones as well. Let the number of questions be *N*. The examiner supposes that the student can successfully answer 0, 1, 2, …, *x*, …, *N* questions, and regards all these {implied} events, $(N + 1)$ in number, as *equally possible*. The probability of each of them is $1/(N + 1)$ so that

$$P_1 = P_2 = \ldots = P_{N+1} = 1/(N + 1).$$

In our case p_x is the probability that the student will answer *l* questions if it is known that he was able to answer *x* of them. In order to determine this probability, we note that *x/N* is the probability that under our condition the student will answer one {more} question.

However, after the card with this question is extracted, there remains $(N - 1)$ more out of which he will answer $(x - 1)$ ones; he can draw any one of these $(N - 1)$ questions, and the probability that he will extract a favorable question is
$(x - 1)\,/\,(N - 1)$. Thus, the probability that the student will answer two questions is $(x/N)\,(x - 1)\,/\,(N - 1)$. Continuing to reason in the same manner, we find that

$$p_x = \frac{x(x-1)(x-2)\ldots(x-l+1)}{N(N-1)(N-2)\ldots(N-l+1)}. \qquad\qquad (ii)$$

Therefore, the probability that he answers *x* questions (denote it by Q_x) will be

$$\frac{x(x-1)(x-2)\ldots(x-l+1)}{N(N-1)(N-2)\ldots(N-l+1)} \Big/ \sum_{0}^{N+1} \frac{x(x-1)(x-2)\ldots(x-l+1)}{N(N-1)(N-2)\ldots(N-l+1)}.$$

Now {formula (24) in Chapt. 2} the numerator of the sum is

$$\frac{(N+1)N(N-1)\ldots(N-l+1)}{l+1}$$

and

$$Q_x = \frac{x(x-1)(x-2)\ldots(x-l+1)(l+1)}{(N+1)N(N-1)\ldots(N-l+1)}$$

so that

$$Q_N = \frac{l+1}{N+1}.$$

Note 1. {As I put on record (*History of theTheory of Probability to the Beginning of the 20th Century*. Berlin, 2004, p. 196), Liapunov remarked that Chebyshev had sometimes wrongly used the singular form instead of the plural.}

Note 2. {Chebyshev considered questions written out on cards in advance and randomly extracted from the pile of cards by the students.}

3.1.7. We are now going over to the fourth law which can be considered as a corollary of the previous ones. It can be formulated in the form of the following theorem.

Theorem 4. *The occurrence of event E is only possible together with one of the events F_1, F_2, ... independent one from another. Then the probability H of event G which takes place after E occurred and which is also only possible together with one of the events F_1, F_2,... is determined by the formula*

$$H = (P_1 p_1 q_1 + P_2 p_2 q_2 + \ldots) / (P_1 p_1 + P_2 p_2 + \ldots) =$$
$$\frac{\sum P_x p_x q_x}{\sum P_x p_x} \qquad \text{(iii)}$$

where P_x and p_x have the same meaning as in Theorem 3 and q_x is the probability of the event G under the hypothesis F_x, i.e., after F_x took place.

Supposing that the event E occurred, the probability that it took place together with event F_x will be, in accord with the preceding theorem,

$$P_x p_x / (P_1 p_1 + P_2 p_2 + \ldots).$$

But event G can only take place either with event F_1, or with event F_2,…, so that the probability of its occurring at all is determined by formula (iii).

Let us consider now the following example. A student draws l questions and answers them successfully. It is inquired what is the probability of his also successfully answering the next extracted question.

Let N be the number of questions. Event E is the drawing of l favorable questions. Events, or, rather, hypotheses F_1, F_2,…, are the suppositions that the student is able to answer 0, 1, 2, …, N questions. It is assumed (certainly wrongly) that all these hypotheses are equally probable so that $P_0 = P_1 = \ldots = P_x = \ldots = P_N = 1/(N + 1)$. The probability p_x that the student, having being able to answer x questions, extracts l questions and successfully answers them, is expressed by the formula (ii).

The event G consists in that the student is able to answer the $(l + 1)$-th question. Its probability under the hypothesis F_x is $q_x = (x - l) / (N - l)$ so that H, the probability sought, will be determined by the formula

$$H = \frac{\sum\limits_{0}^{N+1} \dfrac{x(x-1)\ldots(x-l+1)(x-l)}{(N+1)N(N-1)\ldots(N-l+1)(N-l)}}{\sum\limits_{0}^{N+1} \dfrac{x(x-1)\ldots(x-l+1)}{(N+1)N(N-1)\ldots(N-l+1)}} =$$

$$[1/(N - l)] \frac{\sum x(x-1)\ldots(x-l)}{\sum x(x-1)\ldots(x-l+1)}.$$

But we have {again cf. formula (24) from Chapt.2}

$$\frac{(N+1)N(N-1)\ldots(N-l)}{l+2} \quad \text{and} \quad \frac{(N+1)N(N-1)\ldots(N-l+1)}{l+1}$$

for the numerator and the denominator respectively so that

$$H = [1/(N - l)] \frac{(N+1)N(N-1)\ldots(N-l+1)(l+1)}{(N+1)N(N-1)\ldots(N-l+1)(l+2)} = \frac{l+1}{l+2}.$$

Note that the probability determined by the fourth law agrees with reality (i.e., with our inner belief) as badly as it does when being determined by the third law.

We conclude here the exposition of the laws of probability and go over to its applications, and we shall begin with the prior probabilities; that is, with the applications of the first two laws.

3.2. On Mathematical Expectation

3.2.1. We shall now consider a quantity called *mathematical expectation*. It, as also mathematical probability, presents itself when we determine the chances of some event, but in practice it is more important than probability itself because on its basis we form an opinion about what we may expect before some event takes place.

Suppose that p_1, p_2, ..., p_i are the probabilities of incompatible events E_1, E_2, ..., E_i and that we expect that one of them will take place. Assume now that

$$a_1, \ a_2, \ ..., \ a_i \qquad\qquad\qquad\qquad \text{(iv)}$$

are quantities measuring these events (if, for example, the events are some gains then (iv) are their values); then we shall call the magnitude

$$a_1 p_1 \ + \ a_2 p_2 \ + \ ... \ + \ a_i p_i \ = \ \Sigma \, a_j p_j$$

the mathematical expectation of one of these events taking place. We shall simply call this magnitude the mathematical expectation of quantities (iv).

If we have only one event, E_1, its mathematical expectation will be $a_1 p_1$. This magnitude indeed represents what is usually called the mathematical expectation of quantity a_1.

We shall now turn to the solution of the following problem: Suppose that we have quantities x, y, z, ... the first of which can only take one of the values

$$x_1, \ x_2, \ ..., \ x_\lambda; \qquad\qquad\qquad\qquad \text{(v)}$$

the second one, of the values

$$y_1, \ y_2, \ ..., \ y_\mu; \qquad\qquad\qquad\qquad \text{(vi)}$$

and the third quantity, only one of the values

$$z_1, \ z_2, \ ..., \ z_\nu, \text{ etc,}$$

and assume also that the number of values x_1, x_2, ..., y_1, y_2, ..., z_1, z_2, ..., ... can be indefinitely large. Then, we suppose that p_λ is the probability that x has value x_λ; q_μ is the probability that y has value y_μ; r_ν, that z has value z_ν; etc.

On this basis we shall try to find the probability P that the sum

$$x_\lambda + \ y_\mu \ + \ z_\nu \ + ... \qquad\qquad\qquad\qquad \text{(vii)}$$

is contained between certain boundaries L and M (which, as we shall see below, will not be absolutely arbitrary).

We denote the mathematical expectations of the quantities x, y, z, ... by a, b, c, ... respectively, so that

$$a \ = \ \Sigma \, x_\lambda p_\lambda, \ b \ = \ \Sigma \, y_\mu q_\mu, \ c \ = \ \Sigma \, z_\nu r_\nu, \ ...$$

Then, the mathematical expectations of the squares of these quantities will be a_1, b_1, c_1, ...:

$$a_1 \ = \ \Sigma \, x_\lambda^2 p_\lambda, \ b_1 \ = \ \Sigma \, y_\mu^2 q_\mu, \ c_1 \ = \ \Sigma \, z_\nu^2 r_\nu, \ ...$$

Since, according to our supposition, x certainly ought to take one of the values (v); y, one of the values (vi), etc, then

$$\Sigma\, p_\lambda \;=\; 1, \;\; \Sigma\, q_\mu \;=\; 1, \;\; \Sigma\, r_\nu = 1 \,\ldots$$

We shall consider now the sum

$$S \;=\; \Sigma\, [x_\lambda \,+\, y_\mu \,+\, z_\nu \,+\ldots\, -(a \,+\, b \,+\, c \,+\,\ldots)]^2\, p_\lambda q_\mu r_\nu\ldots \qquad\qquad \text{(viii)}$$

extended over all the values of x, y, z, … indicated above.
   Supposing now that

$$U \;=\; y_\mu \,+\, z_\nu \,+\ldots\, -\, b \,-\, c \,-\,\ldots$$

we have

$$S \;=\; \Sigma\, (x_\lambda \,-\, a \,+\, U)^2\, p_\lambda q_\mu r_\nu\ldots$$

or

$$S \;=\; \Sigma\, (x_\lambda \,-\, a)^2\, p_\lambda q_\mu r_\nu\ldots \;+\; 2\,\Sigma\, (x_\lambda \,-\, a)\, U\, p_\lambda q_\mu r_\nu\ldots \;+\; \Sigma\, U^2\, p_\lambda q_\mu r_\nu\ldots$$

The first term is

$$\Sigma\, (x_\lambda \,-\, a)^2\, p_\lambda\, \Sigma\, q_\mu\, \Sigma\, r_\nu\ldots \;=\; \Sigma\, x_\lambda^2\, p_\lambda \,-\, 2a\, \Sigma\, x_\lambda\, p_\lambda + \,a^2\, \Sigma\, p_\lambda \;=\; a_1 \,-\, a^2.$$

Since U does not depend on λ, the second term is

$$2\,\Sigma\, (x_\lambda \,-\, a)\, p_\lambda\, \Sigma\, U\, q_\mu r_\nu\ldots, \;\; 2\,\Sigma\, (x_\lambda \,-\, a)\, p_\lambda \;=\; 2\,\Sigma\, x_\lambda\, p_\lambda \,-\, 2a\, \Sigma\, \lambda \;=\; 0.$$

Finally, the third term is

$$\Sigma\, p_\lambda\, \Sigma\, U^2\, q_\mu r_\nu\ldots \;=\; \Sigma\, U^2\, q_\mu r_\nu\ldots$$

And so, the sum (viii) is

$$S \;=\; a_1 \,-\, a^2 \,+\, \Sigma\, U^2\, q_\mu r_\nu\ldots=$$
$$a_1 \,-\, a^2 \,+\, \Sigma\, [y_\mu \,+\, z_\nu \,+\ldots\, -\, (b \,+\, c \,+\,\ldots)]^2\, q_\mu r_\nu\ldots$$

 It is not difficult to conclude now that

$$S \;=\; (a_1 \,-\, a^2) \,+\, (b_1 \,-\, b^2) \,+\, (c_1 \,-\, c^2)+\ldots$$

Supposing now that

$$\{[x_\lambda \,+\, y_\mu \,+\, z_\nu \,+\ldots\, -(a \,+\, b \,+\, c \,+\,\ldots)] \,/\, t\sqrt{S}\} \;=\; V_{\lambda\mu\nu}\ldots$$

where t is an absolutely arbitrary quantity, we have

$$\Sigma\, V^2\, p_\lambda q_\mu r_\nu\ldots \;=\; 1/t^2$$

where the sum is extended over all the indicated values of x, y, z, …
   Decompose now this sum into three such sums that the first one, denoted by Σ_1, will extend over all the values of these variables for which $-\infty < V < -1$; the second, Σ_2, over the values for which
   $-1 < V < 1;$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (ix)

and the third one, Σ_3, will extend over those values for which $1 < V < +\infty$. We shall therefore have

$$[\Sigma_1 + \Sigma_2 + \Sigma_3]\, V^2\, p_\lambda q_\mu r_\nu \ldots = 1/t^2.$$

Since V^2 is greater than 1 both in the first and the third sums, and since it is greater than 0 in the second one, we obtain, noting that all the terms are positive,

$$\Sigma_1 + \Sigma_2 + \Sigma_3 > \Sigma_1\, p_\lambda q_\mu r_\nu \ldots + \Sigma_3\, p_\lambda q_\mu r_\nu \ldots$$

and

$$\Sigma_1\, p_\lambda q_\mu r_\nu \ldots + \Sigma_3\, p_\lambda q_\mu r_\nu \ldots < 1/t^2. \qquad\qquad\qquad (\mathrm{x})$$

But $p_\lambda q_\mu r_\nu \ldots$ is the probability of the sum (vii) and the sum of these products extended over all the *possible* values of the variables $x,\ y,\ z,\ \ldots$ is the probability that we will have one of the sums (vii), and this probability is equal to 1. Thus,

$$\Sigma_1\, p_\lambda q_\mu r_\nu \ldots + \Sigma_2\, p_\lambda q_\mu r_\nu \ldots + \Sigma_3\, p_\lambda q_\mu r_\nu \ldots = 1.$$

Subtracting the inequality (x) we have

$$\Sigma_2\, p_\lambda q_\mu r_\nu \ldots > 1 - 1/t^2.$$

However; the left side is the probability that there exists one of the sums (vii) which only includes the quantities $x,\ y,\ z,\ \ldots$ satisfying the inequalities (ix). In other words, this side is the probability P that the sum $(x + y + z + \ldots)$ takes one of the values imparted to it by the quantities $x,\ y,\ z,\ \ldots$ leading to values of V obeying inequalities (ix). We thus have

$$1 > P > 1 - 1/t^2.$$

But the inequalities (ix) provide

$$|x_\lambda + y_\mu + z_\nu + \ldots - (a + b + c + \ldots)| < |t|\sqrt{(a_1 - a^2) + (b_1 - b^2) + \ldots}$$

Supposing now that

$$L = a + b + c + \ldots - t\sqrt{(a_1 - a^2) + (b_1 - b^2) + (c_1 - c^2) + \ldots}$$
$$M = a + b + c + \ldots + t\sqrt{(a_1 - a^2) + (b_1 - b^2) + (c_1 - c^2) + \ldots}$$

we come to the following conclusion: *The probability P that the sum $(x + y + z + \ldots)$ is contained within the boundaries L and M is determined by the equality*

$$P = 1 - \theta/t^2 \qquad\qquad\qquad (\mathrm{xi})$$

where θ is some positive proper fraction.

3.2.2. Suppose now that the number of the quantities $x,\ y,\ z,\ \ldots$ is n so that $(x + y + z + \ldots)/n$ is their arithmetic mean. On the basis of the preceding exposition we conclude that P is the probability that the mean is confined within the boundaries

$$\frac{a + b + c + \ldots}{n} + \frac{t}{\sqrt{n}}\sqrt{\frac{(a_1 - a^2) + (b_1 - b^2) + (c_1 - c^2) + \ldots}{n}},$$

$$\frac{a + b + c + \ldots}{n} - \frac{t}{\sqrt{n}}\sqrt{\frac{(a_1 - a^2) + (b_1 - b^2) + (c_1 - c^2) + \ldots}{n}}.$$

Assuming that x, y, z, ... are repetitions of one and the same event, , we will have $a = b = c = ...,$ $a_1 = b_1 = c_1 = ...$ Consequently, we find that the probability P that the inequalities

$$a - \frac{t}{\sqrt{n}} \sqrt{a_1 - a^2} < \frac{x + y + z + ...}{n} < a + \frac{t}{\sqrt{n}} \sqrt{a_1 - a^2}$$

take place is determined by the formula (xi).

It is seen now that by choosing t we may bring P arbitrarily close to 1 and that, on the other hand, the boundaries within which the mean is confined merge at any t and $n = \infty$ and become equal to a. Thus, if we take $t = n^{1/6}/\alpha$, then the probability P that the mean is contained within the boundaries

$$a - \sqrt{a_1 - a^2} \, / \alpha n^{1/3} \text{ and } a + \sqrt{a_1 - a^2} \, / \alpha n^{1/3}$$

is expressed by the formula

$$P = 1 - \alpha^2 \theta / n^{1/3}.$$

Since t is arbitrary, we may assume that α is some constant quantity not depending on n, and we will have

$$\lim P_{n=\infty} = 1$$

where the left side is the probability that

$$\lim[(x + y + z + ...)/n]_{n=\infty} = a.$$

Since this probability is equal to 1, we indeed infer that the limit of the mean of the quantities x, y, z, ... as their number increases to infinity is a.

Here, we encounter a new concept of limit because we cannot anymore apply to our case the definition of limit which is usually met with in mathematics. In accord with that definition the limit is such a constant quantity the difference between which and the variable can be made less than any given magnitude. As a matter of fact, we cannot here assert this with certainty; we may only say that *the probability*, that the difference between the constant and the variable quantity can be made less than any given magnitude, has 1 as its limit and in this instance we attribute to the word *limit* its usual meaning. We are now accepting this new definition of limit according to which we conclude about the limit of some quantity by its probability having 1 as its limit.

On the grounds of this definition we may state the inferences reached in this section in the form of the following theorem.

Theorem. *The arithmetic mean of a very large number of quantities having the same mathematical expectations has as its limit this mathematical expectation.*

3.2.3. Let us now go over to other corollaries of the results attained in §3.2.1. Especially remarkable is the case in which all depends on whether or not some event takes place. In this instance we consider x, y, z, ... as the magnitudes of such events each of which has two forms: it either does not occur; or occurs. Let us agree to attribute to these magnitudes values equal to 0 in the first case, and to 1 in the second one, so that

$$x_1 = 0, \, x_2 = 1; \, y_1 = 0, \, y_2 = 1; \, z_1 = 0, \, z_2 = 1; \text{ etc.}$$

Here, p_1, q_1, r_1, ...{see above} are the probabilities that the first, the second; the third; ... event does not occur, and p_2, q_2, r_2, ...are the probabilities that these events take place. And, since x, y, z, ... only take values 0 and 1, it follows that

$$a = 0 \, p_1 + 1 \, p_2 = p_2; \, a_1 = 0^2 p_1 + 1^2 p_2 = p_2;$$
$$b = 0 \, q_1 + 1 \, q_2 = q_2; \, b_1 = 0^2 q_1 + 1^2 q_2 = q_2; \text{ etc.}$$

Denoting now p_2, q_2, r_2, ... by p, q, r, ... we have

$a = a_1 = p$, $b = b_1 = q$, $c = c_1 = r$, etc.

Consequently, we find that the expression (xi) is the probability that the arithmetic mean $[(x + y + z + ...]/n]$ is contained within the boundaries

$$\frac{p+q+r+...}{n} - \frac{t}{\sqrt{n}} \sqrt{\frac{(p-p^2)+(q-q^2)+...}{n}},$$

$$\frac{p+q+r+...}{n} + \frac{t}{\sqrt{n}} \sqrt{\frac{(p-p^2)+(q-q^2)+...}{n}}.$$

It is not difficult to see that

$$\frac{x+y+z+...}{n} = \frac{m}{n}$$

with m being the number of cases in which the event[s] took place, so that, like it happened in §3.2.2, we conclude that

$$\lim (m/n) = [(p + q + r + ...]/n].$$

This equality expresses the following theorem.

Theorem. *The limit of the ratio of the number of repetitions of an event to the number of trials is equal to the arithmetic mean of the probabilities of the events.*

This law was discovered by Poisson and represents a generalization of the Bernoulli law which is obtained from the above under the assumption that $p = q = r = ...$ and which therefore can be expressed by the equality

$$\lim (m/n) = p.$$

This equality shows that, *given a very large number of trials performed on some event with the probability of its occurrence remaining the same at each trial, the limit of the ratio of the number of the repetitions of the event to the number of the trials is equal to the probability of the event.*

3.3. On the Repetition of Events

3.3.1. Suppose that n trials are performed on some event E and assume at first that at each definite trial the probabilities of this event are different [1] so that p_i is the probability that the event occurs at the i-th trial. We shall try now to determine the probability $P_{m,n}$ that in these n trials the event took place m times. Noting that this result can happen in very different ways depending on how the trials in which the event occurs, and does not occur, follow one another, we shall determine the probability of the event taking place m times in some definite order and calculate the sum of these probabilities which will indeed be the probability $P_{m,n}$ sought.

Each of the terms of this sum will be composed of the quantities p_1, p_2, ..., p_n in the following way: If the event E occurs m times in the first m trials, and does not take place anymore, then this will represent one of the ways in which E can occur m times in n trials.

This instance may be considered as a compound event because it represents a joint occurrence of several events, i.e., of the event taking place at the first; at the second; ...; at the m-th trial; and not occurring at the $(m + 1)$-th; ...; at the n-th trial. Therefore, the probability of this combination will be expressed as

$p_1 p_2 ... p_m (1 - p_{m+1}) (1 - p_{m+2}) ... (1 - p_n).$

The probabilities of the other combinations will be found in this way {as well}, and the sum of the probabilities of all the different combinations whose number is $C_n{}^m$ will indeed represent, as stated above, the probability sought, $P_{m, n}$.

It is not difficult to see now that $P_{m, n}$ is the coefficient of t^m in the expansion

$$(p_1 t + 1 - p_1)(p_2 t + 1 - p_2) \ldots (p_n t + 1 - p_n) = \Sigma P_{n, k} t^k.$$

This method of determining probability as the coefficient in some expansion was proposed by Laplace [2] and the function whose expansion determines the magnitudes $P_{n, k}$ is called the *fonction génératrice*.

For the case in which the event has one and the same probability p of taking place in all the trials $p_1 = p_2 = \ldots = p_n = p$; consequently, the *generating* function will become $(pt + 1 - p)^n$ so that

$$P_{m, n} = C_n{}^m p^m (1 - p)^{n-m}.$$

Noting that the multiplier $p^m (1 - p)^{n-m}$ represents here the probability that the event E occurs m times in n trials in some definite order, and that the other factor is the number of such orders, we could have also directly determined the probability $P_{m, n}$ in this particular case.

Note 1. {A careless phrase.}
Note 2. {Simpson and Lagrange preceded Laplace.}

3.3.2. We shall now determine the probability $P_{m, n}$ by a third method which will indicate a widely spread and in many cases useful trick for deriving probabilities. The trick {the method} consists in working out and integrating an equation in finite differences which the probability sought should satisfy.

