Det Kgl. Danske Videnskabernes Selskab.
Mathematisk-fysiske Meddelelser. **XVIII,** 3.

ON THE DISTRIBUTION OF RELATIVE ERRORS FROM A NORMAL POPULATION OF ERRORS

A DISCUSSION OF SOME PROBLEMS IN THE THEORY OF ERRORS

BY

NIELS ARLEY

KØBENHAVN
EJNAR MUNKSGAARD
1940

Printed in Denmark.
Bianco Lunos Bogtrykkeri A/S.

INTRODUCTION

The present paper has arisen from the author's considerations in connection with the writing of a textbook of probability and the theory of errors[1].

In the practical applications of the theory of errors, one meets with two important problems. The first problem is how to test whether or not a given sample of measurements has come from a normal population. Such a test is necessary if one wants to apply with reasonable safety the usual theory of errors based on the assumption of the validity of the normal distribution law. If the sample is large, we have only to draw the frequency, or the total frequency, polygon, and compare it with the corresponding normal frequency, or total frequency, curve, possibly by means of the χ^2-test of goodness of fit. A specially elegant and efficient method of comparison is the method of probits[2], whereby the total frequency curve is transformed into a straight line. However, in practice, these methods can seldom be applied, since the samples are as a rule too small containing only few measurements—of the order 10 or less. *The question is, therefore, how to test for normality small samples consisting perhaps of only 4 measurements?*

The next problem is how to test whether or not an unusually large error has to be rejected as being due to

[1] N. Arley and K. R. Buch: Calculus of Probability with Applications on Statistics, Theory of Errors and Theory of Adjustment. Copenhagen 1940. (In Danish).

[2] Cf. e. g. R. A. Fisher and F. Yates: Statistical Tables. London 1938.

some false measurements. It is, of course, very important that the false observations should be rejected, since one false observation can completely vitiate the results of the measurements. This problem is, however, a far more delicate one than the first problem. On the one hand, assuming the population of errors considered to be normal means, in fact, allowing arbitrarily large errors to occur, though with extremely small probabilities. The only safe and legitimate procedure in rejecting unusually outlying observations is, therefore, *to reject them during the observations themselves,* because some peculiarities arouse suspicions as to the constancy of the conditions of the measurements or the like. On the other hand, if e. g. five measurements of the starting velocity of a projectile gave the results 398.6, 442.1, 442.3, 441.8 and 442.4 m/sec., nobody would hesitate, even without any knowledge of the method of measurement used, to suspect the first figure to be obtained under *different* conditions from the other four figures. A closer investigation of the conditions would, therefore, be advisable before the figure could be admitted as true. In fact, it does turn out that the first shot gives a smaller velocity than the following ones, because the gun is heated up during this first shot ("Anwärmeschuss"). In practice one would, consequently, be inclined to cut off artificially the tails of the distribution curve by discarding errors exceeding certain limits. *The question is, therefore, how to obtain such limits?* We already stress here, however, that whatever criterion for false observations we may establish has to be applied with the utmost critique and caution. Otherwise we shall run the risk of discarding many true observations and obtaining a false impression of the accuracy of the measurements.

It is the purpose of this paper to discuss these two problems by deducing the distribution of the relative errors. Thus, we first obtain a method for testing even very small samples for normality and, next an objective criterion for false observations giving us complete control of the risk we run of discarding true, but fortuitously large errors.

I. Deduction of the distribution law for direct and equally good observations.

§ 1. Let x be a normally distributed statistical variable, i. e. with the frequency function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\xi)^2}{2\sigma^2}\right] \tag{1.1}$$

where $f(x)\,dx$ is the probability of finding x in the range between x and $x+dx$, ξ the mean value, and σ the standard deviation[1]. If $x_1, x_2, \cdots, x_n$ is a sample consisting of n ($\geqq 2$) independent and equally good observations of x, then, as is well-known, their average value

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{1.2}$$

is normally distributed with mean value ξ and standard deviation $\frac{\sigma}{\sqrt{n}}$, $\bar{x}$ being the best estimate of the parameter ξ. The n quantities

[1] In this paper the mean value of a statistical variable with frequency function $f(x)$ is denoted by

$$m\{x\} = \int_{-\infty}^{\infty} x\,f(x)\,dx$$

and the standard deviation by

$$\sigma^2\{x\} = m\left\{(x - m\{x\})^2\right\} = \int_{-\infty}^{\infty} (x - m\{x\})^2 f(x)\,dx.$$

$$\boldsymbol{v}_i = \boldsymbol{x}_i - \overline{\boldsymbol{x}} = \sum_{k=1}^{n} \left(\delta_{ik} - \frac{1}{n}\right) \boldsymbol{x}_k\,, \quad i = 1, 2, \cdots, n \tag{1.3}$$

which are called the *residuals* in contrast to the *true errors* $\varepsilon_i = \boldsymbol{x}_i - \xi$, are also, being sums of normally distributed variables, themselves normally distributed with the mean value

$$m\{v_i\} = m\{x_i\} - m\{\overline{x}\} = 0, \quad i = 1, 2, \cdots, n \tag{1.4}$$

and the standard deviation

$$\sigma\{v_i\} = \left[\sum_{k=1}^{n} \left(\delta_{ik} - \frac{1}{n}\right)^2 \sigma^2\right]^{1/2} = \sqrt{\frac{n-1}{n}}\,\sigma, \quad i = 1, 2, \cdots, n. \tag{1.5}$$

The n quantities

$$\rho_i = \frac{\boldsymbol{v}_i}{\sigma\{v_i\}} = \frac{\boldsymbol{x}_i - \overline{\boldsymbol{x}}}{\sqrt{\frac{n-1}{n}}\,\sigma}, \quad i = 1, 2, \cdots, n \tag{1.6}$$

are consequently normally distributed with mean value 0 and standard deviation 1. The probability, $P(\rho)$, of $|\rho| \geqq \rho$ is then given by

$$P(\rho) = 2\int_{\rho}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt. \tag{1.7}$$

It is conventional to regard an observed value x_i as false if the corresponding ρ_i is greater than the value ρ corresponding to some small arbitrarily chosen probability P. If we choose for this probability e. g. the value 0.001 we must reject x_i if $|\rho_i| \geqq \rho\,(0.001) = 3.29$ (cf. Table I with $f = n - 2 = \infty$).

Now the exact value of the parameter σ is not known, but can only be estimated from observations. In the usual case where the n observations $x_1, x_2, \cdots, x_n$ constitute all our information, the mean square error

$$s = \frac{q}{\sqrt{n-1}} \tag{1.8}$$

where

$$q = +\left[\sum_{i=1}^{n} v_i^2\right]^{1/2} = +\left[\sum_{i=1}^{n} (x_i - \bar{x})^2\right]^{1/2} \tag{1.9}$$

is the best estimate of σ. Substituting this value for σ in (1.6) we obtain the n quantities

$$\boldsymbol{r}_i = \frac{\boldsymbol{x}_i - \bar{\boldsymbol{x}}}{\sqrt{\frac{n-1}{n}}\,\boldsymbol{s}} = \frac{\boldsymbol{v}_i}{\boldsymbol{q}}\sqrt{n}, \quad i = 1, 2, \cdots, n \tag{1.10}$$

which are called the *relative errors*. From (1.9) and (1.10) it follows that

$$\sum_{i=1}^{n} r_i = 0 \tag{1.11}$$

and

$$\sum_{i=1}^{n} r_i^2 = n. \tag{1.12}$$

From (1.11) and (1.12) it follows that

$$|r_i| < \sqrt{n}, \quad i = 1, 2, \cdots, n \tag{1.13}$$

which shows that *the relative errors are not normally distributed.* Taking the mean value on both sides of (1.11) and (1.12) we have, because of the symmetry between $r_1, r_2, \cdots, r_n$,

$$m\{r_i\} = 0 \tag{1.14}$$

and $\qquad i = 1, 2, \cdots, n$

$$m\{r_i^2\} = \sigma^2\{r_i\} = 1 \tag{1.15}$$

Thus, each relative error has mean value 0 and standard deviation 1. Squaring and taking the mean value on both sides of eq. (1.11) gives us

$$\sum_{i=1}^{n}\sum_{j=1}^{n} m\{r_i r_j\} = \sum_{i=1}^{n} m\{r_i^2\} + \sum_{i \neq j}\sum m\{r_i r_j\} =$$
$$= n + n(n-1)\, m\{r_i r_j\} = 0. \qquad (i \neq j).$$

Thus, the correlation coefficient between each two of the relative errors is equal to

$$\rho\{r_i, r_j\} = \frac{m\{r_i r_j\}}{\sigma\{r_i\}\,\sigma\{r_j\}} = -\frac{1}{n-1}. \qquad (i \neq j). \qquad (1.16)$$

We shall now deduce the distribution of these relative errors.

§ 2. Since $\boldsymbol{v}$[1] and $\boldsymbol{q}$ are correlated their correlation function is not simply the product of the frequency functions of $\boldsymbol{v}$ and $\boldsymbol{q}$. In order to deduce the frequency function of $\boldsymbol{r}$ we then first write down the probability $S(x_1, x_2, \cdots, x_n)\, dx_1\, dx_2 \cdots dx_n$ of the sample $x_1, x_2, \cdots, x_n$ which from (1.1) is given by

$$\left.\begin{aligned} & S(x_1, x_2, \cdots, x_n)\, dx_1\, dx_2 \cdots dx_n = \\ & = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \xi)^2\right] dx_1\, dx_2 \cdots dx_n \end{aligned}\right\} \quad (2.1)$$

We next introduce in (2.1) instead of $x_1, x_2, \cdots, x_n$ the $n+2$ new variables

$\bar{x}$ (defined in (1.2))
q (— - (1.9))

1) We shall in the following drop the index i.

and $$u_i, \quad i = 1, 2, \cdots, n$$

defined by the equations

$$x_i = \bar{x} + q u_i, \quad i = 1, 2, \cdots, n. \tag{2.2}$$

From (1.2) and (1.9) it follows that

$$\sum_{i=1}^{n} u_i = 0 \tag{2.3}$$

and

$$\sum_{i=1}^{n} u_i^2 = 1 \quad \text{i. e. } |u_i| < 1, \ i = 1, 2, \cdots, n \tag{2.4}$$

so that only $n-2$ of the u-variables are free, the two remaining being functions of the $n-2$ others. For $n = 2$ u_1 and u_2 are consequently constants $\left(= \pm \frac{1}{\sqrt{2}}\right)$ and we shall therefore assume that $n \geq 3$. Further we shall choose $u_1, u_2, \cdots, u_{n-2}$ as the free variables. Using the identity

$$\left.\begin{aligned} \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (x_i - \xi)^2 = \sum_{i=1}^{n} v_i^2 + n(\bar{x} - \xi)^2 = \\ = q^2 + n(\bar{x} - \xi)^2 \end{aligned}\right\} \tag{2.5}$$

we then obtain from (2.1) that the probability $S(x_1, x_2, \cdots, x_n) \times$ $\times\, dx_1\, dx_2 \cdots dx_n$, expressed in the new variables, of the sample $x_1, x_2, \cdots, x_n$ is given by

$$\left.\begin{aligned} & S(x_1, x_2, \cdots, x_n)\, dx_1\, dx_2 \cdots dx_n = \\ & = S(\bar{x}, q, u_1, u_2, \cdots, u_{n-2})\, d\bar{x}\, dq\, du_1\, du_2 \cdots du_{n-2} = \\ & = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2}\left(q^2 + n(\bar{x} - \xi)^2\right)\right] \times \\ & \times \left| \frac{\partial(x_1, x_2, \cdots, x_n)}{\partial(\bar{x}, q, u_1, u_2, \cdots, u_{n-2})} \right| d\bar{x}\, dq\, du_1\, du_2 \cdots du_{n-2}. \end{aligned}\right\} \tag{2.6}$$

Here the Jacobian functional determinant is given by

$$\left.\begin{aligned}\frac{\partial(x_1, x_2, \cdots, x_n)}{\partial(\bar{x}, q, u_1, u_2, \cdots, u_{n-2})} &= \begin{vmatrix} \frac{\partial x_1}{\partial \bar{x}} & \frac{\partial x_1}{\partial q} & \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} & \cdots & \frac{\partial x_1}{\partial u_{n-2}} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial x_n}{\partial \bar{x}} & \frac{\partial x_n}{\partial q} & \frac{\partial x_n}{\partial u_1} & \frac{\partial x_n}{\partial u_2} & \cdots & \frac{\partial x_n}{\partial u_{n-2}} \end{vmatrix} = \\ \begin{vmatrix} 1 & u_1 & q & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & u_{n-2} & 0 & 0 & \cdots & q \\ 1 & u_{n-1} & q\frac{\partial u_{n-1}}{\partial u_1} & q\frac{\partial u_{n-1}}{\partial u_2} & \cdots & q\frac{\partial u_{n-1}}{\partial u_{n-2}} \\ 1 & u_n & q\frac{\partial u_n}{\partial u_1} & q\frac{\partial u_n}{\partial u_2} & \cdots & q\frac{\partial u_n}{\partial u_{n-2}} \end{vmatrix} &= q^{n-2} D(u_1, u_2, \cdots, u_{n-2}; n).\end{aligned}\right\} \tag{2.7}$$

Introducing this in (2.6) we finally obtain

$$\left.\begin{aligned} S(\bar{x}, q, u_1, u_2, \cdots, u_{n-2})\, d\bar{x}\, dq\, du_1\, du_2 \cdots du_{n-2} = \\ = \left\{\frac{\sqrt{n}}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{n(\bar{x}-\xi)^2}{2\sigma^2}\right] d\bar{x}\right\} \left\{\frac{1}{\left(\frac{n-3}{2}\right)!\, 2^{\frac{n-3}{2}}} \left(\frac{q}{\sigma}\right)^{n-2} \exp\left[-\frac{q^2}{2\sigma^2}\right] \frac{dq}{\sigma}\right\} \times \\ \times \left\{\pi^{-\frac{n-1}{2}}\, n^{-\frac{1}{2}} \left(\frac{n-3}{2}\right)!\, D(u_1, u_2, \cdots, u_{n-2}; n)\, du_1\, du_2 \cdots du_{n-2}\right\} \end{aligned}\right\} \tag{2.8}$$

$$-\infty < \bar{x} < \infty$$

$$0 \leq q < \infty$$

$$-1 < u_i < 1, \qquad i = 1, 2, \cdots, n-2$$

where the coefficients have been so chosen that the integrals taken over all possible values of the variables give unity for each of the three factors[1]. (2.8) shows that $\bar{x}$

[1] We note that since q can take on both positive and negative values and since we have chosen q positive, the last factor has been taken with an extra factor 2.

and $\boldsymbol{q}$ are uncorrelated variables, and that the $\boldsymbol{u}_i$'s are uncorrelated to $\bar{\boldsymbol{x}}$ and $\boldsymbol{q}$. Further it is seen, that the last factor does not contain either the parameter ξ og σ, which shows that $\bar{x}$ and q are what FISHER calls "sufficient statistics"[1].

To obtain the frequency function $f(u)$ of one of the $\boldsymbol{u}_i$'s, say $\boldsymbol{u}_1$, we now have to integrate (2.8) over all possible values of $\bar{x}, q, u_2, u_3, \cdots, u_{n-2}$. Having deduced $f(u)$ the distribution of the relative errors is immediately given, since from (1.10) and (2.2)

$$\boldsymbol{r}_i = \boldsymbol{u}_i \sqrt{n}, \qquad i = 1, 2, \cdots, n. \tag{2.9}$$

The factors in (2.8) containing $\bar{x}$ and q integrate immediately to 1 and we next have to evaluate

$$\left.\begin{aligned} & f(u_1) = \\ = \underset{\text{over all possible values}}{\int du_2 \int du_3 \cdots \int du_{n-2}}\, \pi^{-\frac{n-1}{2}}\, n^{-\frac{1}{2}} & \left(\frac{n-3}{2}\right)!\; D(u_1, u_2, \cdots, u_{n-2}). \end{aligned}\right\} \tag{2.10}$$

§ 3. We first have, however, to work out the n-dimensional determinant D which from (2.7) is given by

$$D(u_1, u_2, \cdots, u_{n-2};\, n) = \begin{vmatrix} 1 & u_1 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & u_{n-2} & 0 & 0 & \cdots & 1 \\ 1 & u_{n-1} & \frac{\partial u_{n-1}}{\partial u_1} & \frac{\partial u_{n-1}}{\partial u_2} & \cdots & \frac{\partial u_{n-1}}{\partial u_{n-2}} \\ 1 & u_n & \frac{\partial u_n}{\partial u_1} & \frac{\partial u_n}{\partial u_2} & \cdots & \frac{\partial u_n}{\partial u_{n-2}} \end{vmatrix}. \tag{3.1}$$

[1] R. A. FISHER: Statistical Methods for Research Workers. Chap. 1.

