SKEW FREQUENCY CURVES
IN BIOLOGY AND STATISTICS

2ND PAPER

BY

DR. J. C. KAPTEYN,
PROFESSOR OF ASTRONOMY AT THE UNIVERSITY OF GRONINGEN

AND

DR. M. J. VAN UVEN,
PROFESSOR OF MATHEMATICS AT THE HIGH SCHOOL FOR AGRICULTURE,
HORTICULTURE AND FORESTRY AT WAGENINGEN.

PUBLISHED BY THE ASTRONOMICAL LABORATORY AT GRONINGEN.

HOITSEMA BROTHERS, GRONINGEN 1916.
INTRODUCTION

BY

J. C. KAPTEYN.

When — now more than 11 years ago — I published my paper on "Skew frequency curves in Biology and Statistics" (a paper which further on will be referred to simply as first paper), I felt that I would be, unable, probably for many years to come, "further to prosecute a subject, which lies somewhat far from my usual "studies" and I expressed "the hope that some mathematician may take up the task of developing the theory in a more general way". *)

The necessity of such a more general theory appeared very soon after the publication. Work done at the botanical laboratory of Groningen and elsewhere convinced me that the special form

\[ F(x) = (x + K)^2 \]

which is the only case completely worked out (p. 18 etc.), is really too restricted for the requirements of practice. I even was not long in recognising the fact that no special form whatever would be quite satisfactory and that only a wholly general development promised to be extensively applicable.

Fortunately I recognised at the same time that not only no limitation is necessary, but that the derivation of the complete course of \( F(x) \) is hardly more laborious than the working out of any particular hypothesis.

Besides this there is another point that made a new treatment of the question desirable. The equation (7) is erroneous, because in it the squares of \( \Delta x \), which were neglected, ought to have been retained.

As it did not appear that any other investigator had the intention of taking up the matter, these considerations soon led me to make occasional notes and to collect examples, in the hope that at some future time I might find time myself for a rediscussion of the whole matter. This hope was not fulfilled. Often for months I hardly found a single hour to devote to the subject and after a while I felt compelled, very reluctantly, to lay the subject definitively aside, either forever or at the least until the time of my resignation as a professor.

*) See Preface to 1st paper.
This was the state of affairs when, last year, Prof. van Uven courteously offered his cooperation. I at once gratefully accepted this generous offer.

In order not to endanger the completion of the work any more we agreed to simplify matters by leaving out all mathematical developments which, though they might offer some mathematical interest, would probably be of little importance in the study of the frequency curves offered by nature. On the same ground we resolved, later, when we experienced great difficulties in collecting numerous pregnant examples, rather to limit ourselves to what examples we had already brought together, than to delay much further the time of publication. The scarcity of good examples in literature must, I think, probably be attributed to the way in which the study of frequency curves has been conducted up to the present and I have my hopes that a somewhat extensive trial of the method presented in the following pages will soon put ample materials at our disposal. For as long as we start from the idea that all frequency curves must fall in a very limited number of mathematical types, there will naturally, consciously or unconsciously, be a strong bias in favour of cases which fit into these types, while deviating cases are apt to be neglected or to be attributed to exceptional causes. For the present form of our method this danger will not exist. It will rather encourage the investigation of peculiar forms, which present no greater difficulty than the more common forms.

In both the present and the first papers the main purpose is different from that of other studies on the subject. While the latter only try to find good interpolation formulae, the main purpose of our papers is to learn something about the effects of the causes to which any particular form of frequency curve is due.

Meanwhile the first paper is still in so far in conformity with the writings of Pearson and others, that the attempt was made — be it as a secondary aim — of bringing the observed frequency curve under a mathematical form, that is of finding for it a mathematical interpolation formula. This plan has been almost altogether given up in the present paper. The substitution of a mathematical expression, with a moderate number of constants, for the frequency curve, is necessarily equivalent to a limitation, a limitation wholly unjustified by the nature of the problem.

In the application of the method of the present paper everything is done graphically or numerically. From the graphical representation of the frequency curve, we derive the graphical representation of that function which is normally distributed. From this function we further derive graphically the reaction-curve — if need be the growth-curve. This does not prevent us from finding such quantities as the median and the quartiles.
Quite the contrary. The finding of these quantities — as indeed the whole of the discussion — becomes of extreme simplicity.

Summarising we may say that the present method is distinguished from other methods, including that of our own first paper, by its perfect generality, without loss of simplicity. From other methods, excluding that of our first paper, by its aim to learn something about causes.

The present paper is indeed quite independent of the preceding paper. Still it might be decidedly recommendable to read at least the first 15 pages of the latter before entering on the reading of the former.

For the rest I think that most of the words written in the introduction of the first paper still apply to the present one, even the student of statistics who wishes to apply the method, but finds himself unable to follow the argument of these (mathematical) articles, need not be deterred. The derivation of the theory is necessarily mathematical, its application is absolutely elementary.

The main purpose of both papers — the finding something about causes — is no doubt an ambitious one. Indeed it may be well to warn expressively against too sanguine expectations. The mathematical theory necessarily starts from certain assumptions. These assumptions are probably not or not fully realised in nature. Therefore it is impossible to say a priori in how far our theory will apply to the cases offered by nature. The main ground for not being altogether sceptical lies in the fact that a close approach to the normal curve has already been found to occur frequently. Now our theory is only as it were an extension of the mathematical theory which leads to the normal curve and this extension starts from what is certainly in innumerable cases a "vera causa" viz that the "deviations" are dependent on the size already reached by the individual. A reasoning like that of art. 9 of the first paper, shows this with perfect evidence.

Still the fact remains that the conclusions to which the theory leads must not be taken as well established facts but rather as "working hypotheses".

There is another quite different cause for not being overly sanguine.

For evident reasons a theory for the benefit of biologists would be best worked out by a biologist.

If he cannot do it, because he is but a poor mathematician, the next best thing — still not approximately equally good — would be to have the work entrusted to a biologist working in close cooperation with an expert mathematician.

About the worst possible thing will be to put the task wholly on a mathematician. Now up to a short time ago, the last case, has been that of myself, with the only exception, that I cannot even call myself a
regular mathematician. By the cooperation of Prof. van Uven, this exception at least has been removed, but still we are in the third case, the very worst of all.

In urging this point on the biological investigator, who may happen to give our method a trial, it is not our intention of invoking his clemency in judging about this study. It is rather to invoke his cooperation. If he finds some difficulty, or some point not sufficiently worked out, let him not at once throw the method overboard. It may well be only the consequence of our not being biologists, and of himself not being mathematician enough to judge about the possibility of removing his difficulties. In this way we might come at least a little nearer to the second case, the case of the close cooperation of the biologist and the mathematician.

In order to make my meaning clearer, I may perhaps quote an example of what happened in the case of the first paper.

In this paper the theory was fully worked out only for the special case

$$F(x) = (x + K)^2.$$  

This form was deemed sufficient, because it embraced all the curves tried by myself. Some investigators, however, finding that this form did not cover the facts with which they were dealing, concluded that the theory had to be rejected.

Now this conclusion is unjustified. The fact only proved that a somewhat more general treatment was necessary. As already mentioned, it was one of the main motives for undertaking the present treatment. To my regret I must say that I experienced very little of this sort of cooperation after the publication of my first paper. Criticisms have not been wanting. Quite the contrary. But they were mostly from mathematicians who evidently had studied the matter somewhat carelessly. *) The workers of the Groningen botanical laboratory only have assisted me very materially. To them and particularly to Miss Dr. Tamnes and Prof. Moll I feel deeply indebted for help and encouragement both in writing the paper and afterwards. May they extend their kind interest to the present publication.

*) This carelessness must be my excuse for not replying to most of these criticisms. In proof of it I might quote many instances. One of these may suffice for the present. An objection made either in writing or in print by the greater part of my critics, is, that in my theory only four ordinates of the given frequency curve are used in determining the constants of the best fitting curve, whereas all the ordinates are equally entitled to contribute (see for instance Koopman's inaugural dissertation (Leiden) p. 188 as also his 5th thesis). Yet the most superficial reading of my paper must convince anyone that the objection is completely unfounded.
Nothing now remains but to state the exact part that each of the joint authors took in the work. As indicated already in the heading of the two first chapters, the first treatment of the main problem is by myself; the second quite independent derivation is by Prof. van Uven. The examples given in the subsequent chapters were mostly collected by myself. Their treatment by graphical methods is entirely due to Prof. van Uven.

Groningen June 1915.
DEVELOPMENT OF THE THEORY

BY

J. C. KAPTEYN.

CHAPTER I.

1. The normal curve. Many investigations have been made about the way in which the normal GAUSSIAN frequency-curves are produced. We will simply summarize the results.

Imagine a numerous collection of $N$ individuals who began by all having the same value $x_0$ of $x$. This $x$ may represent the length or the weight or the distance from any determined origin etc. for any individual. On these individuals there come to operate, successively or simultaneously, a great number of causes $C_1, C_2, \ldots, C_n$, tending to change the $x$ of the different individuals in different ways. We will call these causes, causes of deviation.

The result is obtained that the distribution of the frequencies of the several values of $x$, for considerable values of $n$, rapidly converges to a limit, which limit is reached for $n = \infty$. It is this limiting form which is usually applied to the cases of nature, that is, it is assumed that we can, without appreciable error, put $n = \infty$. Presently we will have to consider this supposition more closely.

Adopting it provisionally, we introduce the following notations:

\[
\begin{align*}
&C_h \text{ deviation cause;} \\
&A_{h,k} \text{ deviation caused by } C_h \text{ in the } k^{th} \text{ individual;} \\
&\bar{A}_h \text{ the mean value of all the } A_{h,k}. \quad \text{(1)}
\end{align*}
\]

and let

\[
A_{h,k} = \bar{A}_h + \alpha_{h,k}
\]

as a consequence of the last supposition we have:

\[
(2) \quad \sum_{k=1}^{N} \alpha_{h,k} = 0.
\]

\[
(3) \quad \overline{\alpha_h^2} = \frac{1}{N} [\alpha_{h,1}^2 + \alpha_{h,2}^2 + \ldots + \alpha_{h,N^2}].
\]

\[
\quad \text{In what follows a dash over any quantity will denote the arithmetical mean of the whole of these quantities.}
\]

\textit{Note:}
The result of the investigation then is:

If the causes $C_h$ produce deviations which satisfy the following conditions:

a. that they are independent of each other;

b. that the $a_h$ are of the same order of smallness *) then, after the operation of all the causes, the individuals will be spread in the normal Gaussian curve

$$y = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-x_0-M)^2}$$

where $x_0$ is the size of all the individuals before any deviation has taken place. We will call it the undisturbed value of $x$. Furthermore

$$\begin{align*}
M &= \bar{A}_1 + \bar{A}_2 + \ldots + \bar{A}_n \\
\sigma^2 &= a_1^2 + a_2^2 + \ldots + a_n^2.
\end{align*}$$

In what follows we will call $\bar{A}_h$ the mean growth under the influence of cause $C_h$; similarly $\sigma^2$ will be the corresponding mean fluctuationsquare. $M$ and $\sigma^2$ will be called total growth and total fluctuationsquare.

2. On the order of the quantities $\bar{A}$ and $a$ and on the number $n$ of causes.

As was already mentioned the equation (4) was derived in the supposition of an infinite number of causes. Of course this cannot be the case of nature, but as doubtless the number of causes is generally very great and as further — even for moderate values of $n$ — the form of the frequency curve approaches very rapidly to the limit, it has been generally assumed that there can be no serious objection against putting $n = \infty$. This view, however, shall have to be modified, at least if we wish to extend the theory to the size of plants and animals or of parts thereof. The necessity of the modification is not a consequence of the deviation to be apprehended from the Gaussian exponential form and we will retain this form even where we do not take $n = \infty$. It is a consequence of the observed proportion of the constants $\sigma$ and $M$.

This is easily seen. For let us begin by really taking $n = \infty$. From the equation (4) it appears that, in order that the frequency curve be a real curve, $\sigma^2$ must be finite.

We may exclude the case $\sigma = 0$, for in this case the frequency curve will be reduced to a single point. Now, as we assume that the quantities $a$ are of the same order of magnitude, this order must evidently be that of $\frac{1}{\sqrt{n}}$, so that the quantities $a^2$ will be of the order $\frac{1}{n}$. It is true that

*) Which will not exclude that a part of them may be of a higher order of smallness. These will simply have no appreciable influence on the result.
it is allowable (see footnote preceding page) to admit for some of the \( \alpha^2 \) a value of an order smaller than \( \frac{1}{n} \). Still it is necessary — in order that \( \epsilon \) may remain finite — that the number of the quantities \( \alpha^2 \), which is of the order \( \frac{1}{n} \), remain of the order \( n \).

Of the order of the quantities \( \bar{A} \) little can be said in general. If however they are all of the same sign and of the same order and if \( M \) is finite and not zero, they are evidently of the order of \( \frac{1}{n} \).

In this case therefore we find that — in order that we may have a real frequency curve — the fluctuations must be infinitely greater than the growths.