For making up such an equation in the case under consideration, we note that the event E can only be repeated m times in n trials in two ways: either taking place at the n-th trial, or not. In the first case, the magnitude sought becomes $P_{n-1, m-1}$ which is the probability that the event will be repeated $(m - 1)$ times in the first $(n - 1)$ trials; in the second instance, it becomes $P_{n-1, m}$, i.e., the probability that the event takes place m times in the $(n - 1)$ trials. [1]

Supposing now that the probability of E is constant in each trial, we shall find that the probability of the first assumption is p, whereas the probability of the second, exactly contrary one, is $(1 - p)$. We thus have the equation

$$P_{n, m} = p P_{n-1, m-1} + (1 - p) P_{n-1, m}.$$

It includes two independent variables, n and m, and does not therefore belong to those considered in Chapter 2. In general, the integration of such equations presents serious difficulties, but this one can be integrated by means of *generating* functions. We shall indeed do this now, but at first we note the following. When integrating any equation, there appear arbitrary constants whose determination in each particular case is sometimes very troublesome. In the theory of probability, this determination does not present any difficulties because some particular cases cannot take place in virtue of the point of matter itself. Thus, in the expression $P_{n, m}$ m cannot exceed n, so that $P_{n, m} = 0$ at $m > n$ because such an event is impossible. For the same reason $P_{0, m} = 0$ and $P_{n, 0} = 0$ { ? }; again, $P_{n, m} = 0$ for negative values of m or n.

Incidentally, this fact shows that in probability theory it is not necessary to consider whether some hypothesis is possible or not: its impossibility will lead to the vanishing of its probability and the final result will have such a form as though we had only regarded possible hypotheses.

After these remarks we shall go over to the integration of our equation. Multiplying it by t^m where t is an arbitrary quantity we have

$$P_{n, m} t^m = P_{n-1, m-1} p t^m + P_{n-1, m} (1 - p) t^m.$$

Summing both sides of this equality over all possible values of m, i.e., from 0 to n inclusive, we obtain [2]

$$\sum_{m=0}^{n+1} P_{n, m} t^m = \sum_{m=0}^{n+1} P_{n-1, m-1} p t^m + \sum_{m=0}^{n+1} P_{n-1, m} (1 - p) t^m. \qquad \text{(xii)}$$

Supposing now that

$$U_\lambda = \sum_{m=o}^{\lambda+1} P_{n,m} t^m$$

we obtain [3] U_n for the left side of (xii). The first term on the right side will be

$$pt \sum_{m=0}^{n+1} P_{n-1,m-1} t^{m-1} = pt \sum_{m=1}^{n} P_{n-1,m} t^m =$$

$$pt [P_{n-1,-1} t^{-1} + \sum_{m=0}^{n} P_{n-1,m} t^m].$$

But, as we saw, $P_{n-1,-1} = 0$ so that this term equals ptU_{n-1}. Then, the second term on the right side is

$$(1-p) [\sum_{m=0}^{n} P_{n-1,m} t^m + P_{n-1,n} t^n] = (1-p) \sum_{m=0}^{n} P_{n-1,m} t^m$$

because $P_{n-1,n} = 0$. Therefore the second term is equal to $(1-p)U_{m-1}$ and our equation becomes

$$U_n = ptU_{n-1} + (1-p)U_{n-1}.$$

We have thus the following linear equation

$$U_n - (pt + 1 - p)U_{n-1} = 0$$

whose integral is, as it is not difficult to see,

$$U_n = C(pt + 1 - p)^n. \tag{xiii}$$

Here, C is an arbitrary constant which we shall determine. Note that

$$U_1 = C(pt + 1 - p)$$

which follows from the equation (xiii) and that, at the same time,

$$U_1 = \sum_{m=0}^{2} P_{1,m} t^m = P_{1,0} t^o + P_{1,1} t.$$

But $P_{1,0}$ is the probability that the event does not take place in one trial, and $P_{1,1}$, the probability that it occurs, so that

$$P_{1,0} = 1 - p, \quad P_{1,1} = p \quad \text{and} \quad C = 1.$$

And so, we have the formula

$$\sum_{m=0}^{n+1} P_{n,m} t^m = (pt + 1 - p)^n. \tag{1}$$

Here, t is absolutely arbitrary. The coefficients of the same powers of t in both sides of the equality should be equal; expanding the right side in accord with the Newton binomial, we have

$$P_{n,m} = C_n^m \, p^m \, (1-p)^{n-m}. \qquad\qquad\qquad (2)$$

Note that at the limiting cases of $m = 0$ and $m = n$ this expression should not be considered literally: in these instances the coefficient, as it follows from the expansion, ought to be considered equal to 1, so that $P_{n,0} = 1 - p$, $P_{n,n} = p^n$.

Note 1. {Chebyshev thus changed the notation of §3.3.1. Moreover, even there he wrote either $P_{m,n}$ or $P_{n,m}$ but I had then standardized this.}

Note 2. {Chebyshev did not pay due attention to the boundaries of the sums below; he apparently acted in the spirit of his own remarks formulated just above.}

Note 3. We take the quantity pt out of the sign of summation because it does not depend on the variable m with respect to which the summation is carried out. Indeed, p is constant whereas t is an arbitrary quantity which we also assume independent from m.

3.3.3. Let us derive now the Bernoulli law by issuing from equation (1). Differentiating it with respect to t we find that

$$\sum_{m=0}^{n+1} m P_{n,m} \, t^{m-1} = n \, (pt + 1 - p)^{n-1} p,$$

$$\sum_{m=0}^{n+1} m \, (m-1) P_{n,m} \, t^{m-2} = = n \, (n-1) \, (pt + 1 - p)^{n-2} p^2.$$

Assume now that in equation (1) and in these two equations $t = 1$. We obtain

$$\sum_{m=0}^{n+1} P_{n,m} = 1, \quad \sum_{m=0}^{n+1} m P_{n,m} = np,$$

$$\sum_{m=0}^{n+1} m \, (m-1) \, P_{n,m} = n \, (n-1) \, p^2 = n^2 p^2 - np^2.$$

From the last two equations we find that

$$\sum_{m=0}^{n+1} m^2 \, P_{n,m} = n^2 p^2 + n \, (1-p) \, p$$

and therefore

$$\sum_{m=0}^{n+1} (m - np)^2 \, P_{n,m} = \sum_{m=0}^{n+1} m^2 \, P_{n,m} - 2pn \sum_{m=0}^{n+1} m \, P_{n,m} +$$

$$p^2 n^2 \sum_{m=0}^{n+1} P_{n,m} = n \, (1-p) \, p$$

and

$$\sum_{m=0}^{n+1} P_{n,m} \left(\frac{m - np}{s\sqrt{n(1-p)p}} \right)^2 = \frac{1}{s^2}$$

where s is an arbitrary number. Hence

$$\left(\sum_{m=0}^{\mu+1} + \sum_{m=\mu+1}^{\nu+1} + \sum_{m=\nu+1}^{n+1} \right) P_{n,m} \left(\frac{m - np}{s\sqrt{n(1-p)p}} \right)^2 = \frac{1}{s^2}.$$

Denote the fraction in brackets by A and assume that

$$-\infty < A < -1 \text{ for } 0 \le m < \mu + 1,$$
$$-1 < A < 1 \text{ for } \mu + 1 \le m < \nu + 1,$$
$$1 < A < +\infty \text{ for } \nu + 1 \le m \le n + 1,$$

then

$$\left(\sum_{m=0}^{\mu+1} + \sum_{m=\nu+1}^{n+1} \right) P_{n,m} < \frac{1}{s^2}. \qquad\qquad (\text{xiv})$$

But we have

$$\left(\sum_{m=0}^{\mu+1} + \sum_{m=\mu+1}^{\nu+1} + \sum_{m=\nu+1}^{n+1} \right) P_{n,m} = 1$$

so that, taking into account the inequality (xiv),

$$\sum_{m=\mu+1}^{\nu+1} P_{n,m} > 1 - \frac{1}{s^2}.$$

The left side of this inequality is the probability that in n trials the number of repetitions of the event is contained within the boundaries $(\mu + 1)$ and $[\nu + 1)$; or, the probability that

$$-1 < \frac{m - np}{s\sqrt{n(1-p)p}} < 1. \qquad\qquad (3)$$

Denoting it by Π we obtain

$$1 > \Pi > 1 - \frac{1}{s^2}.$$

The inequality (3) provides

$$pn - s\sqrt{p(1-p)n} < m < pn + s\sqrt{p(1-p)n}$$

so that

$$p - (s/\sqrt{n})\sqrt{p(1-p)} < m/n < p + (s/\sqrt{n})\sqrt{p(1-p)}.$$

It is seen now that the boundaries within which m/n is contained merge at $n = \infty$ and become p whereas the probability that m/n is contained within them is represented by the formula

$$\Pi = 1 - \theta/s^2$$

where θ is a proper fraction and s, an arbitrary number. It is seen that this probability may be made arbitrarily close to 1, and we conclude that

$$\lim (m/n)_{n=\infty} = p.$$

In this case, the limit has that special meaning which we discussed in §3.2.2. We thus arrived at the Bernoulli law. Our derivation was based on the value of the sum

$$\sum_{m=0}^{n+1} (m - np)^2 P_{n,m} = \sum_{m=0}^{n+1} C_n^m p^m (1 - p)^{n-m}(m - np)^2.$$

Note that that sum can also be determined otherwise. Suppose that in equation (1) $t = e^\alpha$ and multiply both its sides by $e^{-\alpha pn}$ to obtain

$$\sum_{m=0}^{n+1} e^{\alpha(m-np)} P_{n,m} = [p\, e^{\alpha(1-p)} + e^{-\alpha p}(1 - p)]^n.$$

However,

$$e^x = 1 + (x/1!) + (x^2/2!) + \dots$$

and

$$e^{\alpha(m-np)} = 1 + (\alpha/1!)(m - np) + (\alpha^2/2!)(m - np)^2 + \dots,$$
$$e^{\alpha(1-p)} = 1 + (\alpha/1!)(1 - p) + (\alpha^2/2!)(1 - p)^2 + \dots,$$
$$e^{-\alpha p} = 1 - (\alpha/1!)p + (\alpha^2/2!)p^2 - \dots$$

Consequently,

$$\sum_{m=0}^{n+1} P_{n,m} + \alpha \sum_{m=0}^{n+1} P_{n,m}(m - np) + (\alpha^2/2) \sum_{m=0}^{n+1} P_{n,m}(m - np)^2 + \dots =$$
$$\{[p + (\alpha/1!)\,p\,(1 - p) + (\alpha^2/2!)\,p\,(1 - p)^2 + \dots] +$$
$$[(1 - p) - (\alpha/1!)\,p\,(1 - p) + (\alpha^2/2!)\,p^2\,(1 - p) - \dots]\}^n =$$
$$[1 + (\alpha^2/2!)\,p\,(1 - p) + \dots]^n.$$

Therefore

$$\sum_{m=0}^{n+1} P_{n,m} = 1, \quad \sum_{m=0}^{n+1} P_{n,m}(m - np) = 0,$$
$$\sum_{m=0}^{n+1} P_{n,m}(m - np)^2 = p\,(1 - p)\,n, \dots$$

3.3.4. Let us now go over to another problem. We shall try to determine the most probable number of repetitions of an event in a definite number of trials. The problem is reduced to the derivation of the value of m at which $P_{n,m}$ becomes maximal for given values of n and p.

Let μ be the sought value of m. For $P_{n,\mu}$ to be maximal, conditions $P_{n,\mu-1} \leq P_{n,\mu}$, $P_{n,\mu+1} \leq P_{n,\mu}$ are necessary. We have however

$$P_{n,\mu-1} = C_n^{\mu-1} p^{\mu-1}(1 - p)^{n-\mu+1}, \quad P_{n,\mu} = C_n^\mu p^\mu (1 - p)^{n-\mu},$$
$$P_{n,\mu+1} = C_n^{\mu+1} p^{\mu+1}(1 - p)^{n-\mu-1}$$

and our conditions become

$$\frac{1-p}{n-\mu+1} \leq \frac{p}{\mu}, \quad \frac{p}{\mu+1} \leq \frac{1-p}{n-\mu}.$$

It follows that

$$(1 - p)\mu \leq p(n - \mu + 1), \quad p(n - \mu) \leq (1 - p)(\mu + 1),$$
$$\mu \leq p(n + 1), \quad p(n + 1) - 1 \leq \mu.$$

We ought to consider now two cases: either $p(n+1)$ is a fraction or an integer. In the first instance equalities are impossible. Issuing then from the inequalities

$$p(n+1) - 1 < \mu < p(n+1)$$

we obtain

$$\mu = E[p(n+1)]. \qquad\qquad\qquad\qquad (xv)$$

In the second case

$$\mu \le p(n+1), \ \mu > p(n+1) - 1, \ \mu \ge p(n+1), \ \mu < p(n+1)$$

so that we get two solutions

$$\mu_1 = p(n+1) - 1, \ \mu_2 = p(n+1)$$

and the corresponding values of $P_{n,\mu}$ will both be maximal. Note that if n, m and $(n-m)$ are very large numbers, we may approximately assume np as the value imparting the maximal value to $P_{n,\mu}$. In the sequel, we shall therefore assume $P_{n,np}$ as the maximal value of $P_{n,m}$.

3.3.5. Applying the Stirling formula we can write the following approximate equality

$$x! = \sqrt{2\pi}\, x^{x+1/2} e^{-x}$$

so that

$$P_{n,m} = \sqrt{2\pi}\ \frac{n^{n+1/2} e^{-n} p^m (1-p)^{n-m}}{2\pi m^{m+1/2} e^{-m} (n-m)^{n-m+1/2} e^{-(n-m)}}.$$

After reductions, we will have

$$P_{n,m} = \sqrt{\frac{n}{2\pi m(n-m)}}$$

$$P_{n,,m} = \sqrt{\frac{n}{2\pi m(n-m)}}\ \frac{n^n p^m (1-p)^{n-m}}{m^m (n-m)^{n-m}} =$$

$$\sqrt{\frac{n}{2\pi m(n-m)}} \left(\frac{np}{m}\right)^m \left(\frac{n(1-p)}{n-m}\right)^{n-m}. \qquad\qquad (4)$$

For the maximal value we have approximately $m = np, \ n - m = (1-p)n$ so that this value will be

$$P_{n,np} = \frac{1}{\sqrt{2\pi\, p(1-p)\, n}}. \qquad\qquad\qquad (5)$$

Issuing from (4), let us find now the most probable number m. Since this formula represents an approximate expression for $P_{n,m}$ for very large values
of n, m and $(n-m)$, and since we neglect proper fractions as compared with these numbers, we may consider m as a quantity changing continuously and, consequently, as being able to take fractional values. Indeed, supposing that m has some very large integral value, and adding to it any proper fraction and calculating the probability $P_{n,m}$ by formula (4) for both cases, we will obtain results very little differing from each other if only m differs from np by a finite magnitude or by an infinite magnitude of an order lower than 1 with respect to n.

Accordingly, when determining the most probable number m which makes $P_{n,,m}$ maximal, we may act in accord with the rules of the differential calculus. From equation (4) we have

$$\ln P_{n,,m} = (1/2)\ln n - (1/2)\ln(2\pi) - (1/2)\ln m - (1/2)\ln(n-m) +$$
$$m \ln np - m \ln m + (n-m)\ln[n(1-p)] - (n-m)\ln(n-m)$$

so that

$$(1/P_{n,,m})\, dP_{n,,m}/dm = -(1/2m) + 1/[2(n-m)] + \ln np - \ln m - 1 -$$
$$\ln[n(1-p)] + \ln(n-m) + 1.$$

In order to determine the most probable m we thus obtain the equation

$$\frac{n-2m}{2m(n-m)} + \ln\frac{m}{n-m} = \ln\frac{p}{1-p}. \qquad (6)$$

It cannot be solved exactly and we will therefore solve it approximately, especially because the strict solution is not here important for us: the equation (4) itself can only provide an approximate value of m.

Noting that for very large n, m and $(n-m)$ of the same order, the magnitude $(n-2m)/[m(n-m)]$ becomes very small (because its numerator is of the first order and its denominator, of the second order with respect to these numbers), we may neglect this fraction in the first approximation and obtain

$$\ln[m/(n-m)] = \ln[p/(1-p)], \quad m = np.$$

It is seen now that in the first approximation we arrived at the same magnitude as before {cf. (xv)}

$$m = E[p(n+1)] = np + \theta$$

where θ is a proper positive or negative fraction.

For the second approximation we shall substitute in equation (6), in its first term on the right side, the just obtained value of m instead of that letter. This will provide the equation

$$\frac{n-2np}{2np(1-p)n} + \ln\frac{m}{n-m} = \ln\frac{p}{1-p}$$

or

$$\frac{1-2p}{2p(1-p)n} + \ln\frac{m}{n-m} = \ln\frac{p}{1-p}$$

so that

$$\frac{m}{n-m} = \frac{p}{1-p}\exp\left[-\frac{1-2p}{2p(1-p)n}\right]$$

and

$$m = np\exp\left[-\frac{1-2p}{2p(1-p)n}\right] : \left\{[1-p+p\exp\left[-\frac{1-2p}{2p(1-p)n}\right]\right\} =$$
$$np:\left\{p+(1-p)\exp\left[\frac{1-2p}{2p(1-p)n}\right]\right\}.$$

Neglecting the higher powers of the expression $(1-2p)/[p(1-p)n]$ beginning with its square, we may substitute

$$\exp\left[-\frac{1-2p}{2p(1-p)n}\right] = 1 + \frac{1-2p}{2p(1-p)n}$$

so that

$$m = np \Big/ \left[1 + \frac{1-2p}{2p(1-p)n}\right] = np \Big/ \left[1 + \frac{1-2p}{2p(1-p)n}\right]^{-1}$$

and approximately

$$m = np - (1 - 2p)/2.$$

3.3.6. Issuing now from formula (4), we shall search for the probability that, for very large values of n, m will very little differ from np. Suppose that

$$m = np + z. \tag{7}$$

We derived formula (4) under the assumption that m, n and $(n - m)$ were very large numbers, and we ought to apply it under the same condition concerning z as well if the last quantity is considered separately and by itself. However, z can be very small as compared with n. We shall suppose that it has the same order as $\sqrt{n}$, then it will indeed possess the properties indicated above.

Applying expression (7), we have

$$\sqrt{2\pi\, m(n-m)} = \sqrt{2\pi\,(np+z)(n-np-z)} =$$
$$\sqrt{2\pi\, n^2[p+(z/n)][1-p-(z/n)]}\,.$$

Neglecting a very small magnitude z/n, we obtain

$$\sqrt{2\pi\, m(n-m)} = \sqrt{2\pi\, n^2 p(1-p)}$$

so that

$$\sqrt{\frac{n}{2\pi\, m(n-m)}} = \sqrt{\frac{1}{2\pi\, p(1-p)n}}\,.$$

We then have

$$\left(\frac{np}{m}\right)^m = \left(\frac{np}{np+z}\right)^{np+z} = \left(\frac{1}{[1+(z/np)]}\right)^{np+z}$$

and

$$\ln\left(\frac{np}{m}\right)^m = -(np+z)\ln[1+(z/np)] =$$
$$-(np+z)[(z/np) - (z^2/2n^2p^2) + (z^3/3n^3p^3) - \ldots] =$$
$$-z + (z^2/2np) - (z^2/np) - (z^3/3n^2p^2) + (z^3/2n^2p^2) + \ldots$$

However, the term (z^3/n^2p^2) and the following ones are very small magnitudes because, assuming that $z = a\sqrt{n}$, we have

$$\frac{z^{\mu}}{(np)^{\mu-1}} \frac{n^{\mu/2}}{n^{\mu-1}} = \frac{\alpha^{\mu} / p^{\mu-1}}{n^{[(\mu/2)-1]}}.$$

It is seen now that for $\mu > 2$ the first fraction in the left side is a very small magnitude (at $\mu = 2$ it is finite). Therefore, neglecting these small terms, we get

$$\ln \left(\frac{np}{m}\right)^{m} = -z - (z^2/2np).$$

Acting in the same way with the other multiplier in formula (4) we obtain it in the form

$$\frac{1}{\{1 - \dfrac{z}{n(1-p)}\}^{n(1-p)-z}}.$$

The logarithm of that multiplier will be

$$-[n(1-p) - z] \ln [1 - \frac{z}{n(1-p)}] =$$

$$[n(1-p) - z] [\frac{1}{n(1-p)} + \frac{z^2}{2n^2(1-p)^2} + \ldots] =$$

$$z + \frac{z^2}{2n(1-p)} - \frac{z^2}{n(1-p)} = z - \frac{z^2}{2n(1-p)}$$

and therefore

$$\ln \left\{ \left(\frac{np}{m}\right)^{m} \left(\frac{n(1-p)}{n-m}\right)^{n-m} \right\} = -(z^2/np) - [z^2/2n(1-p)] =$$

$$- [z^2/2p(1-p)n] \qquad (8)$$

and

$$\left(\frac{np}{m}\right)^{m} \left(\frac{n(1-p)}{n-m}\right)^{n-m} = \exp\left(-\frac{z^2}{2p(1-p)n}\right) = \exp\left\{-\frac{(m-np)^2}{2p(1-p)n}\right\}$$

so that

$$P_{n,m} = \frac{1}{\sqrt{2\pi \, p(1-p)n}} \exp\left\{-\frac{(m-np)^2}{2p(1-p)n}\right\}.$$

We shall now search for the probability that z is contained within some boundaries which is tantamount to determining the similar probability concerning m. Calling these boundaries L and M, denoting the probability sought by Π and supposing that m increases to the upper boundary not inclusively, we obtain

$$\Pi = P_{n,L} + P_{n,L+1} + P_{n,L+2} + \ldots + P_{n,M-1}$$

so that

$$\Pi = \sum_{m=L}^{M} \frac{1}{\sqrt{2\pi \, p(1-p)n}} \exp\left\{-\frac{(m-np)^2}{2p(1-p)n}\right\}.$$

However, we have in general

$$\sum V = \int V dx - (1/2)V + A_1\, dV/dx + A_2\, d^2V/dx^2 + \dots$$

In the case under consideration this will assume the form

$$\frac{1}{\sqrt{2\pi\, p\,(1-p)\,n}} \sum_{m=L}^{M} \exp\left\{-\frac{(m-np)^2}{2p\,(1-p)\,n}\right\} =$$

$$\frac{1}{\sqrt{2\pi\, p\,(1-p)\,n}} \int_{L}^{M} \exp\left\{-\frac{(m-np)^2}{2p\,(1-p)\,n}\right\}dm - \left\{\frac{1}{\sqrt{2\pi\, p\,(1-p)\,n}}\right.$$

$$\left.\exp\left[-\frac{(m-np)^2}{2p\,(1-p)\,n}\right][(1/2) + A_1\frac{m-np}{p(1-p)n} + \dots]\right\}.\begin{matrix}M\\[4pt]L\end{matrix}$$

But, for very large values of n, the quantities

$$\frac{1}{\sqrt{2\pi\, p\,(1-p)\,n}}, \quad \frac{m-np}{p(1-p)n}, \quad \left(\frac{m-np}{p(1-p)n}\right)^2, \quad \dots$$

are very small whereas

$$\exp\left\{-\frac{(m-np)^2}{2p\,(1-p)\,n}\right\} \tag{xvi}$$

is always finite as also is the product of its integral by

$$\frac{1}{\sqrt{2\pi\, p\,(1-p)\,n}},$$

see below. Therefore, neglecting the terms consisting of the product of the indicated magnitudes by (xvi) we have

$$\Pi = \frac{1}{\sqrt{2\pi\, p\,(1-p)\,n}} \int_{L}^{M} \exp\left\{-\frac{(m-np)^2}{2p\,(1-p)\,n}\right\}dm.$$