If we here add to the first row all the $n-1$ other rows we obtain with the help of (2.3)

$$D=\begin{vmatrix} n & 0 & 0 & 0 & \cdots & 0 \\ 1 & u_2 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & u_{n-2} & 0 & 0 & \cdots & 1 \\ 1 & u_{n-1} & \frac{\partial u_{n-1}}{\partial u_1} & \frac{\partial u_{n-1}}{\partial u_2} & \cdots & \frac{\partial u_{n-1}}{\partial u_{n-2}} \\ 1 & u_n & \frac{\partial u_n}{\partial u_1} & \frac{\partial u_n}{\partial u_2} & \cdots & \frac{\partial u_n}{\partial u_{n-2}} \end{vmatrix} = n \begin{vmatrix} u_2 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ u_{n-2} & 0 & 0 & \cdots & 1 \\ u_{n-1} & \frac{\partial u_{n-1}}{\partial u_1} & \frac{\partial u_{n-1}}{\partial u_2} & \cdots & \frac{\partial u_{n-1}}{\partial u_{n-2}} \\ u_n & \frac{\partial u_n}{\partial u_1} & \frac{\partial u_n}{\partial u_2} & \cdots & \frac{\partial u_n}{\partial u_{n-2}} \end{vmatrix} \quad (3.2)$$

Solving the equations (2.3) and (2.4) for u_n and u_{n-1} we obtain

$$\left.\begin{matrix} u_n \\ u_{n-1} \end{matrix}\right\} = \frac{1}{2}\left(-\sum_{i=1}^{n-2} u_i \pm \left[2\left(1-\sum_{i=1}^{n-2} u_i^2\right)-\left(\sum_{i=1}^{n-2} u_i\right)^2\right]^{1/2}\right) \quad (3.3)$$

where

$$0 \leq 2\left(1-\sum_{i=1}^{n-2} u_i^2\right)-\left(\sum_{i=1}^{n-2} u_i\right)^2. \quad (3.4)$$

Obviously

$$u_n - u_{n-1} = \left[2\left(1-\sum_{i=1}^{n-2} u_i^2\right)-\left(\sum_{i=1}^{n-2} u_i\right)^2\right]^{1/2}. \quad (3.5)$$

Differentiating (3.3) with respect to u_i, $i = 1, 2, \cdots, n-2$, gives, using (3.5),

$$\left.\begin{aligned} \frac{\partial u_n}{\partial u_i} &= \frac{u_{n-1}-u_i}{u_n-u_{n-1}} \\ \frac{\partial u_{n-1}}{\partial u_i} &= \frac{u_i-u_n}{u_n-u_{n-1}} \end{aligned}\right. \qquad i = 1, 2, \cdots, n-2 \quad (3.6)$$

which introduced in (3.2) gives

$$D=\frac{n}{u_n-u_{n-1}}\begin{vmatrix} u_2 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ u_{n-2} & 0 & 0 & \cdots & 1 \\ u_{n-1} & (u_1-u_n) & \left(\dfrac{u_2-u_n}{u_n-u_{n-1}}\right) & \cdots & \left(\dfrac{u_{n-2}-u_n}{u_n-u_{n-1}}\right) \\ u_n & (u_{n-1}-u_1) & \left(\dfrac{u_{n-1}-u_2}{u_n-u_{n-1}}\right) & \cdots & \left(\dfrac{u_{n-1}-u_{n-2}}{u_n-u_{n-1}}\right) \end{vmatrix}. \tag{3.7}$$

If we here add the last row to the next last row, we obtain

$$D=\frac{n}{u_n-u_{n-1}}\begin{vmatrix} u_p & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ u_{n-2} & 0 & 0 & \cdots & 1 \\ (u_{n-1}+u_n) & (u_{n-1}-u_n) & -1 & \cdots & -1 \\ u_n & (u_{n-1}-u_1) & \left(\dfrac{u_{n-1}-u_p}{u_n-u_{n-1}}\right) & \cdots & \left(\dfrac{u_{n-1}-u_{n-2}}{u_n-u_{n-1}}\right) \end{vmatrix} \tag{3.8}$$

$$=\frac{n}{u_n-u_{n-1}}E_{n-p}, \quad \text{where } p=2.$$

Developing the $(n-p+1)$-dimensional determinant E_{n-p} according to the first row we obtain the recurrence formula

$$\left.\begin{aligned} E_{n-p} &= u_p\begin{vmatrix} 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \\ (u_{n-1}-u_n) & -1 & -1 & \cdots & -1 \\ (u_{n-1}-u_1) & \left(\dfrac{u_{n-1}-u_p}{u_n-u_{n-1}}\right) & \left(\dfrac{u_{n-1}-u_{p+1}}{u_n-u_{n-1}}\right) & \cdots & \left(\dfrac{u_{n-1}-u_{n-2}}{u_n-u_{n-1}}\right) \end{vmatrix} + \\ &+ \begin{vmatrix} u_{p+1} & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ u_{n-2} & 0 & 0 & \cdots & 1 \\ (u_{n-1}+u_n) & (u_{n-1}-u_n) & -1 & \cdots & -1 \\ u_n & (u_{n-1}-u_1) & \left(\dfrac{u_{n-1}-u_{p+1}}{u_n-u_{n-1}}\right) & \cdots & \left(\dfrac{u_{n-1}-u_{n-2}}{u_n-u_{n-1}}\right) \end{vmatrix} = \\ &= u_p(u_p-u_1)+E_{n-(p+1)}. \end{aligned}\right\} \tag{3.9}$$

Repeatingly using this we obtain, using (2.3) and (2.4)

$$E_{n-2} = u_2(u_2-u_1)+u_3(u_3-u_1)+\cdots+u_n(u_n-u_1) = 1 \qquad (3.10)$$

and thus from (3.5)

$$\left.\begin{aligned} &D(u_1, u_2, \cdots, u_{n-2}; n) = \\ &= \frac{n}{u_n-u_{n-1}} = \frac{n}{\sqrt{2}} \frac{1}{\left[\left(1-\sum_{i=1}^{n-2} u_i^2\right)-\frac{1}{2}\left(\sum_{i=1}^{n-2} u_i\right)^2\right]^{1/2}} \end{aligned}\right\} \qquad (3.11)$$

§ 4. Inserting (3.11) in (2.10) the integral becomes

$$\left.\begin{aligned} &f(u_1) = \\ &= \underset{\text{over all possible values}}{\int du_2 \int du_3 \cdots \int du_{n-2}} \frac{\pi^{-\frac{n-1}{2}} \sqrt{\frac{n}{2}}\left(\frac{n-3}{2}\right)!}{\left[\left(1-\sum_{i=1}^{n-2} u_i^2\right)-\frac{1}{2}\left(\sum_{i=1}^{n-2} u_i\right)^2\right]^{1/2}}. \end{aligned}\right\} \qquad (4.1)$$

We now have the identity, for each $p = 2, 3, \cdots, n-1$,

$$\left.\begin{aligned} a_p^2 &= \left(1-\sum_{i=1}^{n-p} u_i^2\right)-\frac{1}{p}\left(\sum_{i=1}^{n-p} u_i\right)^2 = \\ &\left[\left(1-\sum_{i=1}^{n-(p+1)} u_i^2\right)-\frac{1}{p+1}\left(\sum_{i=1}^{n-(p+1)} u_i\right)^2\right]-\left(\frac{p+1}{p}\right)\left[u_{n-p}+\frac{1}{p+1}\left(\sum_{i=1}^{n-(p+1)} u_i\right)\right]^2 = \\ &= a_{p+1}^2 - b_{p+1}^2 (u_{n-p}+c_{p+1})^2 \geqq 0 \end{aligned}\right\} \qquad (4.2)$$

which last fact follows by induction from (3.4) if we put $p = 2$. Thus u_{n-p} can vary, for fixed values of $u_1, u_2, \cdots, u_{n-(p+1)}$ between the limits

$$-c_{p+1}-\frac{a_{p+1}}{b_{p+1}} \leqq u_{n-p} \leqq -c_{p+1}+\frac{a_{p+1}}{b_{p+1}}. \qquad (4.3)$$

Setting now

$$\left.\begin{aligned} & f_{n-p}(u_1, u_2, \cdots, u_{n-p}) = \\ & = \underset{\text{over all possible values}}{\int du_{n-p+1} \int du_{n-p+2} \cdots \int du_{n-2}} \frac{\pi^{-\frac{n-1}{2}} \sqrt{\frac{n}{2}} \left(\frac{n-3}{2}\right)!}{\left[\left(1-\sum_{i=1}^{n-2} u_i^2\right) - \frac{1}{2}\left(\sum_{i=1}^{n-2} u_i\right)^2\right]^{1/2}} \\ & = \underset{\text{over all possible values}}{\int du_{n-p+1} \int du_{n-p+2} \cdots \int du_{n-2}} \, \pi^{-\frac{n-1}{2}} \sqrt{\frac{n}{2}} \left(\frac{n-3}{2}\right)! \; a_2^{-1} \end{aligned}\right\} \quad (4.4)$$

we can prove by induction that

$$f_{n-p}(u_1, u_2, \cdots, u_{n-p}) = \pi^{-\frac{n-p}{2}} \sqrt{\frac{n}{p}} \frac{\left(\frac{n-3}{2}\right)!}{\left(\frac{p-3}{2}\right)!} a_p^{p-3}. \quad (4.5)$$

In fact we have from (4.4), (4.2) and (4.3) that

$$\left.\begin{aligned} & f_{n-(p+1)}(u_1, u_2, \cdots, u_{n-(p+1)}) = \underset{\text{over all possible values}}{\int du_{n-p}} f_{n-p}(u_1, u_2, \cdots, u_{n-p}) \\ & = \pi^{-\frac{n-p}{2}} \sqrt{\frac{n}{p}} \frac{\left(\frac{n-3}{2}\right)!}{\left(\frac{p-3}{2}\right)!} \int_{-c_{p+1}-\frac{a_{p+1}}{b_{p+1}}}^{-c_{p+1}+\frac{a_{p+1}}{b_{p+1}}} \left[a_{p+1}^2 - b_{p+1}^2 (u_{n-p} + c_{p+1})^2\right]^{\frac{p-3}{2}} du_{n-p} \\ & = \pi^{-\frac{n-p}{2}} \sqrt{\frac{n}{p}} \frac{\left(\frac{n-3}{2}\right)!}{\left(\frac{p-3}{2}\right)!} a_{p+1}^{p-2} \sqrt{\frac{p}{p+1}} \int_{-1}^{+1} (1-x^2)^{\frac{p-3}{2}} dx \\ & = \pi^{-\frac{n-(p+1)}{2}} \sqrt{\frac{n}{p+1}} \frac{\left(\frac{n-3}{2}\right)!}{\left(\frac{(p+1)-3}{2}\right)!} a_{p+1}^{(p+1)-3} \end{aligned}\right\} (4.6)$$

because

$$\int_{-1}^{1}(1-x^2)^{\frac{f}{2}}dx = \int_0^1 t^{-\frac{1}{2}}(1-t)^{\frac{f}{2}}dt = B\left(\frac{1}{2}, \frac{f}{2}+1\right) = \sqrt{\pi}\frac{\left(\frac{f}{2}\right)!}{\left(\frac{f+1}{2}\right)!} \tag{4.7}$$

where $B(p, q)$ is the so-called complete Beta-Function. This proves (4.5) since the integrand in (4.4) is just f_{n-p} as given in (4.5) with $p = 2$. For $p = n-1$ we then finally obtain from (4.1) and (4.5), dropping the index 1,

$$f(u) = f_{n-(n-1)} = \frac{1}{\sqrt{\pi}}\sqrt{\frac{n}{n-1}}\frac{\left(\frac{n-3}{2}\right)!}{\left(\frac{n-4}{2}\right)!}\left(1-\frac{n}{n-1}u^2\right)^{\frac{n-4}{2}} \tag{4.8}$$

$$n \geqq 3 \qquad |u| \leqq \sqrt{\frac{n-1}{n}}.$$

§ 5. From (4.8) and (2.9) the probability of a relative error lying between r and $r+dr$ is given by

$$\boxed{f(r)\,dr = \frac{1}{\sqrt{\pi}}\frac{1}{\sqrt{n-1}}\frac{\left(\frac{n-3}{2}\right)!}{\left(\frac{n-4}{2}\right)!}\left(1-\frac{r^2}{n-1}\right)^{\frac{n-4}{2}}dr \qquad n \geqq 3 \qquad |r| \leqq \sqrt{n-1}.} \tag{5.1}$$

From (4.7) it is easily seen, that

$$\int_{-\sqrt{n-1}}^{\sqrt{n-1}} f(r)\,dr = 1 \tag{5.2}$$

as it should be. Furthermore we have in accordance with (1.14)

$$\int_{-\sqrt{n-1}}^{\sqrt{n-1}} r\, f(r)\, dr = 0, \tag{5.3}$$

because the integrand is an odd function, and in accordance with (1.15)

$$\left.\begin{aligned}\int_{-\sqrt{n-1}}^{\sqrt{n-1}} r^2 f(r)\, dr &= \frac{n-1}{\sqrt{\pi}} \frac{\left(\frac{n-3}{2}\right)!}{\left(\frac{n-4}{2}\right)!} \int_{-1}^{1} x^2 (1-x^2)^{\frac{n-4}{2}} dx = \\ = \frac{n-1}{\sqrt{\pi}} \frac{\left(\frac{n-3}{2}\right)!}{\left(\frac{n-4}{2}\right)!} B\left(\frac{3}{2}, \frac{n-2}{2}\right) &= \frac{n-1}{\sqrt{\pi}} \frac{\left(\frac{n-3}{2}\right)!}{\left(\frac{n-4}{2}\right)!} \cdot \frac{\frac{\sqrt{\pi}}{2}\left(\frac{n-4}{2}\right)!}{\left(\frac{n-1}{2}\right)!} = 1.\end{aligned}\right\} \tag{5.4}$$

This distribution (5.1) turns out to be a familiar one, for $r' = \dfrac{r}{\sqrt{n-1}}$ is distributed in the same way as the estimate

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\left(\sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot \sum_{j=1}^{n} (y_j - \bar{y})^2\right)^{1/2}} \tag{5.5}$$

of the correlation coefficient, ρ, when $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ is a sample of n drawn from a normal population of a two-dimensional statistical variable $(\boldsymbol{x}, \boldsymbol{y})$ for which $\rho = 0$.[1)]

It is to be noticed that *for $n = 3$ the relative errors are the more probable the larger they are*, and that *for $n = 4$ every relative error is equally probable.* For $n \geqq 5$ the curve looks more like a usual error curve with a maximum for $r = 0$ and approaching the r-axis for increasing values of

1) R. A. Fisher, Biometrika, **10** (1915), 507.
P. R. Rider, Annals of Math. **31** (1930), 577.

r. In figs. 1—5 are shown the curve for $n = 3, 4, 5, 6$ and 7, and in the last fig. the normal error curve (1.1) with

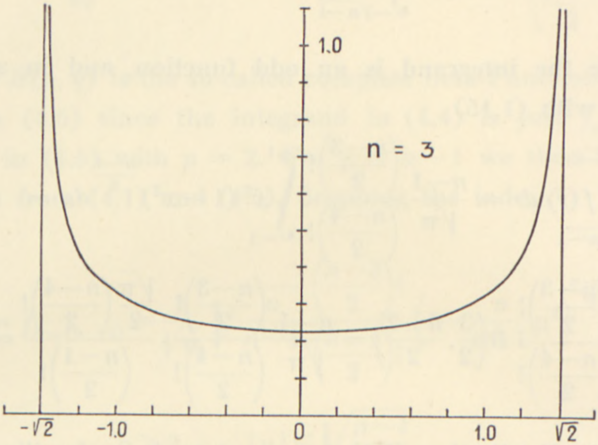


Fig. 1.