The correctness of this at first somewhat startling result is easily illustrated by a particular example. As such a particular example let the \( \bar{A} \) be not only of the same order but also equal. Similarly let all the \( \alpha \) be numerically equal. In this case it must be evident that the growth must increase proportionally with the number of causes. On the other hand the fluctuations, which by definition are as often positive as negative, will grow (according to a well-known rule extensively used in the theory of observation errors) proportionally with the square root of the number of causes.

After the operation of \( n \) causes, therefore, the total growth will be \( n\bar{A} \), the total fluctuationsquare \( n\bar{\alpha^2} \). In order that the frequency curve produced shall be a real curve, both these quantities must be finite. Therefore both the \( \bar{A} \) and the \( \bar{\alpha^2} \) must be of the order \( \frac{1}{n} \), the \( \alpha \) themselves of the order of \( \frac{1}{\sqrt{n}} \). Consequently (if \( n = \infty \)) the fluctuations \( \alpha \) must be infinitely great as compared with the growths \( \bar{A} \).

Meanwhile we thus get into contradiction with nature. In considering the size of plants and animals or parts thereof, we have to do with oned sided deviations, i.e. with deviations all in one sense, usually the sense of growth. For instance: under the influence of certain causes some plants will grow a little, some will grow more, others will grow considerably more, but not a single individual will diminish in size. Wherever this is the case it is impossible that the fluctuations produced by any one cause be very much greater, far less infinitely greater, than the mean growth produced by that same cause.

For, a fluctuation in excess of the mean growth and in the negative sense, means a total deviation which would be negative, a case excluded a priori.
In these cases we are compelled to admit that the fluctuations must be of the same order of magnitude as the growths. But this being admitted and $n$ being still considered to be infinite, $\varepsilon$ must become infinitely small as compared to $M$. The frequency curve would thus necessarily be reduced to a single point, or in other words, all the individuals will finally have the same size.

I conclude that where the deviations are one-sided — if we are sure before hand that the deviations are one-sided — the number $n$ of causes cannot be assumed to be infinite.

We may even go one step further and say that the number of causes must be of the order of $\left(\frac{M}{\varepsilon}\right)^2$. For a finite number of causes this is of course rather a vague expression. In practice it may be taken to mean something like this, that, though this number may be anywhere between one half or double the value of $\left(\frac{M}{\varepsilon}\right)^2$, it will probably not reach one tenth or 10 times this amount.

Such a conclusion has certainly something very surprising. It may not be clear at first sight why we could not for instance, in the case of growing plants, consider every minute of rain or sunshine as a separate cause. If we could, this would of course lead us generally to admit a very high number of causes.

But some reflection will lead us to think otherwise in this matter.

The conclusion that $\varepsilon$ must grow somewhat proportional to $\sqrt{n}$, while $M$ must grow more nearly proportional to $n$, rests on the supposition that the several causes are independent of each other. This implies that if certain individuals $a$ have been benefited by the cause $C_1$ to a smaller degree than certain other individuals $b$, the case may as well be reversed for the next cause $C_2$. I mean that now the individuals $a$ must have as good a chance of being the favoured ones as the individuals $b$. But this will, I think, mostly not be the case if — as assumed just now — every minute of rain or sunshine is considered as a separate cause. If certain individuals are less favoured by one minute of sunshine than certain others, there probably is a reason for this, which will not have ceased in the next minute. When certain plants are in the shadow a first minute, they will mostly still be so during the next minute. If this happens to be the case, then we cannot — in the present theory — consider the one minute of sunshine as a separate cause, but we must take as one cause the whole of all the consecutive minutes during which the effect is constantly favourable to a determined set of individuals, less favourable to another set.
Nobody will of course expect that by considerations like these, we will be enabled to draw a sharp division line between the domains of what we have to consider as different causes.

We will not expect to see the favours of fate distributed over the individuals of a given set of plants in absolutely the same way during a certain number of minutes and then suddenly to find this distribution changed for another. We will rather expect that after some time, while the great majority are still favoured in the same way, certain individuals will begin to gain or lose in favour. That as time proceeds the number of these individuals will gradually increase, till finally we come to a time for which we may say that we have got quite a new distribution of the favourable conditions. We will then know that we have arrived in the domain of a new cause, though we will be unable to assign exactly the division line between this domain and the domain of the preceding cause.

According to this consideration it is even conceivable that, by careful observation at every moment, of the degree to which the several individuals are favoured, we could get a roughly approximate idea about the number of what we have to consider as independent causes. It can hardly be expected that some one will really undertake such observations, which in every case will be extremely long and difficult, in many cases impossible. We have therefore to stop at the rough notion we get at once from the frequency curve, which is, that the number of causes must be of the order of

\[ \left( \frac{M}{\varepsilon} \right)^2. \]

It is to be noted that the number of causes implied by this estimate — though of course far from infinite — is still as a rule not inconsiderable. For just in the case of one-sided deviations here considered, which is the case of plants and animals, we generally find that the divergences from the mean size are small as compared to the mean size itself. So for instance Quetelet finds for the length of adult Italians

\[ M = 59.0 \quad \varepsilon = 2.47. \]

We conclude that the number of causes — in the sense of the present investigation — must be of the order of 600, that is to say it may be easily 1000, but probably not 10 000.\(^*\)

\(^*\) Of course this is in the supposition that we have to do with really homogeneous material. In the case quoted it seems far more probable that we have to do with a mixture and that the size of all the individuals would have been found much more nearly equal, consequently the concluded number of causes much more considerable, if we had had before us, as tacitly assumed, a case of real „reine Linie“. 

\[ (6) \]

A normal distribution is not a rule.

\[ (7) \]

\[ (8) \]
I have been somewhat long in explaining this point, because I think no attention has been drawn to it before. For the present paper it was only necessary to point out that the evident necessity there is, in the case of plants and animals, of admitting that the quantities $\bar{A}$ and $a$ are of the same order, does not exclude them from our theory. If we had confined it to cases in which the deviations are nearly or wholly as often positive as negative, the theory at least for the skew curves would only have been slightly more simple, because in that case — the number of causes being still considered to be very great — the quantities $\bar{A}$ might have been treated as of a higher order of smallness than the quantities $a$. For the normal curves even such a simplification does not exist at all, for in the formulae (4) and (5) no supposition in regard to the order of the quantities $\bar{A}$ is required.

Remark. The condition $a$ (art. 1) that the causes of deviation must be independent, implies that the deviations must be independent of the size $x$. Meanwhile the derivation of the normal curve proves that the deviations experienced by the individuals of different size, need not be identically the same. It is only required that for individuals of different size $x$, the quantities $\bar{A}_x$ and $a_x^2$ be the same.

3. Skew curves. In the first paper p. 10 the remark was made that not only must skew curves occur in nature, but that they must be the rule. The skewness, however, may well be too small for ready detection.

The reason is that, even if certain quantities $z$ are normally distributed, the different functions of $z$ cannot be so distributed. The remark naturally leads to the following

Problem. On certain quantities $Z$ there come to operate the causes $c_a$, producing deviations $\Delta Z$, which satisfy the conditions $a$ and $b$. These deviations consequently are independent of the size $Z$.

Therefore let

$$\Delta Z = A_a = \bar{A}_a + a_a^2.$$  

According to what precedes the frequency curve produced will be normal. Now let the quantities $x$ be dependent on the quantities $Z$ according to the equation

$$Z = F(x).$$  

What will be the frequency curve of the $z$?

Solution. If the $\Delta Z$ and $\Delta x$ are corresponding deviations we will evidently have, neglecting powers higher than the second

$$\Delta Z = F'(x) \Delta x + \frac{1}{2} F''(x) \Delta x^2.$$
Solving this equation we will have, to the same degree of approximation:

\[ \Delta x = \frac{1}{F'(x)} \Delta Z - \frac{1}{2} \left[ \frac{F''(x)}{F'(x)^2} \right] \Delta Z^2. \tag{9} \]

According to what precedes the \( a_{n,k} \) will be, as a rule, of the order of \( \frac{1}{\sqrt{n}} \). The \( \bar{A}_n \) will mostly be considerably smaller. Still, according to what has been said about one-sided deviations, it will be necessary to treat the \( \bar{A} \) as quantities which may be of the same order as the \( a \). At all events we will assume that none of the \( \bar{A} \) is greater than a quantity of the order of \( \frac{1}{\sqrt{n}} \).

Accurate to quantities of the order \( \frac{1}{n} \) we will thus have:

deviation of the \( k \)th individual, at abscissa \( x \),

\[ \Delta x = \frac{A}{F(x)} - \frac{1}{2} \left[ \frac{F''(x)}{F'(x)^2} \right] A^2 \text{ or} \tag{10} \]

\[ \Delta x = \frac{\bar{A}_n + a_{n,k}}{F(x)} - \frac{1}{2} \left[ \frac{F''(x)}{F'(x)^2} \right] \left( \bar{A}_n \right)^2 + 2 \bar{A}_n a_{n,k} + a_{n,k}^2 \text{.} \tag{11} \]

According to the formulae (4) and (5) the equation of the normal frequency curve of \( Z \) is

\[ y = \frac{1}{\epsilon \sqrt{2\pi}} e^{-\frac{1}{2\epsilon^2} (Z - \mu)^2} \text{ in which} \tag{12} \]

\[ M = \Sigma \bar{A}_n = \text{total mean growth of the quantities } Z \]

\[ \epsilon^2 = \Sigma a_{n,k}^2 = \text{total fluctuationsquare of the } Z. \tag{14} \]

Now it is evident that, \( x \) and \( Z \) being corresponding quantities, frequency \( Z \) to \( Z + dZ = \) frequency \( x \) to \( x + dx \).

Therefore, if \( y = \Omega(x) \) represents the frequency curve of the \( x \)

\[ \Omega(x) dx = \frac{1}{\epsilon \sqrt{2\pi}} e^{-\frac{1}{2\epsilon^2} (x - \mu)^2} dZ \]

for which equation, because \( Z = F(x) \) and \( dZ = F'(x) dx \), we may write (dividing by \( dx \))

\[ \Omega(x) = \frac{F'(x)}{\epsilon \sqrt{2\pi}} e^{-\frac{1}{2\epsilon^2} (F(x) - \mu)^2} \text{ in which equation the meaning of } M \text{ and } \epsilon \text{ is still given by (13) and (14).} \tag{15} \]

**Remark 1.** As has already been remarked the term with \( \Delta Z^2 \) in (9) was erroneously neglected in my first paper. In the preceding article it was shown that such a course is inadmissible, the reason being that if the \( \Delta Z^2 \), therefore also the \( a^2 \), were really negligible, then we would find...
...of approxim-
...ly (14) that \( z^2 \) would be zero. The neglect of the term in question, is a single point.

The error may be made apparent in quite another way, by showing indeed in this way that my attention was drawn to it.

As the quantities \( F(a) \) were assumed to be normally distributed, we will have for the medians \( c(y) = \frac{1}{2} \), that is for the value of \( z \), above and below which the numbers are equal:

\[
F(\pm \frac{1}{2}) = 0.6826,
\]

\[
F(\pm 1) = 0.9544.
\]

Consequently, the arithmetical mean of all the \( z \)'s cannot be changed. But at starting the size of all the individuals is \( z_0 \). Therefore, it must be equal to the arithmetical mean, not to the median.

Remarks 2. In general, the deviations \( \Delta z \), as they are assumed to...
have form (10), though of course they may be symmetrical for any particular value of \( x \), cannot be symmetrical for the individuals of every size. As a consequence we cannot expect \( x_0 \) to be equal to the arithmetical mean but in particular cases.

The only particular cases, where there is a possibility of symmetrical deviation for all the individuals are the two following:

\[
F(x) = a + bx \\
F(x) = h \log (a + bx).
\]

For the proof see Appendix I where at the same time it is shown that in these cases we have really \( x_0 = \text{arithm. mean} \).

Remark 3. The solution of the main problem of this article is evidently equivalent with that of finding the frequency curve produced in the case of deviations of the form (10). But in this form the problem may seem to be lacking in plausibility. It would seem to be much more natural to inquire what would be the frequency curve in the case that the deviations were of the form

\[
A \\
F''(x)
\]

This greater naturalness, however, is only apparent as will become evident, if we fix our attention not on the deviations but on the intensity with which the individuals of the size \( x \), react on a given cause. For, where there are at work causes on which the individuals of size \( x \) react with an intensity proportional to

\[
1 \quad F'(x)
\]

there we will in reality get deviations of the form (9). Under the action of such causes the deviations would be of the form

\[
\Delta x = A \\
F''(x)
\]

only in the case that we might neglect terms of the second order. But we know that this is not allowable. If therefore, we admit higher powers, we shall have to consider that, as soon as, by the beginning deviation, the size of an individual, which originally was \( x \), has changed a little, say to size \( x + \Theta \), then the reaction of the cause will no longer be proportional to \( 1/F'(x) \) but to \( 1/F'(x + \Theta) \).

Now, neglecting \( 2^d \) and higher powers of \( \Theta \),

\[
1 \\
F'(x + \Theta) = 1 \\
F'(x) - \Theta \\
F''(x)
\]

which It

In

case the
tant an
case th

or as

higher

b.

negligible

problem

4. single case

Ac

(accurdi

(19) .