Substituting now

$$\frac{m-np}{\sqrt{2p\,(1-p)\,n}} = t$$

we obtain

$$\Pi = (1/\sqrt{\pi})\int_{t_0}^{t_1} \exp(-t^2)\, dt \tag{9}$$

where

$$t_0 = \frac{L-np}{\sqrt{2p\,(1-p)\,n}}, \quad t_1 = \frac{M-np}{2p\,(1-p)\,n}$$

so that

$$L = np + t_0 \sqrt{2p\,(1-p)\,n}\,, \quad M = np + t_1 \sqrt{2p\,(1-p)\,n}\,.$$

And so, the probability that m is contained within L and M is determined by the formula (9). Especially remarkable is the case in which $t_0 = -u$, $t_1 = u$. Then

$$\Pi = \frac{2}{\sqrt{\pi}} \int_0^u \exp(-t^2)\,dt \qquad\qquad\qquad (10)$$

is the probability that m is contained within the boundaries

$$np - u\sqrt{2p\,(1-p)\,n} \quad \text{and} \quad np + u\sqrt{2p\,(1-p)\,n}$$

or that $[(m/n) - p]$ is contained within

$$-u\sqrt{2p\,(1-p)/n} \quad \text{and} \quad u\sqrt{2p\,(1-p)/n}$$

We have however

$$\int_0^\infty \exp(-t^2)\,dt = \sqrt{\pi}/2$$

and therefore the left side of formula (10) tends to 1 as n increases to infinity. Nevertheless, for any u the boundaries, within which $[(m/n) - p]$ is contained, draw together as n increases and vanish {coincide} at $n = \infty$. We conclude therefore, as we did before, that

$$\lim (m/n) = p.$$

3.3.7. We shall now calculate the integral

$$I = \int_0^u \exp(-t^2)\,dt$$

so as to show how rapidly its values approaches $\sqrt{\pi}/2$ as u increases. We have

$$\exp(-t^2) = 1 - t^2/1! + t^4/2! - \dots$$

so that

$$I = u - u^3/(1\cdot 3) + u^5/(1\cdot 2\cdot 5) - u^7/(1\cdot 2\cdot 3\cdot 7) + u^9/(1\cdot 2\cdot 3\cdot 4\cdot 9) - \dots \quad (11)$$

We have thus expressed this integral by a series very rapidly converging at a small u and therefore very convenient for calculations under this condition. However, the case in which u is large is especially important for us whereas this series then converges very slowly and we ought therefore to turn to another method of calculating the integral (11). We note that

$$\int_0^u \exp(-t^2)\,dt = \int_0^\infty \exp(-t^2)\,dt - \int_u^\infty \exp(-t^2)\,dt =$$

$$\sqrt{\pi}/2 \; - \; \int_u^\infty \exp\left(-t^2\right) dt$$

so that

$$\Pi \; = \; 1 - \frac{2}{\sqrt{\pi}} \int_u^\infty \exp\left(-t^2\right) dt \tag{12}$$

and we shall now indeed turn to calculating the integral

$$\int_u^\infty \exp\left(-t^2\right) dt. \tag{xvii}$$

It is equal to

$$- (1/2) \int_u^\infty -2t \exp\left(-t^2\right) dt \, (1/t) \; = \; \left[\exp\left(-t^2\right)/t\right]_u^\infty \quad + $$

$$\int_u^\infty - (1/2) \, t^{-2} \exp\left(-t^2\right) dt \; = \; \left[\exp\left(-u^2\right)/2u\right] - (1/2)\int_u^\infty t^{-2} \exp\left(-t^2\right) dt$$

and

$$\int_u^\infty t^{-2} \exp\left(-t^2\right) dt \; = \; (1/2) - \int_u^\infty -2t \exp\left(-t^2\right) dt \, t^{-3} \; =$$

$$\left\{\left[\exp\left(-t^2\right)/2\right] t^{-3}\right\}_u^\infty \quad - (3/2)\int_u^\infty t^{-4} \exp\left(-t^2\right) dt \; =$$

$$\left[\exp\left(-u^2\right)/2\, u^3\right] - (3/2)\int_u^\infty t^{-4} \exp\left(-t^2\right) dt.$$

Consequently integral (xvii) equals

$$\left[\exp\left(-u^2\right)/2\, u\right] - \left[\exp\left(-u^2\right)/2^2\, u^3\right] + (3/2^2)\int_u^\infty t^{-4} \exp\left(-t^2\right) dt.$$

Since u and the integrals as well are here positive quantities, we conclude that

$$\left[\exp\left(-u^2\right)/2\, u\right] > \int_u^\infty \exp\left(-t^2\right) dt > \left[\exp\left(-u^2\right)/2\, u\right] -$$
$$\left[\exp\left(-u^2\right)/2^2 u^3\right] \tag{xviii}$$

and we can therefore arrive at

$$\int_u^\infty \exp\left(-t^2\right) dt = \theta \left[\exp\left(-u^2\right)/2\, u\right]$$

where θ is some proper fraction. And so,

$$\Pi \; = \; 1 - (\theta/\sqrt{\pi}) \left[\exp\left(-u^2\right)/u\right].$$

Issuing from the inequalities (xiii) we can also obtain a more precise formula because it follows from them that

$$\int_u^\infty \exp(-t^2)\,dt = [\exp(-u^2)/2u] - \theta\,[\exp(-u^2)/2^2 u^3].$$

This expression provides

$$\Pi = 1 - (1/\sqrt{\pi})\,[\exp(-u^2)/u]\,[1 - \theta/2u^2].$$

We see now that the probability sought very rapidly approaches 1 as u increases.

3.3.8. We shall now determine that limiting series by whose means the integral (xvii) can be expressed. Integrating by parts, we get

$$\int_u^\infty t^{-2n} \exp(-t^2)\,dt = \frac{u^{-2n-1}}{2}\exp(-u^2) - \frac{2n+1}{2}\int_u^\infty t^{-2n-2}\exp(-t^2)\,dt.$$

We arrive therefore consecutively at

$$\int_u^\infty \exp(-t^2)\,dt = (u^{-1}/2)\exp(-u^2) - (1/2)\int_u^\infty t^{-2}\exp(-t^2)\,dt,$$

$$\int_u^\infty t^{-2}\exp(-t^2)\,dt = (u^{-3}/2)\exp(-u^2) - (3/2)\int_u^\infty t^{-4}\exp(-t^2)\,dt,$$

$$\int_u^\infty t^{-4}\exp(-t^2)\,dt = (u^{-5}/2)\exp(-u^2) - (5/2)\int_u^\infty t^{-6}\exp(-t^2)\,dt$$

and

$$\int_u^\infty \exp(-t^2)\,dt = (u^{-1}/2)\exp(-u^2) - (u^{-3}/2^2)\exp(-u^2) +$$

$$(1\cdot3/2^3)\,(u^{-5}/2^3)\exp(-u^2) - [(1\cdot3\cdot5)/2^3]\int_u^\infty t^{-6}\exp(-t^2)\,dt.$$

We may now conclude by analogy that

$$\int_u^\infty \exp(-t^2)\,dt = (u^{-1}/2)\exp(-u^2) - (u^{-3}/2^2)\exp(-u^2) +$$
$$(1\cdot3/2^3)(u^{-5}/2^3)\exp(-u^2) - (1\cdot3\cdot5/2^3)(u^{-7}/2^4)\exp(-u^2) + \ldots +$$
$$(-1)^n\,[1\cdot3\cdot5 \ldots (2n-1)/2^{n+1}]\,u^{-2n-1}\exp(-u^2) + \ldots \qquad\qquad \textbf{(13)}$$

Let us substantiate this formula. Substituting

$$v_{n+1} = [1\cdot3\cdot5 \ldots (2n-1)/2^{n+1}]\,u^{-2n-1}\exp(-u^2)$$

and denoting by R_{n+1} the remainder of the series (13) beginning with the term following after v_{n+1}, we arrive at

$$R_{n+1} = -(-1)^n [1\cdot3\cdot5 \ldots (2n-1)(2n+1)/2^{n+1}] \int_u^\infty t^{-2n-2} \exp(-t^2)\, dt$$

$$-(-1)^n [1\cdot3\cdot5 \ldots (2n+1)/2^{n+2}] u^{-2n-3} \exp(-u^2) +$$

$$(-1)^n [1\cdot3\cdot5 \ldots (2n+1)(2n+3)/2^{n+2}] \int_u^\infty t^{-2n-4} \exp(-t^2)\, dt =$$

$$v_{n+2} + R_{n+2}.$$

This indeed justifies the series (13).

3.3.9. Until now, we determined the probability of some number of repetitions of an event in a certain number of trials given the probability of the event occurring in one trial. Now we go over to the solution of the inverse problem: Knowing that in a certain number of trials an event occurred some number of times, we will determine the (posterior) probability that it takes place in one trial. More precisely, we will derive the most probable boundaries within which this probability ought to be contained. We shall have to assume that all the hypotheses that can be made concerning this probability are equally probable.

Suppose that the number of these hypotheses is N and that the probabilities that the event occurs in one trial are $0/N$, $1/N$, $2/N$, …, λ/N, … $(N-1)/N$ respectively. In this case we have $P_1 = P_2 = \ldots = P_N = 1/N$. Denoting the recorded number of the repetitions of the event in n trials by m we obtain

$$p_\lambda = C_n^m (\lambda/N)^m [1 - (\lambda/N)]^{n-m}.$$

In accord with Theorem 3 we have now, for the probability that the event took place under hypothesis λ,

$$Q_\lambda = \frac{p\lambda/N}{\sum\limits_{\lambda=0}^N p\lambda/N} =$$

$$(\lambda/N)^m [1 - (\lambda/N)]^{n-m} (1/N) \div \sum_{\lambda=0}^N (\lambda/N)^m [1 - (\lambda/N)]^{n-m} (1/N).$$

Denoting the sought probability that the event occurs in one trial by p, we shall search for the probability S that p is contained within the boundaries $p_0 = \mu/N$ and $p_1 = \mu_1/N$. Since

$$S = \sum_{\lambda=\mu_0}^{\mu_1} Q_\lambda$$

we arrive at

$$S = \sum_{\lambda=\mu_0}^{\mu_1} (\lambda/N)^m [1 - (\lambda/N)]^{n-m} (1/N) \div \sum_{\lambda=0}^N (\lambda/N)^m [1 - (\lambda/N)]^{n-m} (1/N).$$

$$\textbf{(14)}$$

The most remarkable is the case in which p_λ can assume all possible values. We shall indeed go over to this instance by assuming that $N = \infty$. The equation (14) will accordingly become

$$S = \int_{p_0}^{p_1} x^m (1-x)^{n-m}\, dx \div \int_0^1 x^m (1-x)^{n-m}\, dx. \qquad \textbf{(15)}$$

For example, let us assume that an event occurred once in three trials, and we shall search for the probability that p is contained within the boundaries 0 and $1/2$. In this case we obtain

$$S = \int\limits_0^{1/2} x\,(1\,-\,x)^2\,dx \div \int\limits_0^1 x\,(1\,-\,x)^2\,dx \;=\; 11/16.$$

And so, the probability sought is 11/16. It was possible to foresee that it will be higher than 1/2.

Formula (15) can be modified. To this aim, note that

$$\int\limits_0^1 x^{\lambda-1}\,(1\,-\,x)^{\mu-1}\,dx \;=\; \Gamma(\lambda)\,\Gamma(\mu) \div \Gamma(\lambda\,+\,\mu),$$

but for integer values of λ and μ we have

$$\Gamma(\lambda) = (\lambda\,-\,1\,)!, \;\; \Gamma(\mu) = (\mu\,-\,1)!$$

so that

$$\int\limits_0^1 x^m\,(1\,-\,x)^{n-m}\,dx \;=\; \frac{\Gamma(m+1)\,\Gamma(n-m+1)}{\Gamma(n+2)} \;=\; \frac{m!\,(n-m)!}{(n+1)!}.$$

Equality (15) can therefore be presented in the following form

$$S \;=\; \frac{(n+1)!}{m!\,(n-m)!}\int\limits_{p_0}^{p_1} x^m\,(1\,-\,x)^{n-m}\,dx. \tag{16}$$

3.3.10. We shall now suppose that the difference $(p_1\,-\,p_0)$ is a very small given magnitude and we shall search, under this condition, where should we choose this interval so that the probability S be maximal. Denoting

$$(p_1\,-\,p_0)\;=\;2\omega, \;\;(p_1\,+\,p_0)/2\;=\;\rho$$

with ω assumed to be very small, we obtain

$$p_1 \;=\; \rho\,+\,\omega, \;\; p_0\;=\;\rho\,-\,\omega,$$

$$S \;=\; \int\limits_{\rho-\omega}^{\rho+\omega} x^m\,(1\,-\,x)^{n-m}\,dx \div \int\limits_0^1 x^m\,(1\,-\,x)^{n-m}\,dx.$$

Note now that if $[AB]$ is some mean value of A and B, then, in general,

$$\int\limits_A^B f(x)\,dx \;=\; (B\,-\,A)\,f\,([AB]).$$

Again, since $[\rho\,-\,\omega;\,\rho\,+\,\omega]\;=\;\rho\,\pm\,\theta\,\omega$ where θ is some proper fraction, we obtain

$$S \;=\; 2\,\omega(\rho\,\pm\,\theta\omega)^m\,(1\,-\,\rho\,\mp\,\theta\omega)^{n-m} \div \int\limits_0^1 x^m\,(1\,-\,x)^{n-m}\,dx$$

so that

$$(1/2\omega)\, S \;=\; (\rho \pm \theta\omega)^{\,m}\, (1 - \rho \mp \theta\omega)^{\,n-m} \div \int_0^1 x^m\, (1 - x)^{n-m}\, dx.$$

Going over to the limit as $\omega = 0$, we have

$$\lim\, [(1/2\omega)\, S]_{\,\omega\,=\,0} \;=\; \rho^{\,m}\, (1 - \rho)^{\,n-m} \div \int_0^1 x^m\, (1 - x)^{n-m}\, dx.$$

Let us find now the maximal value of this expression. We will have the equation

$$m\,\rho^{\,m-1}\, (1 - \rho)^{\,n-m} - (n - m)\,\rho^{\,m}\, (1 - \rho)^{\,n-m-1} \;=\; 0$$

which can be reduced to

$$(m - n\,\rho)\,\rho^{\,m-1}\, (1 - \rho)^{\,n-m-1} \;=\; 0.$$

It has roots 0, 1 and m/n. the first two of them transform the expression $\rho^{\,m}\, (1 - \rho)^{\,n-m}$ into zero; and, since it is positive for any ρ, the root m/n will make it maximal so that the expression

$$[(1/2\omega)\, S]$$

will also become maximal.

And so, the most probable boundaries of the probability sought are

$$p_0 \;=\; m/n - \omega; \quad p_1 \;=\; m/n + \omega.$$

3.3.11. We shall now search for the probability that p is contained within the boundaries p_0 and p_1 very little differing from m/n. Let

$$p_0 \;=\; m/n + z_0 \quad \text{and} \quad p_1 \;=\; m/n + z_1$$

where z_0 and z_1 are so small positive or negative quantities that their squares may be neglected. And we will assume that m, n and $(n - m)$ are very large numbers. Applying the Stirling theorem we obtain

$$n! \;=\; \sqrt{2\pi}\; n^{\,n+1/2}\, e^{\,-n}, \quad m! \;=\; \sqrt{2\pi}\; m^{\,m+1/2}\, e^{\,-m},$$
$$(n - m)! \;=\; \sqrt{2\pi}\; (n - m)^{\,n-m+1/2}\, e^{\,-n+m}.$$

Consequently, formula (16) provides

$$S \;=\; (1/\sqrt{2\pi}\,)\; \frac{n^{\,n+1/2}\,(n + 1)}{m^{\,m+1/2}\,(n - m)^{\,n-m+1/2}}\; \int_{p_0}^{p_1} x^m\, (1 - x)^{n-m}\, dx.$$

Noting now that $(n + 1) = n\,[1 + (1/n)]$ and neglecting the very small quantity $1/n$, we obtain

$$S \;=\; (1/\sqrt{2\pi}\,)\; \frac{n^{\,n+3/2}}{m^{\,m+1/2}\,(n - m)^{\,n-m+1/2}}\; \int_{p_0}^{p_1} x^m\, (1 - x)^{n-m}\, dx.$$

Set $x = [(m/n) + z]$, then

$$S \;=\; \frac{n^{\,3/2}}{\sqrt{2\pi\, m\,(n - m)}}\; \int_{z_0}^{z_1} [1 + (nz)/m]^{\,m}\, \{1 - [(nz)/(n - m)]\}^{\,n-m}\, dz.$$

But we have

$$\ln\left([1 + (nz)/m]^{m}\{1 - [(nz)/(n - m)]\}^{n-m}\right) = m\left[\frac{n\,z}{m} - \frac{n^2 z^2}{2m^2} + \ldots\right] -$$

$$(n - m)\left[\frac{nz}{n - m} + (1/2)\frac{n^2 z^2}{(n - m)^2} + \ldots\right] =$$

$$- (1/2)\frac{n^2 z^2}{m} - (1/2)\frac{n^2 z^2}{n - m} + \ldots = -\frac{n^3 z^2}{2m(n - m)} + \ldots$$

Therefore, neglecting terms with higher powers of z, we obtain

$$S = \frac{n^{3/2}}{\sqrt{2\pi\, m(n - m)}} \int_{z_0}^{z_1} \exp\left[-\frac{n^3 z^2}{2m(n - m)}\right] dz.$$

Denoting

$$\frac{n^3 z^2}{2m(n - m)} = t^2$$

we transform this expression to

$$S = (1/\sqrt{\pi}) \int_{t_0}^{t_1} \exp(-t^2)\, dt \tag{17}$$

where

$$t_0 = \frac{n^{3/2}}{\sqrt{2m(n - m)}} z_0, \; t_1 = \frac{n^{3/2}}{\sqrt{2m(n - m)}} z_1.$$

And so, the probability that p is contained within the boundaries

$$(m/n) + t_0\sqrt{2m(n - m)/n^3}\,, \; (m/n) + t_1\sqrt{2m(n - m)/n^3}$$

will be determined by the formula (17). If $t_0 = -u$, $t_1 = u$, this formula becomes

$$S = (2/\sqrt{\pi}) \int_0^u \exp(-t^2)\, dt. \tag{18}$$

When u is somewhat considerable (for example, larger than 2), the value of the integral will be very close to $(\sqrt{\pi}/2)$ and the probability will therefore be very close to 1. But, on the other hand, for very large values of n the boundaries within which p is contained will be very close to each other and very little differ from m/n; we may therefore say that at $n = \infty$ the limit of p is m/n.

3.3.12. We are now going over to the solution of a new problem concerning the repetition of events. Suppose that in n trials an event occurred m times. It is required to form a conclusion about the probability that in k {further} trials it will take place l times. The solution of this problem is based on Theorem 4.

Suppose that the probability p that the event occurs in one trial can only take values $0/N$, $1/N$, $2/N$, …, λ/N, …, $(N - 1)/N$ indeed representimg the different hypotheses under which the studied event can occur. We assume now that all these hypotheses have equal probabilities (an assumption for which there are no grounds)

and since their number is N with one of them certainly taking place, the probability of each will be $1/N$. And so, $P_0 =$
$P_1 = \ldots = P_{N-1} = 1/N$ and

$$p_\lambda = C_n^m \, (\lambda/N)^m \, [1 - (\lambda/N)]^{n-m}, \quad q_\lambda = C_k^l \, (\lambda/N)^l \, [1 - (\lambda/N)]^{k-l}.$$

Therefore, denoting the probability sought by $H_{k,l}$ we have

$$H_{k,l} = (1/N) \sum_{\lambda=0}^{N} p_\lambda \, q_\lambda \div (1/N) \sum_{\lambda=0}^{N} p_\lambda =$$

$$C_k^l \sum_{\lambda=0}^{N} (\lambda/N)^{m+l} \, [1 - (\lambda/N)]^{n-m+k-l} \div \sum_{\lambda=0}^{N} (\lambda/N)^m \, [1 - (\lambda/N)]^{n-m} =$$

$$C_k^l \int_0^1 x^{m+l} \, (1 - x)^{n-m+k-l} \, dx \div \int_0^1 x^m \, (1 - x)^{n-m} \, dx \qquad \textbf{(19)}$$

if p can take all possible values from 0 to 1 for which we should assume that $N = \infty$.

As an example, we shall solve such a problem. In n trials an event occurred all the time; to find the probability that it will take place at the $(n + 1)$-th trial as well. In this case, we should assume that $m = n, \ k = l = 1$ so that

$$H_{1,1} = \int_0^1 x^{m+1} dx \div \int_0^1 x^m dx = (m + 1)/(m + 2).$$

Formula (19) can be modified when noting that

$$\int_0^1 x^{\lambda-1} \, (1 - x)^{\mu-1} \, dx = \Gamma(\lambda) \, \Gamma(\mu) / \Gamma(\lambda + \mu)$$

and that, for an integer n, $\Gamma(n) = (n - 1)!$. We therefore obtain

$$\int_0^1 x^{m+l} \, (1 - x)^{n-m+k-l} \, dx = \frac{(m+l)! \, (n-m+k-l)!}{(n+k+1)!},$$

$$\int_0^1 x^m \, (1 - x)^{n-m} \, dx = \frac{m! \, (n-m)!}{(n+1)!}$$

and consequently

$$H_{k,l} = \frac{k! \, (m+l)! \, (n-m+k-l)! \, (n+1)!}{l! \, (k-l)! \, (n+k+1)! \, m! \, (n-m)!}. \qquad \textbf{(20)}$$

3.3.13. We shall now find the value of l which makes the probability $H_{k,l}$ maximal. Calling this value λ, we ought to have

$$H_{k,\lambda} \geq H_{k,\lambda-1} \text{ and } H_{k,\lambda} \geq H_{k,\lambda+1} \text{ or}$$
$$H_{k,\lambda} / H_{k,\lambda-1} \geq 1 \text{ and } H_{k,\lambda} / H_{k,\lambda+1} \geq 1. \qquad (21)$$

If one of these conditions were satisfied by an equality sign, we would have obtained two values for l: λ and $(\lambda + 1)$ or $(\lambda - 1)$ and λ. Applying formula (20) we get

$$\frac{H_{k,\lambda}}{H_{k,\lambda-1}} = \frac{(m-\lambda)(k-\lambda-1)}{\lambda(n-m+k-\lambda+1)}, \quad \frac{H_{k,\lambda}}{H_{k,\lambda+1}} = \frac{(\lambda+1)(n-m+k-\lambda)}{(m+\lambda+1)((k-\lambda)}$$

and the conditions (21) will become

$$(m + \lambda)(k - \lambda + 1) \geq \lambda(n - m + k - \lambda + 1),$$
$$(\lambda + 1)(n - m + k - \lambda) \geq (m + \lambda + 1)(k - \lambda)$$

so that

$$\lambda \leq m(k + 1)/n, \quad \lambda \geq [m(k + 1)/n] - 1.$$

Consequently, if

$$[m(k + 1)/n] \qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(xix)}$$

is a fraction, we arrive at a single solution [1]

$$\lambda = E[m(k + 1)/n]$$

If, however, (xix) is an integer, we obtain two solutions

$$\lambda_1 = [m(k + 1)/n], \quad \lambda_2 = [m(k + 1)/n] - 1.$$

It is seen now that in general

$$\lambda = (m/n)k + (m/n) - \theta = (m/n)k \pm \theta_1$$

where θ and θ_1 are proper fractions. Assuming that m, n and k are very large numbers we may therefore say that to within 1 the most probable number of the repetitions of the event is mk/n.