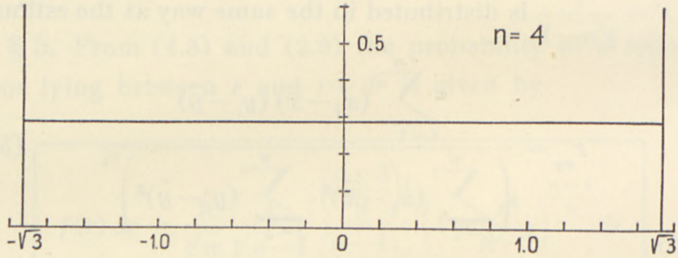


Fig. 2.

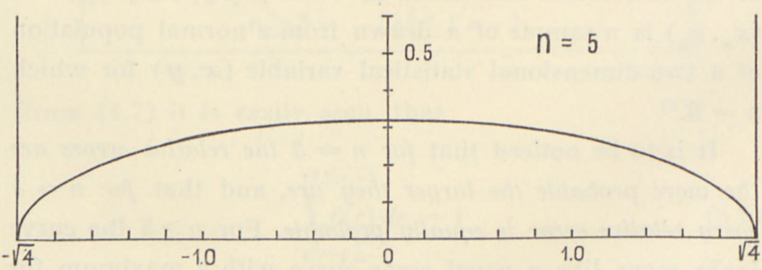


Fig. 3.

$\xi = 0$ and $\sigma = 1$ is plotted for comparison. In fact it is easily proved that for large values of n r is approximately

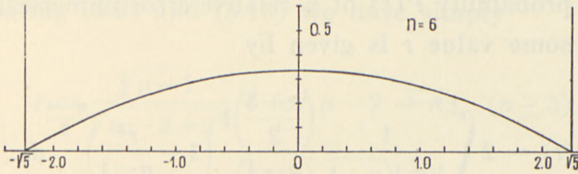


Fig. 4.

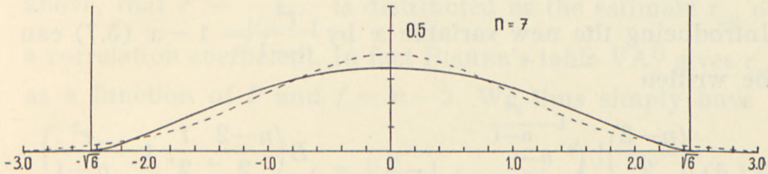


Fig. 5.

normally distributed as we should expect, since in that case the difference between ρ_i (1.6) and r_i (1.10) should be negligible. For any fixed value of r we have, using Stirling's formula for the factorials

$$\left.\begin{aligned}
&\lim_{n\to\infty} \frac{1}{\sqrt{\pi}\sqrt{n-1}} \frac{\left(\frac{n-3}{2}\right)!}{\left(\frac{n-4}{2}\right)!} \left(1-\frac{r^2}{n-1}\right)^{\frac{n-4}{2}} = \\
&\lim_{n\to\infty} \frac{1}{\sqrt{\pi}\sqrt{n-1}} \frac{\sqrt{2\pi}\left(\frac{n-3}{2}\right)^{\frac{n-2}{2}} \exp\left(-\frac{n-3}{2}\right)}{\sqrt{2\pi}\left(\frac{n-4}{2}\right)^{\frac{n-3}{2}} \exp\left(-\frac{n-4}{2}\right)} \left(1-\frac{r^2}{n-1}\right)^{\frac{n-4}{2}} = \\
&\lim_{n\to\infty} \frac{1}{\sqrt{2\pi}} \frac{\left(1-\frac{3}{n}\right)^{\frac{n-2}{2}}}{\left(1-\frac{1}{n}\right)^{1/2}\left(1-\frac{4}{n}\right)^{\frac{n-3}{2}}} \left[\left(1-\frac{r^2}{n-1}\right)^{\frac{n-1}{r^2}}\right]^{\frac{n-4}{n-1}\cdot\frac{r^2}{2}} = \\
&\qquad = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{r^2}{2}\right).
\end{aligned}\right\} \quad (5.6)$$

The probability $P(r)$ of a relative error numerically exceeding some value r is given by

$$P(r) = 2\int_r^{\sqrt{n-1}} \frac{1}{\sqrt{\pi}\sqrt{n-1}} \frac{\left(\frac{n-3}{2}\right)!}{\left(\frac{n-4}{2}\right)!} \left(1-\frac{r^2}{n-1}\right)^{\frac{n-4}{2}} dr. \quad (5.7)$$

Introducing the new variable x by $\frac{r^2}{n-1} = 1-x$ (5.7) can be written

$$P(r) = \frac{1}{\sqrt{\pi}} \frac{\left(\frac{n-3}{2}\right)!}{\left(\frac{n-4}{2}\right)!} \int_0^{1-\frac{r^2}{n-1}} x^{\frac{n-4}{2}} (1-x)^{-\frac{1}{2}} dx = \frac{B\left(\frac{n-2}{2}, \frac{1}{2}; 1-\frac{r^2}{n-1}\right)}{B\left(\frac{n-2}{2}, \frac{1}{2}\right)} \quad (5.8)$$

where the B's are the so-called incomplete and the complete Beta-function respectively. From a table of the incomplete Beta-function[1] one can tabulate P as a function of r or r as a function of P. We can, however, also introduce a new variable y by $1-\frac{r^2}{n-1} = \frac{1}{1+y^2}$ thus obtaining

$$P(r) = \frac{2}{\sqrt{\pi}} \frac{\left(\frac{n-3}{2}\right)!}{\left(\frac{n-4}{2}\right)!} \int_{\frac{r}{\sqrt{n-1-r^2}}}^{\infty} (1+y^2)^{-\frac{n-1}{2}} dy \quad (5.9)$$

which is just the function tabulated by Fisher[2] as the so-called t-distribution. In fact his table IV gives t as a function of P and n_t, where

$$P(t) = \frac{2}{\sqrt{\pi}} \frac{\left(\frac{n_t-1}{2}\right)!}{\left(\frac{n_t-2}{2}\right)!} \int_{\frac{t}{\sqrt{n_t}}}^{\infty} (1+t^2)^{-\frac{n_t+1}{2}} dt. \quad (5.10)$$

1) Karl Pearson: Tables of the incomplete Beta-function. London 1934.
2) Fisher: Statistical Methods for Research Workers.

Comparing (5.9) and (5.10) we have simply

$$r = \frac{\sqrt{n-1}}{\sqrt{n-2+t^2}} \cdot t \text{ for } n-2 = n_t, \quad (n \geqq 3). \tag{5.11}$$

For certain values of P we can also use the fact mentioned above, that $r' = \frac{r}{\sqrt{n-1}}$ is distributed as the estimate r_{xy} of a correlation coefficient. In fact FISHER's table VA[1] gives r_{xy} as a function of P and $f = n-2$. We thus simply have

$$r = \sqrt{n-1}\; r_{xy}. \tag{5.12}$$

In table 1 we give r as a function of P and $f = n-2 \geqq 1$ obtained in these ways, where f is the number of degrees of freedom because of the two equations (1.11) and (1.12). Since $|r| \leqq \sqrt{n-1} = \sqrt{f+1}$ we have also listed $\sqrt{f+1}$. The last row with $f = \infty$ gives simply r for the normal distribution (5.6). From this table one can at once decide whether or not a given relative error is more or less probable than any given value P, e. g. $P = 0.001$.

It will be noted that whereas t is a *monotonic decreasing* function of n_t, r is a *monotonic increasing* function for small values of P, *monotonic decreasing* for larger values of P and non-monotonic for medium values of P.

II. Deduction of the distribution law for indirect and unequally good observations.

§ 6. We shall now show that the same distribution law (5.1) also holds for the relative errors in case the observations are indirect and unequally good, as is e. g. the case in triangulation.

[1] FISHER: Statistical Methods for Research Workers.

We shall first recall the *theory of adjustment* written in a matrixform[1]. Let[2]

$$\underset{n1}{\boldsymbol{L}} = \begin{Bmatrix} l_1 \\ l_2 \\ \vdots \\ l_n \end{Bmatrix} \tag{6.1}$$

be a sample of $n(\geq 2)$ independent observations of n normally distributed statistical variables with true values (i. e. mean values)

$$\underset{n1}{\Lambda} = \begin{Bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{Bmatrix}. \tag{6.2}$$

Connecting these true values we have $m(<n)$ linear equations called the *equations of condition*

$$\underset{n1}{\Lambda} = \underset{n1}{\boldsymbol{A}_0} + \underset{nm}{\boldsymbol{A}} \cdot \underset{m1}{\Xi} \tag{6.3}$$

where

$$\underset{m1}{\Xi} = \begin{Bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_m \end{Bmatrix} \tag{6.4}$$

are the true values (i. e. the mean values) of m free variables called the *elements*, which completely determine the system considered.

We denote by

$$\underset{nn}{\boldsymbol{P}} = \begin{Bmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & p_n \end{Bmatrix} \tag{6.5}$$

[1] Cf. H. Jensen: Herleitung einiger Ergebnisse der Ausgleichungsrechnung mit Hilfe von Matrizen. Publications of the Geodetic Institute, No. 13. Copenhagen 1939.

N. Arley and K. R. Buch loc. cit. chap. 12.

[2] We shall in the following denote matrices by capital clarendon types, true values by greek letters and the best estimates for these by the corresponding latin letters with a bar. The transposed matrix we denote by an asterisk.

the *weight matrix*, the p_i's being n arbitrary constants—the *weights*—satisfying the relations

$$p_1 \sigma_1^2 = p_2 \sigma_2^2 = \cdots = p_n \sigma_n^2 = \sigma^2. \tag{6.6}$$

Here $\sigma_1, \sigma_2, \cdots, \sigma_n$ are the standard deviations of the observations $l_1, l_2, \cdots, l_n$ and σ is called the *standard deviation of the weight unit*. As is well known, the best estimates

$$\underset{n1}{\overline{\boldsymbol{L}}} = \begin{Bmatrix} \bar{l}_1 \\ \bar{l}_2 \\ \vdots \\ \bar{l}_n \end{Bmatrix} = \underset{n1}{\boldsymbol{L}} + \underset{n1}{\boldsymbol{V}} \tag{6.7}$$

and

$$\underset{m1}{\overline{\boldsymbol{X}}} = \begin{Bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_m \end{Bmatrix} \tag{6.8}$$

for the true values Λ and Ξ are those values, which, satisfying the equations of condition

$$\overline{\boldsymbol{L}} = \boldsymbol{L} + \boldsymbol{V} = \boldsymbol{A}_0 + \boldsymbol{A} \cdot \overline{\boldsymbol{X}}, \tag{6.9}$$

make the weighted sum of the squares of the *errors* $\boldsymbol{V}$[1])

$$\sum_{i=1}^{n} p_i v_i^2 = [pvv] = \boldsymbol{V}^* \cdot \boldsymbol{P} \cdot \boldsymbol{V} \tag{6.10}$$

as small as possible. The condition for this to be satisfied is that all the partial derivatives of $[pvv]$ with respect to $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_m$ shall vanish, which gives us the m equations

$$\underset{mn}{\boldsymbol{A}^*} \cdot \underset{nn}{\boldsymbol{P}} \cdot \underset{n1}{\boldsymbol{V}} = \underset{m1}{\boldsymbol{O}}. \tag{6.11}$$

Eliminating $\boldsymbol{V}$ between (6.9) and (6.11) gives us the *normal equations*

$$\boldsymbol{B} \cdot \overline{\boldsymbol{X}} = \boldsymbol{A}^* \cdot \boldsymbol{P} \cdot \boldsymbol{N} \tag{6.12}$$

where $\boldsymbol{B}$ and $\boldsymbol{N}$ are abbreviations for

1) Comparing (6.9) with (1.3) it is seen that the errors, or residuals, used here have the opposite sign. It would, therefore, be more correct to call them *corrections*, but the notation "errors" is now commonly accepted in the theory of adjustment.

$$\underset{mm}{\boldsymbol{B}} = \underset{mn}{\boldsymbol{A}^*} \cdot \underset{nn}{\boldsymbol{P}} \cdot \underset{nm}{\boldsymbol{A}}$$

and (6.13)

$$\underset{n1}{\boldsymbol{N}} = \underset{n1}{\boldsymbol{L}} - \underset{n1}{\boldsymbol{A}_0}$$

respectively. Since the elements are assumed to be free variables the (*symmetric*) matrix $\boldsymbol{B}$ has an inverse, and the solution of the normal equations is, therefore, uniquely given by

$$\overline{\boldsymbol{X}} = \boldsymbol{C} \cdot \boldsymbol{N} \tag{6.14}$$

where $\boldsymbol{C}$ is an abbreviation for

$$\underset{mn}{\boldsymbol{C}} = \underset{mm}{\boldsymbol{B}^{-1}} \cdot \underset{mn}{\boldsymbol{A}^*} \cdot \underset{nn}{\boldsymbol{P}}. \tag{6.15}$$

We note that $\boldsymbol{C}$, because of (6.13) satisfies the important equation

$$\underset{mn}{\boldsymbol{C}} \cdot \underset{nm}{\boldsymbol{A}} = \boldsymbol{B}^{-1} \cdot \boldsymbol{A}^* \cdot \boldsymbol{P} \cdot \boldsymbol{A} = \underset{mm}{\boldsymbol{E}}. \tag{6.16}$$

From (6.14) and (6.9) we have

$$\overline{\boldsymbol{L}} = (\boldsymbol{E} - \boldsymbol{A} \cdot \boldsymbol{C}) \cdot \boldsymbol{A}_0 + \boldsymbol{A} \cdot \boldsymbol{C} \cdot \boldsymbol{L}$$

and (6.17)

$$\boldsymbol{V} = \overline{\boldsymbol{L}} - \boldsymbol{L} = (\boldsymbol{A} \cdot \boldsymbol{C} - \boldsymbol{E}) \cdot (\boldsymbol{L} - \boldsymbol{A}_0)$$

where $\boldsymbol{E}$ is a unit matrix. It is easily shown that the $\boldsymbol{V}$ so obtained actually makes $[pvv]$ as small as possible. Let $\boldsymbol{V}'$ be another set of errors corresponding to a set of elements $\boldsymbol{X}'$, i. e.

$$\boldsymbol{V}' = -\boldsymbol{N} + \boldsymbol{A} \cdot \boldsymbol{X}'. \tag{6.18}$$

Subtracting (6.9) from (6.18) we have

$$\boldsymbol{V}' - \boldsymbol{V} = \boldsymbol{A} \cdot (\boldsymbol{X}' - \overline{\boldsymbol{X}}) \tag{6.19}$$

and, therefore, using (6.11)

$$\begin{aligned}[pv'v'] = \boldsymbol{V}'^* \cdot \boldsymbol{P} \cdot \boldsymbol{V}' &= \boldsymbol{V}^* \cdot \boldsymbol{P} \cdot \boldsymbol{V} + (\boldsymbol{V}' - \boldsymbol{V})^* \cdot \boldsymbol{P} \cdot (\boldsymbol{V}' - \boldsymbol{V}) + \\ &\boldsymbol{V}^* \cdot \boldsymbol{P} \cdot \boldsymbol{A} \cdot (\boldsymbol{X}' - \overline{\boldsymbol{X}}) + (\boldsymbol{X}' - \overline{\boldsymbol{X}})^* \cdot \boldsymbol{A}^* \cdot \boldsymbol{P} \cdot \boldsymbol{V} = \\ &[pvv] + [p(v' - v)^2].\end{aligned} \tag{6.20}$$

This shows immediately that $[pv'v']$ is as small as possible for $\boldsymbol{V}' = \boldsymbol{V}$.