Fur
In order to find the total deviation we will have to find the average value of this expression. This is obtained by putting \( \Theta = \frac{1}{2} \Delta x \), or as (17) gives the value of \( \Delta x \) accurate to first powers of \( A \),

\[
\Theta = \frac{1}{2} \frac{A}{F'(x)} + \text{terms in } A^2, A^3, \ldots \text{ so that finally, neglecting } 3^4 \text{ and higher powers of } A \text{ we have}
\]

\[
\Delta x = \frac{A}{F'(x)} - \frac{1}{2} \frac{F''(x)}{[F'(x)]^3} A^2
\]

which is just the form (10).

It thus becomes clear that the finding of the frequency curve for the case that the deviations have the form (10) is really a much more important and natural problem than the finding of the frequency curve for the case that the deviation would have the form (17), a fortunate circumstance because the latter problem must be much the more difficult of the two. At least some trials made by myself have not been successful.

For the rest it may be well to remark:

a. that the very frequent occurrence of normal, or at least approximately normal curves, leads very naturally to the suspicion that in those cases where we meet with decidedly skew distributions, this skewness may be attributable to the fact that we did not measure the most suitable quantities; that the normal curve would duly have made its appearance, had we measured other quantities, which are functionally connected with the observed ones, — if for instance we had measured surfaces or volumes instead of diameters. Considerations like these lead of course immediately just to the problem investigated in what precedes.

b. As will appear further on, the second term of (10) must be almost negligible in many cases of nature. In such cases of course the two problems become identical.

4. Mean growth and mean fluctuationsquare under the influence of a single cause \( C_a \).

According to (11) we have — the mean value of \( a_{h,k} \) being zero (according to (21)) —

\[
\bar{A}_x = \frac{A_1}{F'(x)} - \frac{1}{2} \frac{F''(x)}{[F'(x)]^3} \left\{ (\bar{A}_h)^2 + \bar{a}_{h}^2 \right\}.
\]

Furthermore, the divergence of any arbitrary \( \Delta x \) from the mean \( \bar{x} \) of all, is

\[
\Delta x - \bar{x} = \frac{a_{h,k}}{F'(x)} - \frac{1}{2} \frac{F''(x)}{[F'(x)]^3} \left\{ 2 \bar{A}_h a_{h,k} + a_{h,k}^2 - \bar{a}_{h}^2 \right\}
\]
Therefore, to the same degree of approximation as before,

\[(\Delta z - \Delta x)^2 = \frac{a_{h,k}^2}{[F'(x)]^2}.\]

Consequently

\[(20) \quad \text{Mean fluctuationsquare at abscissa } x, \text{ cause } C_h = \frac{a_{h,k}^2}{[F'(x)]^2}.\]

5. **Mean growth and mean fluctuationsquare under the influence of the whole of all the causes.**

If now we take the mean of all the expressions (19) and (20), for the whole of the causes \(C_h (h = 1, 2, 3 \ldots n)\) we get by (5), if for the sake of brevity we put

\[(21) \quad \Sigma (\bar{A}_h)^2 = B \varepsilon^2,\]

Mean growth at abscissa \(x = \frac{1}{n^2} \frac{M}{F'(x)} - \frac{(1 + B)^2}{2n} \frac{F''(x)}{[F'(x)]^3},\)

Mean fluctuationsquare = \(\varepsilon^2 \frac{1}{n} \frac{M}{[F'(x)]^3} \)

or, if we put

\[(22) \quad \frac{M}{n} = K,\]

\[(23) \quad \text{Mean growth at abscissa } x = \frac{1}{M} \frac{1}{[F'(x)]^3} - \frac{(1 + B)^2}{2} \frac{F''(x)}{[F'(x)]^3} K.\]

\[(24) \quad \text{Mean fluctuationsq. abscissa } x = \frac{\varepsilon^2}{M} \frac{1}{[F'(x)]^3} K.\]

For the case \(M = 0\), we will rather put

\[(25) \quad \frac{\varepsilon^2}{n} = H.\]

For this case therefore we get

\[(26) \quad \text{Mean growth at abscissa } x = -\frac{1}{2} H (1 + B) \frac{F''(x)}{[F'(x)]^3},\]

\[(27) \quad \text{Mean fluctuationsq. abscissa } x = \frac{H}{[F'(x)]^3}.\]

**Remark.** According to (5) and (21)

\[\varepsilon^2 = \Sigma \alpha^2, \quad B \varepsilon^2 = \Sigma (\bar{A})^2.\]

According to what has been said in art. 2, wherever the number of causes is very great, the \(\bar{A}\) must be small as compared to the \(\alpha\). In general, therefore, \(B\) will probably be a very small positive quantity. In the case of one-sided deviations, however, we are compelled to admit that the \(\alpha\) and the \(\bar{A}\) are of the same order. Still it seems natural to think that
even here the $\lambda$ must mostly be rather smaller than the $\alpha$. Particularly so in the case that $\frac{M}{\varepsilon}$ is unusually small. For this quantity can become small for two reasons: 1st because the number of causes, which according to the same article must be of the order of $\left(\frac{M}{\varepsilon}\right)^{1/3}$, is small; 2nd because the growths $\lambda$ are somewhat small as compared to the $\alpha$. — If now we find a case in which $\frac{M}{\varepsilon}$ is exceptionally small, whereas there seems no a priori reason to think that the number of causes is particularly small, we will be led to suspect the existence of the second cause. — In short there seems to be every reason to admit that the quantity $B$, which must generally be very small, must, even in the case of one-sided deviations be mostly smaller than unity. It must be expected to be particularly small wherever the value of $\frac{M}{\varepsilon}$ is little considerable.

6. Inverse Problem.

Given that for certain quantities $x$ we have found by observation the frequency curve

$$y = \Omega(x).$$

Required 1st to find a function $F(x)$ which is normally distributed.

2nd, the growth-curve, fluctuation-curve and reaction-curve.

We call growth — resp. fluctuation — and reaction-curve, the curves of which the ordinates are proportional with the mean growth resp. the square root of the mean fluctuationsquare and the intensity with which the individuals react on the causes of deviation.

Solution. From what precedes it appears that if the deviations of certain quantities $x$ have the form (10), with which corresponds an intensity of reaction proportional to (18), and if, as a consequence thereof the mean growth and the mean fluctuationsquare have the values (23) and (24), resp. (26) and (27), that then the deviations of the quantities $Z = F(x)$ become independent of $x$, as a consequence whereof the $Z$ will be distributed in a normal curve, whereas the $x$ will be distributed in a frequency curve of the form (15).

May we conclude that the inverse holds too, i.e. may we conclude that if the equation of the given frequency curve has been brought under the form (15), the quantities $Z = F(x)$ will be distributed in a normal curve?

Such would be the case if we might conclude that the deviations $\Delta Z$ are independent of the $Z$. But cannot other deviations than those which are independent of the $Z$ also produce a normal frequency curve?

This possibility really exists, at least if we assume that growth and
fluctuation are independent of each other. So for instance it is easily proved that a normal curve will be produced in the case that the growth (but not the fluctuation) is a linear function of the $Z$, and even this does not seem to be the most general case. In the above assumption therefore, if we find the $Z$ distributed in a normal curve, we cannot conclude that the $\Delta Z$ are independent of the $Z$ and, as a consequence thereof, we have to admit that with any given skew frequency curve may correspond more than one form of growth — and fluctuation — curve.

Meanwhile it is hardly conceivable that our assumption holds in nature. The subdivision of the deviation into a mean growth and a fluctuation is purely artificial and merely introduced for the convenience of the mathematical discussion. Their independence therefore seems inadmissible. What we have to expect is that the intensity of the reaction of the individuals of different size $x$, will be a function of $x$. With this one function there will correspond a determined mean growth and a determined mean fluctuation (a constant factor being disregarded). With this one function there will also correspond a single solution for the frequency curve. This appears from the solution given in the second chapter by Prof. van Uven.

Admitting therefore that with one reaction curve there corresponds but one frequency curve, the solution of the inverse problem, now under consideration becomes evident.

If by observation, we have found for the quantities $x$, the frequency curve

$$y = \Omega(x)$$

and if we have determined $F(x)$ from the equation

$$\Omega(x) = \frac{F'(x)}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{F(x) - F(x_0)}{\sigma^2} \right)^2}$$

— a determination which we will have to consider presently — then the quantities $F(x)$ will be distributed normally and the mean growth and mean fluctuationsquare will be determined by (23) and (24) resp. by (26) and (27), whereas the intensity of the reaction will be proportional to (16). The equations of the growth — fluctuation — and reaction-curve will be:

$$
\begin{align*}
\text{growth-curve} & \quad y = \frac{1}{F'(x)} - \frac{1}{2} \frac{(1 + B) e^x}{M} F''(x) \\
\text{fluctuation and reaction-curve} & \quad y = \frac{1}{F'(x)} \\
\text{if } M & \neq 0
\end{align*}
$$

respectively

$$
\begin{align*}
\text{growth-curve} & \quad y = -\frac{F''(x)}{[F'(x)]^2} \\
\text{fluctuation and reaction-curve} & \quad y = \frac{1}{F'(x)} \\
\text{if } M & = 0,
\end{align*}
$$

in which

$$
(32)\quad
$$

It is

neglected

fluctuations

have to

in mind the

frequencies

we may

For the

integration

which

A

By its

It

for the

directly

If

(35),

we

$$
(37)\quad
$$
in which we have always, according to (12) and (13)

\[ M = \sum \bar{A}_h = \text{total growth of } F(x) \]

\[ E^2 = \sum a_h^2 = \text{total fluctuationsquare of } F(x). \]

It is to be noted that both in the equations (30) and in (31) we have neglected a constant factor which is not the same for the growth and the fluctuation. If we wish to have the true proportion of the two we shall have to go back to the equations (22) to (26).

7. Derivation of \( F(x) \) from \( \Omega(x) \).

Let \( x = \tau \) and \( x = \nu \) represent the lower and upper limit of the given frequency curve. In order that the quantities \( F(x) \) be normally distributed we must have, to begin with:

\[ F(\tau) = -\infty; \quad F(\nu) = +\infty. \]

For other values of \( x \) we will find \( F(x) \) if we multiply (29) by \( dx \) and integrate between the limits \( \tau \) to \( x \). We get

\[
\int_{\tau}^{x} \frac{\Omega(x) \, dx}{\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}} \int_{\tau}^{x} F'(x) e^{-\frac{1}{2}(F(x) - F(x_0) - M)^2} \, dx
\]

which reduces to

\[ \int_{\tau}^{x} \Omega(x) \, dx = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2} dt. \]

A table for the integral \( \frac{1}{\sqrt{\pi}} \int_{-\infty}^{R} e^{-t^2} \, dt \) has been given in the first paper.

By its aid we get at once, for every value of \( x \),

\[ f(x) = \frac{1}{\sqrt{2}} \left[ F(x) - F(x_0) - M \right]. \]

It is clear that since the \( F(x) \) are normally distributed, the same holds for the \( f(x) \). It will be convenient therefore to take \( f(x) \), which is directly and completely given by the observed frequency curve \( y = \Omega(x) \), for the required function, which is normally distributed.

If we act in this way and if we remark that by putting \( x = x_0 \) in (35), we get

\[ \frac{M}{\sqrt{2}} = -f(x_0), \]

we find that finally the solution of the present question comes to this:

From the observed function \( \Omega(x) \) derive \( f(x) \) by

\[ \frac{1}{\sqrt{\pi}} \int_{-\infty}^{f(x)} e^{-t^2} \, dt = \int_{\tau}^{x} \Omega(x) \, dx. \]
The \( f(x) \) will be distributed in the normal frequency curve

\[
(38) \quad y = \frac{1}{\sqrt{\pi}} e^{-x^2}.
\]

According to (30) and (31) we will further have, expressed in \( f(x) \) and leaving out constant factors:

\[
(39) \quad \begin{align*}
\text{Growth-curve } y &= \frac{1}{f'(x)} + \frac{1 + B}{4 f(x_0) [f'(x_0)]^3} f(x_0) \not= 0 \\
\text{fluct. & react. curve } y &= \frac{1}{f'(x)}
\end{align*}
\]

\[
(40) \quad \begin{align*}
\text{Growth-curve } y &= -\frac{f''(x)}{[f'(x)]^3} \quad f(x_0) = 0 \\
\text{fluct. & react. curve } y &= \frac{1}{f'(x)}
\end{align*}
\]

Remark. The second term in the growth-curve (39) may cause trouble on account of the generally unknown factor \( \frac{1 + B}{f(x_0)} \).

Let us first assume that \( x_0 \) is known from another source, and consider separately the cases of one-sided and not one-sided deviations.

a. Where, as with plants and animals, the deviations are one-sided and in the sense of positive growth, \( x_0 \) cannot exceed \( r \), it may at most be equal to \( r \). As a rule the limit of an observed frequency-curve cannot be assigned with any precision. In these cases it seems advisable to me to take \( r = x_0 \). It will follow, according to (35), that \( f(r) \) is not \( -\infty \), but

\[
(41) \quad f(r) = f(x_0) = -\frac{M}{\epsilon V^{\frac{1}{2}}}
\]

This being so, the value of the integral

\[
\int_r^v f'(x) e^{-\eta^2 x^2} \, dx
\]

will no longer be equal to unity, but will differ from it by the amount

\[
(42) \quad \frac{1}{\sqrt{\pi}} \int_{-\infty}^{-\infty} e^{-\epsilon^2 \xi} \, d\xi.
\]

This amount will be necessarily accumulated at the lower limit. We will have to revert to this question of accumulation at the limits of the frequency-curve.