3.3.14. We shall now assume that m, n, k and l, $(n - m)$ and $(k - l)$ are very large numbers. Under this assumption we shall find the probability $H_{k,\lambda}$ of the most probable number of repetitions of an event in k trials given the number m of its occurrences in n trials. To this aim we shall apply the formula (20) which we transform by the Stirling formula

$$x! = \sqrt{2\pi x}\, x^x\, e^{-x}.$$

We have {Chebyshev writes out this formula for $k!$, $(m + l)!$, $(n - m + k - l)!$, $(n + 1)!$, $l!$, $(k - l)!$, $(n + k + 1)!$, $m!$ and $(n - m)!$}. Therefore

$$H_{k,l} = \sqrt{\frac{n(m+l)(n-m+k-l)n}{2\pi l(k-l)(n+k)m(n-m)}}\ \frac{n+1}{n+k+1}\ \cdot$$
$$\frac{k^k(m+l)^{m+l}(n-m+k-l)^{n-m+k-l}n^n}{l^l(k-l)^{k-l}(n+k)^{n+k}m^m(n-m)^{n-m}}\ \cdot$$

Noting that $k = l + (k - l)$ and $n = m + (n - m)$, and that n and k are very large, so that instead of $(n + 1)/(n + k + 1)$ we may take $n/(n + k)$, we have

$$H_{k,l} = \sqrt{\frac{n^3(m+l)k(n-m+k-l)}{2\pi l(k-l)(n+k)^3 m(n-m)}}\left(\frac{m+l}{n+k}\right)^{m+l}\cdot$$
$$\left(\frac{n-m+k-l}{n+k}\right)^{n-m+k-l}\left(\frac{k}{l}\right)^l\left(\frac{k}{k-l}\right)^{k-l}\left(\frac{n}{m}\right)^m\left(\frac{n}{n-m}\right)^{n-m}. \qquad\qquad \text{(xx)}$$

In this formula, we may attribute to l not only integer, but fractional values as well. Therefore, substituting the most probable number mk/n determined above instead of l, we will indeed obtain $H_{k,\lambda}$, the probability sought. But we have

$$\left(\frac{m+\lambda}{n+k}\right)^{m+\lambda} = \left(\frac{[m+(mk/n)]}{n+k}\right)^{m+mk/n} = \left(\frac{m}{n}\right)^{m(n+k)/n},$$

$$\left(\frac{n-m+k-\lambda}{n+k}\right)^{n-m+k-\lambda} = \left(\frac{n-m+k-mk/n}{n+k}\right)^{n-m+k-mk/n} = \left(\frac{n-m}{n}\right)^{(n-m)(n+k)/n},$$

$$\left(\frac{k}{\lambda}\right)^{\lambda} = \left(\frac{k}{mk/n}\right)^{mk/n} = \left(\frac{n}{m}\right)^{mk/n}, \quad \left(\frac{k}{k-\lambda}\right)^{k-\lambda} = \left(\frac{k}{k-mk/n}\right)^{k-mk/n} =$$

$$\left(\frac{n}{n-m}\right)^{(n-m)k/n}.$$

Therefore,

$$H_{k,\lambda} = \sqrt{\frac{n^3(m+mk/n)\,k\,(n-m+k-mk/n)}{2\pi\,(m/n)\,k\,(k-mk/n)(n+k)^3 m\,(n-m)}} =$$

$$\sqrt{\frac{n^3}{2\pi\,k\,(n+k)\,m\,(n-m)}}.$$

The numerator of this expression is of degree $3/2$ with respect to the letters n, k, m and l, and the denominator, of degree 2. It follows that for very large values of these magnitudes the considered probability is very low.

3.3.15. We shall now search for the probability that l very little differs from $\lambda = mk/n$. Let $l = mk/n + z$ where z is assumed very small as compared with numbers m, n, k and λ so that it can be neglected with regard to them. Denote also $m/n = \rho$, $k/n = \nu$, then

$$m = n\rho, \ l = \rho n\nu + z, \ n - m = n(1-\rho), \ k - l = n\nu(1-\rho) - z,$$
$$m + l = n\rho(1+\nu) + z, \ n + k = n(1+\nu), \ n - m + k - l =$$
$$n(1+\nu)(1-\rho) - z.$$

Therefore {cf. (xx)}

$$\sqrt{\frac{n^3(m+l)\,k\,(n-m+k-l)}{2\pi\,l\,(k-l)(n+k)^3 m\,(n-m)}} =$$

$$\sqrt{\frac{\nu\,[n\rho(1+\nu)+z][n(1+\nu)(1-\rho)-z]}{2\pi\,(n\rho\nu+z)[n\nu(1-\rho)-z]\,\rho(1-\rho)(1+\nu)^2 n}}.$$

Noting now that z is very small as compared with n, we may neglect it in the multipliers under the radical sign of this expression. Accordingly, after reductions the expression will become

$$\frac{1}{\sqrt{2\pi\,\nu\,\rho(1-\rho)\,n(1+\nu)}}.$$

Denoting the other multiplier in the expression for $H_{k,l}$ by K we obtain

$$H_{k,l} = \frac{1}{\sqrt{2\pi\,\nu\,\rho(1-\rho)\,n(1+\nu)}}\,K,$$

$$K = \left(\frac{n\nu}{\rho\, n\nu + z}\right)^{\rho\, n\nu + z} \left(\frac{n\nu}{n\nu(1-\rho) - z}\right)^{n\nu(1-\rho) - z} \cdot \left(\frac{n\rho\,(1+\nu) + z}{n(1+\nu)}\right)^{n\rho(1+\nu)+z} \cdot$$

$$\left(\frac{n\,(1-\rho)(1+\nu) - z}{n\,(1+\nu)}\right)^{n\,(1-\rho)(1+\nu)-z} \frac{1}{\rho^{n\rho}\,(1-\rho)^{n(1-\rho)}} = \left(\frac{1}{\rho + z/n\nu}\right)^{n\rho\nu+z} \cdot$$

$$\left(\frac{1}{1 - \rho - z/n\nu}\right)^{n\nu(1-\rho)-z} \left(\rho + \frac{z}{n(1+\nu)}\right)^{n\rho(1+\nu)+z} \cdot$$

$$\left(1 - \rho - \frac{z}{n\,(1+\nu)}\right)^{n(1-\rho)\,(1+\nu)-z} \frac{1}{\rho^{n\rho}\,(1-\rho)^{n(1-\rho)}} =$$

$$\frac{\rho^{n\rho(1+\nu)+z}\{1 + z/[n\rho\,(1+\nu)]\}^{n\rho(1+\nu)+z}}{\rho^{n\rho}\,(1-\rho)^{n(1-\rho)}\,\rho^{n\rho\nu+z}\{1 + [z/n\rho\nu]\}^{n\rho\nu+z}} \cdot$$

$$\frac{(1-\rho)^{n(1-\rho)(1+\nu)-z}\{1 - [z/n(1-\rho)(1+\nu)]\}^{n(1-\rho)(1+\nu)-z}}{(1-\rho)^{n(1-\rho)\nu-z}\{1 - [z/n(1-\rho)\nu]\}^{n(1-\rho)\nu-z}} \cdot$$

And so,

$$K = \frac{\{1 + [z/n\rho(1+\nu)]\}^{n\rho(1+\nu)+z}\{1 - [z/n(1-\rho)(1+\nu)]\}^{n(1-\rho)(1+\nu)-z}}{\{1 + [z/n\rho\nu]\}^{n\rho\nu+z}\{1 - [z/n(1-\rho)\nu]\}^{n(1-\rho)\nu-z}} \cdot$$

But then

$$\ln[\text{denominator}] = (n\rho\nu + z)\{[z/n\rho\nu] - [z^2/2n^2\rho^2\nu^2] + \ldots\} + [n(1-\rho)\nu - z]\{[z/n(1-\rho)\nu] - [z^2/2n^2(1-\rho)\nu^2] - \ldots\} =$$
$$z + z^2[(1/n\rho\nu) - (1/2n\rho\nu) + \ldots] - z + z^2[1/n(1-\rho)\nu] - 1/[2n(1-\rho)\nu]\} + \ldots = z^2/[2n(1-\rho)\nu\rho].$$

In the same way

$$\ln[\text{numerator}] = z^2/[2n(1+\nu)\rho(1-\rho)].$$

Therefore, neglecting the other terms which are very small when the order of z is not higher than $\sqrt{n}$, we arrive at

$$K = \exp\left(-\frac{z^2}{2n\nu(1+\nu)\rho\,(1-\rho)}\right)$$

and, consequently,

$$H_{k,l} = \exp\left(-\frac{z^2}{2n\nu(1+\nu)\rho\,(1-\rho)}\right) \div \sqrt{2\pi\, n\nu\,(1+\nu)\,\rho\,(1-\rho)}.$$

The probability that l is contained within the known boundaries L and M (denoted by Π) will be expressed through the sum of the values of $H_{k,l}$ extended between these boundaries. Acting in the same way as we did in §3.3.6, we will find that this sum can be replaced by an integral (given the degree of precision with which we make all our calculations). Thus, we obtain

$$\Pi = \int_L^M H_{k,l}\, dz.$$

Substituting now

$$[z/\sqrt{2n\nu(1+\nu)\rho(1-\rho)}\,] \; = \; \{[l \; - \; (mk/n)]\,/\sqrt{2n\nu(1+\nu)\rho(1-\rho)}\,\} \; = \; u$$

we shall have

$$\Pi \; = \; (1/\sqrt{\pi})\int_{u_0}^{u_1} \exp(-u^2)\,du \tag{22}$$

where

$$u_0 \; = \; [L \; - \; (mk/n)] \div \sqrt{2n\nu(1+\nu)\rho(1-\rho)}\,,$$
$$u_1 \; = \; [M \; - \; (mk/n)] \div \sqrt{2n\nu(1+\nu)\rho(1-\rho)}\,.$$

Consequently, substituting the values of ν and ρ, we get

$$L \; = \; (mk/n) \; + \; u_0\sqrt{\frac{2m(n-m)\,k\,(n+k)}{n^3}}\,,$$
$$M \; = \; (mk/n) \; + \; u_1\sqrt{\frac{2m(n-m)\,k\,(n+k)}{n^3}}\,.$$

The most remarkable case is the one in which $u_0 = -t$ and $u_1 = t$. Then, denoting

$$\sqrt{\frac{2m(n-m)\,k\,(n+k)}{n^3}} \; = \; c,$$

we have

$$L \; = \; (mk/n) \; - \; tc, \; M \; = \; (mk/n) \; + \; tc$$

and

$$\Pi \; = \; (2/\sqrt{\pi})\int_0^t \exp(-t^2)\,dt. \tag{23}$$

For a somewhat considerable t the right side of (23) differs little from 1 so that it is very probable that l is contained within the boundaries

$$(mk/n) \; + \; t\sqrt{\frac{2m(n-m)\,k\,(n+k)}{n^3}}\,, \; (mk/n) \; - \; t\sqrt{\frac{2m(n-m)\,k\,(n+k)}{n^3}}\,.$$

This conclusion can be expressed in the form of the following theorem:

Theorem. *Formula* (23) *determines the probability of the existence of the inequalities*

$$(m/n) \; - \; t\sqrt{(2m/n)[1-(m/n)][(1/n)+(1/k)]} \; < \; l/k \; <$$
$$(m/n) \; + \; t\sqrt{(2m/n)[1-(m/n)][(1/n)+(1/k)]}\,. \tag{24}$$

Issuing from here, we can arrive at the results already reached before and contained in (24) as particular cases. Assuming that $n = \infty$ and noting that $\lim (m/n) = p$, we find that

$$p \; - \; t\sqrt{\frac{2p(1-p)}{k}} \; < \; l/k \; < \; p \; + \; t\sqrt{\frac{2p(1-p)}{k}}\,.$$

We thus obtain the result achieved in §3.3.6. Supposing that $k = \infty$ we will find that

$$(m/n) - t\sqrt{(2m/n^2)[1-(m/n)]} < p < (m/n) + t\sqrt{(2m/n^2)[1-(m/n)]}$$

which is a result gotten in §3.3.11.

3.3.16. We now go over to considering the case in which the probability p that an event occurs in one trial is different in different trials. Suppose that p_1, p_2, …, p_n are the probabilities that the event takes place in the first, the second, …, the n-th trial. In this case, as we saw, the probability $P_{n,m}$ that the event occurs m times in n trials will be the coefficient of t^m in the expansion

$$(p_1 t + 1 - p_1)(p_2 t + 1 - p_2) \ldots (p_n t + 1 - p_n) = \sum_{m=0}^{n+1} P_{n,m} t^m.$$

Issuing from this equation[1] we may express the probability $P_{n,m}$ by a definite integral. Indeed, as we know from Chapt. 1 {§1.4.1}, if a function $f(x)$ can be expanded into a series

$$f(x) = A_o + A_1 x + A_2 x^2 + \ldots + A_m x^m + \ldots$$

the coefficient A_m of this series will be determined by the formula

$$A_m = (1/2\pi) \int_{-\pi}^{\pi} f(e^{\varphi i}) e^{m\varphi i} \, d\varphi.$$

Therefore, substituting $m = n + 1$, we have

$$f(t) = \sum_{m=0}^{n+1} P_{n,m} t^m$$

and $P_{n,m}$ will be equal to

$$(1/2\pi) \int_{-\pi}^{\pi} (p_1 e^{\varphi i} + 1 - p_1)(p_2 e^{\varphi i} + 1 - p_2) \ldots (p_n e^{\varphi i} + 1 - p_n) e^{-m\varphi i} \, d\varphi. \quad \textbf{(25)}$$

We shall show now that the integrand is a noticeable quantity at all only at values of φ close to 0. To this end note that

$$p_1 e^{\varphi i} + 1 - p_1 = p_1 \cos\varphi + 1 - p_1 + i p_1 \sin\varphi,$$

$$[\mathrm{mod}\,(p_1 e^{\varphi i} + 1 - p_1)]^2 = [p_1 \cos\varphi + 1 - p_1]^2 + p_1 \sin^2\varphi =$$
$$p_1^2 + (1 - p_1)^2 + 2 p_1(1 - p_1) - 2 p_1(1 - p_1) + 2 p_1(1 - p_1)\cos\varphi =$$

$$1 - 2 p_1(1 - p_1)(1 - \cos\varphi)$$

so that

$$[\mathrm{mod}\,(p_1 e^{\varphi i} + 1 - p_1)]^2 \le 1$$

where the sign of equality corresponds to the value $\varphi = 0$. The number n is supposed to be very large; consequently, only those elements of the integral (25) will be significant whose modulus is {whose moduli are} very close to unity, i.e. those which correspond to the values of φ close to zero. Therefore, when approximately calculating the probability $P_{n,m}$, we shall neglect the powers higher than the second one in the power series of φ. We have however

$\ln (p_1 e^{\varphi i} + 1 - p_1) = \ln [p_1(1 + \varphi i - \varphi^2/2 - \ldots) + 1 - p_1] =$
$\ln [1 + p_1\varphi i - p_i\varphi^2/ 2 + \ldots] = p_1\varphi i - p_1\varphi^2/ 2 - [p_1\varphi i - (p_1\varphi^2/2)]^2/2 \ldots$
$= p_1\varphi i - (p_1/2) (1 - p_1)\varphi^2 + \ldots$

so that denoting

$$p_1 + p_2 + \ldots + p_n = nq, \quad p_1(1 - p_1) + \ldots + p_n(1 - p_n) = nQ,$$

we can replace formula (25) by the following approximate expression

$$P_{n,m} = (1/2\pi) \int_{-\pi}^{\pi} \exp [nq\varphi i - (nq/2)\varphi^2 - m\varphi i]d\varphi =$$

$$(1/2\pi) \int_{-\pi}^{\pi} \exp [- (nq/2) \varphi^2] \{\cos [nq - m) \varphi] + i\sin [nq - m) \varphi]\} \, d\varphi.$$

And since

$$\int_{-\pi}^{\pi} \exp [- (nq/2) \varphi^2] \sin [nq - m) \varphi] \, d\varphi = 0$$

it follows that

$$P_{n,m} = (1/2\pi) \int_{-\pi}^{\pi} \exp [- (nq/2) \varphi^2] \cos [nq - m) \varphi] \, d\varphi =$$

$$(1/\pi) \int_{0}^{\pi} \exp [- (nq/2) \varphi^2] \cos [nq - m) \varphi] \, d\varphi.$$

We assumed that n was very large; therefore, neglecting the magnitude of the second integral taken from π to ∞, we get

$$P_{n,m} = (1/\pi) \int_{0}^{\infty} \exp [- (nq/2) \varphi^2] \cos [nq - m) \varphi] \, d\varphi.$$

But we have

$$\int_{0}^{\infty} \exp (- ax^2) \cos bx \, dx = (1/2) \sqrt{\pi / a} \exp [- b^2/4a]$$

and therefore

$$P_{n,m} = (1/2\pi) \sqrt{2\pi / nQ} \exp [- (nq - m)^2/2nQ] =$$

$$\frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{nQ}} \exp [- (nq - m)^2/2nQ]. \qquad\qquad (25')$$

It is seen now that the maximal probability corresponds to the case in which

$$m/n = q = (p_1 + p_2 + \ldots + p_n)/n.$$

Denoting the probability that m is contained within the boundaries L and M by Π and supposing that the numbers n, m and $(n - m)$ are very large, we obtain the following approximate formula:

$$\Pi = \int_{L}^{M} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{nQ}} \exp[-(m-nq)^2/2nQ] \, dm.$$

Introducing

$$(m-nq)/\sqrt{2nQ} = t, \quad (L-nq)/\sqrt{2nQ} = t_o, \quad (M-nq)/\sqrt{2nQ} = t_1,$$

we transform this formula thus:

$$\Pi = (1/\sqrt{\pi}) \int_{t_0}^{t_1} \exp(-t^2) \, dt. \tag{26}$$

If $p_1 = p_2 = \ldots = p_n = q$ formula (9) is derived from here as a particular case. When $t_o = -u$ and $t_1 = u$ formula (26) becomes

$$\Pi = (2/\sqrt{\pi}) \int_{0}^{u} \exp(-t^2) \, dt \tag{27}$$

which is the probability of the existence of the inequalities

$$q + u\sqrt{2Q/n} > m/n > q - u\sqrt{2Q/n}.$$

Since their probability for somewhat considerable values of u is very close to 1, and, on the other hand, since for a very large n the boundaries within which m/n is contained differ very little from q and become equal to it at $n = \infty$, we may say that

$$\lim (m/n)_{n=\infty} = q = (p_1 + p_2 + \ldots + p_n)/n.$$

This is the essence of the law of large numbers first proved and formulated by Poisson {cf. §3.2.3}.

Note 1. {Concerning the boundaries of the sum above and the one below see Note 2 in §3.3.2.}

3.3.17. Let us go over now to the issue about the repetition of several events. Let A_1, A_2, …, A_l be different events one of which certainly takes place in each trial. Therefore, denoting their probabilities by

$$p_1, p_2, \ldots, p_l \tag{xxi}$$

respectively, we have $p_1 + p_2 + \ldots + p_l = 1$. We assume that two different events cannot occur in the same trial and that the probabilities of each given event are the same in each trial so that the probability of event A_i is p_i both in the first, and in any k-th trial.

Suppose that in n trials the event A_i occurred m_i times, then

$$m_1 + m_2 + \ldots + m_l = n.$$

We will search for the probability P that in n trials the event A_1 occurs m_1 times; the event A_2, m_2 times; …; and the event A_l, m_l times. The probability sought may be considered as the probability of an event having several incompatible forms; it is therefore equal to the sum of the probabilities of each of these forms. One of them is that in the first m_1 trials the event A_1 was repeated m_1 times but the events A_2, A_3, …, A_l did not then take place; that in the following m_2 trials the event A_2 occurred m_2 times, but the events A_1, A_3, …, A_l did not then occur; and, finally, that in the last m_l trials the event A_l was repeated m_l times but the events A_1, A_2, …, A_{l-1} did not take place. Since we supposed that the probabilities (xxi) were constant, the probability of the considered form will be

$$p_1{}^{m_1} p_2{}^{m_2} ... p_l{}^{m_l}.$$

However, owing to the same condition, this will also be the probability of each of the other forms. And there will be as many forms as there are possible combinations of n elements containing l groups of identical elements, m_1 of them in one group, m_2 of them in another one, etc. Therefore, the probability sought will be expressed in the following way

$$P = \frac{n!}{m_1! m_2! ... m_l!} \, p_1{}^{m_1} p_2{}^{m_2} ... p_l{}^{m_l}. \tag{28}$$

Neither is it difficult to convince ourselves that the probability sought can be determined as the coefficient of

$$t_1{}^{m_1} t_2{}^{m_2} ... t_l{}^{m_l}$$

in the expansion

$$(p_1 t_1 + p_2 t_2 + ... + p_l t)^n$$

so that we may assume that

$$(p_1 t_1 + p_2 t_2 + ... + p_l t)^n = \sum P \, t_1{}^{m_1} t_2{}^{m_2} ... t_l{}^{m_l}. \tag{29}$$

From this equality we can also obtain the expression (28) for the probability sought.

If the events A_1, A_3, ..., A_l are determined by some numerical magnitudes, so that the event A_i is determined, for example, by the magnitude of the function $\theta\,(i)$, then formula (29) can serve for solving the problem about the probability that in n trials the sum of these magnitudes will take a value s given beforehand. Thus, if the event A_i is the drawing of a card with number i, the problem might consist in determining the probability that the sum of the numbers on the cards extracted in n trials equals s.