(6.17) shows that $\boldsymbol{V}$ is a linear function of $\boldsymbol{L}$. $\boldsymbol{V}$ is, therefore, normally distributed, having mean value $\boldsymbol{O}$ because, from (6.3) and (6.16),

$$\left.\begin{aligned} m\{\boldsymbol{V}\} &= (\boldsymbol{A}\cdot\boldsymbol{C}-\boldsymbol{E})\cdot(\Lambda-\boldsymbol{A}_0) = (\boldsymbol{A}\cdot\boldsymbol{C}\cdot\boldsymbol{A}-\boldsymbol{A})\cdot\Xi = \\ &= (\boldsymbol{A}-\boldsymbol{A})\cdot\Xi = \boldsymbol{O}. \end{aligned}\right\} \quad (6.21)$$

To obtain the standard deviation of $\boldsymbol{V}$ we first evaluate the *momentmatrix* of $\boldsymbol{V}$. Quite generally, if we have k statistical variables $y_1, y_2, \cdots, y_k$ which are linear functions of the observations $\boldsymbol{L}$,

$$\underset{k1}{\boldsymbol{Y}} = \underset{k1}{\boldsymbol{F}_0} + \underset{kn}{\boldsymbol{F}}\cdot\underset{n1}{\boldsymbol{L}} \quad (6.22)$$

we denote by the momentmatrix $\boldsymbol{M}^{(Y)}$ of $\boldsymbol{Y}$ the matrix

$$\left.\begin{aligned} \underset{kk}{\boldsymbol{M}^{(Y)}} &= \{\mu_{rs}\} = \{m\{(y_r-\mu_r)(y_s-\mu_s)\}\} = \{m\{y_r y_s\} - m\{y_r\}m\{y_s\}\} = \\ &= \{m\{(y_r-f_{or})(y_s-f_{os}\} - m\{y_r-f_{or}\}m\{y_s-f_{os}\}\} \end{aligned}\right\} (6.23)$$

where $\mu_i = m\{y_i\}$ is the mean value of y_i. The elements of this matrix give the standard deviations and correlation coefficients of $y_1, y_2, \cdots, y_k$

$$\begin{aligned} \sigma^2\{y_r\} &= \sigma_r^2 = \mu_{rr} \\ \rho\{y_r, y_s\} &= \rho_{rs} = \frac{\mu_{rs}}{\sigma_r\,\sigma_s} \qquad (r \neq s) \end{aligned} \quad (6.24)$$

(In case $\boldsymbol{M}^{(Y)}$ is a diagonal matrix, $\boldsymbol{Y}$ is said to be *free functions.*) Introducing (6.22) into (6.23) we obtain, because *the observations* $\boldsymbol{L}$ *are mutually independent,*

$$\begin{aligned} \mu_{rs} &= m\left\{\left(\sum_{i=1}^{n} f_{ri}\,l_i\right)\left(\sum_{j=1}^{n} f_{sj}\,l_j\right)\right\} - m\left\{\sum_{i=1}^{n} f_{ri}\,l_i\right\} m\left\{\sum_{j=1}^{n} f_{sj}\,l_j\right\} = \\ &= \sum_{i=1}^{n} f_{ri} f_{si}\,\sigma_i^2 = \sigma^2 \sum_{i=1}^{n} f_{ri} f_{si}\frac{1}{p_i} = \sigma^2(\boldsymbol{F}\cdot\boldsymbol{P}^{-1}\cdot\boldsymbol{F}^*)_{rs} \end{aligned}$$

i. e.

$$\boldsymbol{M}^{(Y)} = \sigma^2\,\boldsymbol{F}\cdot\boldsymbol{P}^{-1}\cdot\boldsymbol{F}^*. \quad (6.25)$$

As special cases we obtain at oncė from (6.14) and (6.17)

$$\underset{mm}{\boldsymbol{M}^{(\bar{X})}} = \sigma^2\,\boldsymbol{C}\cdot\boldsymbol{P}^{-1}\cdot\boldsymbol{C}^* = \sigma^2\,\boldsymbol{B}^{-1} \quad (6.26)$$

$$\underset{nn}{\boldsymbol{M}^{(\bar{L})}} = \sigma^2\,\boldsymbol{A}\cdot\boldsymbol{C}\cdot\boldsymbol{P}^{-1}\cdot\boldsymbol{C}^*\cdot\boldsymbol{A}^* = \sigma^2\,\boldsymbol{A}\cdot\boldsymbol{B}^{-1}\cdot\boldsymbol{A}^* \quad (6.27)$$

$$\underset{nn}{\boldsymbol{M}^{(V)}} = \sigma^2(\boldsymbol{A}\cdot\boldsymbol{C}-\boldsymbol{E})\cdot\boldsymbol{P}^{-1}(\boldsymbol{C}^*\cdot\boldsymbol{A}^*-\boldsymbol{E}) = \sigma^2\underset{nn}{\boldsymbol{T}}. \quad (6.28)$$

(We note that (6.26) and (6.27) show that in general $\bar{\boldsymbol{X}}$ and $\bar{\boldsymbol{L}}$ are *not* free functions).

The n quantities, analogous to (1.6)

$$\rho_i = \frac{\boldsymbol{v}_i}{\sigma\{v_i\}} = \frac{\bar{\boldsymbol{l}}_i - \boldsymbol{l}_i}{\sqrt{t_{ii}}\,\sigma} \qquad i = 1, 2, \cdots, n \qquad (6.29)$$

where t_{ii} is the i'th diagonalelement of $\boldsymbol{T}$ given in (6.28), are consequently normally distributed with mean value 0 and standard deviation 1. Now the exact value of the parameter σ is not known, but can only be estimated from the observations $l_1, l_2, \cdots, l_n$. In the usual case where these quantities constitute all our information, the mean square error, analogous to (1.8)

$$s = \frac{q}{\sqrt{n-m}} \qquad (6.30)$$

where, analogously to (1.9)

$$q = +\left(\sum_{i=1}^{n} p_i v_i^2\right)^{1/2} = +[pvv]^{1/2} = +\left(\sum_{i=1}^{n} p_i (\bar{l}_i - l_i)^2\right)^{1/2} \qquad (6.31)$$

is the best estimate of σ. Substituting this value for σ in (6.26) and (6.27) gives us the expressions for the mean errors of the best estimates $\bar{\boldsymbol{X}}$ and $\bar{\boldsymbol{L}}$ for the true values Ξ and Λ. We note that as a control of the computations one can use the relation

$$\left.\begin{aligned} &\sum_{i=1}^{n} p_i \sigma^2\{\bar{l}_i\} = \sigma^2 \operatorname{Spur} \boldsymbol{P}\cdot\boldsymbol{A}\cdot\boldsymbol{B}^{-1}\cdot\boldsymbol{A}^* = \\ &= \sigma^2 \operatorname{Spur} \boldsymbol{B}^{-1}\cdot\boldsymbol{A}^*\cdot\boldsymbol{P}\cdot\boldsymbol{A} = \sigma^2 \operatorname{Spur} \boldsymbol{E}_{mm} = m\sigma^2. \end{aligned}\right\} \qquad (6.32)$$

Substituting (6.30) for σ in (6.29) we obtain the n quantities, analogous to (1.10)

$$\boldsymbol{r}_i = \frac{\bar{\boldsymbol{l}}_i - \boldsymbol{l}_i}{\sqrt{t_{ii}}\, s} = \frac{\boldsymbol{v}_i}{\boldsymbol{q}} \sqrt{\frac{n-m}{t_{ii}}} \qquad (6.33)$$

which are called the *relative errors*. From (6.11) it follows that they satisfy the m linear equations, analogous to (1.11)

$$\boldsymbol{A}^*_{mn} \cdot \boldsymbol{P}_{nn} \cdot \boldsymbol{T}'^{1/2}_{nn} \cdot \boldsymbol{R}_{n1} = \boldsymbol{O}_{n1} \tag{6.34}$$

where $\boldsymbol{T}'$ is given by

$$(\boldsymbol{T}')_{rs} = t_{rr}\,\delta_{rs}. \tag{6.35}$$

Because of (6.31) we have furthermore the equation, analogous to (1.12)

$$\sum_{i=1}^{n} p_i\, t_{ii}\, r_i^2 = [ptr^2] = \boldsymbol{R}^* \cdot \boldsymbol{T}' \cdot \boldsymbol{P} \cdot \boldsymbol{R} = n - m. \tag{6.36}$$

The number of degrees of freedom of the relative errors is thus given by

$$f = n - m - 1. \tag{6.37}$$

§ 7. As in § 2 we first write down the probability $S(l_1, l_2, \cdots, l_n)\, dl_1\, dl_2 \cdots dl_n$ of the sample $l_1, l_2, \cdots, l_n$, which, analogously to (2.1), is given by

$$\left.\begin{aligned} & S(l_1, l_2, \cdots, l_n)\, dl_1\, dl_2 \cdots dl_n = \\ = & \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n (p_1\, p_2 \cdots p_n)^{1/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} p_i\, (\lambda_i - l_i)^2\right] dl_1\, dl_2 \cdots dl_n. \end{aligned}\right\} \tag{7.1}$$

Taking in (6.19) and (6.20) $\boldsymbol{X}' = \boldsymbol{\Xi}$ we have, using (6.31) and denoting by ε_i the i'th *true error*, $\varepsilon_i = \lambda_i - l_i$,

$$\sum_{i=1}^{n} p_i (\lambda_i - l_i)^2 = [p\varepsilon\varepsilon] = q^2 + (\bar{\boldsymbol{X}} - \boldsymbol{\Xi})^* \cdot \boldsymbol{B} \cdot (\bar{\boldsymbol{X}} - \boldsymbol{\Xi}). \tag{7.2}$$

We next introduce, as in § 2, in (7.1) instead of $l_1, l_2, \cdots, l_n$ the n new variables

$$\bar{\boldsymbol{X}} \quad (\text{given in } (6.14))$$

$$q \quad (\text{ — } \text{ - } (6.31))$$

and

$$u_i \qquad i = 1, 2, \cdots, n$$

defined by the equations, analogous to (2.2)

$$l_i = \bar{l}_i - \frac{q}{\sqrt{p_i}} u_i. \tag{7.3}$$

Comparing with (6.33) we see that

$$\boldsymbol{R} = \sqrt{n-m}\, \boldsymbol{T}'^{-1/2} \cdot \boldsymbol{P}^{-1/2} \cdot \boldsymbol{U}. \tag{7.4}$$

From (6.34) and (6.36) respectively we have, analogously to (2.3) and (2.4)

$$\left.\begin{aligned} &\boldsymbol{A}^* \cdot \boldsymbol{P}^{1/2} \cdot \boldsymbol{U} = \boldsymbol{O} \\ \text{and} \\ &\boldsymbol{U}^* \cdot \boldsymbol{U} = \sum_{i=1}^{n} u_i^2 = 1. \end{aligned}\right\} \tag{7.5}$$

Taking as free variables the first $n-m-1 = f$ u_i's we have, using (7.2), that the probability $S(l_1, \cdots, l_n)\, dl_1 \cdots dl_n$ of the sample $l_1, \cdots, l_n$ expressed in the new variables is given by the expression, analogous to (2.8),

$$\left.\begin{aligned} & S(l_1, \cdots, l_n)\, dl_1 \cdots dl_n = \\ & = S(\bar{x}_1, \cdots, \bar{x}_m; q; u_1, \cdots, u_f)\, d\bar{x}_1 \cdots d\bar{x}_m\, dq\, du_1 \cdots du_f = \\ & \left\{ \frac{\sqrt{|\boldsymbol{B}|}}{\left(\sqrt{2\pi}\,\sigma\right)^m} \exp\left[-\frac{1}{2\sigma^2} (\bar{\boldsymbol{X}} - \boldsymbol{\Xi})^* \cdot \boldsymbol{B} \cdot (\bar{\boldsymbol{X}} - \boldsymbol{\Xi}) \right] d\bar{x}_1 \cdots d\bar{x}_m \right\} \times \\ & \times \left\{ \frac{1}{\left(\frac{f'-2}{2}\right)!\, 2^{\frac{f'-2}{2}}} \left(\frac{q}{\sigma}\right)^{f'-1} \exp\left[-\frac{q^2}{2\sigma^2} \right] \frac{dq}{\sigma} \right\} \times \\ & \times \left\{ \pi^{-\frac{n-m}{2}} |\boldsymbol{B}|^{-1/2} \left(\frac{f-1}{2}\right)!\, |D(u_1, \cdots, u_f; n)|\, du_1 \cdots du_f \right\}. \end{aligned}\right\} \tag{7.6}$$

Here
$$-\infty < \bar{x}_i < \infty \qquad i = 1, 2, \cdots, m$$
$$0 \leqq q < \infty$$
$$-1 < u_i < 1 \qquad i = 1, 2, \cdots, f$$
$$f' = f + 1$$ [1]

and

$$\left.\begin{aligned} &(p_1 \cdots p_n)^{1/2} \frac{\partial (l_1, \cdots\cdots\cdots\cdots\cdots\cdots, l_n)}{\partial (\bar{x}_1, \cdots, \bar{x}_m; q; u_1, \cdots, u_f)} = \\ &= q^f (-1)^{f+1} \begin{vmatrix} \boldsymbol{P}^{1/2}_{nn} \boldsymbol{A}_{nm} & \vdots & \boldsymbol{U}_{n1} & \vdots & \begin{matrix} \boldsymbol{E}_{ff} \\ \cdots\cdots \\ \boldsymbol{W}_{n-f,f} \end{matrix} \end{vmatrix} \\ &= q^f (-1)^{f+1} D (u_1, \cdots, u_f; n) \end{aligned}\right\} \tag{7.7}$$

with

$$\boldsymbol{W}_{n-f,f} = \left\{\frac{\partial u_r}{\partial u_s}\right\} \qquad \begin{aligned} r &= f+1, \cdots, n \\ s &= 1, \cdots, f \end{aligned} \tag{7.8}$$

The coefficients in (7.6) have been so chosen that the integrals taken over all possible values of the variables give unity for each of the three factors.[2] (7.6) shows that $\bar{\boldsymbol{X}}$ and q are uncorrelated variables, and that $\boldsymbol{U}$ is uncorrelated to $\bar{\boldsymbol{X}}$ and q. Further it is seen that the last factor does not contain either the parameter ξ or σ, which shows that $\bar{\boldsymbol{X}}$ and q are what Fisher calls "sufficient statistics".[3] (We now also see that the three notions "free functions", "uncorrelated variables" and "mutually independent variables" are identical for normally distributed variables since if $\boldsymbol{B}^{-1}$ is diagonal, $\boldsymbol{B}$ is also and vice versa).

Squaring the determinant D in (7.7) we have, using (7.5),

[1] f' has been introduced here, because q has $f+1$ degrees of freedom.

[2] Regarding the coefficient in the first factor see e. g. Cramér: Random Variables 1937, p. 109.

[3] Fisher: Statistical Methods for Research Workers.

$$\left.\begin{aligned} D^2 &= \left|\begin{array}{c} \boldsymbol{A}^* \cdot \boldsymbol{P}^{1/2} \\ \cdots\cdots \\ \boldsymbol{U}^* \\ \cdots\cdots \\ \boldsymbol{E} \vdots \boldsymbol{W}^* \end{array}\right| \left|\begin{array}{c:c:c} \boldsymbol{P}^{1/2} \cdot \boldsymbol{A} & \boldsymbol{U} & \begin{array}{c} \boldsymbol{E} \\ \cdots \\ \boldsymbol{W} \end{array} \end{array}\right| = \left|\begin{array}{c:c:c} \boldsymbol{B}_{mm} & \boldsymbol{O}_{m1} & \boldsymbol{O}_{mf} \\ \cdots & \cdots & \cdots \\ \boldsymbol{O}_{1m} & \boldsymbol{E}_{11} & \boldsymbol{O}_{1f} \\ \cdots & \cdots & \cdots \\ \boldsymbol{O}_{fm} & \boldsymbol{O}_{f1} & \boldsymbol{E}_{ff} + \boldsymbol{W}^* \cdot \boldsymbol{W} \end{array}\right| \\ &= \left|\boldsymbol{B}_{mm}\right| \left|\boldsymbol{E}_{ff} + \boldsymbol{W}^*_{f,n-f} \cdot \boldsymbol{W}_{n-f,f}\right| . \end{aligned}\right\} \quad (7.9)$$

To evaluate the second determinant in (7.9) would, however, be rather unpleasant. We therefore proceed in another way, generalizing a method due to Cramér.[1)]

§ 8. Let us for the sake of simplicity assume that $l_1, l_2, \cdots, l_n$ each have mean value 0 and standard deviation 1, i. e. that $\sigma = 1$ and $\boldsymbol{P} = \boldsymbol{E}$. This assumption leaves the relative errors and their distribution unchanged, since it only means changing the zero-point and the unit of length for each of the observations. From (6.20) we then obtain, if we put $v'_i = 0 - l_i$,

$$q^2 = [vv] = [ll] - [\bar{l}\bar{l}] = \boldsymbol{L}^* \cdot \boldsymbol{L} - \bar{\boldsymbol{L}}^* \cdot \bar{\boldsymbol{L}} \qquad (8.1)$$

and each r_i can, therefore, be written in the form

$$\frac{r_i^2}{n-m} = \frac{\frac{1}{t_{ii}} (\bar{l}_i - l_i)^2}{\boldsymbol{L}^* \cdot \boldsymbol{L} - \bar{\boldsymbol{L}}^* \cdot \bar{\boldsymbol{L}}} . \qquad (8.2)$$

Since $\bar{\boldsymbol{L}}$ is a function of $\boldsymbol{L}$, the expression on the right side is a function of the n mutually independent variables $\boldsymbol{L}$. We shall now show that the expression is, in fact, only a function of $n - m$ mutually independent variables. Let us choose the m elements in such a way, that in (6.9) $\boldsymbol{A}_0 = \boldsymbol{O}$ i. e.