It will be sufficient to remark in connection with the case under
consideration, that such an accumulation is readily conceivable. The whole of the individuals begin by being accumulated at \( x = x_0 \). Where the deviations are one-sided they may have any values between zero and any small positive value. We are even compelled to admit values down to zero, in the case, here assumed, that \( \tau = x_0 \).

The consequence must be that after the operation of a certain number of causes, some individuals will still have size \( x_0 \). The number of individuals being sufficiently great, such must even be the case after the operation of a very great number of causes.

Still we do not deny the difficulty, or even in many cases impossibility, both of the assumption \( \tau = x_0 \) and of a finite accumulation at this limit. The difficulty however, is of the same nature as that which we meet in the case of the normal curve which extends from \(-\infty\) to \(+\infty\), so that a certain probability is attributed to any size whatever, both positive and negative though in reality sizes beyond a certain amount are never observed. The difficulty arises from the fact that this curve presumes an infinity of causes, whereas in nature this number is necessarily limited. It is not felt as an objection, however, because the chances found for the extreme sizes, are generally so small that they must escape notice.

The indeterminateness in the growth-curve as a consequence of the unknown factor
\[
\frac{1 + B}{4f(x_0)}
\]
of the second term, will mostly not be very serious. For as \( f(x_0) \), though perhaps not \(-\infty\), will doubtlessly generally have a considerable negative value and as \( B \) will mostly be moderate, particularly where \( \frac{M}{e} \) therefore \(-f(x_0)\) is smaller than usual (see Remark to art. 5), the total neglect of the second term in (39) must not, as a rule, change the growth curve to such an extent that its main features would be obliterated. There is one circumstance which still further reassures us about the conclusions to be drawn if we use
\[
y(x) = \frac{1}{f'(x)}
\]
instead of the more complete equation (39) as the equation of the growth-curve. It is this. The complete equation may be written in the form
\[
y(x) = \frac{1}{f'(x)} \left[ 1 - \frac{1 + B}{4f(x_0)} \frac{d}{dx} \frac{1}{f'(x)} \right].
\]

Therefore the second term vanishes absolutely in the points where the curve (41) has its maxima and minima. These points, therefore,
which in most cases are the really interesting points of the curve, would not be changed even were it possible to use the rigorous form. (The abscissa of the maxima and minima may of course be slightly changed).

b. In the case of not one-sided deviations, \( x_0 \) will lie between the limits \( r \) and \( v \) of the frequency curve, \( x_0 \) being known from other sources, \( f(x) \) will therefore be determined at the same time with the whole course of \( f'(x) \). The constant \( B \) will be negligible in this case, at least if we may admit that the number of causes is very great. In conclusion, therefore, the approximate determination of the growth curve will also not present any difficulty in this case.

If \( x_0 \) be not given a priori, the growth curve cannot be determined with any precision in this case. We shall have to rest content with the reaction curve. Meanwhile we may here refer to the theory of the proportional curves further below, from which it appears that in some cases at least we might get indications about the value of \( x_0 \).

8. Conclusions to be drawn from the observed frequency curves.

The equations (36) to (40) enable us from given frequency curves to draw some conclusions about the intensity with which individuals of different size \( x \) react on the growth-causes, and about their mean growth itself.

Of course we have not to forget that the hypotheses which lie at the foundation of the theory may partly or wholly not be realised in the cases of nature under investigation. For this reason the conclusions we will draw will have no absolute cogency. They ought to be taken as working hypotheses, hypotheses which draw the attention to certain more or less probable facts. The present theory claims no other advantage than this. How great the probability of the results is, what therefore must be the value of the working hypotheses at which we arrive, must appear from long experience.

The nature of the conclusions to which we may be led has already been set forth in art. 23 of our first paper on skew curves. It seems reasonable to expect that — if the theory gets a much more extensive trial than we were able to give it — results in other domains and of another nature will be found. Meanwhile there is one sort of conclusions, well illustrated by the application given below, to which we wish here to draw particular attention.

If for certain values of \( x \) the reaction becomes small, then, according to what precedes, \( \frac{1}{f'(x)} \) will be small, and \( f(x) \) will be considerable. If, therefore, \( \xi \) be a small quantity \( f(x + \xi) \) will differ relatively much from \( f(x) \).
$f(x)$. Therefore the number of individuals which, in the frequency-curve, will have a size between $x$ and $x + \xi$, that is

$$\frac{1}{\sqrt{\pi}} \int_{x+\xi}^{f(x)+D} e^{-t^2} dt$$

will also be relatively high.

We thus reach the conclusion that, wherever the reaction on the growth causes becomes small, there we will find accumulation in the frequency curve and conversely, wherever in the frequency-curve we have exceptional accumulation, there we must conclude to relative rest in the growth.

A fine example of this phenomenon is furnished by the case of the spores of Mucor Mucedo. The enormous accumulation of individuals not far from the middle of the frequency curve indicates at once a stagnation in the growth of the individuals near the time at which the size, determined by the minimum of the reaction curve, is reached. Shortly after this conclusion was reached, my attention was drawn by Prof. H. de Vries to the investigation of Prof. Errera (Botan. Zeitung Vol 42 (1884) p. 497) who found a period of rest in the growth of the sporangia of some of the fungi of the same Genus.

The probability therefore seems to be very great that direct observation, not of the sporangia but of the spores, will confirm our conclusions.

If we find in the frequency curve not an accumulation but a depression, we will similarly conclude to a high degree of reaction for the individuals having the size at which the depression occurs. A good example is that of the ear-length of wheat under scanty feeding, given by Dr. C. de Bruyker (Handelingen van het 13de Vlaamsche Natuur-genootschap, p. 172). If we neglect a pretty insignificant top, the curve has two very decided maxima. Of such two-topped curves, there are at present a fair number of examples in botanical literature. They are considered as an indication that we have to do with hybrids, descended from two parental forms having very different frequency curves.

The present theory furnishes another "working hypothesis", which may be valuable in those cases where it may be considered probable or certain that we are not concerned with hybrids.

In the present instance we are led to the conclusion of a much accelerated growth of our individuals at about the time that a size of 45 to 85 millimeter is reached. Observations made for the express purpose of testing this conclusion shall have to decide whether our explanation is valid or not.

---

1) Perfectly analogous results were found for the spores of Mucor Mucedo. I owe both series of observation to Mr. G. Postma, who at the time worked as a student in the botanical Laboratory of Groningen.
9. Accumulation at the limits.

Accumulation may well occur at or in the immediate neighbourhood of the limits. It will do so if at these points the reaction curve stops very suddenly, that is, if the ordinates of that curve, from being moderately great very near the limits become zero at these points.

This is easily seen. Between the limits $\tau$ and $v$ and up to any finite distance from them $\frac{1}{f'(x)}$ can never be zero. For $\frac{1}{f'(x)}$ is proportional to the reaction. Therefore, as all individuals begin by having size $x_0$, as soon as by continued deviation they reach a point for which $\frac{1}{f'(x)} = 0$, all further deviation stops, so that no individual can pass that point, which thus of necessity becomes a limit of the frequency-curve. Therefore $\frac{1}{f'(x)}$ must be finitely different from zero for all values of $x$ at finite distance from the limits. This being so we have $f'(x)$ and, as a consequence thereof, also $f(x)$ finite for all values of $x$ at all finite distances from the limits. All this will hold in every case.

Now let us assume that we have the case of a reaction curve which comes suddenly to a stop. Our theory will still hold provided we admit that the vanishing of $\frac{1}{f'(x)}$ from a finite value to zero does not occur with absolute suddenness but that the change occurs gradually and in a way satisfying a certain condition, within a small interval $x = \tau$ to $x = \tau + \xi$ resp. $x = v - \theta$ to $x = v$, which we will assume to be infinitesimally small.

In this case therefore we have: $\frac{1}{f'(x)}$ finitely different from zero for $\tau + \xi \leq x \leq v - \theta$

therefore $f'(x)$, consequently also $f(x)$, finite between these same limits.

Therefore, finally, we will have, among a total number $N$ of individuals:

\[
\begin{align*}
\text{numb. of indiv. between } & x = \tau \text{ and } x = \tau + \xi \\
& = \frac{N}{\sqrt{\pi}} \int_{\tau}^{\tau + \xi} f'(x) e^{-U(x)} dx = \frac{N}{\sqrt{\pi}} \int_{-\infty}^{(\tau + \xi)} e^{-t} dt
\end{align*}
\]

\[
\begin{align*}
\text{numb. of indiv. between } & x = v - \theta \text{ and } x = v \\
& = \frac{N}{\sqrt{\pi}} \int_{v - \theta}^{v} f'(x) e^{-U(x)} dx = \frac{N}{\sqrt{\pi}} \int_{f(v - \theta)}^{\infty} e^{-t} dt
\end{align*}
\]

1) The modification required by the theory for the case of an absolutely sudden breaking off of the reaction curve is easily made in a particular case. I have not succeeded in solving it generally. In nature the case can, I think, hardly be expected to exist.

2) The condition is that the ordinates must diminish in such a way that the greatest deviation for any individual of size $x$ remains constantly smaller than the distance which still separates it from the limit, or in other words that every individual continually approaches the limit without ever reaching it.
in which expressions \( f(r + \xi) \) and \( f(v - \theta) \) are both finite. The integrals therefore have also finite values, that is we have finite accumulation of individuals within the infinitesimal intervals 

\[ r \text{ to } r + \xi, \quad v - \theta \text{ to } v. \]

To my regret I have not found in literature any case of such accumulations at the limit. Still there seems to be no doubt but that such cases must exist. Imagine a number of plants of one species growing in a flat topped greenhouse. As soon as the plants have reached a size equal to the height of the green-house, further growth becomes impossible. The reaction curve comes suddenly to a stop. Corresponding therewith we will find an accumulation of individuals with a size just equal to that of the green-house.

It seems probable that many cases of impediments against growth beyond a certain size must exist in nature — though generally the limit may not be so sharply determined as in the preceding instance. In such cases attention will be called to such impediments by more or less evident accumulations at the limit of the frequency curve.


What becomes of the frequency curves:

a. if the reaction on every one of the acting causes becomes \( \lambda \) fold.

b. if — the average reaction or deviation remaining equal — the number of causes grows in the proportion of 1 : \( \lambda \)?

We will call the curves in respect to the original one, proportional curves of the first resp. the second kind. In regard to those of the first kind the reaction, which originally was \( F'(x) \), now becomes \( \lambda \frac{\text{const.}}{F'(x)} \). In order to pass from the original curve to the proportional one we have therefore only to substitute \( \frac{1}{\lambda} F'(x) \) to \( F'(x) \). According to (11) this comes to the same as if — leaving \( F'(x) \) unchanged — we put \( \lambda A_h \) instead of \( A_h \); \( \lambda a_h \) instead of \( a_h \).

According to (13) and (14) the consequence of such a change will be that in the equation of the frequency-curve

\[
M \text{ changes to } \lambda M \\
\epsilon^2 \text{ to } \frac{\lambda}{\epsilon^2} F'(x) \]

so that it becomes

\[
y = \frac{F'(x)}{\lambda \epsilon^2 V 2\pi} e^{-\frac{1}{2\lambda \epsilon^2} [F(v) - F(v_0) - \lambda M]},
\]
which, expressed in terms of \( f(x) \), according to (35) and (36) becomes

\[
y_1 = \frac{1}{\lambda} V_{\pi} f'(x) e^{-\frac{1}{\lambda} \left[ f(x) + (\lambda - 1)f(x_0) \right]^2}.
\]

As for the proportional curves of the second kind, we suppose the average deviations to remain the same, the \( \bar{A}_k \) and \( \bar{a}_k^2 \) will remain the same in the average. As, however, their number is supposed to increase in the proportion of \( 1 : \lambda \) it follows from (13) and (14) that

\[
M \text{ will change to } \lambda M
\]

\[
e_2 = \lambda \varepsilon^2.
\]

The equation of the proportional curve of the second kind will therefore become:

\[
y_2 = \frac{F(x)}{eV_{2\pi}} e^{-\frac{1}{\lambda} \left[ F(x) - F(x_0) - \lambda M \right]^2}.
\]

or in terms of \( f(x) \)

\[(45) \quad y = \frac{1}{\sqrt{\lambda} \pi} f'(x) e^{-\frac{1}{\lambda} \left( f(x) + (\lambda - 1)f(x_0) \right)^2}.
\]

We may summarise these results as follows: A frequency curve

\[(46) \quad y = \frac{1}{\sqrt{\lambda} \pi} \varphi'(x) e^{-\left( \varphi(x) \right)^2}.
\]

will, in regard to another frequency curve,

\[(47) \quad y = \frac{1}{\sqrt{\lambda} \pi} f'(x) e^{-\left( f(x) \right)^2}.
\]

(48) be proportional of the first kind if \( \varphi(x) = \frac{1}{\lambda} \left[ (\lambda - 1)f(x_0) + f(x) \right] \)

\[(49) \quad \varphi(x) = \frac{1}{\sqrt{\lambda} \pi} \left[ (\lambda - 1)f(x_0) + f(x) \right].
\]

In both cases therefore the functions \( \varphi(x) \) and \( f(x) \) will be linear functions of each other. Wherever we find two frequency curves which show such a linear relation, there thus exists the possibility of their being proportional curves. If neither \( \lambda \) nor \( f(x_0) \) are known \textit{a priori}, we will be unable to decide the kind of proportionality. If \( f(x_0) \) is known \textit{a priori} the decision becomes possible. It deserves attention that in any case where we find a linear relation between \( \varphi(x) \) and \( f(x) \), if we have reason to assume that the proportionality must be of a determined kind, we can determine both \( \lambda \) and \( f(x_0) \) and consequently \( \varphi_0 \). So in the case, treated further below, of the summer and winter barometerheights. We assume that we have to do with a proportionality of the first kind. If this is really so then the undisturbed barometerheight at den Helder must be 761.2 mm.
In the theory of observation errors similar cases would offer the possibility of finding the correct value of the unknown $x_0$ notwithstanding the presence of unknown systematic errors.