For solving this problem we assume that $t_i = t^{\theta\,(i)}$ so that formula (29) is transformed into

$$\sum P \, t^{m_1\theta(1)+m_2\theta(2)+...+m_l\theta(l)} = [p_1 t^{\theta\,(1)} + p_2 t^{\theta\,(2)} + ... + p_l t^{\theta\,(l)}]^n.$$

Since the probability sought is equal to the sum of all the probabilities for which m_i satisfy the conditions

$$m_1 + m_2 + ... + m_l = n, \; m_1 \theta(1) + m_2 \theta(2) + ... + m_l \theta(l) = s,$$

the preceding equality shows that the probability sought is the coefficient of t^s in the expansion of the expression

$$[p_1 t^{\theta\,(1)} + p_2 t^{\theta\,(2)} + ... + p_l t^{\theta\,(l)}]^n.$$

Here, the most remarkable is the particular case in which $\theta\,(x) = x$ so that, consequently, the probability sought that

$$m_1 1 + m_2 2 + ... + m_l \cdot l = s$$

is determined as the coefficient of t^s in the expansion of the expression

$$[p_1 t + p_2 t^2 + ... + p_l t^l]^n.$$

If we denote the probability considered by P_s, then, in the general case, we will have

$$\sum P_s t^s = [p_1 t^{\theta\,(1)} + p_2 t^{\theta\,(2)} + ... + p_l t^{\theta\,(l)}]^n \tag{30}$$

whereas in the particular case indicated above this formula becomes

$$\sum P_s\, t^s \; = \; [p_1 t \; + \; p_2 t^2 \; + \; \ldots \; + \; p_l\, t^l]^n. \tag{xxii}$$

Suppose now that the events A_i are equally probable, i.e., that $p_1 \; = \; p_2 \; = \; \ldots \; = \; p_l$. Then $p_i \; = \; 1/l$ for any i. In this case the formula derived (xxii) will be

$$\sum P_s\, t^s \; = \; (t^n/l^n)\, [1 \; + \; t \; + \; t^2 \; + \; \ldots \; + \; t^{l-1}]^n \; =$$
$$(t^n/l^n)\, [(t^l \; - \; 1)/(t \; - \; 1)]^n. \tag{31}$$

Issuing from formula (31) we can indeed calculate the probability P_s that the sum of the magnitudes determining the events occurring in n trials (in this case, the sum of the numbers corresponding to the events taking place in these trials) is equal to a given magnitude s.

3.3.18. We shall now show how to obtain, when issuing from formula (31), the expression for the probability P_s as some series. The problem is reduced to the determination of the coefficient of t^{s-n} in the expansion of the expression

$$(t^l \; - \; 1)^n/(t \; - \; 1)]^{-n} \tag{xxiii}$$

in powers of t. But we have

$$(t^l \; - \; 1)^n \; = \; t^{ln} \; - \; (n/1!)\, t^{l\,(n-1)} \; + \; [n(n \; - \; 1)/2!]\, t^{l\,(n-2)} \; - \; \ldots, \tag{xxiv}$$
$$(t \; - \; 1)]^{-n} \; = \; t^{-n} \; + \; (n/1!)\, t^{-n-1} \; + \; [n(n \; + \; 1)/2!]\, t^{-n-2} \; + \; \ldots \; +$$
$$[n\,(n \; + \; 1) \ldots (n \; + \; \lambda \; - \; 1)/\lambda!]\, t^{-n-\lambda} \; + \; \ldots \tag{xxv}$$

Denoting $l(n \; - \; i) \; - \; (n \; + \; \lambda) \; = \; s \; - \; n$ we obtain $\lambda \; = \; l(n \; - \; i) \; - \; s$. Inserting this value of λ in the expression for the general term of the series (xxv) and making i consecutively equal to 0, 1, 2, … we will obtain in this expansion all the possible terms, which, being multiplied by the first, the second, the third, … term of the series (xxiv), will provide terms including t^{s-n}. We thus multiply the term number i in (xxv) by the $(i \; + \; 1)$-th term in (xxiv).

It is seen now that the sum of the terms of the product of (xxiv) and (xxv) containing t^{s-n} will have the form

$$t^{s-n}\, \sum\, (-\,1)^i\, \frac{n(n-1)\ldots(n-i+1)}{i!} \cdot \frac{n(n+1)\ldots[n+l(n-i)-s-1]}{[l(n-i)-s]!}$$

where we ought to take all the values of i from 0 to n inclusively although under the conditions that the first factor be replaced by 1 at $i \; = \; 0$ and the second factor, by 1 at $i \; = \; (nl \; - \; s)/l$ and by 0 at larger values of i. It ought to be remarked that nl is the maximal value that s can take; this follows from the fact that the sum of the numbers {on the cards} cannot exceed the maximal number {on a card} taken as many times as there were trials. On the grounds of the above, we conclude that the probability sought can be represented as

$$P_s \; = \; (1/l^n) \sum_{i=0}^{K}\, (-\,1)^i\, \frac{n(n-1)\ldots(n-i+1)}{i!}\, \frac{n(n+1)\ldots[n+l(n-i)-s-1]}{[l(n-i)-s]!} \tag{32}$$

with

$$K \; = \; E\,[(nl \; - \; s)/l] \; + \; 1.$$

Issuing from here, we find the following expression for the probability P_s in the form of a series

$$P_s \; = \; (1/l^n)\, \{\frac{n(n+1)\ldots(n+nl-s-1)}{(n-s)!} \; - \; C_n^1\, \frac{n(n+1)\ldots[n+l(n-1)-s-1)}{[l(n-1)-s]!}$$

$$+ \ C_n{}^2 \ \frac{n(n+1)...[n+l(n-2)-s-1]}{[l(n-2)-s]!} \ -$$

$$C_n{}^3 \ \frac{n(n+1)...[n+l(n-3)-s-1]}{[l(n-3)-s]!} \ + ...\}$$

which breaks off as soon as we arrive at a term equal to 1 or 0.

To illustrate, we shall apply this conclusion to dice playing. This {particular} game consists in throwing six dice having the form of cubes {of a cube}; numbers 1, 2, …, 6 are cut in their faces and the gain or loss according to the condition of the game depends on the sum of the numbers turned up. Since it makes no difference whether to throw one die six times or to toss six identical dice only once, we may consider the throwing of each die as a trial so that in the case under consideration $n = 6$.

The number appearing on the upper face of the fallen die may be assumed to be the magnitude measuring the event taking place in some {in the corresponding} trial, so that it is seen that here $l = 6$. Let us determine now the probability P_s that the sum of the appeared numbers is s. In the case under consideration formula (32) becomes

$$P_s = (1/6^6) \sum_{i=0}^{K} (-1)^i \ \frac{6 \cdot 5...(7-i)}{i!} \cdot \frac{6 \cdot 7 \cdot 8...(41-6i-s)}{(36-6i-s)!} \qquad (33)$$

with $K = E[(36 - s)/6] + 1$. The magnitude s cannot exceed 36. Formula (33) provides $P_{36} = 1/6^6 = 1/46\,656$. For $s = 35$ and 34, $K = 1$ in both cases and $P = (1/6^6) \cdot 6 = 1/7776$ and $(1/6^6) \cdot 6 \cdot 7/2 = (7/2) \cdot (1/7776) = 7/15\,552$ respectively.

In the same way we find that

$$P_{33} = \frac{7 \cdot 8}{15\,552 \cdot 2} = 7/5832; \ P_{32} = \frac{7 \cdot 9}{5832 \cdot 4} = 7/2592;$$

$$P_{31} = \frac{7 \cdot 10}{2592 \cdot 5} = 7/1296; \ ...$$

Supposing now that $s = 30$ we have $K = E[(36 - 30)/6] + 1 = 2$ and

$$P_{30} = \frac{7 \cdot 11}{1296 \cdot 6} - \frac{6}{6^6} = (1/7776)(77 - 1) = 19/1944, \ ...$$

We see thus that the probability P_s increases with a decreasing s. It is not difficult to confirm that

$$P_{36} = P_6; \ P_{35} = P_7; \ P_{34} = P_9 \ \text{etc.} \qquad (34)$$

Therefore, the probability P_s increases with an increasing s beginning with $s = 6$ and this increase continues until s does not take the mean value between 6 and 36, i.e., until $s = 21$. And so, P_{21} is the maximum of P_s. At $s = 21$ $K = E[(36 - 21)/6] + 1 = 2 + 1 = 3$. Therefore

$$P_{21} = (1/6^6) \sum_{i=0}^{3} (-1)^i \ \frac{6 \cdot 5 (7-i)}{i!} \frac{6 \cdot 7 \cdot 8...(20-6i)}{(15-6i)!} =$$

$$(1/6^6) \{ \frac{6 \cdot 7 \cdot 8...18 \cdot 19 \cdot 20}{15!} - 6 \frac{6 \cdot 7 \cdot 8...14}{9!} + \frac{6 \cdot 5}{1 \cdot 2} \cdot \frac{6 \cdot 7 \cdot 8}{3!} \} =$$

$$\frac{3 \cdot 16 \cdot 17 \cdot 19 - 6 \cdot 11 \cdot 13 \cdot 14 + 3 \cdot 5 \cdot 7 \cdot 8}{6^6} = 4332/6^6 = 361/2592.$$

We have written the equalities (34) as though they were obvious but we shall prove them analytically. To this aim, we shall represent the formula expressing P_s in a somewhat different form. We have the formula {cf. (31)}

$$\sum P_s \, t^s = (t^n / l^n) \left(\frac{t^l - 1}{t - 1} \right)^n \qquad\qquad \text{(xxvi)}$$

and, consequently,

$$\sum P_s \, t^s = (1/l^{\,n}) \cdot t^{\,n} \cdot (1 - t^{\,l})^{\,n} \cdot (1 - t)^{-n} =$$
$$(1/l^{\,n}) \, t^{\,n} [1 - C_n^1 t^{\,l} + C_n^2 t^{\,2l} - \dots + (-1)^i C_n^i t^{\,il} + \dots] \cdot$$
$$\{ 1 + nt + [n(n + 1)/2!] t^2 + \dots + [n(n + 1) \dots (n + \lambda - 1)/\lambda!] \, t^\lambda + \dots \}.$$

Denoting now $\lambda + il = s - n$ so that $\lambda = s - n - il$, we obtain the following expression for the general term of the coefficient of t^{s-n}:

$$(-1)^i \, C_n^i \, \frac{n(n+1)\dots(s-il-1)}{(s-il-n)!}.$$

The coefficient itself will therefore be the sum of such expressions where i takes all the values from 0 to a value (not inclusively) at which $(s - n - il)$ becomes negative, i.e., to

$$i = \mathrm{E}\,[(s - n)/l] + 1 = H.$$

However, C_n^i at $i = 0$, and the fraction at $(s - n - il) = 0$ should be replaced by unities. Under this condition we obtain

$$P_s = (1/l^n) \sum_{i=0}^{H} (-1)^i \, C_n^i \, \frac{n(n+1)\dots(s-il-1)}{(s-il-n)!}. \qquad\qquad \textbf{(35)}$$

When applying this formula to the game of dice we shall reduce it to the following form

$$P_s = (1/6^6) \sum_{i=0}^{G} (-1)^i \, \frac{6 \cdot 5 \dots (7-i)}{i!} \, \frac{6 \cdot 7 \cdot 8 \dots (s-6i-1)}{(s-6i-6)!}$$

with $G = \mathrm{E}\,[(s - 6)/6] + 1$. Substituting now $s = 36 - k$ into formula (33) we obtain

$$P_{36-k} = (1/6^6) \sum_{i=0}^{\mathrm{E}[k/6]+1} (-1)^i \, \frac{6 \cdot 5 \dots (7-i)}{i!} \, \frac{6 \cdot 7 \cdot 8 \dots (5+k-6i)}{(k-6i)!}.$$

When however assuming that $s = 6 + k$ in formula (34), we get an identical expression and it is seen now that indeed $P_{36-k} = P_{6+k}$. This, however, is a particular case of a general theorem because, when issuing from formulas (32) and (35), it is not difficult to prove that, in general, $P_{ln-k} = P_{n+k}$. And so, we find that

$$P_6 = P_{36} = = (1/6^6) = 1/46\,656; \; P_7 = P_{35} = (6/6^6) = 1/7776;$$
$$P_8 = P_{34} = (21/6^6) = 7/15\,552; \; P_9 = P_{33} = (56/6^6) = 7/5832; \dots;$$
$$P_{21} = 361/2592.$$

Let us study now several versions of a lottery where gains depend on the appearance of some sums of numbers when tossing six dice. We have approximately

$$P_6 = P_{36} = 1/46\,656; \; P_7 = P_{35} = 1/7776; \; P_8 = P_{34} = 1/2222; \dots$$

Suppose now that the lottery is such that a gain of $46\,656$ corresponds to the sums equal to 6 and 36; a gain of 7776 to sums 7 and 35; and a gain of 2222, to the sums 8 and 34, and assume also that no gains are attached to the other sums. This version of the lottery can be described by the following table (Table 1). It is

seen that the expectation {of gain} in this lottery is 6; therefore, if desiring that it is fair, it is necessary that the stake be equal to 6. Usually {however} it is 10

Table 1

Probabilities of gains	Gains	Expectations
$P_6 = 1/46\ 656$	46 656	1
$P_7 = 1/7776$	7776	1
$P_8 = 1/2222$	2222	1
$P_{34} = 1/2222$	2222	1
$P_{35} = 1/7776$	7776	1
$P_{36} = 1/46\ 656$	46 656	1

Table 2

Probabilities of gains	Gains in lotteries (rubles)		
	No. 1	No. 2	No. 3
P_6 and P_{36}	7776	11 664	23 328
P_7 and P_{35}	1296	1944	0
P_8 and P_{34}	370.3	0	0

copecks so that for being fair the gains in the lottery corresponding to the sums 6 and 36, 7 and 35, and 8 and 34 should be

(46 656/600)·10 = 777.6 *rubles*; (7776/600)·10 = 129.6*r*;
(2222/600)·10 = 37.0*r* respectively

or, roughly, 780; 130; and 40 *rubles.*

Actually, however, the person licensed to organize a lottery assigns far lower gains so that the lottery is profitable for him and very disadvantageous for its participants. Thus, instead of the 780, 130 and 40*r*, 78, 13 and 4*r* were for example assigned so that 9/10 of the stakes went to the organizer. Suchlike lotteries were therefore abolished in every nation.

Let us compare now three lotteries: the first one, as described above; another lottery which includes four gains, two of 46 656*r* each and two other ones of 7776*r* each; and a third one with {only} two gains of 46 656*r* each. We assume that the lotteries are fair. The expectation {of gain} in the first lottery, as we have seen, is 6. For the second lottery, it is 4; and the expectation in the third lottery is 2. Supposing that the stakes are identical and equal to 1*r*, we obtain the following comparative table (Table 2) for these lotteries.

We see that the gains in these {improved} lotteries considerably differ one from another. The largest ones are in the lottery having the least number of gains, and although the probability of winning at least something is different for each lottery, all of them are equally fair because the expectations are the same and {moreover} equal to the stake[1].

Note 1. {According to Buffon's reasonable advice, a layman should rather ignore low probabilities (of gain) regarding them as non-existent. The same conclusion follows from the concept of moral certainty (of loss) which goes back to Descartes, Huygens and Jakob Bernoulli.}

3.3.19. Let us now consider the case in which l in formula (31) becomes very large and finally goes to infinity, that is, the case in which we are engaged when determining the probability that a sum of a very large or an infinite number of quantities having identical probabilities has a given value. We have {cf. (31)}

$$\sum P_s\, t^s = (t^n/l^n)\, [(t^l - 1)^n/(t - 1)]^n. \qquad\qquad (\text{xxvii})$$

We know however that if a function $f(t)$ can be expanded into a series

$$A_0 + A_1 t + A_2 t^2 + \dots + A_N t^N + \dots$$

then

$$A_N = (1/2\pi) \int_{-\pi}^{\pi} f(e^{\varphi i}) \, e^{-N\varphi i} d\varphi.$$

Therefore, denoting the right side of (xxvii) by $f(t)$, we shall find {since $P_s = A_s$} that

$$l^n P_N = (1/2\pi) \int_{-\pi}^{\pi} e^{\varphi i (n-N)} \left(\frac{e^{l\varphi i} - 1}{e^{\varphi i} - 1} \right)^n d\varphi.$$

Noting now that

$$\left(\frac{e^{l\varphi i} - 1}{e^{\varphi i} - 1} \right)^n = e^{(n/2)(l-1)\varphi i} \left(\frac{\sin(l\varphi/2)}{\sin(\varphi/2)} \right)^n$$

we will find that

$$P_N = (1/2\pi) \int_{-\pi}^{\pi} \exp\{[\frac{n(l-1)}{2} - N]\varphi i\} \left(\frac{\sin(l\varphi/2)}{l\sin(\varphi/2)} \right)^n d\varphi.$$

But

$$\exp\{[\frac{n(l-1)}{2} - N]\varphi i\} = \cos\{[\frac{n(l-1)}{2} - N]\varphi\} +$$
$$i\sin\{[\frac{n(l-1)}{2} - N]\varphi\}$$

and consequently

$$P_N = (1/2\pi) \int_{0}^{\pi} \cos\{[\frac{n(l-1)}{2} - N]\varphi\} \left(\frac{\sin(l\varphi/2)}{l\sin(\varphi/2)} \right)^n d\varphi.$$

Substituting however $l\varphi/2 = \theta$ we obtain

$$P_N = (1/\pi) \int_{0}^{\pi l/2} \cos\{[\frac{n(l-1)}{2} - N](2\theta/l)\} \left(\frac{\sin\theta}{l\sin(\theta/l)} \right)^n (2/l)\, d\theta. \qquad \textbf{(36)}$$

The probability P_N is thus expressed by a definite integral. Comparing this formula with formulas (32) amd (35) we can calculate the value of the integral

$$\int_{0}^{\pi l/2} \cos\{[\frac{n(l-1)}{2} - N](2\theta/l)\} \left(\frac{\sin\theta}{\sin(\theta/l)} \right)^n d\theta.$$

We shall now search for the probability that N is contained within boundaries N_1 and N_2. Calling this probability Π, we obtain

$$\prod_{N_1}^{N_2} = \sum_{N_1}^{N_2} P_N =$$

$$(1/\pi) \int_{0}^{\pi l/2} \sum_{N_1}^{N_2} \cos\{[\frac{n(l-1)}{2} - N](2\theta/l)\} \left(\frac{\sin\theta}{l\sin(\theta/l)} \right)^n (2/l)\, d\theta.$$

This is an exact expression of the probability sought. We shall now derive its approximate expression for the case in which l and n are very large. We have

$$P_N = (2/\pi\, l) \int_0^{\pi l/2} \cos\{[\frac{n(l-1)}{2} - \frac{2N}{l}]\theta\} \left((\sin\theta/\theta)\frac{\theta/l}{\sin(\theta/l)}\right)^n d\theta.$$

It is seen therefore that for a very large l we may approximately suppose that

$$P_N = (2/\pi\, l) \int_0^{\pi l/2} \cos\{[n - (2N/l)]\,\theta\}\,(\sin\theta/\theta)^n\, d\theta$$

because for very large values of n only those elements of the integral influence its value whose θ is very small. Indeed, considering here the integral as the limit of a sum in which the variable θ varies from $=$ to π $l/2$, we see that for very large values of n those terms of this sum in which θ has a considerable value will be very small because the multiplier, $(\sin\theta/\theta)^n$, will be very small. It follows that only those terms in which θ is very small will influence the value of the integral.

On these grounds, when approximately determining the probability P_N, we may consider θ to be very small; and in this case, noting that

$$(\sin\theta/\theta) = 1 - \theta^2/6 + \ldots$$

and

$$\ln(\sin\theta/\theta)^n = n\ln[1 - (n\,\theta^2/6)] = -(n\,\theta^2/6) - \ldots$$

we have

$$\cdot\ (\sin\theta/\theta)^n = \exp[-(n\,\theta^2/6)]$$

and

$$P_N = (2/\pi\, l) \int_0^{\pi l/2} \exp[-(n\,\theta^2/6)]\cos\{[n - (2N/l)]\,\theta\}\, d\theta.$$

Neglecting now the same integral taken between the limits $\pi\, l/2$ and ∞, we obtain

$$P_N = (2/\pi\, l) \int_0^{\infty} \exp[-(n\,\theta^2/6)]\cos\{[n - (2N/l)]\,\theta\}\, d\theta.$$

Issuing from the formula

$$\int_0^{\infty} \cos ax\, \exp(-bx^2)\, dx = (\sqrt{\pi}/2\sqrt{b})\exp(-a^2/4b)$$

we shall find that

$$P_N = (\sqrt{6}/l\sqrt{\pi\, n})\exp\left(\frac{-6[N-(nl/2)]^2}{nl^2}\right)$$

and that therefore

$$\prod_{N_1}^{N_2} = \sum_{N_1}^{N_2} (\sqrt{6}/l\sqrt{\pi\, n}) \exp\left(\frac{-6[N-(nl/2)]^2}{nl^2}\right).$$

Expressing this sum through an integral, as we also did before, and noting that for very large values of n all the terms except for the integral {?} will be very small, we may assume that

$$\prod_{N_1}^{N_2} = (\sqrt{6}/l\sqrt{\pi\, n}) \int_{N_1}^{N_2} \exp\left(\frac{-6[N-(nl/2)]^2}{nl^2}\right) dN.$$

Substituting now $(\sqrt{6}/l\sqrt{n})\,[N - (nl/2)] = t$, we will obtain

$$\prod_{N_1}^{N_2} = (1/\sqrt{\pi}) \int_{t_1}^{t_2} \exp(-t^2)\, dt \qquad\qquad (37)$$

where

$t_1 = (\sqrt{6}/l\sqrt{n})\,[N_1 - (nl/2)], \quad t_2 = (\sqrt{6}/l\sqrt{n})\,[N_2 - (nl/2)].$

Especially remarkable is the case in which $t_1 = -u,\ t_2 = u,$ so that

$$\Pi = (2/\sqrt{\pi}) \int_{0}^{u} \exp(-t^2)\, dt. \qquad\qquad (38)$$

It is seen now that it is very probable that N satisfies the inequalities

$(nl/2) + (l\sqrt{n}/\sqrt{6})\, u > N > (nl/2) - (l\sqrt{n}/\sqrt{6})\, u$

or

$$(l/2) + (lu/\sqrt{6n}) > N/n > (l/2) - (lu/\sqrt{6n}). \qquad\qquad (39)$$

Here, N is the sum of the values of all the events {cf. §3.3.17} occurring in n trials and we know that only one of these events can take place in any separate trial and that, on the other hand, one of them certainly ought to occur in each trial.

The formula (38) therefore determines the probability that the arithmetic mean of a very large number of magnitudes having equal probabilities of taking place in a very large number of trials is contained within the boundaries

$(l/2) + (lu/\sqrt{6n}),\ (l/2) - (lu/\sqrt{6n}).$

Since this probability can be made arbitrarily close to 1, we might say that

$\lim (N/n)_{n=\infty} = l/2,$

i.e., that the arithmetic mean of a very large number of quantities having equal probabilities tends to the limit, as the number of trials increases to infinity, equal to half the number of all the quantities; or, to half of the maximal quantity.

When desiring to solve the problem of whether an event is random or should it be attributed to certain causes, scientists base some of their physico-mathematical investigations on the conclusion to which we have arrived.

Suppose that event A_i is measured by magnitude ih so that the events under consideration have magnitudes $h,\ 2h,\ \ldots,\ lh$ with lh being their maximal value which we shall denote by a. Multiplying the inequality (39) by h we have

$(a/2) + (au/\sqrt{6n}) > hN/n > (a/2) - (au/\sqrt{6n})$.