1) Private communication. I wish to thank prof. Cramér very much for kindly indicating his method to me.

$$\bar{\boldsymbol{L}} = \boldsymbol{A} \cdot \bar{\boldsymbol{X}}. \tag{8.3}$$

Since $m\{\boldsymbol{L}\} = \boldsymbol{O}$ we have from (6.14)

$$m\{\bar{\boldsymbol{X}}\} = \boldsymbol{\mathfrak{z}} = \boldsymbol{O}. \tag{8.4}$$

Introducing (8.3) into (8.2) we have

$$\frac{r_i^2}{n-m} = \frac{\frac{1}{t_{ii}}(\bar{l}_i - l_i)^2}{\boldsymbol{L}^* \cdot \boldsymbol{L} - \bar{\boldsymbol{X}}^* \cdot \boldsymbol{A}^* \cdot \boldsymbol{A} \cdot \bar{\boldsymbol{X}}}. \tag{8.5}$$

As a consequence of our assumption $\sigma = 1$ and of the definition of t_{ii} the variable

$$y_{m+1} = \frac{v_i}{\sqrt{t_{ii}}} = \frac{\bar{l}_i - l_i}{\sqrt{t_{ii}}} \tag{8.6}$$

is normally distributed with mean value 0 and standard deviation 1. The linear form expressing y_{m+1} as a function of $\boldsymbol{L}$ is, therefore, a normalized linear form. From (7.6) it is seen, that v_i and thus y_{m+1} is independent of $\bar{\boldsymbol{X}}$. The linear form y_{m+1} is, therefore, orthogonal to each of the linear forms expressing $\bar{x}_i$ as a function of $\boldsymbol{L}$. Now $\boldsymbol{A}^* \cdot \boldsymbol{A} = \boldsymbol{B}$ is *symmetric*, and consequently we can bring the bilinear form $\bar{\boldsymbol{X}}^* \cdot \boldsymbol{B} \cdot \bar{\boldsymbol{X}}$ on diagonalform by an orthogonal substitution. Furthermore, since it is a *positive definite* form, we can, by suitably choosing the scale for each of the new variables, bring it on unity form. Consequently there exist m new variables

$$\underset{m1}{\boldsymbol{Y}} = \underset{mm}{\boldsymbol{D}} \cdot \underset{m1}{\bar{\boldsymbol{X}}} \tag{8.7}$$

with the property, that

$$\begin{aligned} &\bar{\boldsymbol{X}}^* \cdot \boldsymbol{A}^* \cdot \boldsymbol{A} \cdot \bar{\boldsymbol{X}} = \\ &= \bar{\boldsymbol{X}}^* \cdot \boldsymbol{B} \cdot \bar{\boldsymbol{X}} = \boldsymbol{Y}^* \cdot \boldsymbol{D}^* \cdot \boldsymbol{B} \cdot \boldsymbol{D} \cdot \boldsymbol{Y} = \boldsymbol{Y}^* \cdot \boldsymbol{Y}. \end{aligned} \tag{8.8}$$

From (8.4), (8.7), (8.8) and (7.6) with $\sigma = 1$, we see, that these m variables are mutually independent, normally distributed with mean values 0 and standard deviations 1. $\boldsymbol{Y}$ is therefore given by m normalized and mutually orthogonal linear forms in $\boldsymbol{L}$, which are orthogonal to the linear form y_{m+1}. We can now construct $n-m-1$ other normalized and orthogonal linear forms, which are orthogonal to the $m+1$ first ones, and we thus obtain n new variables

$$\underset{n1}{\boldsymbol{Y}} = \left\{ \begin{matrix} y_1 \\ \vdots \\ y_m \\ y_{m+1} \\ y_{m+2} \\ \vdots \\ y_n \end{matrix} \right\} \tag{8.9}$$

which are mutually independent and normally distributed with mean values 0 and standard deviations 1, and which, therefore, have the property that

$$\boldsymbol{L}^* \cdot \boldsymbol{L} = \boldsymbol{Y}^* \cdot \boldsymbol{Y} = \sum_{j=1}^{n} y_j^2. \tag{8.10}$$

Introducing $\underset{n1}{\boldsymbol{Y}}$ in (8.5) we now have, using (8.8) and dropping the index i,

$$\frac{r^2}{n-m} = \frac{y_{m+1}^2}{\sum_{j=1}^{n} y_j^2 - \sum_{j=1}^{m} y_j^2} = \frac{y_{m+1}^2}{\sum_{j=m+1}^{n} y_j^2} \tag{8.11}$$

which shows that r^2 depends only on $n-m$ variables, and that $0 \leqq r^2 \leqq n-m$.

Let us quite generally have a statistical variable of the form

$$\left.\begin{aligned} z &= \frac{\sum\limits_{j=1}^{f_1} y_j^2}{\sum\limits_{j=1}^{f_1+f_2} y_j^2} = \frac{\chi_1^2}{\chi_1^2+\chi_2^2} \\ \chi_1^2 &= \sum_{j=1}^{f_1} y_j^2, \quad \chi_2^2 = \sum_{j=f_1+1}^{f_1+f_2} y_j^2 \end{aligned}\right\} \tag{8.12}$$

where $y_1, y_2, \cdots, y_{f_1+f_2}$ are f_1+f_2 mutually independent, normally distributed statistical variables with mean values 0 and standard deviations 1. As is easily proved[1)], χ_1 and χ_2 have both the same distribution as $\frac{q}{\sigma}$, given by the second factor in (7.6), with f' equal to f_1 and f_2, respectively. From (7.6) we find, since χ_1^2 and χ_2^2 are mutually independent, that their correlation function is given by

$$\left.\begin{aligned} h(\chi_1^2, \chi_2^2)\, d\chi_1^2\, d\chi_2^2 &= \frac{1}{\left(\frac{f_1-2}{2}\right)!\left(\frac{f_2-2}{2}\right)!\, 2^{\frac{f_1+f_2}{2}}} \times \\ \times (\chi_1^2)^{\frac{f_1-2}{2}} (\chi_2^2)^{\frac{f_2-2}{2}} &\exp\left[-\frac{1}{2}(\chi_1^2+\chi_2^2)\right] d\chi_1^2\, d\chi_2^2. \end{aligned}\right\} \tag{8.13}$$

Introducing z from (8.12) instead of χ_1^2 we have

$$\left.\begin{aligned} \chi_1^2 &= \frac{z}{1-z}\chi_2^2 \\ \begin{vmatrix} \frac{\partial \chi_1^2}{\partial z} & \frac{\partial \chi_1^2}{\partial \chi_2^2} \\ \frac{\partial \chi_2^2}{\partial z} & \frac{\partial \chi_2^2}{\partial \chi_2^2} \end{vmatrix} &= \frac{\chi_2^2}{(1-z)^2} \end{aligned}\right\} \tag{8.14}$$

and thus

1) For $f' = 1$ the statement is clear. It has then only to be proved that the sum of two variables with this distribution, corresponding to the degrees of freedom f_1 and f_2 respectively, has the same distribution with $f = f_1 + f_2$. This lemma is, however, easily proved by a method analogous to that used later in the text.

$$\left.\begin{aligned} h(z,\chi_2^2)\,dz\,d\chi_2^2 &= \frac{1}{\left(\frac{f_1-2}{2}\right)!\left(\frac{f_2-2}{2}\right)!\,2^{\frac{f_1+f_2}{2}}}\times \\ \times\, z^{\frac{f_1-2}{2}}(1-z)^{-\frac{f_1+2}{2}}&(\chi_2^2)^{\frac{f_1+f_2-2}{2}}\exp\left[-\frac{1}{2}\frac{\chi_2^2}{1-z}\right]dz\,d\chi_2^2. \end{aligned}\right\} \quad (8.15)$$

Integrating over χ_2^2 from 0 to ∞ we obtain the distribution of z

$$\left.\begin{aligned} f(z)\,dz = dz\int_0^\infty h(z,\chi_2^2)\,d\chi_2^2 &= \\ \frac{\left(\frac{f-2}{2}\right)!}{\left(\frac{f_1-2}{2}\right)!\left(\frac{f_2-2}{2}\right)!}\,&z^{\frac{f_1-2}{2}}(1-z)^{\frac{f_2-2}{2}}\,dz \\ f = f_1+f_2&. \end{aligned}\right\} \quad (8.16)$$

In our case, we have $f_1 = 1$, $f_2 = n-m-1$ and $z = \frac{r^2}{n-m}$. Taking into account that the distribution of r is symmetric in positive and negative values, we finally obtain from (8.16) that r has the distribution

$$\boxed{\begin{gathered} f(r)\,dr = \frac{1}{\sqrt{\pi}}\,\frac{1}{\sqrt{f+1}}\,\frac{\left(\frac{f-1}{2}\right)!}{\left(\frac{f-2}{2}\right)!}\left(1-\frac{r^2}{f+1}\right)^{\frac{f-2}{2}}dr \\ |r| \leq \sqrt{f+1} \\ f = n-m-1 \geq 1 \end{gathered}} \quad (8.17)$$

which is just the distribution (5.1). For in the case of direct and equally good observations we have that the number of elements m is equal to one.

It should be noted, that the distribution of the relative errors is independent of the equations of condition, i. e. of the matrices $\boldsymbol{A}_0$ and $\boldsymbol{A}$.

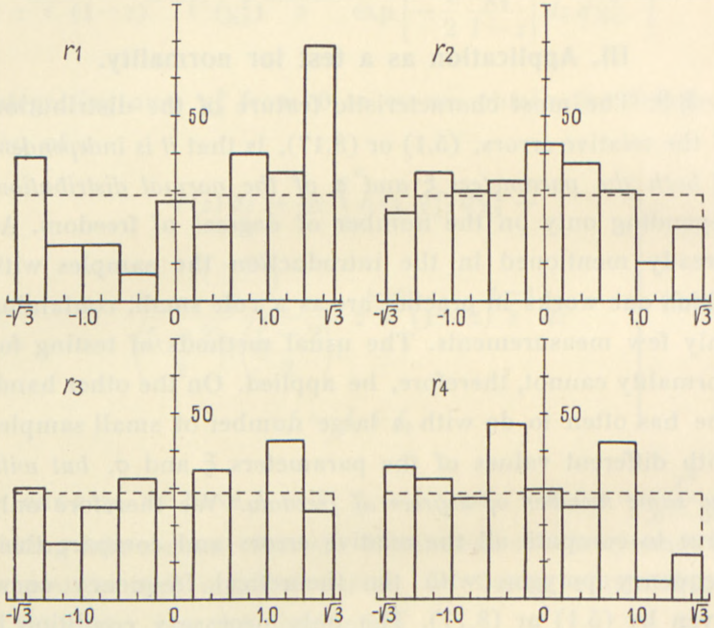
III. Application as a test for normality.

§ 9. The most characteristic feature of the distribution of the relative errors, (5.1) or (8.17), is that *it is independent of both the parameters* ξ *and* σ *of the normal distribution*, depending only on the number of degrees of freedom. As already mentioned in the introduction the samples with which one works in practice are as a rule small, containing only few measurements. The usual methods of testing for normality cannot, therefore, be applied. On the other hand, one has often to do with a large number of small samples with different values of the parameters ξ and σ, *but with the same number of degrees of freedom.* We therefore only have to compute all the relative errors and compare their frequency polygon with the theoretical frequency curve given by (5.1) or (8.17). The only necessary condition is that the number, n, of measurements in the sample is greater than or equal to 3. A more detailed comparison is obtained by comparing the total frequency polygon with the total frequency curve given by

$$F(r) = \int_{-\sqrt{f+1}}^{r} f(r)\,dr = \begin{cases} 1-\frac{1}{2}P(r) & \text{for} \quad r \geqq 0 \\ \frac{1}{2}P(r) & \text{-} \quad r \leqq 0 \end{cases} \tag{9.1}$$

where $P(r)$ is given in (5.7) and tabulated in table 1.

As a numerical example we shall consider 100 samples, each consisting of 4 measurements of the positions of spectral lines, measured by E. Rasmussen.[1] In tables 2—5, figs. 6—9 and figs. 11—14 we give the results for r_1, r_2, r_3 and r_4 respectively. The figures in the second columns,



Figs. 6—9.

p_i, give the number of relative errors in the intervals $t_i - \frac{\Delta t}{2} < t \leqq t_i + \frac{\Delta t}{2}$ with $\Delta t = 0.4$. (For the two endintervals $-\sqrt{3} \leqq t \leqq -1.4$ and $1.4 < t \leqq \sqrt{3}$ since $|r| \leqq \sqrt{4-1} = \sqrt{3}$.) The figures in the third columns, π_i, give the numbers expected from the distribution (5.1), which in our case, $n = 4$, reduces to

$$f(r)\,dr = \frac{dr}{2\sqrt{3}}. \tag{9.2}$$

[1] I wish to thank Dr. Rasmussen very much for kindly placing his measurements at my disposal.

The sum of the figures in the fourth columns is the quantity χ^2 which measures the goodness of fit.[1] In figs. 6—9 we have plotted the frequency polygon $\frac{p_i}{\Delta t}$ and the theoretical frequency curve

$$100\, f(r) = \frac{100}{2\sqrt{3}} = 28.87. \tag{9.3}$$

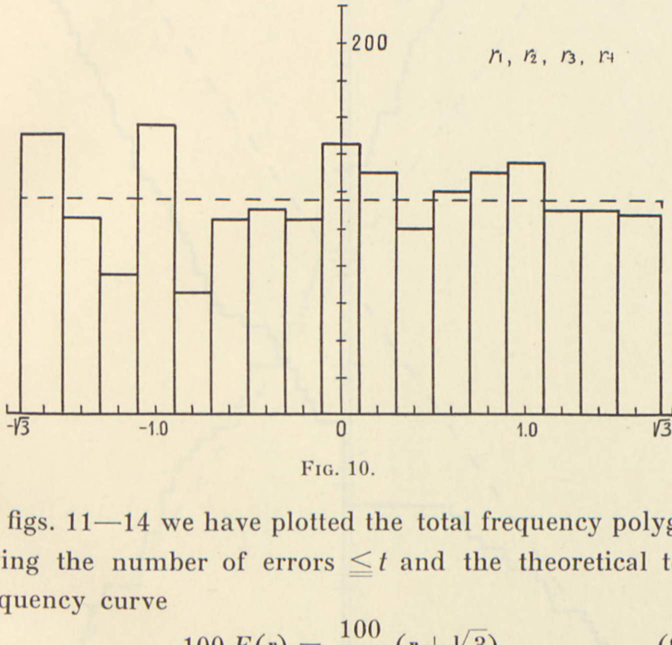


Fig. 10.

In figs. 11—14 we have plotted the total frequency polygon, giving the number of errors $\leqq t$ and the theoretical total frequency curve

$$100\, F(r) = \frac{100}{2\sqrt{3}}(r+\sqrt{3}) \tag{9.4}$$

which is here a straight line.