11. Medians and quartiles.

Let $q_{-25}$, $x_m$, $q_{-75}$ represent the abscissae corresponding to the ordinates which divide the area of the frequency curve in four equal parts; $x_m$ will be what is generally called the median,

$$x_m - q_{-25} \text{ will be the first quartile;}$$

$$q_{-75} - x_m \text{ = second quartile.}$$

The determination of these quantities is extremely simple. If the ordinate of the frequency curve corresponding to the abscissa $x$ be called $y$ and if $\tau$ be the lower limit of the curve then $q_{-25}$, $x_m$, $q_{-75}$ will be respectively determined by

$$\int_{\tau}^{x_m} y\,dx = \frac{1}{4}; \text{ resp. } \int_{\tau}^{x_m} y\,dx = \frac{1}{2}; \int_{\tau}^{q_{-75}} y\,dx = \frac{3}{4}.$$ 

Thus, for instance, the median of the frequency curve (38) will be determined by

$$\frac{1}{V_{-\infty}^{x_m}} f(x) e^{-f(x)\lambda^2} \,dx = \frac{1}{2},$$

which by putting $f(x) = z$ reduces to

$$\frac{1}{V_{-\infty}^{x_m}} e^{-z\lambda^2} \,dz = \frac{1}{2}.$$ 

Therefore

$$\int_{-\infty}^{-0.47694} \int_{-\infty}^{+0.47694} e^{-z\lambda^2} \,dz = 1/4 \text{ and } \int_{-\infty}^{+0.47694} e^{-z\lambda^2} \,dz = \frac{3}{4}.$$ 

In a similar way we get the other quantities. Remembering that

$$\frac{1}{V_{-\infty}^{x_m}} e^{-\lambda \int_{-\infty}^{0} \phi(t)\,dt} = \frac{1}{4} \text{ and } \frac{1}{V_{-\infty}^{x_m}} e^{-\lambda \int_{-\infty}^{\infty} \phi(t)\,dt} = \frac{3}{4},$$

we get, if we include the results for the proportional curves

<table>
<thead>
<tr>
<th>$f(x_m)$</th>
<th>$f(q_{-25})$</th>
<th>$f(q_{-75})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequ. curve (38)</td>
<td>0</td>
<td>-0.47694</td>
</tr>
<tr>
<td>prop. curve 1st kind</td>
<td>$(1-\lambda) f(x_0)$</td>
<td>$\lambda + (1-\lambda) f(x_0)$</td>
</tr>
<tr>
<td>&quot; 2nd</td>
<td>$(1-\lambda) f(x_0)$</td>
<td>$\sqrt{\lambda} + (1-\lambda) f(x_0)$</td>
</tr>
</tbody>
</table>

For any given frequency curve therefore the median and the quartiles are at once read off from the curve $y = f(x)$. 

----------
DEVELOPMENT OF THE THEORY

BY

M. J. van Uven.

CHAPTER II.

1. Introductory. When measuring some quantity appertaining to some organism (length of ears, weight of fruits, sugar percentage of beetroots), the different values obtained are usually of very different frequency. The distribution of the different values among the individuals resembles that of the different results when observing some physical or astronomical quantity. If systematic errors may be left out of consideration and the causes of error are very numerous and independent from each other, the results of observation are symmetrically spread round their arithmetical mean, in accordance with a definite law, the so-called "exponential" law of error. This law expresses how the probability of a certain deviation (or error) from the arithmetical mean is determined by the amount of that deviation. Values so distributed are said to have "normal frequency" or to be represented by a normal frequency-curve (which is the graph of the "exponential" relation between the amount of the error and its probability). According to this law of error the smaller deviations are more numerous than the larger ones, as was to be expected.

The frequency-table of some quantity measured in a great number of individuals sometimes agrees with the normal law of error. In such a case we may be sure that the different causes of deviation, or rather the different causes of growth, are very numerous and independent from each other.

As soon however as the causes of growth are no longer mutually independent, the frequency-table ceases to agree with the exponential relation.

When a certain quantity $x$ is spread according to the law of error, some other quantity $z$, connected with $x$ by some relation, will not be thus distributed (only when $x$ is a mere multiple of $z$, to which a constant has been added, $z$ is spread normally together with $z$).

Now, if this quantity $x$ is measured, the frequency-table given by observation will not be normal. In this case there will be a certain quantity does follow the distributed quantity.

In this case, the frequency will be independent of the magnitude, that the relation between the measured quantity and the mean value $z$ is the exponential function.

In this relation the causation of the frequency to be observed will be frequent reaction of the reaction.
quantity \( z \), connected with \( x \) by some relation, the frequency of which does follow the normal law; and it may be an interesting problem to find out the relation between the measured quantity \( z \) and the normally distributed quantity \( z \).

In the next paragraph it will be proved, that „abnormal” frequency occurs when the effects of the causes of deviation (or of growth) cease to be independent from each other, but, on the contrary, depend on the magnitude of the growing quantity \( x \), which undergoes the deviations, so that the difference between the deviations is due not only to mere chance, but also to a divergence of the values to which they refer. Moreover the relation between the effect of the cause and the amount of the deviating quantity \( z \) will be shown to be connected with the relation existing between the measured quantity \( x \) and the normally distributed quantity \( z \).

The relation between \( z \) and \( x \), mathematically expressed by: \( z \) is a function of \( x \) \((z = f(x))\), or by: \( z \) is a function of \( z \) \((z = \varphi(z))\), may be geometrically represented on squared paper by a „graph” or curve, which is the whole of the points having \( x \) for abscissa and the corresponding value \( z \) for ordinate. We will in what follows speak of the „curve” \( z = f(x) \).

In the same manner, the effect of the cause being denoted by \( \eta \), the relation \( \eta = \psi(x) \) between \( \eta \) and \( x \) may be illustrated by another graph, the curve \( \eta = \psi(x) \). The function \( \psi(x) \) will be called the „reaction-function”, whence the curve \( \eta = \psi(x) \) will be called the „reaction-curve”. The problem to be solved consists in finding the relation (or curve) \( z = f(x) \) from the frequency-table given by observation, and afterwards in deducing the reaction-function (or curve) from the relation (or curve) \( z = f(x) \).

In the next paragraph the mathematical treatment of this problem will be given.

Those readers who have no taste for mathematical analysis may proceed to the following paragraph, containing matter of a more practical kind.

In order to make the practical rules easier to understand, the results of analysis are summarised and translated into easy geometrical language; after this the method of determining the curve \( z = f(x) \) and of deriving the reaction-curve is expounded with the utmost simplicity; finally several practical hints are added.

2. Mathematical treatment. The successive increments of some growing organism may be attributed either to the continuous, though variable, action of a single cause or set of causes, or to the cumulative effects of several causes, each of which acts during a period, small in comparison with the whole duration of the growth.

The rate of increase of some organism under the influence of some
cause depends not only on the intensity of that cause, but also on the
degree in which the organism reacts upon it. For example in some cases
the rate of increase of a plant may be considered to be proportional to
the rate of taking up food, which, in its turn, may be proportional to
the area of some organs of the plant. By measuring the diameter $x$ of
the organ concerned we find the rate of increase proportional not only to
the intensity of the cause of growth (rain, solar heat, etc.) but also to the
second power of the measured quantity $x$. In this case we call $x^2$ the
“reaction-function”. Thus the increment of the plant, particularly
of the measured quantity $x$, is proportional as well to the intensity of the
cause as to the particular value of the reaction-function $\psi(x)$.

In order to avoid difficult intricacies, we shall suppose the several
causes to have nearly the same reaction-functions, so that the notion
“mean reaction-function” may not be deprived of sense.

The increment within some period of growth may be considered as
a multiple $m$ of an elementary increment. This multiple being proportional
to the intensity of the cause, the elementary increment is, in its turn,
proportional to the reaction-function.

Calling the elementary increment $\alpha^1$, the increment within some
period of growth may be given by

$$\Delta x = m \alpha,$$

the dependence of $\alpha$ from the reaction-function being expressed by

$$\alpha = \beta \psi(x),$$

where $\beta$ represents a constant quantity of such an order of magnitude,
that $m\beta$ becomes of the same order as $\Delta x$.

Hence in

(1) \hspace{1cm} \Delta x = m\beta \psi(x)

we may suppose the reaction-function to assume values of normal finite
amount.

In the same growth-period the different individuals, even if they have
equal values of $x$, will possess different values of $m$.

Here we introduce the essential supposition, that these different values
of $m$, representing the different intensities of the same cause, are due to
pure chance.

In consequence, denoting by $\bar{m}$ the arithmetical mean of the individual
values of $m$, the deviations

$$\mu = m - \bar{m}$$

$^1$) This quantity $\alpha$ is not to be confounded with the symbol $\alpha$ of the preceding
chapter, where $\alpha$ was used to indicate fluctuations, whereas in the present chapter it
means the elementary increment.
from this arithmetical mean are supposed to be distributed according to the exponential law of error. So the probability that \( \mu \) may be found between the limits \( \gamma - \frac{\Delta \gamma}{2} \) and \( \gamma + \frac{\Delta \gamma}{2} \) equals

\[
(2) \quad \Delta W = \frac{1}{V2\pi \varepsilon^2} e^{-\frac{\varepsilon^2}{2\varepsilon^2}} \Delta \gamma,
\]

where \( \varepsilon \) is the "mean error" or "standard-deviation".

In order to develop more methodically the theory of the general case of a variable elementary increment \( \alpha \), we will make a preliminary study of the case of a constant elementary increment.

The increments \( \Delta x = m \alpha \) having from their arithmetical mean \( m \alpha \) the deviations \( \mu \alpha \), these latter, being the product of the constant \( \alpha \) and the quantity \( \mu \) distributed according to (2), are also spread according to the normal law of frequency.

The whole time of growing may be composed of a great number of successive growth-periods \( P_1, P_2, \ldots, P_n \).

The initial value of \( x \) being denoted by \( x_0 \), the successive values of \( x \) are

after \( P_1 \), \( x_1 = x_0 + \Delta x_0 = x_0 + m_1 \alpha = x_0 + m \alpha + \mu_1 \alpha \)

\[P_2 \] \( x_2 = x_1 + \Delta x_1 = x_1 + m_2 \alpha = x_1 + m \alpha + \mu_2 \alpha = x_0 + (m_1 + m_2 + \mu_1 + \mu_2) \alpha \)

\[ \vdots \]

\[P_n \] \( x_n = x_{n-1} + \Delta x_{n-1} = x_{n-1} + m_\alpha = x_{n-1} + m \alpha + \mu_\alpha = \]

\( = x_0 + (m_1 + m_2 + \ldots + m_n) \alpha + (\mu_1 + \mu_2 + \ldots + \mu_n) \alpha \)

Putting generally

\[ \Sigma m_k = m_1 + m_2 + \ldots = m, \quad \Sigma m_k = m, \quad \Sigma \mu_k = \mu, \]

we have finally

\[ x = x_0 + m \alpha = x_0 + m \alpha + \mu \alpha, \]

or putting

\[ x_0 + m \alpha = \bar{x}, \quad \mu \alpha = \xi, \quad x = \bar{x} + \xi. \]

The different values of \( x \) obviously have the arithmetical mean \( \bar{x} \); and the deviations \( \xi \) from that mean, being the product of the constant \( \alpha \) and the sum \( \mu = \Sigma \mu_k \) — each term of which follows the exponential law —, are also distributed according to this law, so that the probability that \( \xi \) will lie between \( \bar{\xi} - \frac{\Delta \xi}{2} \) and \( \bar{\xi} + \frac{\Delta \xi}{2} \) amounts to

\[
\Delta W = \frac{e^{-\frac{\bar{\xi}^2}{2\bar{\xi}^2}}}{V2\pi \varepsilon^2} \Delta \xi.
\]

Hence: the elementary increment being constant and the values of
its multiples, contained in one growth-period, being purely accidental, the
different lengths measured after a finite lapse of time are still found normally
distributed round their arithmetical mean.