It is seen now that at $n = \infty$ the arithmetic mean of all the magnitudes, hN/n, has half of the maximal one of them, $a/2$, as its limit. Let us apply this result to a problem in astronomy. If the inclinations of the planes of the planetary orbits to the plane of the ecliptic were absolutely random; in other words, if the probability that the inclination θ did not depend on θ, we would have found that the arithmetic mean of all the inclinations approximately equalled $90°$ (one half of the maximal inclination, of $180°$), i.e., the mean plane of the planetary orbits would have been perpendicular to the plane of the ecliptic. It occurs however that the planetary orbits make very small angles with the ecliptic so that the arithmetic mean of all the inclinations very little differs from zero. On these grounds it is concluded that the inclination of the planetary orbits is not {inclinations are not} random; that there existed some causes which imparted an approximately the same small inclination to all of them[1].

Note 1. {See a description of the pertinent work of Laplace in my paper (*Arch. Hist. Ex. Sci.*, vol. 9, 1973).}

3.3.20. Let us go over to a new and the last issue of the theory of probability. Although it does not concern repetitions of events, we insert it in this section because it is very closely linked with the problem considered in §3.3.19.
And so, we are going over to the determination of the probability that a sum of quantities varying due to random causes is contained within given boundaries. Suppose that we have several quantities x, y, z, etc, and assume that x can only have the values x_1, x_2, …; y, only the values y_1, y_2, …; z, only the values z_1, z_2, …

Denote the probabilities that x has value x_i, by p_i; that y has value y_i, by q_i; that z has value z_i, by r_i; etc. It is assumed that x, y, z, etc certainly have one of their values as stipulated above, so that

$p_1 + p_2 + \dots = 1$, $q_1 + q_2 + \dots = 1$, $r_1 + r_2 + \dots = 1$.

Then, let

$$p_1 x_1 + p_2 x_2 + \dots = a, \quad q_1 y_1 + q_2 y_2 + \dots = b, \quad r_1 z_1 + r_2 z_2 + \dots = c;$$
$$p_1 x_1^2 + p_2 x_2^2 + \dots = a_1, \quad q_1 y_1^2 + q_2 y_2^2 + \dots = b_1,$$
$$r_1 z_1^2 + r_2 z_2^2 + \dots = c_1 \tag{40}$$

so that a, b, c, … are the expectations of the {considered} quantities, and a_1, b_1, c_1, … the expectations of their squares. Let us now search for the probability that

$x + y + z + \dots = s$.

It is not difficult to see that if the probability sought is P_s, then

$$\sum P_s t^s = \left(p_1 t^{x_1} + p_2 t^{x_2} + \dots\right)\cdot\left(q_1 t^{y_1} + q_2 t^{y_2} + \dots\right)\cdot\left(r_1 t^{z_1} + r_2 t^{z_2} + \dots\right)\dots \tag{41}$$

Now, so as to simplify the derivation, we suppose that x, y, z, … can only have integer values. Later on, it will be possible to remove this restriction by assuming that these quantities are expressed in very small fractions of that unit in which we suppose them to be expressed during our deductions. Of course, such an approach presumes that x, y, z, … are rational quantities, whereas, for a general solution of our problem, we ought to assume them as arbitrary quantities. Nevertheless, we adopt our restriction because we shall only search for an approximate value of the probability P_s.
Under the restrictions made we shall find that

$$P_s = (1/2\pi) \int_{-\pi}^{\pi} [p_1 \exp(x_1\varphi i) + p_2 \exp(x_2\varphi i) + \dots]\cdot$$

$[q_1 \exp(y_1\varphi i) + q_2 \exp(y_2\varphi i) + \dots]\cdot[r_1 \exp(z_1\varphi i) + r_2 \exp(z_2\varphi i) + \dots]\cdot$
$e^{-s\varphi i} d\varphi$.

We shall now make approximate conclusions supposing that the number of the quantities $x, y, z, \ldots$ is very large. Guiding ourselves by the considerations developed in §3.3.16, we may, when approximately calculating this integral, neglect the powers of φ exceeding the second one. Consequently, noting that

$$p_1 \exp(x_1 \varphi i) + p_2 \exp(x_2 \varphi i) + \ldots = p_1 \{1 + [(x_1 \varphi i)/1!] - [(x_1^2 \varphi^2)/2!] + \ldots\} + p_2 \{1 + [(x_2 \varphi i)/1!] - [(x_2^2 \varphi^2)/2!] + \ldots\} + \ldots,$$

we obtain, on the strength of the equalities (40),

$$P_s = (1/2\pi) \int_{-\pi}^{\pi} [1 + a\varphi i - (a_1 \varphi^2)/2] \cdot [1 + b\varphi i - (b_1 \varphi^2)/2] \cdot$$
$$[1 + c\varphi i - (c_1 \varphi^2)/2] \ldots e^{-s\varphi i} \, d\varphi.$$

But we have

$$\ln \{[1 + a\varphi i - (a_1 \varphi^2)/2] \cdot [1 + b\varphi i - (b_1 \varphi^2)/2] \ldots\} =$$
$$a\varphi i - (a_1 \varphi^2)/2 - (1/2)[a\varphi i - (a_1 \varphi^2)/2]^2 + \ldots$$
$$b\varphi i - (b_1 \varphi^2)/2 - (1/2)[b\varphi i - (b_1 \varphi^2)/2]^2 + \ldots = (a + b + c)\varphi i +$$
$$[(a^2 - a_1) + (b^2 - b_1) + (c^2 - c_1) + \ldots](\varphi^2/2) = A\varphi i - B(\varphi^2/2)$$

where

$$a + b + c + \ldots = A, \quad -[(a^2 - a_1) + (b^2 - b_1) + (c^2 - c_1) + \ldots] = B$$
$$\text{(xxviii)}$$

We thus get

$$P_s = (1/2\pi) \int_{-\pi}^{\pi} \exp[-B(\varphi^2/2)] \exp(A\varphi i) \exp(-s\varphi i) \, d\varphi =$$

$$(1/\pi) \int_0^{\pi} \exp[-B(\varphi^2/2)] \cos[(A - s)\varphi] \, d\varphi.$$

Neglecting the magnitude of the last integral taken within π and ∞, which is allowable for sufficiently large and positive values of B, we arrive at

$$P_s = (1/\pi) \int_0^{\infty} \exp[-B(\varphi^2/2)] \cos[(A - s)\varphi] \, d\varphi =$$

$$\frac{1}{\sqrt{2\pi B}} \exp[-(A - s)^2/2B]. \tag{42}$$

In order to justify the approximation made just above, it ought to be proved that $B > 0$ and that, when the number of the quantities $x, y, z, \ldots$ increases, B increases as well. To this end, it is sufficient to prove that all the magnitudes
$(a_1 - a^2), (b_1 - b^2), \ldots$ are positive. Let us take one of these differences, and what will be said about it will also hold for the other ones. We have

$$a_1 - a^2 = a_1 - 2a^2 + a^2 = p_1 x_1^2 + p_2 x_2^2 + \ldots - 2a(p_1 x_1 + p_2 x_2 + \ldots) +$$
$$a^2(p_1 + p_2 + \ldots) = p_1(x_1^2 - 2ax_1 + a^2) + p_2(x_2^2 - 2ax_2 + a^2) + \ldots + =$$
$$p_1(x_1 - a)^2 + p_2(x_2 - a)^2 + \ldots$$

This shows that the magnitude $(a_1 - a^2)$ is always positive.

We turn now to formula (42) that determines the probability P_s that the sum $(x + y + z + \ldots)$ has a given value s. Denoting the probability that s is contained within the boundaries s_o and s_1 by Π and

noting that in virtue of our assumption B is very large, so that the sum might be without a perceptive error replaced by an integral, we obtain

$$\Pi = \int_{s_0}^{s_1} P_s \, ds. \qquad\qquad\qquad\qquad (\text{xxix})$$

where P_s is given by expression (42).

Considering this formula, we note that it determines the probability in the form of a homogeneous function of degree zero with respect to the quantities $x, \ y, \ z, \ \ldots$ Therefore, it will not change if we change the unit in which these quantities are expressed. Consequently, the adopted restriction concerning these quantities may be abandoned and we shall now suppose that s in the formula (xxviii) is arbitrary[1].

Denoting now

$$[(s - A)/\sqrt{2B}\,] = t, \ s_0 = A + t_0\sqrt{2B}, \ s_1 = A + t_1\sqrt{2B}$$

we obtain

$$\prod_{A+t_0\sqrt{2B}}^{A+t_1\sqrt{2B}} = (1/\sqrt{\pi}) \int_{t_0}^{t_1} \exp(-t^2) \, dt.$$

If $t_0 = -u$ and $t_1 = u$ we get

$$\prod_{A-u\sqrt{2B}}^{A+u\sqrt{2B}} = (2/\sqrt{\pi}) \int_0^u \exp(-t^2) \, dt. \qquad\qquad\qquad (\mathbf{43})$$

This is the formula that indeed determines the probability that

$$A + u\sqrt{2B} > s > A - u\sqrt{2B}$$

with A and B introduced in formulas (xxviii). This probability tends to 1 with an increasing u, but, on the other hand, the interval between the boundaries for s will {then} become wider. These boundaries depend also on the magnitude B, which, in turn, depends on the number of the quantities $x, \ y, \ z, \ \ldots$ Denoting this number by n, we find that the formula (43) determines the probability that

$$(A/n) - (u/\sqrt{n})\sqrt{2B/n} < s/n < (A/n) + (u/\sqrt{n})\sqrt{2B/n}.$$

We have already obtained these inequalities and their probability in §§3.2.1 and 3.2.2 where it was shown that

$$\prod_{A-u\sqrt{2B}}^{A+u\sqrt{2B}} = 1 - (\theta/2u^2) \qquad\qquad\qquad\qquad (\text{xxx})$$

where θ was a proper fraction. Formula (xxx) was proved there absolutely rigorously[2]. This formula ought to be therefore applied in theoretical investigations although it does not allow the {actual} calculation of the probability. Formula (43) that provides such a possibility was however derived in a non-rigorous way. The lack of rigor in the derivation consisted in that we made various assumptions without determining the boundary of the ensuing errors. In its present state, mathematical analysis cannot derive this boundary in any satisfactory fashion[3]. In spite of this, we shall apply formula (43) when expounding the method of least squares to which we are now indeed going over.

Note 1. This already follows from the replacement of the sum by an integral so that ds was introduced assuming that s varied continuously.

Note 2. Formula (xxx) can be obtained by substituting $t = u\sqrt{2}$ in the final formula of §3.2.1.

Note 3. Liapunov […] provided quite a rigorous general proof of this so-called {central} "limit theorem of the theory of probability". It is possible that the preceding words of his celebrated teacher had indeed prompted him to consider this problem. A. Krylov.

3.4. Applications of the Theory of Probability to the Treatment of Observations

3.4.1. In the sequel, we shall base our considerations on formula (43) that determines the probability that the sum $(x + y + z + …)$ of quantities varying due to random circumstances is contained within the boundaries

$$A + u\sqrt{2B} \quad \text{and} \quad A - u\sqrt{2B}.$$

We shall choose such a value for u that this probability determined by the formula

$$(2/\sqrt{\pi})\int_0^u \exp(-t^2)\, dt \qquad\qquad (\text{xxxi})$$

will be equal to $1/2$. This value is approximately 0.48. In this case, we shall call the boundaries indicated above *probable* because the probable boundaries for the sum $(x + y + z + …)$ are such that it is contained with equal probability either within or beyond them. The "width" of these boundaries is approximately $2 \cdot 0.48\sqrt{2B} = 0.96\sqrt{2B}$. We know however that for somewhat considerable values of u, for example for $u = 3$, the magnitude (xxxi) is very close to 1 [1]; therefore, if the "width" of the probable boundaries be increased six- or seven-fold, we shall already obtain such boundaries for which we may say with a very high probability that the sum $(x + y + z + …)$ is contained within them.

When considering errors of observation, we shall call them *probable* if they are contained within probable boundaries; that is, between the boundaries $A - 0.48\sqrt{2B}$ and $A + 0.48\sqrt{2B}$, and we shall always assume in the sequel that $u = 0.48$.

Before going ahead, let us agree about one more term; we shall say that the observations do not include a "constant error" if positive and negative errors are equally probable, i.e., if the expectation of the errors is zero. We shall assume that this condition is fulfilled; that is, we shall consider the expectation of the errors equal to zero. If however we shall have to study observations corrupted by constant errors we shall make the appropriate reservation. It ought to be noted that the equally probable errors are supposed to be those having equal numerical values.

Note 1. For the sake of clearness we append a short table of the values of this integral {omitted}. A. Krylov.

3.4.2. Suppose that we are concerned with measuring some quantity V and that the observations provided its following values: $L_1, L_2, …, L_n$. In this case their arithmetic mean, i.e., $(L_1 + L_2 + … + L_n)/n$, is usually taken as the value of V; and, the larger is the number of observations, n, the closer we consider it to be to V. We shall now show the grounds on which such opinions are based.

Let $\varepsilon_1, \varepsilon_2, …, \varepsilon_n$ be the errors of the first, the second, .., the n-th observation. We shall consider those errors which increase the real value of V as positive and regard those that decrease it as negative. Under these conditions we have

$$V = L_1 - \varepsilon_1,\ V = L_2 - \varepsilon_2, …,\ V = L_n - \varepsilon_n \qquad\qquad (\text{xxxii})$$

so that

$$nV = (L_1 + L_2 + … + L_n) - (\varepsilon_1 + \varepsilon_2 + … + \varepsilon_n),$$
$$V = [(L_1 + L_2 + … + L_n)/n] - [(\varepsilon_1 + \varepsilon_2 + … + \varepsilon_n)/n].$$

It is seen therefore that, when assuming the arithmetic mean of the observational values as V, we make an error equal to

$$\varepsilon = (\varepsilon_1 + \varepsilon_2 + … + \varepsilon_n)/n.$$

Let us find the probable boundaries of this error. Applying formula (43) to this case, and denoting $x = \varepsilon_1/n$, $y = \varepsilon_2/n$, $z = \varepsilon_3/n$, ... we have in this case

$$a = (1/n)\sum \varepsilon_{1(i)} p_i = 0; \quad b = (1/n)\sum \varepsilon_{2(i)} q_i = 0; \ldots$$

where $\varepsilon_{1(i)}$ is one of the possible errors of the first observation and p_i, its probability; $\varepsilon_{2(i)}$, one of the possible errors of the second observation and q_i, its probability, etc.

Consequently, we have $A = a + b + c + \ldots = 0$. Then

$$a_1 = \sum (\varepsilon_{1(i)}/n)^2 p_i = (1/n^2) \sum (\varepsilon_{1(i)})^2 p_i = k/n^2$$

where

$$k = \sum (\varepsilon_{1(i)})^2 p_i.$$

The value of k depends on the quality of the observations and it is not difficult to see that the less it is the better are the observations because k can only be small when the errors are small in numerical value. If all the observations are equally good, then k is the same for all of them and we will have

$$a_1 = b_1 = c_1 = \ldots = k/n^2$$

so that

$$B = (a_1 - a^2) + (b_1 - b^2) + (c_1 - c^2) + \ldots = nk/n^2 = k/n.$$

We thus see that the magnitudes

$$-u\sqrt{2k/n} \quad \text{and} \quad u\sqrt{2k/n}$$

will be the probable boundaries for ε. They show that the "width" of the probable boundaries decreases with the increase in the number of observations provided that all of them are equally good. This indeed is the basis for increasing this number when it is desired to determine the quantity sought "more precisely" through the arithmetic mean of the observed values. Let us now go over to another issue.

3.4.3. Suppose that we are again concerned with determining the quantity V for which the observations provided the values L_1, L_2, ..., L_n. It is required to find such a combination of these observations as would have furnished the *most probable* value of V. This problem can be formulated either as finding the best combination out of all the possible ones; or, out of all combinations of a given form.

In the first case, the problem is of course much more general, and it can only be solved by applying the law of hypotheses [1] whereas no deductions made on its basis have adequate rigor. We shall therefore solve this problem in its second version choosing the following form of the combinations

$$(\lambda_1 L_1 + \lambda_2 L_2 + \ldots + \lambda_n L_n)/(\lambda_1 + \lambda_2 + \ldots + \lambda_n). \tag{xxxiii}$$

Here, we shall try to determine λ_1, λ_2, ..., λ_n in conformity with the condition that this combination expresses V in the best way, i.e., that the probable boundaries of the error be the tightest {the narrowest}. Note that the arithmetic mean is a particular case of this combination; namely, the case in which $\lambda_1 = \lambda_2 = \ldots = \lambda_n$.

From equalities (xxxii) {multiplying them by λ_i respectively, etc} we have

$$V(\lambda_1 + \lambda_2 + \ldots + \lambda_n) = (\lambda_1 L_1 + \lambda_2 L_2 + \ldots + \lambda_n L_n) - (\varepsilon_1 \lambda_1 + \varepsilon_2 \lambda_2 + \ldots + \varepsilon_n \lambda_n),$$

$$V = [(\lambda_1 L_1 + \lambda_2 L_2 + \ldots + \lambda_n L_n)/(\lambda_1 + \lambda_2 + \ldots + \lambda_n)] - [(\varepsilon_1 \lambda_1 + \varepsilon_2 \lambda_2 + \ldots + \varepsilon_n \lambda_n)/(\lambda_1 + \lambda_2 + \ldots + \lambda_n)]$$

so that, when assuming the quantity (xxxiii) as V, we make an error equal to

$$\varepsilon = (\varepsilon_1\lambda_1 + \varepsilon_2\lambda_2 + \ldots + \varepsilon_n\lambda_n)/(\lambda_1 + \lambda_2 + \ldots + \lambda_n).$$

We shall now find the probable boundaries of this error. Let

$$x = \frac{\varepsilon_1\lambda_1}{\lambda_1 + \lambda_2 + \ldots + \lambda_n}, \quad y = \frac{\varepsilon_2\lambda_2}{\lambda_1 + \lambda_2 + \ldots + \lambda_n}, \quad z = \frac{\varepsilon_3\lambda_3}{\lambda_1 + \lambda_2 + \ldots + \lambda_3} \quad \text{etc.}$$

In this case we have

$$a = \sum \frac{\varepsilon_{1(i)}\lambda_1}{\lambda_1 + \lambda_2 + \ldots + \lambda_n} p_i = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \ldots + \lambda_n} \sum \varepsilon_{1(i)} p_i = 0$$

so that $a = b = c = \ldots = 0$ and, consequently, $A = 0$. We then have

$$a_1 = \sum \frac{(\varepsilon_{1(i)})^2 \lambda_1^2}{(\lambda_1 + \lambda_2 + \ldots + \lambda_n)^2} p_i = [\frac{\lambda_1}{\lambda_1 + \lambda_2 + \ldots + \lambda_n}]^2 \sum (\varepsilon_{1(i)})^2 p_i =$$

$$k [\frac{\lambda_1}{\lambda_1 + \lambda_2 + \ldots + \lambda_n}]^2.$$

Supposing that all the observations are equally good, we obtain, in an absolutely similar way,

$$b_1 = \frac{k\lambda_2^2}{(\lambda_1 + \lambda_2 + \ldots + \lambda_n)^2}, \quad c_1 = \frac{k\lambda_3^2}{(\lambda_1 + \lambda_2 + \ldots + \lambda_n)^2}, \ldots$$

and therefore

$$B = \frac{k(\lambda_1^2 + \lambda_2^2 + \ldots + \lambda_n^2)}{(\lambda_1 + \lambda_2 + \ldots + \lambda_n)^2}.$$

so that the probable boundaries are

$$-u\sqrt{2B}, \quad u\sqrt{2B}.$$

As already stated, the best combination is that for which the "width" of the boundaries is least so that the problem is reduced to the determination of $\lambda_1, \lambda_2, \ldots, \lambda_n$ in accord with the condition that the expression

$$W = (\lambda_1^2 + \lambda_2^2 + \ldots + \lambda_n^2)/(\lambda_1 + \lambda + \ldots + \lambda_n)^2 \qquad \text{(xxxiv)}$$

is minimal.

It is easy to see however that in such a form the problem is indefinite because the magnitudes $\lambda_1, \lambda_2, \ldots, \lambda_n$ themselves will not thus be determined, only the ratios between them will be derived. We may therefore lay down any {additional} condition between them {connecting them} expressed by one equation. The simplest result will be provided by stipulating that

$$\lambda_1 + \lambda + \ldots + \lambda_n = 1 \qquad (44)$$

so that

$$W = \lambda_1^2 + \lambda_2^2 + \ldots + \lambda_n^2. \qquad \text{(xxxv)}$$

The condition that W be minimal furnishes the equations

$$\lambda_1 \, d\lambda_1 \; + \; \lambda_2 \, d\lambda_2 \; + \ldots + \; \lambda_n \, d\lambda_n \; = \; 0, \; d\lambda_1 \; + \; d\lambda_2 \; + \ldots + \; d\lambda_n \; = \; 0.$$

Multiplying the second one by an indefinite factor ρ, subtracting {the product} from the first one and equating then the coefficients of $d\lambda_1, \; d\lambda_2, \; \ldots, \; d\lambda_n$ to zero, we obtain

$$\lambda_1 \; = \; \rho, \; \lambda_2 \; = \; \rho, \; \ldots, \; \lambda_n \; = \; \rho.$$

Together with equation (44) these equations provide

$$\lambda_1 \; = \; \lambda_2 \; = \; \ldots \; = \; \lambda_n \; = \; (1/n).$$

We thus see that the minimum of W takes place when the magnitudes $\lambda_1, \; \lambda_2, \; \ldots, \; \lambda_n$ are equal one to another.

It is not amiss to derive this {the same result} in another, elementary way. Suppose that

$$\lambda_1 \; + \; \lambda_2 \; + \ldots + \; \lambda_n \; = \; s.$$

We have

$$[\lambda_1 \; - \; (s/n)]^2 \; + \; [\lambda_2 \; - \; (s/n)]^2 \; + \ldots + \; [\lambda_n \; - \; (s/n)]^2 \; =$$
$$\lambda_1{}^2 \; + \; \lambda_2{}^2 \; + \ldots + \lambda_n{}^2 \; - \; 2(s/n) \, (\lambda_1 \; + \; \lambda_2 \; + \ldots + \; \lambda_n) \; + \; (ns^2/n^2)$$

so that

$$\lambda_1{}^2 \; + \; \lambda_2{}^2 \; + \ldots + \lambda_n{}^2 \; = \; (s^2/n) \; +$$
$$[\lambda_1 \; - \; (s/n)]^2 \; + \; [\lambda_2 \; - \; (s/n)]^2 \; + \ldots + \; [\lambda_n \; - \; (s/n)]^2$$

and consequently [2]

$$W \; = \; (1/n) \; + \; \{[\lambda_1 \; - \; (s/n)]^2 \; + \; [\lambda_2 \; - \; (s/n)]^2 \; + \ldots + \; [\lambda_n \; - \; (s/n)]^2\}/s^2.$$

Since the second term on the right side of this equality is always positive, we see that the minimum of W will only take place when this term vanishes; that is, when

$$\lambda_1 \; = \; \lambda_2 \; = \; \ldots \; = \; \lambda_n \; = \; (s/n)$$

or, in accord with the condition that the magnitudes $\lambda_1, \; \lambda_2, \; \ldots, \; \lambda_n$ are equal one to another. We thus conclude that the magnitude $(L_1 \; + \; L_2 \; + \ldots + \; L_n)/n,$
and the arithmetic mean of the observed magnitudes, ought to be taken as V.