It is seen both from the tables and from the figs., that the agreement is satisfactory for r_2 and r_3, less satisfactory for r_4 and not satisfactory for r_1. This is also seen from the values of χ^2 and their probabilities given in table 6. It will, however, be seen that this discrepancy is due to

[1] Cf. e. g. Fisher: Statistical Methods for Research Workers. chap. IV.

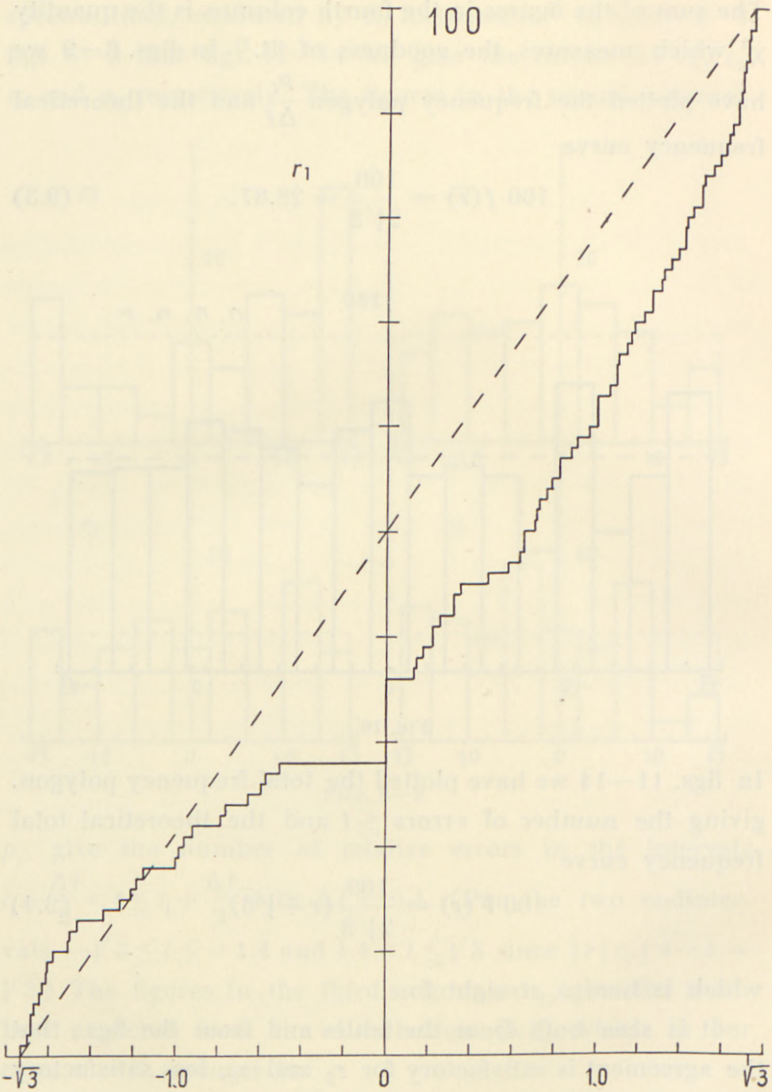


Fig. 11.

the fact, that the first measurement gives far too many *positive* errors, and the fourth one too many *negative* errors. This circumstance indicates that we have to do with a

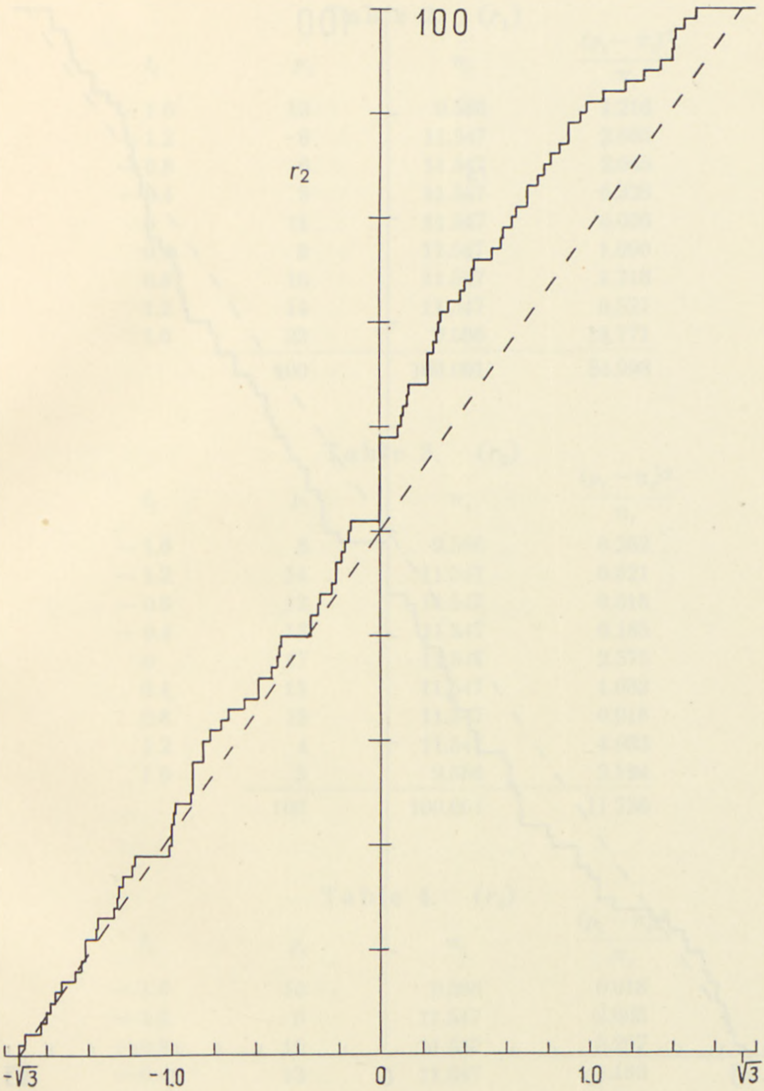


Fig. 12.

systematic error, which is easy to explain. The measurements were, namely, performed so that all the *first* figures, r_1, from each sample were obtained consecutively, then

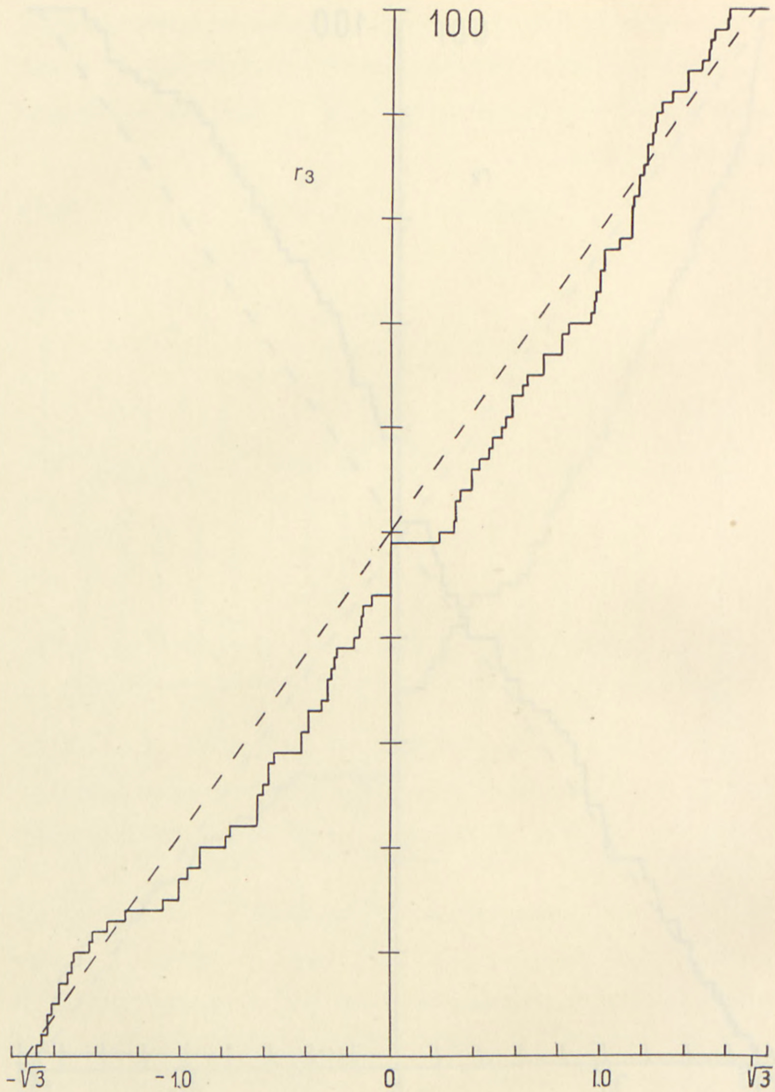


Fig. 13.

all the *second* ones, r_2, and so on. It is quite plausible that the temperature of the measuring apparatus may have changed during this process by an amount which would

Table 2. (r_1)

t_i	p_i	π_i	$\frac{(p_i-\pi_i)^2}{\pi_i}$
−1.6	13	9.586	1.216
−1.2	6	11.547	2.665
−0.8	6	11.547	2.665
−0.4	3	11.547	6.326
0	11	11.547	0.026
0.4	8	11.547	1.090
0.8	16	11.547	1.718
1.2	14	11.547	0.521
1.6	23	9.586	18.771
	100	100.001	34.998

Table 3. (r_2)

t_i	p_i	π_i	$\frac{(p_i-\pi_i)^2}{\pi_i}$
−1.6	8	9.586	0.262
−1.2	14	11.547	0.521
−0.8	12	11.547	0.018
−0.4	13	11.547	0.183
0	17	11.547	2.575
0.4	15	11.547	1.032
0.8	12	11.547	0.018
1.2	4	11.547	4.933
1.6	5	9.586	2.194
	100	100.001	11.736

Table 4. (r_3)

t_i	p_i	π_i	$\frac{(p_i-\pi_i)^2}{\pi_i}$
−1.6	10	9.586	0.018
−1.2	6	11.547	2.665
−0.8	10	11.547	0.207
−0.4	13	11.547	0.183
0	10	11.547	0.207
0.4	14	11.547	0.521
0.8	12	11.547	0.018
1.2	17	11.547	2.575
1.6	8	9.586	0.262
	100	100.001	6.656

Table 5. (r_4)

t_i	p_i	π_i	$\frac{(p_i-\pi_i)^2}{\pi_i}$
— 1.6	12	9.586	0.608
— 1.2	13	11.547	0.183
— 0.8	11	11.547	0.026
— 0.4	19	11.547	4.811
0	12	11.547	0.018
0.4	10	11.547	0.207
0.8	17	11.547	2.575
1.2	2	11.547	7.893
1.6	4	9.586	3.255
	100	100.001	19.576

Table 6.

	χ^2	$P(\chi^2)$	No. of degrees of freedom
r_1	34.998	$P < 0.001$	
r_2	11.736	$0.1 < P < 0.2$	$9 - 1 = 8$
r_3	6.656	$0.6 < P < 0.7$	
r_4	19.576	$0.01 < P < 0.02$	
r_1, r_2, r_3, r_4	16.662	$0.3 < P < 0.5$	$17 - 1 = 16$

Table 7. (r_1, r_2, r_3, r_4)

t_i	p_i	π_i	$\frac{(p_i-\pi_i)^2}{\pi_i}$
— 1.6	35	26.795	2.512
— 1.4	21	23.094	0.190
— 1.2	15	23.094	2.836
— 1.0	31	23.094	2.706
— 0.8	13	23.094	4.412
— 0.6	21	23.094	0.190
— 0.4	22	23.094	0.052
— 0.2	21	23.094	0.190
0	29	23.094	1.510
0.2	26	23.094	0.365
0.4	20	23.094	0.414
0.6	24	23.094	0.035
0.8	26	23.094	0.365
1.0	27	23.094	0.661
1.2	22	23.094	0.052
1.4	22	23.094	0.052
1.6	25	26.795	0.120
	400	400.000	16.662

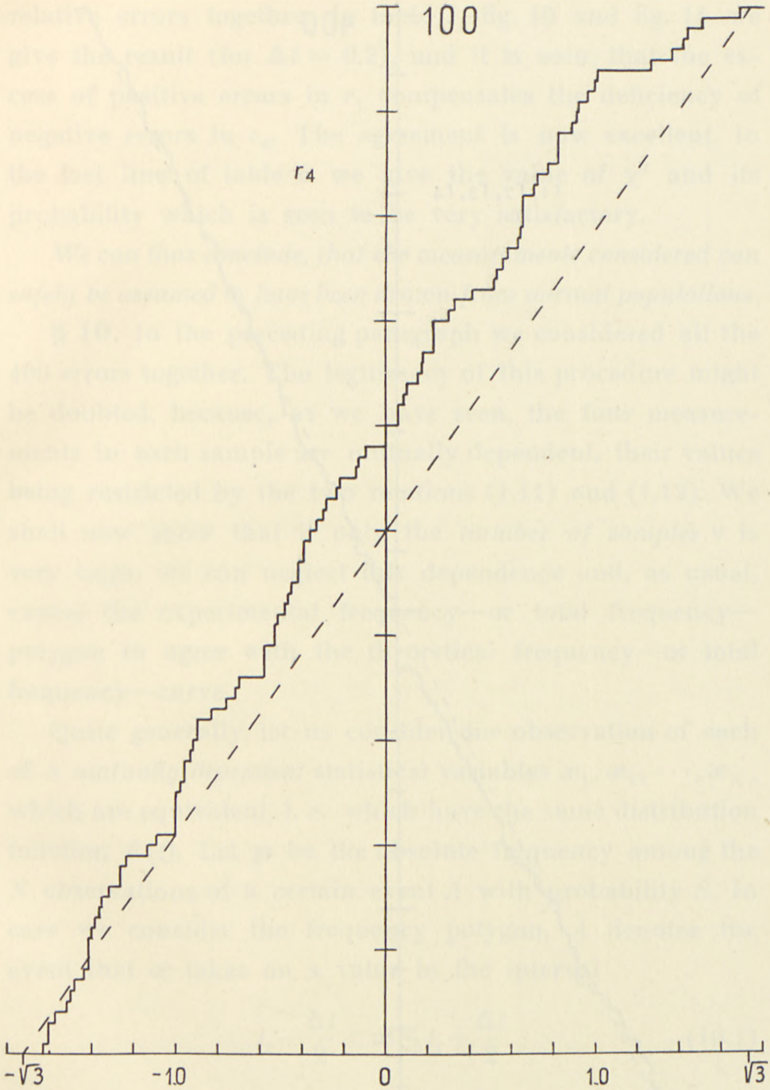


FIG. 14.

just account for the systematic error found. We conclude therefore, that the disagreement regarding r_1 is not significant. This fact is also shown if we consider all the 400

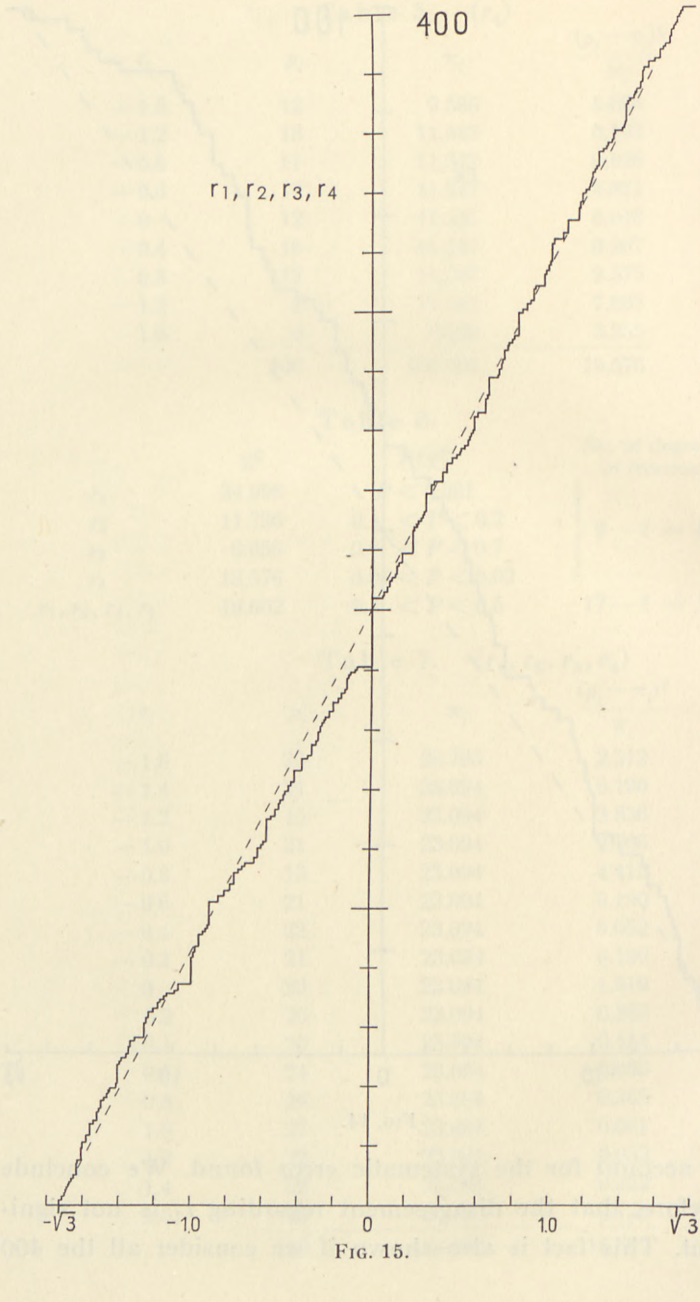


Fig. 15.

relative errors together. In table 7, fig. 10 and fig. 15 we give the result (for $\Delta t = 0.2$), and it is seen, that the excess of positive errors in r_1 compensates the deficiency of negative errors in r_4. The agreement is now excellent. In the last line of table 6 we give the value of χ^2 and its probability which is seen to be very satisfactory.