We have expressly supposed, that the values of the multiples \( m \) were
merely due to chance. The meaning of this is that the causes of growth,
the intensity of which is as it were measured by \( m \), are wholly independent
from each other and built up from a great number of small agents.

Thus far we have restricted ourselves to the case, that a certain cause,
when operating with the same intensity, also has the same effect — purely
accidental deviations left out of consideration — upon the growth, whatever
may be the value of the length \( x \) undergoing the increment.

Next we shall suppose the elementary increment \( a \) to be variable,
that is to say, to depend on the value of \( x \), or to be a function of \( x \). In
agreement with the above notation we put

\[ a = \beta \psi(x), \]

\( \beta \) being a constant of the same order of magnitude as \( a = \frac{\Delta x}{m} \), so that
\( \psi(x) \) will assume normal finite values.

This supposition evidently implies, that the organism reacts upon
the cause in such a way that equal intensities of the cause need not have equal
effects on the growth; that, on the contrary, the increment due to that cause
depends, besides on its intensity, also on the value of \( x \), on which it acts.

When several causes cooperate, we shall assume, that the reaction-
functions are nearly identical, or that a single cause is preponderant, so
that it is sufficient to consider one single reaction-function.

By taking the growth-period so small, that the value of \( x \) and also
of \( \psi(x) \) may be supposed constant, we obtain:

During \( P_1 \ldots x_1 - x_0 = \Delta x_0 = m_1 \beta \psi(x_0) \)

\[ \cdots \]

\[ P_h \ldots x_h - x_{h-1} = \Delta x_{h-1} = m_h \beta \psi(x_{h-1}), \]

or, in general,

\[ \frac{\Delta x_{h-1}}{\psi(x_{h-1})} = m_h \beta. \]

Hence, starting with \( x_0 \) and terminating with \( x \),

\[ \sum_{x_0}^{x} \frac{\Delta x_{h-1}}{\psi(x_{h-1})} = \beta \sum_{h} m_h = m \beta = \bar{m} \beta + \mu \beta. \]

Proceeding to the limit \( \Delta x = 0 \) we find

\[ (3) \ldots \ldots \ldots \ldots \int_{x_0}^{x} \frac{dx}{\psi(x)} = \bar{m} \beta + \mu \beta. \]
Putting
\[ \int \frac{dx}{\psi(x)} = F(x) = Z, \quad \bar{m} \beta = M, \quad \mu \beta = \zeta, \]
the equation (3) passes into
\[ Z_{0} = F(x) - F(x_{0}) = M + \zeta. \]

Here the quantity \( \zeta \), product of the constant factor \( \beta \) with the normally distributed number \( \mu \), follows the exponential law.

The quantity \( z \) itself is no longer, as in the former case, normally spread.

The quantity
\[ (4) \quad \zeta = F(x) - F(x_{0}) - M \]
follows the normal law
\[ \Delta W = \frac{h}{\sqrt{\pi}} e^{-\frac{h^{2}}{2} \Delta z}, \quad h = \frac{1}{e \sqrt{2}}. \]

Introducing
\[ (5) \quad z = h \zeta = hF(x) - hF(x_{0}) - hM = f(x) \]
we obtain the quantity \( z \), distributed round the mean value zero according to the law
\[ (6) \quad \Delta W = \frac{1}{\sqrt{\pi}} e^{-z^{2}} \Delta z. \]

When it is possible to determine the form \( z = f(x) \), the reaction-function \( \psi(x) \) can be deduced from the relation
\[ \psi(x) = \frac{1}{F'(x)} = \frac{h}{f'(x)}, \]
where \( F'(x) \) and \( f'(x) \) are the derivatives of \( F(x) \) and \( f(x) \) resp.

Thus the function \( \psi(x) \) is determined but for a constant, this latter circumstance being a consequence of the indefiniteness of the factor \( \beta \).

In what follows we shall put
\[ \eta = \frac{1}{f'(x)}. \]

\textit{Determination of the functions} \( Z = F(x) \) and \( z = f(x) \).

The distribution of the different values of \( x \) among the individuals examined may be arranged in a frequency-table.

If for every value of \( x \) the probability of occurrence were known, viz.
\[ y = \Omega(x), \]
then this relation would be the equation of the continuous frequency-curve. The probability that \( x \) may lie between \( x_{1} \) and \( x_{2} \) would be
\[ W_{x_{2}}^{x_{1}} = \int_{x_{1}}^{x_{2}} y \, dx = \int_{x_{1}}^{x_{2}} \Omega(x) \, dx. \]
Of course the integral between the extreme limits must be unity. In a lot of $N$ individuals, the number

$$Y_{x_1}^N = N \cdot W_x^{x_2} = N \int_{x_1}^{x_2} \Omega(x) \, dx$$

may be expected to lie between $x_1$ and $x_2$.

Now the observations never furnish the function $\Omega(x)$ itself, only some discrete values of the quantity $Y_{x_1}^N$.

Geometrically spoken: the observations furnish finite parts of the area of the frequency-curve.

Let the values $\xi_1, \xi_2, \ldots, \xi_n$ *) (rising by equal amounts $c$) be observed resp. $Y_1, Y_2, \ldots, Y_n$ times. Thus the constant class-range is $c$, so that

$$\xi_2 - \xi_1 = \xi_3 - \xi_2 = \ldots = \xi_n - \xi_{n-1} = c.$$  

Then it has in fact been settled, that for $Y_k$ individuals $x$ is found between $\xi_k - \frac{c}{2}$ and $\xi_k + \frac{c}{2}$.

Putting

$$(7) \quad \xi_k + \frac{c}{2} = x_k,$$

the observations tell us, that for $Y_k$ individuals $x_{k-1} < x < x_k$.

In what follows we will denote the lower limit of $x$ by $x_0$ and the upper limit by $x_n$. Hence the symbol $x_0$ will no longer indicate the initial or undisturbed value of $x$.

So between $x_0$ and $x_1$ $Y_1$ individuals are found, between $x_0$ and $x_k$ $Y_1 + Y_2 + \ldots + Y_k$ individuals; finally between $x_0$ and $x_n$ the total number $\sum_{i=1}^{n} Y_k = N$ of the individuals is found.

Hence the probability a posteriori amounts

for $x_{k-1} < x < x_k$ to $\frac{Y_k}{N}$,

$$I_k = \frac{Y_k}{N}$$

for $x_0 < x < x_1$ (or $x < x_1$) to $I_1 = \frac{Y_1}{N}$

$$I_2 = \frac{Y_1 + Y_2}{N}$$

$$\vdots$$

$$I_k = \frac{Y_1 + Y_2 + \ldots + Y_k}{N}$$

$$I_n = \frac{Y_1 + Y_2 + \ldots + Y_n}{N} = 1.$$  

*) The symbol $n$ will henceforth indicate the number of distinct values observed for $x$ (in stead of the number of causes, as in Ch. I).
The probability \( I_k \) is obviously represented by the area of the frequency-curve contained within the axis of \( x \), the frequency-curve \( y = \Omega(x) \) and the ordinate-line of \( x_k \). Usually the ordinate of the lower limit \( x_0 \) is zero. Hence

\[
I_k = \int_{x_0}^{x_k} \Omega(x) \, dx.
\]

In order to determine the function \( z = f(x) \) we are guided by the following principle:

*Corresponding values of \( x \) and \( z \) have equal probabilities.*

We can only verify that the probability of \( x \) being between \( x_1 \) and \( x_2 \) equals the probability of \( z \) lying between the conjugate values \( z_1 \) and \( z_2 \). Whether \( x_1 < x < x_2 \) corresponds with \( z_1 < z < z_2 \) or with \( z_2 < z < z_1 \) has not yet been settled.

The elementary increment was found above to be

\[
a = \beta \cdot \psi(x) = \frac{\beta}{F(x)} = \frac{\beta h}{f'(x)}.
\]

Excluding infinite values for the reaction, we postulate

\[
f'(x) \neq 0;
\]

the meaning of this is that the function \( z = f(x) \) may not have maxima or minima in the real domain.

By taking \( h \) and \( \beta \) positive we assume the elementary increment \( a \) and the derived function \( f'(x) \) to have always the same sign. As a rule \( a \) and \( f'(x) \) will be positive. When the elementary increment \( a \) is negative for some values of \( x \), then also \( f'(x) < 0 \), and the function \( f'(x) \), in passing from positive to negative values or inversely, must become infinite.

For the present we make the simplifying supposition, that the elementary increment shall always be positive. Any negative increment is then due to a negative value of the multiple \( m \) (see above).

So we have

*First simplification:*

\[
f'(x) > 0.
\]

The variable \( z \) ranges from \(-\infty\) to \(+\infty\). Unless particular circumstances compel us to admit infinite values for \( z = f(x) \) corresponding with values \( x \) within the limits \( x_0 \) and \( x_n \), we shall for convenience' sake suppose, that \( z \) becomes infinite only at the limits of the real domain, viz. \( x_0 \) and \( x_n \). In consequence of the first simplification the lower limit \( x = x_0 \) is conjugate to \( z = -\infty \), and the upper limit \( x = x_n \) to \( z = +\infty \).

So the

*Second simplification*

\[
f(x) \neq \pm \infty \quad \text{for} \quad x_0 < x < x_n
\]
gives us two pairs of conjugate values \((x, z)\), or, geometrically, two points \((x, z)\), viz. \((x_0, -\infty)\) and \((x_n, +\infty)\).

At present we are able to completely to determine the correspondence \((x, z)\):
\[
W_{x_0}^x = \int_{x_0}^x y \, dx = \frac{1}{V_n} \int_{-\infty}^z e^{-t} \, dt = \Theta(z).
\]

Now for \(W_{x_0}^x\), i.e., the probability of \(x < x_0\), only the a-posterioric value can be given. It amounts to \(I_k = \frac{Y_1 + Y_2 + \ldots + Y_k}{N}\). So we have as an approximate value for \(\Theta(z_k)\) (which is the value a priori)
\[
(9a) \quad \Theta(z_k) = I_k - I(z_k) = \frac{Y_1 + Y_2 + \ldots + Y_k}{N}.
\]

The most probable value of the probability a priori is the probability a posteriori \(p\). According to the reversed theorem of Bernoulli (2^4 theorem of Bayes) the probable error \(q_p\) of the probability a posteriori \(p\) considered as an approximate value of the probability a priori, is given by
\[
(10) \quad q_p = q \sqrt{\frac{2p(1 - p)}{N}}
\]
where \(q = 0.476936\ldots\), and \(N\) is the whole number of trials, the fraction \(p\) of which has a favorable result.

So the chances are equal that the true probability a priori lies between
\[
p - q \sqrt{\frac{2p(1 - p)}{N}} \quad \text{and} \quad p + q \sqrt{\frac{2p(1 - p)}{N}}.
\]

Since the probability a priori is not absolutely certain, the quantity \(z\) is not determinate either. So the correspondence \((x, z)\) always has an element of uncertainty, which may be expressed in numerical value by the probable error \(q_z\) of \(z\) itself. In the Appendix to Ch. II, I A (p. 62) we shall prove that for \(q_z\) the following approximate values may be taken:

1º. in the neighbourhood of \(z = 0\):
\[
(11) \quad q_z = q \sqrt{\frac{\pi}{2N}} = \frac{0.6}{V_N}
\]

2º. for great positive or negative values \(\zeta\) of \(z\)
\[
(11a) \quad q_{\zeta} = q \sqrt{\frac{\pi}{2N}} \times \frac{e^{\zeta^2}}{V_{\zeta}}
\]

By operating with a sufficiently great number \(N\) of individuals the value of \(q_z\) round the centre of the domain \((z = 0)\) is small, and accordingly there is but a slight uncertainty in the correspondence \((x, z)\).

On the contrary, the error in \(z\) at the extremities of the domain \((z = -\infty\) and \(z = +\infty)\) is very important; the formula shows
\[
\lim_{z=\pm \infty} q_z = \infty.
\]
So the probable error of \( z \) increases together with the absolute arithmetical value of \( z \). At the limits the value of \( z \) is absolutely uncertain; the meaning of this is that it is impossible to decide whether \( z = -\infty \) or \( z = -\xi \) (finite) must be made to correspond to a provisional value of \( x_0 \). Inversely it is absolutely uncertain, which value \( x_0 \) answers to \( x = -\infty \), or which value \( x_n \) must be made to correspond to \( z = +\infty \).

Hence the correspondence \((x, z)\) is nearly exact at the centre of the domain, doubtful at the values \( x_1 \) and \( x_{n-1} \) preceding the extreme limits and absolutely uncertain at the limits \( x_0 \) and \( x_n \) themselves.

The limits \( x_0 \) and \( x_n \) of the domain of correspondence, which are conjugate to \( z = -\infty \) and \( z = +\infty \) are essentially absolutely indeterminate.

Now it is obvious in what way the function \( z = f(x) \) may be determined.