Note 1. {Chebyshev did not use this term in the relevant sections (§§3.1.6 – 3.1.7) but he obviously thought here about the Bayesian approach with an arbitrary choice of the prior distribution. In §3.4.9 he wrongly attributed to Gauss the justification of the method of least squares by this law of hypotheses. However, the postulate of the arithmetic mean adopted by Gauss in 1809 made the Bayesian approach superfluous; E.T.Whittaker & G. Robinson (*Calculus of Observations*. London, 1924, p. 219n) were the first to indicate this point (overlooked by Gauss), and even their remark was forgotten. Also note that Gauss subsequently abandoned his initial substantiation of the method. In general, Chebyshev was poorly acquainted with the work of the creator of the method of least squares as especially witnessed by his wrong reasoning in §3.4.9, see my papers of 1994 in the *Arch. Hist. Ex. Sci.*, vols 46 and 48.}

Note 2. {Chebyshev issues here from formula (xxxiv).}

3.4.4. Having shown that out of all the combinations of the type of (xxxiii) the best one is the arithmetic mean of the observed magnitudes, we shall show now on what grounds should we search for the best of all the possible combinations.

Suppose that we know the function $\varphi(\varepsilon)$ determining the probability that the error has magnitude ε. In this case we can easily determine the best combination of the observations. Indeed, let us assume that V can take the values V_0, V_1, V_2, ..., V_λ, ... (which can be arbitrarily close one to another). In order to determine V we made n observations providing the values L_1, L_2, ..., L_n. This represents event E.

The various hypotheses under which this event can happen are $V = V_0$, $V = V_1$, $V = V_2$, ..., $V = V_\lambda$, ... Knowing nothing about the appropriate probabilities, we suppose that they are equal one to another and we denote their common value by P. Then $P_0 = P_1 = P_2 = ... = P_\lambda = ... = P$. This assumption is arbitrary and none of the subsequent deductions is therefore rigorous. The probability of the event E under the hypothesis that $V = V_\lambda$, i.e., that, when determining some quantity V_λ by observations, we got the values L_1, L_2, ..., L_n, is

$$p_\lambda = \varphi(L_1 - V_\lambda) \cdot \varphi(L_2 - V_\lambda) ... \varphi(L_n - V_\lambda).$$

Therefore, the probability that the event E took place under the hypothesis $V = V_\lambda$ will be expressed in the following way

$$\frac{P_\lambda p_\lambda}{\sum P_\lambda p_\lambda} = \frac{\varphi(L_1 - V_\lambda)\varphi(L_2 - V_\lambda)...\varphi(L_n - V_\lambda)}{\sum \varphi(L_1 - V_\lambda)\varphi(L_2 - V_\lambda)...\varphi(L_n - V_\lambda)}$$

where the sum is extended over all the values of λ and is therefore a constant magnitude.

The search for the maximum of this probability is thus reduced to the search for the maximum of the expression

$$W = \varphi(L_1 - V_\lambda) \cdot \varphi(L_2 - V_\lambda) ... \varphi(L_n - V_\lambda). \qquad \text{(xxxvi)}$$

In the case in which the function $\varphi(z)$ is known, this condition will indeed provide an equation for determining V_λ as a function of the observed magnitudes L_1, L_2, ..., L_n. This latter function will indeed furnish the best combination of the observations because the probability that V_λ is equal to that combination is maximal. We thus see that everything consists in determining the type of the function $\varphi(z)$. Some theoretical considerations which we will discuss below lead to the conclusion that

$$\varphi(z) = F \exp(-gz^2) \qquad \text{(xxxvii)}$$

where F and g are some constant magnitudes. Experimental corroboration of this formula was being attempted and it was found out that (xxxvii) rather well expresses the law of the probability of error as a function of the change in the value of this error.

Assuming the function (xxxvii) as $\varphi(z)$ we shall find that

$$W = F^n \exp(-gt^2),$$

$$t^2 = (L_1 - V_\lambda)^2 + \cdot (L_2 - V_\lambda)^2 + ... + (L_n - V_\lambda)^2. \qquad \text{(xxxviii)}$$

It is therefore seen that the maximum of W will take place when the right side of (xxxviii) where V_λ is considered as the independent variable is maximal. The value of V_λ making the expression (xxxviii) maximal is determined by the equation

$$-2(L_1 - V_\lambda) - 2(L_2 - V_\lambda) - ... - 2(L_n - V_\lambda) = 0$$

which provides

$$V_\lambda = (L_1 + L_2 + ... + L_n)/n.$$

The assumption that $\varphi(z)$ is expressed by the formula (xxxvii) leads to the conclusion that the best of all the possible combinations is the arithmetic mean of the observed magnitudes.

3.4.5. We shall now show on what theoretical grounds is the determination of the type of the function $\varphi(z)$ founded. Note that when two observations are available it might be assumed as an obvious condition that their best combination is the arithmetic mean because in this case nothing empowers us to prefer one of the observed magnitudes to the other one. We therefore really ought to decide in favor of the magnitude which would be equally distant from each of the observed magnitudes.

But the same can not be said about three or more observations. Indeed; suppose that we made three observations and that two of them furnished one and the same value for the quantity sought, V. In such a case we ought to prefer the magnitude which was repeated twice, but should our preference be expressed in taking 2/3 of the repeated magnitude and 1/3 of the magnitude that occurred {only} once, and in assuming that the sum thus obtained is the quantity V? Obviously we have no right to assert this.

We shall now show that, if the arithmetic mean is taken as the best combination out of three observations, it is possible to find the type of the function $\varphi(z)$. We saw that the best combination is found from the condition of the maximum of the expression (xxxvi). For three observations this will be

$$W = \varphi(L_1 - V_\lambda)\cdot\varphi(L_2 - V_\lambda)\cdot\varphi(L_3 - V_\lambda).$$

The search for its maximum leads to equation

$$\frac{\varphi'(L_1 - V_\lambda)}{\varphi(L_1 - V_\lambda)} + \frac{\varphi'(L_2 - V_\lambda)}{\varphi(L_2 - V_\lambda)} + \frac{\varphi'(L_3 - V_\lambda)}{\varphi(L_3 - V_\lambda)} = 0.$$

Denote $\varphi'(z)/\varphi(z) = \psi(z)$, then this equation will take the form

$$\psi(L_1 - V_\lambda) + \psi(L_2 - V_\lambda) + \psi(L_3 - V_\lambda) = 0.$$

Now we ought to express the demand that this equation be satisfied if

$$V_\lambda = (L_1 + L_2 + L_3)/3.$$

Assuming this value, we obviously have

$$(L_1 - V_\lambda) + (L_2 - V_\lambda) + (L_3 - V_\lambda) = 0.$$

Therefore, denoting $(L_1 - V_\lambda) = y$ and $(L_2 - V_\lambda) = y_1$, we obtain

$$(L_3 - V_\lambda) = -(y + y_1)$$

so that we ought to have the equality

$$\psi(y) + \psi(y_1) + \psi[-(y + y_1)] = 0 \qquad\qquad\qquad (45)$$

valid for any y and y_1. We may thus consider it as an equation determining the function $\psi(y)$.

Here we encounter a new mathematical problem: A function is determined not by a differential equation, not by an equation in finite differences, but by an equation connecting the values of the function sought corresponding to various values of the independent variables somehow connected one with another. Such equations are called functional. Some mathematicians concerned themselves with their solution (Abel among others), and some are engaged in this problem also at present. However, until now there exist no general methods for solving them whereas the existing methods actually consist in that, out of a given functional equation, a differential equation is made up by differentiation and elimination of the unknown magnitudes. We shall show one of these methods while studying the particular case under consideration.

Differentiating the equation (45) with respect to y we find out that

$$\psi'(y) - \psi'(-y - y_1) = 0.$$

Differentiating this equation, now with respect to y_1, we get

$\psi''(-y - y_1) = 0.$

Denoting $-y - y_1 = z$, we obtain $\psi''(z) = 0$ so that

$\psi(z) = C z + C_1.$

Making use now of the determined type of the function $\psi(z)$, we shall reduce the equation (45) to the form

$C y + C_1 + C y_1 + C_1 - C(-y - y_1) + C_1 = 0$

so that $C_1 = 0$. It follows that

$\psi(z) = C z = \varphi'(z)/\varphi(z)$

and, consequently, that

$(1/F) \ln \varphi(z) = C z^2/2, \quad \varphi(z) = F \exp(C z^2/2).$

Therefore

$W = F^3 \exp\{(C/2)[(L_1 - V_\lambda)^2 + (L_2 - V_\lambda)^2 + (L_3 - V_\lambda)^2]\}$

and

$dW/dV_\lambda = CW[-(L_1 - V_\lambda) - (L_2 - V_\lambda) - (L_3 - V_\lambda)],$

$d^2W/dV_\lambda^2 = C(dW/dV_\lambda)[-(L_1 - V_\lambda) - (L_2 - V_\lambda) - (L_3 - V_\lambda)] + 3CW.$

Assuming that $3V = L_1 + L_2 + L_3$, we find that

$d^2W/dV_\lambda^2 = 3CW.$

Since W, at the considered values of V_λ, should be maximal, the expression obtained for d^2W/dV_λ^2 ought to be negative. And since W is positive (as being a product of three positive multipliers), C is negative. Therefore, denoting $C = -2g$, we obtain (xxxvii) where F and g are some positive magnitudes.

Thus, assuming that the arithmetic mean is the best combination of three observations, we found the type of the function $\varphi(z)$. However, as we said already, this supposition is arbitrary and not caused by necessity. Note that the assumption that the arithmetic mean is the best combination for two observations is necessary but not sufficient for determining the type of the function $\varphi(z)$. Indeed, in this case the equation determining the best combination will be

$$\frac{\varphi'(L_1 - V_\lambda)}{\varphi(L_1 - V_\lambda)} + \frac{\varphi'(L_2 - V_\lambda)}{\varphi(L_2 - V_\lambda)} = 0.$$

Denoting as before $\varphi'(z)/\varphi(z) = \psi(z)$ we shall reduce it to

$\psi(L_1 - V_\lambda) + \psi(L_2 - V_\lambda) = 0.$

This equation should be satisfied by $V_\lambda = (L_1 + L_2)/2$; expressing this {condition} and denoting $L_1 + L_2 = 2y$, we obtain

$\psi(y) + \psi(-y) = 0.$

This equation, however, does not determine the function $\psi(y)$ because any odd function satisfies it.

If we assume that indeed (xxxvii) holds, we shall easily arrive at the method of least squares. To this end denote the error of the i-th observation by x_i. We have seen however that the probability of some totality of errors is determined thus:

$$\frac{\varphi(x_1)\,\varphi(x_2)...\varphi(x_n)}{\sum[\varphi(x_1)\,\varphi(x_2)...\varphi(x_n)]}.$$

Therefore, we ought to assign such magnitudes to the errors [1] that the expression

$$\varphi(x_1)\cdot\varphi(x_2)\,...\,\varphi(x_n) \qquad\qquad\qquad (47)$$

be maximal because the denominator in the preceding expression is a constant magnitude. Given the existence of the equation above, this product will be

$$F^{n}\exp[-g(x_1{}^2 + x_2{}^2) + ... + x_n{}^2)].$$

Consequently, we should assign such values to the errors that the expression

$$x_1{}^2 + x_2{}^2 + ... + x_n{}^2$$

be minimal; that is, in order to find the most probable errors we ought to find the minimum of the sum of their squares.

Note 1. {Chebyshev did not expressly distinguish between errors and residuals.}

3.4.6. Very often the observations directly provide not the quantity sought, V, but quantities connected with it by some equations. We shall consider the case in which the observations furnish the quantities

$$\alpha_1 V,\ \alpha_2 V,\ ...,\ \alpha_n V.$$

Here, α_1, α_2, ..., α_n are known numbers and $\alpha_i V$ is a quantity determined by the i-th observation. Suppose that for these quantities the n observations gave such values: L_1, L_2, ..., L_n. the problem consists in finding the most reliable magnitude for V. When solving this problem we will not search for the best of all the possible combinations of the observations, because, as it was shown on a simpler case, this cannot be done in a rigorous way. We will {rather} determine the most favorable combination out of those of the type

$$\frac{\lambda_1 L_1 + \lambda_2 L_2 + ... + \lambda_n L_n}{\lambda_1 \alpha_1 + \lambda_2 \alpha_2 + ... + \lambda_n \alpha_n}.$$

Denoting the error of the i-th observation by ε_i, we have

$$a_1 V = L_1. - \varepsilon_1,\ a_2 V = L_2. - \varepsilon_2,\ ...,\ a_n V = L_n. - \varepsilon_n.$$

Multiplying these equations by λ_1, λ_2, ..., λ_n respectively, and adding up the results, we obtain

$$(a_1 \lambda_1 + a_2 \lambda_2 + ... + a_n \lambda_n) V =$$
$$(\lambda_1 L_1 + \lambda_2 L_2 + ... + \lambda_n L_n) - (\lambda_1 \varepsilon_1 + \lambda_2 \varepsilon_2 + ... + \lambda_n \varepsilon_n)$$

and

$$V = \frac{\lambda_1 L_1 + \lambda_2 L_2 + ... + \lambda_n L_n}{\lambda_1 \alpha_1 + \lambda_2 \alpha_2 + ... + \lambda_n \alpha_n} - \frac{\lambda_1 \varepsilon_1 + \lambda_2 \varepsilon_2 + ... + \lambda_n \varepsilon_n}{\lambda_1 \alpha_1 + \lambda_2 + ... + \lambda_n \alpha_n}. \qquad (xl)$$

It is seen therefore that, when assuming formula (xxxix), we make an error equal to the second term on the right side of (xl).

We shall now search for the probable boundaries for ε and determine $\lambda_1, \lambda_2, \ldots, \lambda_n$ from the condition that these boundaries are as close as possible to each other. However, if we do not {additionally} lay down any condition concerning these magnitudes, the issue will be indefinite. We shall therefore assume that

$$\alpha_1 \lambda_1 + \alpha_2 \lambda_2 + \ldots + \alpha_n \lambda_n = 1 \qquad\qquad\qquad (xl1)$$

from which the generality of the solution certainly will not suffer. And so, we have

$$V = \lambda_1 L_1 + \lambda_2 L_2 + \ldots + \lambda_n L_n, \quad \varepsilon = \lambda_1 \varepsilon_1 + \lambda_2 \varepsilon_2 + \ldots + \lambda_n \varepsilon_n.$$

When determining the probable boundaries for ε, we shall apply formula (43). To this end we denote

$$x = \lambda_1 \varepsilon_1, \ y = \lambda_2 \varepsilon_2, \ z = \lambda_3 \varepsilon_3, \ \ldots$$

and we shall understand $\varepsilon_{1(i)}$ as one of the possible values of ε_1, – that is, as one of the possible errors of the first observation, – $\varepsilon_{2(i)}$ as one of the possible values of ε_2, etc.

Supposing now that the observations have no constant errors, we have

$$a = \sum \lambda_1 \varepsilon_{1(i)} p_i = \lambda_1 \sum \varepsilon_{1(i)} p_i = 0, \ \ldots, \ b = 0, \ c = 0, \ \ldots$$

Assuming in addition that all the observations are of an equally high quality, we obtain

$$a_1 = \sum [\lambda_1 \varepsilon_{1(i)}]^2 p_i = \lambda_1^2 \sum [\varepsilon_{1(i)}]^2 p_i = \lambda_1^2 k$$

and, in the same way,

$$b_1 = \lambda_2^2 k, \ c_1 = \lambda_3^2 k, \ \ldots$$

Thus, we find that

$$A = 0, \ B = k (\lambda_1^2 + \lambda_2^2 + \ldots + \lambda_n^2).$$

The probable boundaries for ε will be

$$-u\sqrt{2B} \ \text{ and } \ u\sqrt{2B}.$$

The problem is now being reduced to the determination of the minimum of

$$\lambda_1^2 + \lambda_2^2 + \ldots + \lambda_n^2$$

under the condition (xli). To this end, we derive the equation

$$(\lambda_1 - \alpha_1 \rho) \, d\lambda_1 + (\lambda_2 - \alpha_2 \rho) \, d\lambda_2 + \ldots + (\lambda_n - \alpha_n \rho) \, d\lambda_n = 0$$

so that

$$\lambda_1 = \alpha_1 \rho, \ \lambda_2 = \alpha_2 \rho, \ \ldots, \ \lambda_n = \alpha_n \rho.$$

Then, from equation (xli),

$$\rho = 1/(\alpha_1^2 + \alpha_2^2 + \ldots + \alpha_n^2)$$

and

$$\lambda_i = \frac{\alpha_i}{\alpha_1{}^2 + \alpha_2{}^2 \ldots + \alpha_n{}^2}.$$

Consequently, the most reliable combination will be

$$V_1 = \frac{\alpha_1 L_1 + \alpha_2 L_2 + \ldots + \alpha_n L_n}{\alpha_1{}^2 + \alpha_2{}^2 + \ldots + \alpha_n{}^2}. \qquad\qquad\text{(xlii)}$$

If α_1, α_2, ..., α_n are integers, it might be said about this expression of V, that, Supposing that we, after making $\alpha_i{}^2$ observations, obtained in each of them the magnitude L_i/α_i for V; and supposing also that the total number of observations was $(\alpha_1{}^2 + \alpha_2{}^2 + \ldots + \alpha_n{}^2)$, the best combination of the observations will be, as we saw, the arithmetic mean of the observed values, hence (xlii).

The described method of determining V, expressed for the first time by Legendre, is indeed called the method of least squares, – not because we search for the minimum of the sum of the squares of the factors λ_1, λ_2, ..., λ_n, but because, when applying it, the value of V imparts the minimal value, as we shall prove now, to the sum of the squares of the errors.

To prove this, we shall find the value of V for which the sum

$$\varepsilon_1{}^2 + \varepsilon_2{}^2 + \ldots + \varepsilon_n{}^2$$

becomes minimal. We have

$$\varepsilon_1{}^2 + \varepsilon_2{}^2 + \ldots + \varepsilon_n = (L_1 - \alpha_1 V)^2 + (L_2 - \alpha_2 V)^2 + \ldots + (L_n - \alpha_n V)^2.$$

For determining its minimum we obtain the equation

$$\alpha_1 (L_1 - \alpha_1 V) + \alpha_2 (L_2 - \alpha_2 V) + \ldots + \alpha_n (L_n - \alpha_n V) = 0$$

from which we indeed arrive at (xlii)

Further on we shall expound the method of finding the best combinations for determining several unknowns from observations; and we shall show that in this case as well the most reliable combinations are those for which the sum of the squares of the errors is the least. We note right here that the method of least squares is not the only one applied for the treatment of observations. There exists one more method according to which the unknown quantity is determined from the observations in such a way that the maximal error is minimal and this method is in some cases more advantageous than the method of least squares, but in general the latter should be preferred.

3.4.7. Suppose now that it is required to determine from observations quantities U and V and that in the i-th observation we search for the value of the expression $(\alpha_i U + \beta_i V)$ where α_i and β_i are given numbers and that this observation provides the magnitude L_i. Denoting the error of the i-th observation by ε_i, we have

$$\alpha_1 U + \beta_1 V = L_1 - \varepsilon_1, \; \alpha_2 U + \beta_2 V = L_2 - \varepsilon_2, \ldots, \alpha_n U + \beta_n V = L_n - \varepsilon_n.$$
$$\textbf{(48)}$$

We shall find the combinations by whose means U and V are determined with least errors. And, as before, we shall only consider linear combinations (with respect to L_1, L_2, ...). To this end we ought to proceed as follows.

Multiply each of the equations (48) by some indefinite factor λ_i and add up the results; this will provide one equation for determining U and V, namely

$$(\alpha_1\lambda_1 + \alpha_2\lambda_2 + \ldots + \alpha_n\lambda_n)U + (\beta_1\lambda_1 + \beta_2\lambda_2 + \ldots + \beta_n\lambda_n)V =$$
$$(L_1\lambda_1 + L_2\lambda_2 + \ldots + L_n\lambda_n) - (\varepsilon_1\lambda_1 + \varepsilon_2\lambda_2 + \ldots + \varepsilon_n\lambda_n).$$

We then subject the factors λ_1, λ_2, ..., λ_n to the conditions

$$\alpha_1 \lambda_1 + \alpha_2 \lambda_2 + \dots + \alpha_n \lambda_n = 1, \quad \beta_1 \lambda_1 + \beta_2 \lambda_2 + \dots + \beta_n \lambda_n = 0. \tag{49}$$

Consequently, we shall directly obtain an expression for U

$$U = (L_1 \lambda_1 + L_2 \lambda_2 + \dots + L_n \lambda_n) - (\varepsilon_1 \lambda_1 + \varepsilon_2 \lambda_2 + \dots + \varepsilon_n \lambda_n).$$

Knowing nothing about the magnitude of the errors, we assume for U the expression

$$U = L_1 \lambda_1 + L_2 \lambda_2 + \dots + L_n \lambda_n \tag{50}$$

thus making an error equal to

$$\varepsilon = \varepsilon_1 \lambda_1 + \varepsilon_2 \lambda_2 + \dots + \varepsilon_n \lambda_n. \tag{51}$$

Had we subjected the factors to conditions

$$\alpha_1 \lambda_1 + \alpha_2 \lambda_2 + \dots + \alpha_n \lambda_n = 0, \quad \beta_1 \lambda_1 + \beta_2 \lambda_2 + \dots + \beta_n \lambda_n = 1, \tag{52}$$

we would have obtained an expression for V.

It is seen now that, after finding a final expression for U, we can directly write down a final expression for V as well by replacing the letters α_i by β_i and vice versa. We shall therefore only discuss now the determination of U, and everything said about it will also apply to the determination of V.

And so, the issue is reduced to the determination of the best combination of the type

$$L_1 \lambda_1 + L_2 \lambda_2 + \dots + L_n \lambda_n$$

where $\lambda_1, \lambda_2, \dots, \lambda_n$ satisfy conditions (49) and we will indeed assume this best combination as U. To this end let us find the probable boundaries for the error (51).