We can thus conclude, that the measurements considered can safely be assumed to have been drawn from normal populations.

§ 10. In the preceding paragraph we considered all the 400 errors together. The legitimacy of this procedure might be doubted, because, as we have seen, the four measurements in each sample are mutually dependent, their values being restricted by the two relations (1.11) and (1.12). We shall now show that if only the *number of samples* ν is very large, we can neglect this dependence and, as usual, expect the experimental frequency—or total frequency—polygon to agree with the theoretical frequency—or total frequency—curve.

Quite generally, let us consider one observation of each of N *mutually dependent* statistical variables $x_1, x_2, \cdots, x_N$, which are equivalent, i. e. which have the same distribution function $F(t)$. Let p be the absolute frequency among the N observations of a certain event A with probability S. In case we consider the frequency polygon, A denotes the event that x takes on a value in the interval

$$t_i - \frac{\Delta t}{2} < x \leq t_i + \frac{\Delta t}{2} \tag{10.1}$$

and we have, therefore,

$$S = \int_{t_i - \frac{\Delta t}{2}}^{t_i + \frac{\Delta t}{2}} dF(t). \tag{10.2}$$

In case we consider the total frequency polygon, A denotes the event that x takes on a value

$$x \leqq t \tag{10.3}$$

and we have, therefore,

$$S = \int_{-\infty}^{t} dF(t). \tag{10.4}$$

The statistical variable p can be written as

$$p = \sum_{i=1}^{N} \delta_i, \tag{10.5}$$

where $\delta_1, \delta_2, \cdots, \delta_N$ are N equivalent statistical variables, δ_i being 1 if the event A happens at the i'th observation, 0 if A does not happen. Irrespectively of whether our variables $x_1, x_2, \cdots, x_N$ are mutually independent or not, we have

$$\left.\begin{aligned} m\{p\} = m\left\{\sum_{i=1}^{N} \delta_i\right\} = \sum_{i=1}^{N} m\{\delta_i\} = \\ = N\big(1 \cdot S + 0 \cdot (1 - S)\big) = NS. \end{aligned}\right\} \tag{10.6}$$

In both cases, independence or dependence, we thus have, that the mean value of the *relative frequency* is

$$m\left\{\frac{p}{N}\right\} = S, \tag{10.7}$$

S being given by (10.2) or (10.4). On the other hand, the standard deviation of the relative frequency will not be the same in the two cases. We have

$$m\{\delta_i^2\} = S \tag{10.8}$$

and

$$m\{\delta_i \delta_k\} = S S_{ik} \tag{10.9}$$

where S_{ik} is the conditioned probability of $\delta_k = 1$ under the condition that $\delta_i = 1$. From (10.7)—(10.9) it follows, that

$$\left.\begin{aligned}\sigma^2\left\{\frac{\boldsymbol{p}}{N}\right\} = m\left\{\left(\frac{\boldsymbol{p}}{N}\right)^2\right\} - m^2\left\{\frac{\boldsymbol{p}}{N}\right\} = \frac{1}{N^2}\, m\left\{\left(\sum_{i=1}^{N} \delta_i\right)^2\right\} - S^2 = \\ = \frac{S}{N} + \frac{S}{N^2}\sum\sum_{i \neq k} S_{ik} - S^2.\end{aligned}\right\} \quad (10.10)$$

In case $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N$ are independent, we have $S_{ik} = S$ and (10.10) thus reduces to the wellknown formula

$$\sigma^2\left\{\frac{\boldsymbol{p}}{N}\right\} = \frac{S(1-S)}{N} \underset{N \to \infty}{\to} 0\,. \quad (10.11)$$

(10.11) shows, using TSCHEBYSCHEFFS inequality, that

$$\frac{\boldsymbol{p}}{N} \underset{\substack{\text{in prob} \\ N \to \infty}}{\to} S, \quad (10.12)$$

the convergence being "in probability".[1] For large values of N we can, therefore, expect agreement between experimental and theoretical curves. In case $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_N$ are dependent, the necessary and sufficient condition for (10.12) still to hold is, from (10.10), that

$$\frac{1}{N^2}\sum\sum_{i \neq k} S_{ik} \underset{N \to \infty}{\to} S. \quad (10.13)$$

This condition is, however, fulfilled in our case. We have here ν samples, each containing n observations, so that

$$N = n\nu. \quad (10.14)$$

Since each two errors from two *different* samples are independent, we have in these cases $S_{ik} = S$. Only in case we

[1] Cf. e. g. CRAMÉR: Random Variables. chap. V.

have two different errors from the *same* sample, we have $S_{ik} = S' \neq S$. Of these last cases we have in all $n(n-1)\nu$ and thus, using (10.14)

$$\left.\begin{aligned} \frac{1}{N^2}\sum_{i \neq k}\sum S_{ik} &= \frac{1}{N^2}\left(n^2\nu(\nu-1)S+n(n-1)\nu S'\right) = \\ &= \left(1-\frac{n}{N}\right)S+\frac{n-1}{N}S' \underset{N\to\infty}{\to} S. \end{aligned}\right\} \quad (10.15)$$

(10.15) shows, that (10.12) is valid also in our case, as we wanted to show.

§ 11. Theoretically the distribution of the relative errors can be used as a test for normality also in those cases, where we consider only *one* sample, containing, however, *many* measurements, i. e.

$$N = n \text{ is large.} \quad (11.1)$$

We calculate the n relative errors and expect their frequency—or total frequency—polygon to agree with the theoretical frequency—or total frequency—curve.[1] We have, namely, again the equation (10.7). In (10.10) we have now, that S_{ik} is constant for all $i \neq k$, i. e.

$$S_{ik} = S'. \quad (11.2)$$

(10.10) thus reduces to

$$\sigma^2\left\{\frac{\boldsymbol{\nu}}{n}\right\} = \frac{S}{n}+\frac{n(n-1)}{n^2}SS'-S^2 = S\left(S'-S+\frac{1-S'}{n}\right). \quad (11.3)$$

The necessary and sufficient condition for $\sigma^2 \underset{n\to\infty}{\to} 0$ is then

$$S' \to S. \quad (11.4)$$

1) For a numerical example cf. e. g. Cramér: Sannolikhetskalkylen p. 132.

This is, however, fulfilled in our case. From (4.5) (with $p = n-2$) and (2.9) we have, that the correlation function of two arbitrary relative errors, say r_1 and r_2, is given by

$$\left.\begin{aligned} h(r_1, r_2) &= \frac{1}{2\pi} \frac{n-3}{\sqrt{n(n-2)}} \left[1 - \frac{1}{n} \frac{n-1}{n-2} \left(r_1^2 + \frac{2}{n-1} r_1 r_2 + r_2^2\right)\right]^{\frac{n-5}{2}} \\ &\underset{n \to \infty}{\to} \frac{1}{2\pi} \exp\left[-\frac{1}{2}(r_1^2 + r_2^2)\right] \end{aligned}\right\} \quad (11.5)$$

(11.5) shows, that each two relative errors will be more and more independent for $n \to \infty$ and thus (11.4) follows. Since a two-dimensional normal distribution is given by

$$\frac{1}{2\pi} \frac{1}{\sigma_1 \sigma_2 \sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{r_1^2}{\sigma_1^2} - \frac{2\rho r_1 r_2}{\sigma_1 \sigma_2} + \frac{r_2^2}{\sigma_2^2}\right)\right] dr_1 dr_2 \quad (11.6)$$

(11.5) shows further, together with (1.15) and (1.16), that for large values of n, we have approximately

$$h(r_1, r_2) \cong \frac{1}{2\pi} \frac{n-1}{\sqrt{n(n-2)}} \exp\left[-\frac{1}{2} \frac{(n-1)^2}{n(n-2)}\left(r_1^2 + \frac{2}{n-1} r_1 r_2 + r_2^2\right)\right]. \quad (11.7)$$

IV. Application as a test for outlying observations.

§ 12. As already discussed in the introduction it is very important in any application of the theory of errors that false observations be rejected. *Admitting the normal law to be appropriate means, however, that the only legitimate procedure in rejecting certain observations as false is to reject them during the observations themselves,* because some peculiarities arouse suspicions as to the constancy of the conditions of the measurements or the like. A closer investigation of the conditions is, therefore, necessary in order to decide whether the figure obtained can be admitted as true or not. In cases where the observation material is schematicly treated by non-scientifically trained persons, as e. g. is often the case in ordnance, such an analysis may be difficult or even impossible to carry out in practice. Since the peculiarities mentioned will, as a rule, consist in the observations being *outlying*, i. e. that the corresponding residuals (1.3) are larger than those corresponding to the other observations, it is, thus, in practice tempting to take the *magnitude* of the residual as the only criterion for whether the observation has to be rejected or not. This procedure means, of course, artificially cutting off the tails of the distribution curve.

The question of how to obtain the limits which the errors are not allowed to exceed has, because of the arbitrariness of the whole problem, puzzled many investigators through the times, e. g. Bertrand, Peirce, Chauvenet, Stone, Vallier, Heydenreich, Mazzuoli, Rohne and many others.[1] Such schematic rules are especially employed in ordnance,[2] though their problematic nature is sometimes recognised. For instance Cranz writes:[2]

"Da das Gausssche Gesetz erst unendlich grosse Abweichungen ausschliesst, so ist von vornherein zu erwarten, dass es auf dem Standpunkt dieses Gesetzes bei der Aufstellung einer Ausschliessungsregel nicht ohne eine gewisse Willkür abgehen wird. Manche Forscher wollen auch von der Annahme jeder Regel zur nachträglichen Ausscheidung einer Beobachtung abgesehen wissen, z. B. Airy, Bessel, Faye. Manche wollen nur dann eine Beobachtung ausschliessen, wenn schon während des Versuches Verdachtsgründe sich zeigten. Indessen scheint es, dass

[1] For the history, cf. e. g. E. Czuber: Jahresber. d. deutschen Math. Ver. 7, 1899, 212 and f.

[2] "Ausreisserregeln". Cf. e. g. Kritzinger-Stuhlmann: Artillerie und Ballistik in Stichworten, p. 20 and C. Cranz: Lehrbuch der Ballistik Bd. I 5. Aufl. p. 420 and f.

speziel für die schiesstechnischen Fragen Ausreisserregeln nicht entbehrt werden können".

And KRITZINGER-STUHLMANN[1] writes:

"Man spricht von echten Ausreissern, bei denen offenbar ein Fehler während des Bestimmungsvorganges der betr. Grösse (z. B. Schussweite) gemacht wurde, und unechten Ausreissern, die nur durch eine allzustrenge Ausreisserregel von einer Verwendung bei der Bildung des Mittelwertes ausgeschlossen wurden. Viele Treffbilder sind seit Jahrzehnten auf diese Weise verfälscht worden". — — — "Dadurch werden eine Menge von Werten als Ausreisser—zu Unrecht—gebrandmarkt."

We can only agree with these remarks and again stress, as already done in the introduction, that any schematic rules have to be applied with the utmost critique and caution. Otherwise they involve the risk of discarding actually true observations and thereby giving a more or less false impression of the accuracy of the measurements. That this falsification may be dangerous, especially in small samples, is obvious.

§ 13. We shall now discuss some of the rules most often used for discarding outlying observations.[2]

From (1.1) we have that the probability $S(\rho)$ of an observation x falling in the interval

$$\xi - \rho\sigma \leqq x \leqq \xi + \rho\sigma \tag{13.1}$$

is given by

$$S(\rho) = \frac{1}{\sqrt{2\pi}} \int_{-\rho}^{\rho} \exp\left(-\frac{t^2}{2}\right) dt = \Theta\left(\frac{\rho}{\sqrt{2}}\right) = 1 - P(\rho) \tag{13.2}$$

where $\Theta(t)$ is the probability integral

1) KRITZINGER-STUHLMANN loc. cit.

2) Cf. CZUBER loc. cit. and CRANZ loc. cit.

$$\Theta(t) = \frac{2}{\sqrt{\pi}} \int_0^t \exp(-t^2)\, dt \tag{13.3}$$

and $P(\rho)$ is given by (1.7). The simplest rule for discarding outlying observations is the following which is already mentioned in § 1.

I. *An observed value is regarded as false if the corresponding* ρ *is greater than the value* ρ *corresponding to some small arbitrarily chosen probability, e. g.*

$$S = 0.999 \text{ i. e. } P = 0.001 \text{ i. e. } \rho = 3.29 \tag{13.4}$$

(cf. table 1 with $f = \infty$).