The observations furnish

\[
\begin{align*}
Y_1 & \text{ times the value } \xi_1, \\
Y_2 & \text{ between } \frac{c}{2} = x_1, \\
& \ldots \\
Y_n & \text{ above } \frac{c}{2} = x_{n-1},
\end{align*}
\]

So the total number of observations amounts to

\[ N = \sum_1^n Y_k. \]

The observations really show, that \( x \) is found

\[
\begin{align*}
Y_1 & \text{ times below } \xi_1 + \frac{c}{2} = x_1, \\
Y_2 & \text{ between } \xi_1 + \frac{c}{2} = \xi_2 - \frac{c}{2} = x_1 \text{ and } \xi_2 + \frac{c}{2} = x_2, \\
& \ldots \\
Y_k & \text{ between } \xi_{k-1} + \frac{c}{2} = \xi_k - \frac{c}{2} = x_{k-1} \text{ and } \xi_k + \frac{c}{2} = x_k, \\
& \ldots \\
Y_n & \text{ above } \xi_{n-1} + \frac{c}{2} = \xi_n - \frac{c}{2} = x_{n-1},
\end{align*}
\]

or, in other words, that \( x \) lies

\[
\begin{align*}
Y_1 & \text{ times between } x_0 \text{ and } x_1, \\
Y_1 + Y_2 & \text{ between } x_0 \text{ and } x_2, \\
& \ldots \\
Y_1 + Y_2 + \ldots + Y_k & \text{ between } x_0 \text{ and } x_k, \\
& \ldots \\
Y_1 + Y_2 + \ldots + Y_n &= N \text{ } \text{ between } x_0 \text{ and } x_n.
\end{align*}
\]

The probability a priori of \( x_0 < x < x_k \) is expressed by

\[
p \pm \xi_p = p \pm \frac{2p(1-p)}{N},
\]
where

\[ p = \frac{Y_1 + Y_2 + \ldots + Y_k}{N} = I(x_k) \]

and

\[ \epsilon_p = \text{probable error of } p \ (\epsilon = 0.476936 \ldots). \]

Then from

\[ p = \Theta(z_k) = I(x_k) \]

we derive the most probable value \( z_k \) of the variable \( z \), which is conjugate to \( x_k \).

In this way we obtain \( n - 1 \) pairs \((x_k, z_k)\), viz. \((x_1, z_1), \ldots, (x_{n-1}, z_{n-1})\).

Marking these pairs by points with coordinates \((x, z)\) we get \( n - 1 \) points of the curve which represents the function \( z = f(x) \).

The situation of these points is most certain at the centre of the domain (round \( z = 0 \)). It has been shown above that the smallest value of the probable error \( \epsilon_z \) of \( z \) is \( \frac{0.6}{\sqrt{N}} \). When moving away from the centre the uncertainty increases with \( z \) itself.

A continuous curve through the \( n - 1 \) marked points is most sharply determined at the centre \( z = 0 \).

The uncertainty in the shape of the curve \( z = f(x) \) may be illustrated by drawing two curves at both sides of the original one, viz.

\[ z = f(x) - \epsilon_z \] and \[ z = f(x) + \epsilon_z. \]

We thus obtain a strip round the most probable curve \( z = f(x) \); this strip is very narrow near the centre, but rather wide near the extremities and even infinite at \( z = +\infty \) and \( z = -\infty \).

From the curve \( z = f(x) \) the reaction-function may be deduced either by calculation or graphically.

3. Practical proceeding.

Summary of the results of the preceding paragraph.

Let the measured quantity \((x)\) have the following values

\[ Y_1 \text{ times } x = \xi_1, \]
\[ Y_2 \text{ times } x = \xi_2, \]
\[ \vdots \]
\[ Y_n \text{ times } x = \xi_n. \]

The whole number of individuals is therefore

\[ N = \sum_{1}^{n} Y_k. \]
The values $\xi_k$ may have a constant difference $c$, so that
\[
c = \xi_2 - \xi_1 = \xi_3 - \xi_2 = \ldots = \xi_n - \xi_{n-1}.
\]
The value $\xi_k$ is considered as the centre of a class which extends to $\frac{c}{2}$ at both sides of the centre $\xi_k$. Hence the class-limits are $x_{k-1} = \xi_k - \frac{c}{2}$ and $x_k = \xi_k + \frac{c}{2}$.

The extreme limits $x_0$ and $x_n$, these being the limits which $x$ cannot exceed a priori, are generally supposed to be different from $\xi_1 - \frac{c}{2}$ and $\xi_n + \frac{c}{2}$.

So the observations furnish the following data for $x$:

\[
\begin{align*}
Y_1 &= x_0 < x < x_1, \\
Y_2 &= x_1 < x < x_2, \\
&\vdots \\
Y_{n-1} &= x_{n-2} < x < x_{n-1}, \\
Y_n &= x_{n-1} < x < x_n,
\end{align*}
\]
where $x_k = \xi_k + \frac{c}{2}$ for $k = 1, 2, \ldots, n - 1$.

Now form the fractions
\[
p_1 = \frac{Y_1}{N}, \quad p_2 = \frac{Y_1 + Y_2}{N}, \ldots, \quad p_{n-1} = \frac{Y_1 + Y_2 + \ldots + Y_{n-1}}{N}
\]
and determine the values of $z$ corresponding to $p$ by the relation
\[
\Theta (z) = p,
\]
where $\Theta (z)$ is a function, tabulated at the end of this book.

The value $z_k$ which is found with $p_k$, is to be joined as ordinate to the abscissa $x_k$.

In this way $n - 1$ points $(x_k, z_k)$ ($k = 1, 2, \ldots, n - 1$) are obtained belonging to the curve $z = f(x)$, which must be traced through these points as exactly as possible. Particularly in the neighbourhood of $z = 0$ the coincidence must be very close.

The value of the reaction-function corresponding to $x$ is, save a constant factor, equal to the trigonometrical tangent of the angle inclosed by the axis of $z$ and the tangent-line to the curve at the point with abscissa $x$. In this manner any number of points of the reaction-curve may be plotted, through which the curve itself is to be traced.

The values of $p$ are obtained by dividing the sums $Y_1, Y_1 + Y_2, \ldots, Y_1 + Y_2 + \ldots + Y_k, \ldots Y_1 + Y_2 + \ldots + Y_{n-1}$ by the total number of indi-

---

1) In what follows $x_0$ will no longer designate the initial value, but the lower limit of $x$. 

---
viduals $N = \sum_{1}^{n} Y_k$. This algebraic operation may be quickly performed with
the aid of calculating-tables or slide rules. Using a slide rule of about
15 cm. length, after some practice an approximation within 0,001 of the
value may be attained, which is usually sufficient. The fractions which
surpass 0,5 must be taken from unity, since the value of $1 - p$ must also
determine within 0,001 of its amount.

Employing squared paper in sheets of $20 \times 26$ cm. (SCHLICHER &
SCHÜLL, No. 332I/9) the axis of $z$ should be taken parallel to the longer side.
The values of $z$ occurring in practice rarely exceed the interval from
$-2,6$ to $+2,6$ [$\Theta (-2,63) = 0,0001$]. The unit of $z$ may therefore be
represented by a length of 5 cm.

The length of the axis of $z$ (which lies in the middle of the sheet)
amounts to 20 cm. For $x$ such a unit is preferable that the class-range
corresponds to a whole number of mms and that all the class-limits $x_1, \ldots
x_{n-1}$ fall inside the sheet.

In order to plot the points $(x_1, z_1) \ldots (x_{n-1}, z_{n-1})$ of the curve $z = f(x)$
either: the values of $z$ conjugate to the class-limits $x_k$ may be taken from
the table of the function $p = \Theta(z)$ at the end of this book, or: instead of
this table we may directly use a scale on which the number $p = \Theta(z)$
corresponding to the value of $z$ is marked at a distance of $z \times 5$ cm. from
the zero-point. Such a scale has at the zero-point itself the number
$0,5 = \Theta(0)$, and at a distance of $4,55$ cm. $= 0,906 \times 5$ cm. from the zero-point
at one side ($z = -0,906$) the number $0,100 = \Theta(-0,906)$ and at the other
side ($z = +0,906$) the number $0,900 = \Theta(+0,906)$. Using this scale
the interpolation may be performed graphically.

When in this way $n - 1$ points of the curve $z = f(x)$ have been plotted,
a smooth line is drawn through them; care must be taken that near the
centre the line passes through the points as exactly as possible. Near the
extremities greater deviations from the given points are allowed in order
to avoid irregularities in shape.

The curve $z = f(x)$ having been drawn, the reaction-fonction
\[ \eta = \psi(x) = \frac{1}{f'(x)} = \frac{dx}{dz} \]
must be determined.

Sometimes it is fairly easy to find the analytic expression $z = f(x)$
corresponding to the plotted curve. Then this function $f(x)$ may be

---

1) Printed on non-shrinking card-board and published by Arnaud Pistoer, 's Hertogen-
bosch, Holland. A reproduction of this scale is found at the end of this book.
differentiated, and the quantity \( \eta \) is determined as the set of reciprocal values of \( \frac{dz}{dx} = f'(x) \).

Usually however the equation \( z = f(x) \), represented by the given curve, is very difficult to deduce. In this case we may have recourse to graphical differentiation, which dispenses with the equation of the curve; this advantage however is diminished by the drawback that the accuracy with which the different values of \( f'(x) \) or \( \frac{1}{f'(x)} \) are determined, is very small.

A slight roughness in the plotted curve immediately has its full effect on the slope of the tangent. It is therefore very necessary to draw the curve as carefully and thinly as possible.

In the graphical determination of \( \frac{dx}{dz} = \frac{1}{f'(x)} \) we have, for some values of \( x \) (for instance for the class-limits \( x_k \), or for the class-centres \( \xi_k \)), to calculate the trigonometrical tangent of the angle inclosed by the axis of \( z \) and the tangent-line to the curve at the corresponding point.

A good plan is to copy the smooth curve \( z = f(x) \) with a fine pen on transparent squared tracing-paper (SCHLEICHER & SCHÜLL No. 3071/4). A sheet of clear white paper, on which a sharp narrow straight line is drawn, is then put under the transparent paper. The sheets are shifted relatively to each other until the straight line of the lower sheet coincides as well as possible with the tangent-line to the curve \( z = f(x) \) at the desired point. A solid ruler should not be used, because it covers one side of the surroundings of the line. If the ruled paper on which the curve is drawn is not transparent, the straight auxiliary line is traced on transparent paper, which is placed on the squared paper.

Now the points are marked where this straight line meets two lines parallel to the axis of \( z \) and the mutual distance of which is 10 cm., or — if the sheet is of sufficient size to get the intersections on it — 20 cm. If the distance of the two marked points in the direction of \( x \) is \( l \) cm., the quotient \( \frac{l}{10} \) (or \( \frac{l}{20} \)) is equal to the trigonometrical tangent of the angle inclosed by the axis of \( z \) and the tangent-line. This trigonometrical tangent \( q \) however is not equal to \( \frac{dx}{dz} \), because, in general, the units of \( x \) and \( z \) are not the same.

Supposing that the unit of \( x \) is represented by \( a \) cm., and that of \( z \) by \( b \) cm. (in our case \( b = 5 \)), then

\[
q = \frac{a}{b} \cdot \frac{dx}{dz} = \frac{a}{b} \cdot \frac{1}{f'(x)},
\]
whence

\[
\frac{1}{f'(x)} = \frac{dx}{dz} = \frac{b}{a} \cdot q.
\]

The different values of \(\frac{1}{f'(x)}\) as well as those of \(q\) must be considered as the corresponding values of the reaction-function, which is determined but for an (essentially unknown) constant factor. The multiplier \(\frac{b}{a}\) of the tangent \(q\) is of no consequence, in fact.

Plotting the values of \(\frac{dx}{dz}\) for the corresponding values of \(x\), the points so marked belong to the reaction-function. The reaction-curve itself may be obtained as the smooth curve passing as exactly as possible through the given points.

This entirely graphical method may be replaced by a „semi-graphical“ one, in which a set of equidistant ordinates of the smoothed curve \(z = f(x)\) is measured. The reciprocal values of the differences of consecutive ordinates are considered as nearly proportional to the values of \(\frac{dx}{dz}\).

When the entirely-graphical method is carried out with the highest possible precision, it is to preferred to the semi-graphical one, which is essentially less exact.

4. Analytic expression for the relation \(z = f(x)\).

Sometimes the curve \(z = f(x)\) has so simple a shape, that it is easy to guess the equation represented by it.

For the present we will treat only two cases. In the appendix a third somewhat more intricate case will be discussed.

I. The points \((x_k, z_k)\) are nearly collinear.

Let the equation of the straight line passing through them be

(12) \[ z = \lambda (x - x_m). \]

The auxiliary straight line put under the transparent squared paper (or, when itself drawn on transparent paper, placed on the squared paper) must be so shifted that it passes as exactly as possible through the points, particularly through the middle points. The axis of \(x\) (\(z = 0\)) is cut in a point, the abscissa of which is called the „median“ and is denoted by \(x_m\).

Since \(z = 0\) corresponds to \(p = \frac{1}{2}\), there are as many individuals for which \(x < x_m\) as for which \(x > x_m\). The median value of \(x\) is that which is passed over with the probability \(\frac{1}{2}\).
ERRATUM.