Denoting

$$k = \sum [\varepsilon_{1(i)}]^2 p_i \tag{53}$$

and supposing that all the observations are of an equally high quality, we shall find, as we did before, that these boundaries are

$$-u\sqrt{2k(\lambda_1^2 + \lambda_2^2 + \dots + \lambda_n^2)}, \quad u\sqrt{2k(\lambda_1^2 + \lambda_2^2 + \dots + \lambda_n^2)}$$

so that everything is reduced to the determination of the minimum of the expression

$$\lambda_1^2 + \lambda_2^2 + \dots + \lambda_n^2 \tag{xliii}$$

under the conditions (48). We have the equations

$$\lambda_1 \, d\lambda_1 + \lambda_2 \, d\lambda_2 + \dots + \lambda_n \, d\lambda_n = 0, \quad \alpha_1 d\lambda_1 + \alpha_2 \, d\lambda_2 + \dots + \alpha_n \, d\lambda_n = 0,$$
$$\beta_1 \, d\lambda_1 + \beta_2 \, d\lambda_2 + \dots + \beta_n \, d\lambda_n = 0.$$

Multiplying the second of these by $-\rho$ and the third one by $-\sigma$; adding up the results obtained with the first equation; and equating the coefficients of $d\lambda_1, d\lambda_2, \dots, d\lambda_n$ to zero, we get equations of the type

$$\lambda_i = \alpha_i \rho + \beta_i \sigma, \quad i = 1, 2, \dots, n. \tag{54}$$

For determining ρ and σ we will have on the strength of (49) the equations

$$\alpha_1(\alpha_1 \rho + \beta_1 \sigma) + \alpha_2(\alpha_2 \rho + \beta_2 \sigma) + \ldots + \alpha_n(\alpha_n \rho + \beta_n \sigma) = 1, \quad \textbf{(55a)}$$

$$\beta_1(\alpha_1 \rho + \beta_1 \sigma) + \beta_2(\alpha_2 \rho + \beta_2 \sigma) + \ldots + \beta_n(\alpha_n \rho + \beta_n \sigma) = 0, \quad \textbf{(55b)}$$

or, in another form,

$$(\alpha_1{}^2 + \alpha_2{}^2 + \ldots + \alpha_n{}^2)\,\rho + (\alpha_1\beta_1 + \alpha_2\beta_2 + \ldots + \alpha_n\beta_n)\,\sigma = 1, \quad \textbf{(56a)}$$

$$(\alpha_1\beta_1 + \alpha_2\beta_2 + \ldots + \alpha_n\beta_n)\,\rho + (\beta_1{}^2 + \beta_2{}^2 + \ldots + \beta_n{}^2)\,\sigma = 0. \quad \textbf{(56b)}$$

Having determined ρ and σ from these equations, we will find $\lambda_1, \lambda_2, \ldots, \lambda_n$ from equations (54). Consequently, we will also find both U and the probable boundaries of the error ε if only we will know the magnitude k whose determination is explained below.

We shall now show that the thus found value of U is identical with the one obtained from the condition of minimum of the sum of the squares of the errors. We have

$$\varepsilon_1 = L_1 - \alpha_1 U - \beta_1 V, \quad \varepsilon_2 = L_2 - \alpha_2 U - \beta_2 V, \ldots,$$
$$\varepsilon_n = L_n - \alpha_n U - \beta_n V.$$

If we assume here that U and V are their real values, then $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ will be constant magnitudes. We will however consider them as variable quantities assigning to U and V not their real values, unknown to us, but all possible variable values; and, among these, we will search for such that provide the minimum of

$$\varepsilon_1{}^2 + \varepsilon_2{}^2 + \ldots + \varepsilon_n{}^2. \quad \text{(xliv)}$$

We have

$$\varepsilon_1{}^2 + \varepsilon_2{}^2 + \ldots + \varepsilon_n{}^2 = (L_1 - \alpha_1 U - \beta_1 V)^2 + (L_2 - \alpha_2 U - \beta_2 V)^2 + \ldots + (L_n - \alpha_n U - \beta_n V)^2.$$

The condition of minimum of this sum leads to the equations

$$\alpha_1(L_1 - \alpha_1 U - \beta_1 V) + \alpha_2(L_2 - \alpha_2 U - \beta_2 V) + \ldots + \alpha_n(L_n - \alpha_n U - \beta_n V) = 0,$$

$$\beta_1(L_1 - \alpha_1 U - \beta_1 V) + \beta_2(L_2 - \alpha_2 U - \beta_2 V) + \ldots + \beta_n(L_n - \alpha_n U - \beta_n V) = 0$$

which are reduced to the form

$$(\alpha_1{}^2 + \alpha_2{}^2 + \ldots + \alpha_n{}^2)U + (\alpha_1\beta_1 + \alpha_2\beta_2 + \ldots + \alpha_n\beta_n)V = \alpha_1 L_1 + \alpha_2 L_2 + \ldots + \alpha_n L_n,$$

$$(\alpha_1\beta_1 + \alpha_2\beta_2 + \ldots + \alpha_n\beta_n)U + (\beta_1{}^2 + \beta_2{}^2 + \ldots + \beta_n{}^2)V = \beta_1 L_1 + \beta_2 L_2 + \ldots + \beta_n L_n.$$

Multiplying the first of these equations by ρ, the second one by σ, and adding up the results we get

$$U\,[(\alpha_1{}^2 + \alpha_2{}^2 + \ldots + \alpha_n{}^2)\,\rho + (\alpha_1\beta_1 + \alpha_2\beta_2 + \ldots + \alpha_n\beta_n)\,\sigma] +$$
$$V\,[(\alpha_1\beta_1 + \alpha_2\beta_2 + \ldots + \alpha_n\beta_n)\,\rho + (\beta_1{}^2 + \beta_2{}^2 + \ldots + \beta_n{}^2)\,\sigma] =$$
$$L_1(\alpha_1\rho + \beta_1\sigma) + L_2(\alpha_2\rho + \beta_2\sigma) + \ldots + L_n(\alpha_n\rho + \beta_n\sigma).$$

Supposing now that ρ and σ satisfy the equations (56) we will find that

$$U = L_1(\alpha_1\rho + \beta_1\sigma) + L_2(\alpha_2\rho + \beta_2\sigma) + \ldots + L_n(\alpha_n\rho + \beta_n\sigma)$$

which coincides with the expression (50) if the factors λ_1, λ_2, ..., λ_n have the values (54) as found above. And so, U and V determined by the best linear combination at the same time make the sum of the squares of the errors minimal.

3.4.8. Let us now consider the general case. Suppose that it is required to determine from observations quantities U, V, W, ... whose number we will assume to be arbitrary. Suppose that we have the equations

$$\alpha_1 U + \beta_1 V + \gamma_1 W + ... = L_1 - \varepsilon_1, \quad \alpha_2 U + \beta_2 V + \gamma_2 W + ... = L_2 - \varepsilon_2,$$
$$..., \alpha_n U + \beta_n V + \gamma_n W + ... = L_n - \varepsilon_n, \tag{57}$$

whose right sides are the directly observed magnitudes L_1, L_2, ..., L_n; ε_1, ε_2, ..., ε_n are their errors whereas α_1, α_2, ..., α_n, β_1, β_2, ..., β_n, γ_1, γ_2, ..., γ_n, ... are given numbers not depending on the observations and not exposed to error. Let us take a number of factors

$$\lambda_1, \ \lambda_2, \ ..., \ \lambda_n \tag{xlv}$$

and subject them to conditions

$$\alpha_1\lambda_1 + \alpha_2\lambda_2 + ... + \alpha_n\lambda_n = 1, \quad \beta_1\lambda_1 + \beta_2\lambda_2 + ... + \beta_n\lambda_n = 0,$$
$$\gamma_1\lambda_1 + \gamma_2\lambda_2 + ... + \gamma_n\lambda_n = 0, \ ... \tag{58}$$

We multiply the equations (57) by λ_1, λ_2, ... respectively; adding up the results obtained we get, in virtue of these equations,

$$U = L_1\lambda_1 + L_2\lambda_2 + ... + L_n\lambda_n - (\varepsilon_1\lambda_1 + \varepsilon_2\lambda_2 + ... + \varepsilon_n\lambda_n),$$

an equation which we will write as

$$U = \sum L_i\lambda_i - \sum \varepsilon_i\lambda_i.$$

It is seen therefore that, when assuming the expression

$$L_1\lambda_1 + L_2\lambda_2 + ... + L_n\lambda_n = \sum L_i\lambda_i \tag{59}$$

as U, we make an error equal to

$$\varepsilon = \varepsilon_1\lambda_1 + \varepsilon_2\lambda_2 + ... + \varepsilon_n\lambda_n = \sum \varepsilon_i\lambda_i. \tag{60}$$

In order to find probable boundaries of this error we assume, as we did before, that all the n observations are of an equally high quality and {cf. (53)} denoting

$$k = \sum [\varepsilon_{1(i)}]^2 p_i \tag{61}$$

we obtain the probable boundaries

$$-u\sqrt{2k(\lambda_1^2 + \lambda_2^2 + ... + \lambda_n^2)}, \quad u\sqrt{2k(\lambda_1^2 + \lambda_2^2 + ... + \lambda_n^2)} \tag{62}$$

The most probable combination for U will be that for which the sum (xliii) will be minimal. In virtue of the conditions (58) we have for determining (xlv) the equations

$$\lambda_1\, d\lambda_1 + \lambda_2\, d\lambda_2 + ... + \lambda_n\, d\lambda_n = 0, \quad \alpha_1 d\lambda_1 + \alpha_2\, d\lambda_2 + ... + \alpha_n\, d\lambda_n = 0,$$
$$\beta_1\, d\lambda_1 + \beta_2\, d\lambda_2 + ... + \beta_n\, d\lambda_n = 0, \quad \gamma_1 d\lambda_1 + \gamma_2\, d\lambda_2 + ... + \gamma_n\, d\lambda_n = 0, \ ...$$
(63)

from which we find by the known method the expressions of the type

$$\lambda_i = \alpha_i \rho + \beta_i \sigma + \gamma_i \tau + ..., \quad i = 1, 2, ..., n \tag{64}$$

with magnitudes ρ, σ, τ, ... being determined from the equations

$$\sum \alpha_i(\alpha_i\rho + \beta_i\sigma + \gamma_i\tau + ...) = 1, \sum \beta_i(\alpha_i\rho + \beta_i\sigma + \gamma_i\tau + ...) = 0,$$
$$\sum \gamma_1(\alpha_1\rho + \beta_1\sigma + \gamma_i\tau + ...) = 0, ... \tag{65}$$

These equations might be represented in the following form

$$\rho\sum \alpha_i^2 + \sigma\sum \alpha_i\beta_i + \tau\sum \alpha_i\gamma_i + ... = 1,$$
$$\rho\sum \alpha_i\beta_i + \sigma\sum \beta_i^2 + \tau\sum \beta_i\gamma_i + ... = 0, \tag{66}$$
$$\rho\sum \alpha_i\gamma_i + \sigma\sum \gamma_i\beta_i + \tau\sum \gamma_i^2 + ... = 0, ...$$

Their number is equal to the number of the magnitudes ρ, σ, τ, ... We thus find that

$$U = \sum L_i(\alpha_i\rho + \beta_i\sigma + \gamma_i\tau + ...) \tag{67}$$

with ρ, σ, τ, ... being determined from the equations (66).

We shall now show that the expression (67) is identical with that which is obtained for U under the condition of the minimal sum of the squares of the errors. We have

$$\sum \varepsilon_i^2 = \sum (L_i - \alpha_i U - \beta_i V - \gamma_i W - ...)^2.$$

The search for the minimum of this sum leads to the solution of the equations

$$\sum \alpha_i(L_i - \alpha_i U - \beta_i V - \gamma_i W - ...) = 0,$$
$$\sum \beta_i(L_i - \alpha_i U - \beta_i V - \gamma_i W - ...) = 0,$$
$$\sum \gamma_i(L_i - \alpha_i U - \beta_i V - \gamma_i W - ...) = 0, ...$$

or, otherwise, of the equations

$$U\sum \alpha_i^2 + V\sum \alpha_i\beta_i + W\sum \alpha_i\gamma_i + ... = \sum \alpha_i L_i,$$
$$U\sum \alpha_i\beta_i + V\sum \beta_i^2 + W\sum \beta_i\gamma_i + ... = \sum \beta_i L_i, \tag{68}$$
$$U\sum \alpha_i\gamma_i + V\sum \beta_i\gamma_i + W\sum \gamma_i^2 + ... = \sum \gamma_i L_i, ...$$

Multiplying the first of these by ρ; the second one, by σ; the third equation, by τ; etc, and adding up the results, we obtain on the strength of the equations (66), which these quantities satisfy,

$$U = \rho\sum \alpha_i L_i + \sigma\sum \beta_i L_i + \tau\sum \gamma_i L_i + ..., \tag{69}$$

i.e., the expression (67), and the theorem is proved.

This is indeed the essence of the method of least squares. We shall show further on how to determine the magnitude k, but now {but first} we offer an example.

Suppose that it is required to determine three quantities U, V and W, and that we made four observations which provided quantities 14, 10, 9, 17 respectively for

$3U + 5V + 7W, \ 4U + 11V - 2W,$
$5U + 13V - 7W, \ 4U - 11V - 13W.$

We search for $U, \ V, \ W$ under the condition that the expression

$(3U + 5V + 7W - 14)^2 + (4U + 11V - 2W - 10)^2 +$
$(5U + 13V - 7W - 9)^2 + (4U - 11V - 13W - 17)^2$

is minimal. It provides equations

$66U + 80V - 74W = 195, \ 80U + 436V + 65W = 110,$
$74U + 65V + 271W = -206.$

From these we will indeed find $U, \ V, \ W$ but we shall not dwell on the numerical calculations.
   Note that the approach to the exposition of the method of least squares as shown in the preceding sections assumes that the number n of observations is very large because only under this condition it is possible to apply the formula (43).

3.4.9. It only remains to show how to determine the magnitude k. Assuming that all the observations deserve the same confidence, we have

$$k = \sum \ [\varepsilon_{1(i)}]^2 \, p_i = \sum \ [\varepsilon_{2(i)}]^2 \, q_i = \dots$$

Let us apply formula (43) to the case under consideration and assume the magnitude $[\varepsilon_{1(i)}]^2$ as x; the magnitude $[\varepsilon_{2(i)}]^2$ as y; etc. Then (still supposing that the number of observations is n) this formula will represent the probability that the sum (xliv) is contained within the boundaries
$A + u\sqrt{2B}$ and $A - u\sqrt{2B}$.

In our case
$A = a + b + c + \dots = \sum \ [\varepsilon_{1(i)}]^2 \, p_i + \sum \ [\varepsilon_{2(i)}]^2 \, q_i + \dots = nk,$
$B = (a_1 - a^2) + (b_1 - b^2) + \dots =$
$\quad \sum \ [\varepsilon_{1(i)}]^4 \, p_i - k^2 + \sum \ [\varepsilon_{2(i)}]^4 \, q_i - k^2 + \dots$

Therefore, denoting

$$\sum \ [\varepsilon_{1(i)}]^4 \, p_i = \sum \ [\varepsilon_{2(i)}]^4 \, q_i = \ \dots \ = k_1,$$

we obtain $B = n(k_1 - k^2)$.
   And so, formula (43) will determine the probability of the existence of the inequalities

$$nk - u\sqrt{2n(k_1 - k^2)} \ < \ \varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_n^2 \ < \ nk + u\sqrt{2n(k_1 - k^2)}$$

or

$k - u\sqrt{(2/n)(k_1 - k^2)} \ < \ (1/n)[\varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_n^2] \ <$
$k + u\sqrt{(2/n)(k_1 - k^2)}.$

It is seen now that

$$k = \lim \ \{(1/n)[\varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_n^2]\}_{\ n = \infty}$$

so that, for a very large number of observations, we may assume that

$$k = (1/n)[\varepsilon_1{}^2 + \varepsilon_2{}^2 + \ldots + \varepsilon_n{}^2]. \qquad\qquad \textbf{(70)}$$

As to the magnitudes of the errors of observations ε_1, ε_2, ..., ε_n which are here included, they may be gotten from a series of observations of known quantities by comparing their real and observed values. However, it is not always possible to carry out such special observations solely intended for the determination of k. Therefore, this magnitude is usually derived from the same observations from which we determine the unknowns U, V, W,... In this case, instead of the real errors unknown to us we take those which turn out after the determination of U, V, W,... by the method of least squares when comparing the results of calculation and observation.

In other words, considering ε_1, ε_2, ..., ε_n as variable quantities we determine them in this case under the condition that the sum (xliv) is minimal and assume their thus obtained values as errors included in the expression for k. A justification of this procedure is that, if we denote the unknown to us real errors by ε_1, ε_2, ..., ε_n, and the values {of errors} obtained by the method of least squares by $\varepsilon_1 + \eta_1$, $\varepsilon_2 + \eta_2$, ..., $\varepsilon_n + \eta_n$, then η_1, η_2, ..., η_n will be very small as compared with ε_1, ε_2, ..., ε_n; or at least we ought to consider them as such because the method provides the most probable result and we may therefore neglect the magnitudes η_1, η_2, ..., η_n.

The magnitude k determining the merit of observations is found in this way. The less it is, the better are the observations. The magnitude $1/k$ which is called the *weight* of the observations may therefore serve as the measure of their merit. It will be said below how this term is justified, now, however, we note that recently some authors have begun to assume the expression

$$[1/(n - l)] [\varepsilon_1{}^2 + \varepsilon_2{}^2 + \ldots + \varepsilon_n{}^2] \qquad\qquad (71)$$

as k[1]. Here, n is the number of observations and l, the number of quantities determined, $- U$, V, W, ... It was stated, in favor of this formula, that when the number of observations is equal to the number of quantities determined, we may exactly satisfy the conditional equations [2] so that, issuing from them, it will occur that $\varepsilon_1 = \varepsilon_2 = \ldots = \varepsilon_n = 0$ (here, the ε's are the calculated and not the real errors) and the previous formula would have provided $k = 0$. This cannot be admitted because it would have indicated that the observations were absolutely precise whereas formula (71) furnishes here an indefinite expression $0/0$ for k.

But the point is that when $n = l$ we cannot apply those conclusions on which the derivation of k was founded because we assumed that n was a very large number whereas the number of unknowns l is always supposed to be restricted so that we ought to reject the case in which $n = l$. If, however, we assume that n is considerably larger than l, it will not matter which formula is being used for determining k because we may neglect the terms beginning with l/n^2 in the expression

$$[1/(n - l)] = (1/n) + l/n^2 + (l^3/n^3) + \ldots$$

Note that the Gauss method based on the law of hypotheses does not demand that n be certainly very large. When assuming it as the foundation {of the method of least squares} we may also consider the case in which $n = l$. Then it is therefore more opportune to determine k by the formula (71). We saw, however, that this method is not really reliable, and it is preferable to determine k by the formula (70).

Note 1. {Indeed, *some authors*,– beginning with Gauss!}
Note 2. {Chebyshev did not use this term in the exposition above; he apparently bore in mind equations (50) which are, however, called *observational*. Conditional equations appear in another version of treating observations (still by the method of least squares or otherwise).}

3.4.10. In concluding, we shall show the influence of k on the quantity determined while assuming for the sake of greater generality that k is different for different observations.
We had the equations

$$U = L_1 - \varepsilon_1, \quad U = L_2 - \varepsilon_2, \ldots, \quad U = L_n - \varepsilon_n.$$

Assuming that U is defined as

$$U = L_1\lambda_1 + L_2\lambda_2 + \ldots + L_n\lambda_n$$

under the condition

$$\lambda_1 + \lambda_2 + \ldots + \lambda_n = 1 \qquad\qquad\qquad (72)$$

we make an error equal to

$$\varepsilon = \varepsilon_1\lambda_1 + \varepsilon_2\lambda_2 + \ldots + \varepsilon_n\lambda_n$$

whose probable boundaries are, in the general case, the magnitudes

$$u\sqrt{2\sum[(\varepsilon_{1(i)})^2 p_i]\lambda_1^{\,2} + 2\sum[(\varepsilon_{2(i)})^2 q_i]\lambda_2^{\,2} + \ldots}\,,$$

$$-u\sqrt{2\sum[(\varepsilon_{1(i)})^2 p_i]\lambda_1^{\,2} + 2\sum[(\varepsilon_{2(i)})^2 q_i]\lambda_2^{\,2} + \ldots}\,.$$

Denoting

$$\sum [\varepsilon_{1(i)}]^2 p_i = k_1, \quad \sum [\varepsilon_{2(i)}]^2 q_i = k_2, \ldots$$

we shall impart the following form to these boundaries:

$$u\sqrt{2(k_1\lambda_1^{\,2} + k_2\lambda_2^{\,2} + \ldots + k_n\lambda_n^{\,2})}, \quad -u\sqrt{2(k_1\lambda_1^{\,2} + k_2\lambda_2^{\,2} + \ldots + k_n\lambda_n^{\,2})}.$$

The problem is thus reduced to the determination of λ_1, λ_2, ..., λ_n from the condition that the expression under the signs of the radicals is minimal and we obtain the equations

$$k_1\lambda_1\, d\lambda_1 + k_2\lambda_2\, d\lambda_2 + \ldots + k_n\lambda_n\, d\lambda_n = 0, \ d\lambda_1 + d\lambda_2 + \ldots + d\lambda_n = 0.$$

Together with the equation (72) they provide

$$k_1\lambda_1 = \rho, \ k_2\lambda_2 = \rho, \ldots, \ k_n\lambda_n = \rho, \ (\rho/k_1) + (\rho/k_2) + \ldots + (\rho/k_n) = 1$$

so that

$$\lambda_i = (1/k_i)/[(1/k_1) + (1/k_2) + \ldots + (1/k_n)]$$

and

$$U = \frac{(1/k_1)L_1 + (1/k_2)L_2 + \ldots + (1/k_n)L_n}{(1/k_1) + (1/k_2) + \ldots + (1/k_n)}.$$

This formula is remarkably analogous with the one serving for the determination of {one of the} coordinate{s} of the center of gravity if the magnitudes $(1/k_1)$, $(1/k_2)$, ... are likened to the weights of material points, and L_1, L_2, ..., to the values of their coordinate. This is the very reason why $(1/k)$ is called the weight of the observation.

It is not difficult to convince ourselves that the obtained expression for U is identical with the one arrived at under the condition of the minimum of the sum

$$(\varepsilon_1^{\,2}/k_1) + (\varepsilon_2^{\,2}/k_2) + \ldots + (\varepsilon_n^{\,2}/k_n) =$$
$$(1/k_1)(U - L_1)^2 + (1/k_2)(U - L_2)^2 + \ldots + (1/k_n)(U - L_n)^2$$

because the value of U making this sum minimal is determined from the equation

$$(1/k_1)(U - L_1) + (1/k_2)(U - L_2) + \ldots + (1/k_n)(U - L_n) = 0.$$

It can be shown that the same will happen as well in the general case in which the quantities $U,\ V,\ W,\ \ldots$ are determined by the observations, i.e., that here also such magnitudes are obtained as make the sum

$$(1/k_1)\varepsilon_1^2\ +\ (1/k_2)\varepsilon_2^2\ +\ \ldots\ +\ (1/k_n)\varepsilon_n^2$$

a minimum. Thus, the most probable magnitudes are always gotten from the condition that the sum of the squares of the errors multiplied by the weights of the corresponding observations is minimal. This represents a generalization of the method of least squares.

The approach to the method of least squares shown in the preceding sections is due to Laplace.

Index of Names

Covering only the main text; the numbers indicate sections