The probability of an observation falling outside the interval (13.1) is $1 - S(\rho) = P(\rho)$. Among n observations, the average number of such observations is given by

$$m\{p\} = n\left(1 - S(\rho)\right) = n\,P(\rho). \tag{13.5}$$

Chauvenet starts from the principle that for true observations

$$m\{p\} \leqq \frac{1}{2} \tag{13.6}$$

leading to the rule:

II. *An observed value is regarded as false if the corresponding* ρ *is greater than the value* ρ *given by the equation*

$$m\{p\} = \frac{1}{2} \text{ i. e. } S(\rho) = 1 - \frac{1}{2n} \text{ i. e. } P(\rho) = \frac{1}{2n} \quad (\text{Chauvenet}) \tag{13.7}$$

Vallier starts from the principle that for true observations

$$m\{p\} \leqq \frac{1}{n} \tag{13.8}$$

leading to the rule:

III. *An observed value is regarded as false if the corresponding* ρ *is greater than the value* ρ *given by the equation*

$$m\{\boldsymbol{p}\} = \frac{1}{n} \text{ i. e. } S(\rho) = 1 - \frac{1}{n^2} \text{ i. e. } P(\rho) = \frac{1}{n^2}. \quad \text{(Vallier)} \quad (13.9)$$

(Only for $n = 4$ and $n = 5$ Vallier uses the rule of Chauvenet.) Heydenreich starts from the principle that for n true observations the average value belonging to $2(n-1)$ observations

$$m_{2(n-1)}\{\boldsymbol{p}\} = 1 \quad (13.10)$$

leading to the rule:

IV. *An observed value is regarded as false if the corresponding* ρ *is greater than the value* ρ *given by the equation*

$$m_{2(n-1)}\{\boldsymbol{p}\} = 1 \text{ i.e. } S(\rho) = 1 - \frac{1}{2(n-1)} \text{ i.e. } P(\rho) = \frac{1}{2(n-1)}.$$

(Heydenreich) (13.11)

Mazzuoli starts from the principle that for true observations

$$m\{\boldsymbol{p}\} = 1 \quad (13.12)$$

leading to the rule:

V. *An observed value is regarded as false if the corresponding* ρ *is greater than the value* ρ *given by the equation*

$$m\{\boldsymbol{p}\} = 1 \text{ i. e. } S(\rho) = 1 - \frac{1}{n} \text{ i. e. } P(\rho) = \frac{1}{n}. \quad \text{(Mazzuoli)} \quad (13.13)$$

Rohne starts from the principle that an observed value must be regarded as false if its omission changes the average value by an amount greater than the probable error of the average value. If

$$\bar{x}_1 = \frac{x_1 + \cdots + x_n}{n} = \frac{v_1^{(1)} + \cdots + v_n^{(1)}}{n} + \bar{x}_1$$

and $$(v_i^{(1)} = x_i - \bar{x}_1)$$

$$\bar{x}_2 = \frac{x_1 + \cdots + x_{n-1}}{n-1} = \frac{v_1^{(1)} + \cdots + v_{n-1}^{(1)}}{n-1} + \bar{x}_1$$

are the average values before and after the omission of x_n we have, using the fact that

$$v_1^{(1)} + \cdots + v_n^{(1)} = 0,$$

$$\bar{x}_1 - \bar{x}_2 = \left(v_1^{(1)} + \cdots + v_{n-1}^{(1)}\right)\left(\frac{1}{n} - \frac{1}{n-1}\right) + \frac{v_n^{(1)}}{n} = \frac{v_n^{(1)}}{n-1}.$$

Since the probable error of the average value is equal to

$$0.67449 \cdot \frac{\sigma}{\sqrt{n}}$$

Rohnes principle leads to the rule:

VI. *An observed value is regarded as false if the corresponding* ρ *is greater than the value* ρ *given by the equation*

$$\frac{v}{n-1} = \frac{\rho\sigma}{n-1} = \frac{0.67449\,\sigma}{\sqrt{n}}, \text{ i. e. } \rho = 0.67449\,\frac{n-1}{\sqrt{n}}. \quad \text{(Rohne)} \quad (13.14)$$

Topsøe-Jensen[1] starts from the principle that for true observations the probability of either the smallest or the greatest observation falling outside the interval (13.1) shall be smaller than $\frac{1}{2}$. Since this probability is equal to the probability that at least one observation falls outside the interval (13.1), which probability, because of (13.2), is given by

$$1 - S^n(\rho), \quad (13.15)$$

this principle leads to the rule:

[1] A. G. Topsøe-Jensen: Textbook in ordnance (in Danish) § 34 d.

VII. *An observed value is regarded as false if the corresponding* ρ *is greater than the value* ρ *given by the equation*

$$1-S^n(\rho)=\frac{1}{2} \quad \text{i. e.} \quad S(\rho)=\sqrt[n]{\frac{1}{2}}. \qquad \text{(Topsøe-Jensen)} \qquad (13.16)$$

§ **14.** We shall now criticize the rules I—VII described in the preceding paragraph. Apart from the smaller or greater arbitrariness, which as pointed out above is inherent in all such schematic rules, *the most serious objection is, that they all assume the true values of the parameters* ξ *and* σ *of the normal distribution to be known.* In practice, however, we only know some *estimates*, given in (1.2) and (1.8) respectively, of these parameters, and this fact has three consequences:

1. We do not know the errors, but only the *residuals*, and these quantities have not the standard deviation σ, but the standard deviation $\sqrt{\frac{n-1}{n}}\,\sigma$ (1.5).
2. The residuals are not mutually independent, since their sum is equal to 0.
3. The estimate s of σ is itself a statistical variable, and the relative errors are, therefore, not normally distributed, but have the distribution (5.1).

Especially it follows that the numerical value of a relative error can never exceed $\sqrt{n-1}$. In table 8 we give the limits of the relative errors

$$r=\frac{x-\bar{x}}{\sqrt{\frac{n-1}{n}}\,s}=\sqrt{\frac{n}{n-1}}\,\rho \qquad (14.1)$$

given by the rules I—VII (under the assumption that $\bar{x}=\xi$ and $s=\sigma$). Comparing with the second column giving the maximum values it is seen, that in certain cases—set off in clarendon types—the rules even give limits exceeding these values!

Table 8.

n	$\sqrt{n-1}$	I	II (Chauvenet)	III (Vallier)	IV (Heydenreich)	V (Mazzuoli)	VI (Rohne)	VII (Topsøe-Jensen)
3	1.41	**3.29**	**1.69**	**1.69**	**1.41**	1.21	0.96	**1.54**
4	1.73	**3.29**	**1.77**	**1.77**	1.60	1.35	1.17	1.63
5	2.00	**3.29**	1.84	1.84	2.08	1.44	1.35	1.70
6	2.24	**3.29**	1.90	**2.40**	2.15	1.51	1.51	1.75
7	2.45	**3.29**	1.94	**2.52**	2.21	1.59	1.65	1.81
8	2.65	**3.29**	2.00	2.60	2.25	1.64	1.79	1.85
9	2.83	**3.29**	2.03	2.64	2.28	1.69	1.91	1.89
10	3.00	**3.29**	2.08	2.73	2.33	1.74	2.02	1.93
12	3.32	3.29	2.13	2.82	2.37	1.82	2.24	1.99
20	4.36	3.29	2.31	3.31	2.52	2.01	2.94	2.17

It is now clear that any reasonable and to some degree theoretically justifiable rule must take its starting point in the relative errors (1.10) and their correlation function given by (2.9), (3.11) and the last factor in (2.8). But of course one can still deduce many different rules. We think, however, that *since any such rule actually means cutting off the tails of the normal distribution it is reasonable to impose the condition that the critical limits which the rules give shall for large samples converge towards the limits given by the rule I for some value of P*. In fact, the larger the sample, the better is our knowledge about the true values of the parameters ξ and σ and the less do the residuals differ from the true errors for which the rule I is deduced. Next we think it reasonable to impose the condition that *the rules shall give us as complete a control as possible of the risk of discarding actually true observations having only fortuitously large errors*. The simplest rule fulfilling these conditions is the rule analogous to I:

An observed value is regarded as false if the corresponding relative error is greater than the value r corresponding to some small arbitrarily chosen probability, e. g.

$$S = 0.999 \text{ i. e. } P = 0.001.$$

These values of r are given in the last column of table 1 for various values of f, the number of degrees of freedom.

Comparing with the second column giving the maximum values of the relative errors, we see that this rule does not work for the two smallest numbers of degrees of freedom, 1 and 2, since in these cases the critical limits are too near the maximum ones. We think that this feature is just a sign of the soundness of the rule. For we think it impossible to draw any conclusions from such small samples whether an observation is true or false unless we have some further knowledge from previous samples of the same nature about the values of the measurements to be expected. For such small samples it is even not improbable from time to time to find relative errors *equal* to the maximum values. For instance we have 3 such values among the 400 relative errors treated in part III. We have e.g. in one case the four measurements 21790, 21789, 21789 and 21789, and certainly nobody would reject the first measurement though its relative error is equal to the maximum value $\sqrt{3}$. This fact reminds us, as already stressed in the introduction, that any schematic rule must be applied with the utmost critique and caution.

It is of course also possible to deduce rules analogous to the rules described in § 13, especially rule V and VII. In both these cases the first condition would, however, not be fulfilled, and we therefore think that the rule suggested is both the most simple and the most reasonable rule.[1)]

Summary.

In the present paper we discuss two problems in the theory of errors. The first problem is how to test whether or not a given sample of measurements has come from a normal population. The second problem is how to test whether or not an unusually large error has to be rejected as being due to some false measurements. It is pointed out that especially for small samples the distribution of the relative errors—the ratios between the deviations from the average value and the mean square error of the deviations—furnishes such tests.

In part I we first treat the case of direct and equally good observations, and next in part II the case of indirect and unequally good observations. In part I we deduce the correlation function for the free relative errors. From this function we then evaluate the frequency function of one relative error. In part II we first recall the theory of adjustment written in a matrixform. We next deduce the correlation function of the free relative errors in a way analogous to part I. Because of the complexity of the expression obtained, we deduce the frequency function of one relative error in a different way than in part I. It is shown that independently of the form of the equations of

1) After the completion of this paper my attention has been drawn to a paper by E. S. Pearson and C. Ch. Sekar: Biometrika, 28 (1936), 308 discussing this rule. They prefer a rule analogous to VII, but it seems that the conditions stated above are so reasonable as to be necessarily fulfilled. From the paper quoted it appears that the distribution (5.1) has already been deduced by W. R. Thompson: Ann. of Math. Statistics VI (1935), 214. Since this periodical is not found in any Danish libraries, we have not, however, been able to see how it is deduced there.

condition, the relative errors have the same frequency function in the two cases. This distribution deviates considerably from the normal distribution for small samples, but approaches the normal one for larger samples. The distribution is shown graphically in figs. 1—5 and is tabulated in table 1. This table gives $r = r\ (P, f)$. Here P is the probability of a relative error—shown always to be numerically smaller than $\sqrt{f+1}$ —numerically exceeding the value r. Further $f = n - m - 1$ is the number of degrees of freedom of the relative errors, n and m being the number of observations and free elements, respectively.

In part III the distribution obtained is applied as a test for normality. As a numerical example we treat 100 samples, each consisting of 4 measurements of the positions of spectral lines. The result is given in tables 2—7 and figs. 6—15. The agreement with the theoretical distribution —being in this case uniform—is shown, by means of the χ^2-test of goodness of fit, to be excellent apart from a small discrepancy, interpreted in terms of a certain systematic error. In connection with this example we deduce the conditions for the legitimacy of using observations which are mutually dependent. It is shown that if the number of the samples considered is large, the dependency is irrelevant in our case. Furthermore it is shown in this connection that the correlation function of each two relative errors approaches for large samples the normal correlation function with correlation coefficient equal to zero.

Finally in part IV we discuss the second problem, that of rejecting outlying observations. It is pointed out that admitting the normal law to be appropriate means that the only legitimate procedure in rejecting certain observations as false is to reject them during the observations themselves, on the basis of an analysis of the constancy of the conditions of the measurements. It is, however, agreed that in certain cases, as e. g. in ordnance, such an analysis may be impossible in practice and that, consequently, recourse must be had to schematic rules using the magnitude of the relative errors as the only criterion. A number of such schematic rules hitherto used, especially in ordnance, are described and criticized. Certain plausible conditions for the rules to be admitted is discussed, and finally it is shown that the distribution of the relative errors furnishes such a rule, fulfilling these conditions.

Table 1. $$P(r) = 2\int_r^{\sqrt{f+1}} \frac{1}{\sqrt{\pi}\sqrt{f+1}}\,\frac{\left(\frac{f-1}{2}\right)!}{\left(\frac{f-2}{2}\right)!}\left(1-\frac{r^2}{f+1}\right)^{\frac{f-2}{2}} dr.$$

$f = n - m - 1$. n: No. of observations. m: No. of free elements (= 1 for direct observations).

f =	$\sqrt{f+1}$	P = 0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.05	0.02	0.01	0.001
1	1.4142	0.221	0.437	0.643	0.832	1.000	1.144	1.260	1.345	1.397	1.409	1.414	1.414	1.414
2	1.7321	.173	.346	.520	.693	0.866	1.039	1.212	1.386	1.559	1.645	1.697	1.715	1.730
3	2.0000	.158	.316	.476	.639	.808	0.983	1.170	1.374	1.611	1.757	1.869	1.918	1.982
4	2.2361	.150	.300	.453	.612	.777	.952	1.144	1.360	1.631	1.814	1.973	2.051	2.178
5	2.4495	.145	.291	.440	.594	.757	.932	1.125	1.349	1.640	1.848	2.040	2.142	2.329
6	2.6458	.141	.285	.431	.583	.744	.918	1.112	1.341	1.644	1.870	2.087	2.208	2.447
7	2.8284	.139	.280	.424	.575	.734	.907	1.102	1.334	1.647	1.885	2.121	2.256	2.540
8	3.0000	.137	.277	.419	.569	.727	.899	1.094	1.329	1.648	1.895	2.146	2.294	2.616
9	3.1623	.136	.274	.416	.564	.721	.893	1.088	1.324	1.649	1.903	2.166	2.324	2.678
10	3.3166	.135	.272	.413	.560	.716	.888	1.083	1.320	1.649	1.910	2.182	2.348	2.730
11	3.4641	.134	.270	.410	.557	.712	.884	1.079	1.317	1.649	1.916	2.195	2.368	2.774
12	3.6056	.133	.269	.408	.554	.709	.881	1.076	1.314	1.649	1.920	2.206	2.385	2.812
13	3.7417	.133	.268	.406	.550	.707	.878	1.073	1.312	1.649	1.923	2.216	2.399	2.845
14	3.8730	.132	.267	.405	.550	.705	.875	1.070	1.310	1.649	1.926	2.224	2.412	2.874
15	4.0000	.132	.266	.404	.548	.703	.873	1.068	1.309	1.649	1.928	2.231	2.423	2.899
16	4.1231	.132	.265	.403	.547	.701	.871	1.066	1.307	1.649	1.931	2.237	2.432	2.921
17	4.2426	.131	.264	.402	.545	.699	.869	1.065	1.305	1.649	1.933	2.242	2.440	2.941
18	4.3589	.130	.264	.401	.544	.698	.868	1.063	1.304	1.649	1.935	2.247	2.447	2.959
19	4.4721	.130	.263	.400	.543	.697	.867	1.062	1.303	1.649	1.936	2.251	2.454	2.975
20	4.5826	.130	.263	.399	.542	.696	.865	1.061	1.302	1.649	1.937	2.255	2.460	2.990
21	4.6904	.130	.262	.398	.541	.695	.864	1.060	1.301	1.649	1.938	2.258	2.465	3.003
22	4.7958	.130	.261	.397	.541	.694	.863	1.058	1.300	1.648	1.940	2.261	2.470	3.015
23	4.8990	.130	.261	.397	.540	.693	.863	1.057	1.299	1.648	1.941	2.264	2.475	3.026
24	5.0000	.130	.261	.397	.539	.692	.862	1.056	1.299	1.648	1.941	2.267	2.479	3.037
25	5.0990	.129	.261	.396	.538	.691	.861	1.056	1.298	1.648	1.942	2.270	2.483	3.047
26	5.1962	.129	.261	.396	.538	.691	.860	1.056	1.298	1.648	1.943	2.272	2.487	3.056
27	5.2915	.129	.260	.395	.538	.690	.859	1.055	1.297	1.648	1.943	2.274	2.490	3.064
28	5.3852	.129	.260	.395	.537	.689	.859	1.054	1.296	1.648	1.944	2.275	2.492	3.071
29	5.4772	.129	.260	.395	.536	.689	.858	1.053	1.295	1.648	1.945	2.277	2.495	3.078
30	5.5678	.129	.260	.394	.536	.689	.858	1.053	1.295	1.648	1.945	2.279	2.498	3.085
35	6.0000	...	...	...	...	...	...	...	...	1.648	1.948	2.286	2.509	3.113
40	6.4031	.128	.258	.392	.534	.685	.854	1.049	1.292	1.648	1.949	2.291	2.518	3.134
45	6.7823	...	...	...	...	...	...	...	...	1.647	1.950	2.295	2.524	3.152
50	7.1414	...	...	...	...	...	...	...	...	1.647	1.951	2.298	2.529	3.166
60	7.8103	.127	.256	.390	.530	.682	.850	1.045	1.289	1.646	1.953	2.302	2.537	3.186
70	8.4262	...	...	...	...	...	...	...	...	1.646	1.954	2.306	2.542	3.201
80	9.0000	...	...	...	...	...	...	...	...	1.646	1.955	2.309	2.547	3.211
90	9.5394	...	...	...	...	...	...	...	...	1.646	1.956	2.310	2.550	3.220
100	10.0498	...	...	...	...	...	...	...	...	1.646	1.956	2.312	2.553	3.227
120	11.0000	.126	.255	.387	.528	.679	.846	1.041	1.285	1.646	1.957	2.315	2.556	3.237
∞	∞	.126	.253	.385	.524	.674	.842	1.036	1.282	1.645	1 960	2.326	2.576	3.291

TABLE OF CONTENTS

Indleveret til Selskabet den 10. Maj 1940.
Færdig fra Trykkeriet den 6. December 1940.