The more intricate case of logarithmic distribution (p. 44) will not be treated in the appendix but in a continuing article mentioned at pag. 52.
Now
\[ f'(x) = \frac{dz}{dx} = \lambda. \]
and
\[ \frac{1}{f'(x)} = \frac{b}{a} \times q. \]
(see p. 44) where \( q \) is the trigonometrical tangent of the angle between the line and the axis of \( z \), \( a \) the length in cm. of the unit of \( x \) and \( b \) that of the unit of \( z \) (usually \( b = 5 \)).

Hence
\[ \lambda = \frac{a}{b} \times \frac{1}{q}, \]
and, putting
\[ \frac{1}{q} = k, \]
\( k \) being the trigonometrical tangent of the angle between the line and the axis of \( x \), we have
\[ \lambda = \frac{a}{b} \times k. \]

So \( \lambda \) may be computed from the numbers \( a \) and \( b \) which have been chosen in advance, and from the quantity \( k \) which is to be measured.

In this way both the constants of the equation are determined. The case just treated is that of normal frequency. The median \( x_m \) here coincides with the arithmetical mean \(^1\).

Only in the case of normal distribution the arithmetical mean may be considered as representative of the different values, as a typical value. In the case of abnormal frequency this mean is of far less importance.

On account of the linear relation between \( x \) and \( z \), the standard-deviation of \( x \) [viz. the square root of the mean square of \( x - x_m \)] corresponds to the standard-deviation of \( z \) [viz. the square root of the mean square of \( z \)], so that
\[ \varepsilon_x = \lambda \varepsilon_z. \]

Now
\[ \varepsilon_z = \frac{1}{\sqrt{2}} \] (see p. 33 and p. 35).

hence
\[ \varepsilon_x = \frac{1}{\lambda \sqrt{2}}. \]

\(^1\) since
\[ \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-z^2} dz = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-\lambda(z-x_m)^2} dz = x_m. \]
which result also follows from the law of distribution
\[ \Delta W = \frac{\lambda}{\sqrt{\pi}} e^{-\lambda x_m^2} \Delta x. \]

So our conclusion runs:

*Normal distribution of the values of x is indicated by a rectilinear disposition of the points \((x_k, z_k)\).*

In this case the arithmetical mean is the abscissa of the point at which the line meets the axis of \(x\), and the standard-deviation is either to be computed from
\[ (13) \quad \varepsilon_z = \frac{1}{\lambda \sqrt{2}} \frac{b}{a \sqrt{2}} = \frac{b}{a \sqrt{2}} \times q = \frac{b}{a \sqrt{2}} \times \frac{1}{k}, \]
or to be read from the figure, viz. as the difference between \(x_m\) and the abscissa of the point, the \(z\) of which amounts to \(\frac{1}{\sqrt{2}} = 0.707\).

Here the reaction-function is a *constant*; so the elementary increment is independent from the value of \(x\), as was to be expected from the preliminary study of the reaction-function.

The case just treated may be illustrated by Example I: Circumference of the chest of recruits, measured by A. Quetelet (see p. 54).

IIa. The points \((x_k, z_k)\) lie in a curve, which sinks rather rapidly to the left — with a tendency to remain to the right of a certain vertical line \(x = x_0\) — and ascends gradually to the right with a decreasing slope (fig. 2).

This shape suggests the equation
\[ (14) \quad z = \lambda \log \frac{x - x_0}{x_m - x_0}, \quad (x_m > x_0, \quad \lambda > 0). \]

\[ \text{Fig. 2} \]

As the numerus of the logarithm becomes negative for \(x < x_0\), the value \(x_0\) is the lower limit for \(x\), at which \(z = \lambda \log 0 = -\infty\). So the line \(x = x_0\) is an asymptote.

We shall say that the quantity \(x\) has "logarithmic distribution."

First of all the value \(x_0\) must be estimated.

Let \(x_0 = 0\) be the supposed lower limit. Then the equation reduces to
\[ (14a) \quad \ldots \quad z = \lambda \log \frac{x}{x_m} = \lambda \log x - \lambda \log x_m. \]

Now we make use of logarithmic ruled paper, and so operate with the coordinates
\[ u = \log x \quad \text{and} \quad z. \]

The equation in the coordinates \(u, z\) runs
\[ z = \lambda (u - u_m), \]
whence the points \((u, z)\) plotted on logarithmic paper must be in a *straight line.*
Inversely, if the points \((u, z)\) (particularly in the vicinity of \(z = 0\)) are nearly in a straight line, this is an indication, that the relation between \(x\) and \(z\) is tolerably well approximated by the logarithmic equation (14).

When we use logarithmic paper of Schleicher & Schüll No. 376/25, the unit of \(u\) is 1 dm., and taking, as before, for the unit of \(z 5\) cm., we find for the slope of the straight line

\[ k = \frac{\lambda}{2}. \]

The point of intersection with the axis of \(x (z = 0)\) has for abscissa \(u = u_a = \log x_a\) and is marked at the margin of the paper by the number \(x_a\) itself, which evidently is the median value.

So, in the case \(x_0 = 0\), both the constants of the equation are immediately found by analysing the curve on logarithmic paper.

The reaction-function is

\[ \eta = \frac{1}{f'(x)} = 1 : \frac{x}{\lambda}. \]

The elementary increment is therefore proportional to the value of \(x\) itself. This case very often occurs in nature. For an illustration of it we may refer to Example II: Threshold of sensation, measured by Prof. G. Heymans. (see p. 55.)

When the curve on ordinary squared paper has the form indicated above, but the points \((u, z)\) plotted on logarithmic paper do not lie in a straight line, this may be due either to an erroneous estimate of \(x_0\), or to the fact, that the relation between \(x\) and \(z\) is not logarithmic at all.

First we may try to bring the points \((u, z)\) in a straight line by correcting the value of \(x_0\).

Instead of putting \(u = \log x\), we now put

\[ u = \log (x - x_0) \]

that is: we subtract the assumed value \(x_0\) from \(x\) and operate with the ordinate-line which has the number \(x - x_0\).

In correcting \(x_0\) graphically we may take the following into consideration:

If \(x_0\) is estimated too small, that is to say: if we operate with \(u' = \log (x - x_0')\) instead of \(u = \log (x - x_0)\), \(x_0\) being larger than \(x_0'\), then

\[ z = \lambda \log \frac{x - x_0}{x_a - x_0} = \lambda [\log (x - x_0) - \log (x_a - x_0)] = \lambda (u - u_a) \]

is conjugate not to \(u\) but to \(u'\).

Now the difference

\[ u' - u = \log (x - x_0') - \log (x - x_0) = \log \left(1 + \frac{x_0 - x_0'}{x - x_0}\right) \]

is positive and the smaller, the larger \(x\). So we join to a certain value
of \( z \) too large an abscissa \( u' \). The figure built up of the points \((u', z)\) therefore lies to the right of the figure corresponding to the coordinates \((u, z)\), which is the straight line in question \( z = \lambda (u - u_w) \).

This line, having a positive direction-tangent \( \lambda \), tends from left below to right above, and the deviations from it are left below larger than right above. So the curve \( V(u', z) \) obtained by too low an estimate of \( x_0 \), is concave downwards, and its curvature decreases towards the top (fig. 3a). (See for the rigorous proof the appendix to Ch. II, I B p. 64).

Replacing \( x_0 \) by a larger value \( x_0' \), we make all differences \( x - x_0' \) smaller, and the difference

\[
\frac{u' - u''}{x - x_0'} = \log (x - x_0') - \log (x - x_0'') = \log \left( 1 + \frac{x_0'' - x_0'}{x - x_0''} \right)
\]

is left below larger than right above. The course of the curve retains the same character, but the curvature has become fainter.

When, at last, we have hit the exact value \( x_0 \), the line is wholly straightened.

If, on the contrary, \( x_0 \) is estimated too large, say \( x_0' > x_0 \), then we operate with \( u' = \log (x - x_0') \) instead of \( u = \log (x - x_0) \), \( u' - u = \log \left( 1 - \frac{x_0' - x_0}{x - x_0} \right) \)

being now negative and in absolute value the smaller, the larger \( x \).

So we join \( z = \lambda (u - u_w) \) to \( u' < u \). Hence the abscissae are taken too small, and left below more so than right above.

The curve \( W(u', z) \), obtained by estimating \( x_0 \) too large, is therefore convex downwards, and its curvature decreases towards the top (fig. 3b). (See appendix to Ch. II, I B p. 64.)

When \( x_0' \) is replaced by a smaller value \( x_0'' \), still larger than \( x_0 \), all differences \( x - x_0' \) become larger, and the difference

\[
\frac{u'' - u'}{x - x_0'} = \log (x - x_0'') - \log (x - x_0') = \log \left( 1 + \frac{x_0'' - x_0'}{x - x_0'} \right)
\]

is left below larger than right above. So the line \( W \) becomes less curved.

By reducing \( x_0' \) too much, the curve \( W \) passes into a curve of the type \( V \).
In the same manner we obtain, in the first case, by increasing \( x_0' \)
too much, a curve of the type \( W \) instead of a straight line.

Usually the exact value of \( x_0 \) may be determined by interpolation. If
the distribution is not strictly logarithmic, a rather great uncertainty
remains in the determination of \( x_0 \). On the other hand \( \lambda \) and \( x_m \)
may be determined pretty accurately.

When it is not possible to straighten the curve by altering the
estimated value of \( x_0 \), the distribution is not really logarithmic.

The relation

\[
z = \lambda \log \frac{x - x_0}{x_m - x_0}
\]

generates the reaction-function

\[\eta = \frac{1}{f(x)} = 1 : \frac{\lambda}{x - x_0} = \frac{x - x_0}{\lambda} = Px + Q.\]

The elementary increment consists of a part proportional to the attained value \( x \) (which is positive for positive values of \( x \)) and of another part independent from \( x \) (which is negative for a positive value of \( x_0 \)). An illustration of this more general case of logarithmic distribution will be given in Example III: Valuation of House Property in England and Wales, by Prof. K. Pearson (see p. 55).

IIb. Sometimes the smooth curve drawn through the points \( (x_n, \; z_n) \) seems to have a vertical asymptote \( x = x_n \) on the right, and rises there with an increasing slope (fig. 4). The relation between \( x \) and \( z \) is now likely to be approximated by

\[z = \lambda \log \frac{x_n - x}{x_n - x_m} \quad (x_m < x_n, \; \lambda > 0).\]

Since the numerus of the logarithm becomes negative for \( x > x_n \), the value \( z_n \) is the upper limit for \( z \), at which \( z = \lambda \log \infty = + \infty \).

This case is treated in the same manner as IIa.

After having found the mentioned shape of the curve on ordinary squared paper, we pass to logarithmic paper; and after estimating the value \( x_n \) of the upper limit, we plot the abscissa \( u = \log (x_n - x) \).

The equation of the line on logarithmic paper will be

\[z = \lambda [\log (x_n - x_m) - \log (x_n - x)] = \lambda (u_m - u) = - \lambda u + \lambda u_m.\]

So we obtain, if operating with the exact \( x_n \), on logarithmic paper a straight line with a negative slope. If \( x_n \) is estimated too small, we operate with \( u' = \log (x_n' - x) \) instead of \( u = \log (x_n - x) \), \( x_n' \) being
smaller than \( x_n \), and we join \( z = \lambda (u_m - u) \) to \( u' \) instead of \( u \). The

difference

\[ u' - u = \log (x_{n'} - x) - \log (x_n - x) = \log \left( 1 - \frac{x_n - x_{n'}}{x_{n'} - x} \right) \]

is negative and the larger, the larger \( x \). So we connect a certain value \( z \) with too small an abscissa \( u' \). The

provisional figure thus lies to the left of the required rectilinear locus \( z = \lambda (u_m - u) \), which tends from right below to left above. The deviations from it will at the left end be larger than at the right. Hence the curve

\[ V'(u', z) \]

obtained by estimating \( x_n \) too small is concave downwards and its curvature increases towards the top (fig. 5a) (see appendix to Ch.II, IB p. 65).

A greater value of \( x_n \) produces a line of fainter curvature.

By estimating \( x_n \) too large \( (x_n' > x_n) \) we get, by a similar reasoning, a curve \( W' \) to the right of the required straight line. This curve \( W' \) deviates more on the left side than on the right. So the curve is convex downwards and its curvature increases towards the top (fig. 5b).

Here too \( x_n \) may be determined, be it roughly, by interpolation.

Now the reaction-function is

\[ \eta = \frac{1}{f'(x)} = 1: \frac{\lambda}{x_n' - x} = \frac{x_n - x}{\lambda} = Q - P \alpha. \]

The elementary increment consists of a part \( Q \) independent from \( x \) (which is positive for positive values of \( x_n \)) and of another part proportional to \( x \) (which is negative for positive values of \( x \)). Hence there is besides a constant element of growth a counteracting or inhibitory cause, proportional to \( x \).

5. Irregularities in the frequency distribution.

I. Domains of very small frequency.

When it appears from the observations that a certain set of successive class-intervals (with centres \( \xi_k, \xi_{k+1}, \ldots \xi_{k+i} \)) is not occupied by individuals, we have \( Y_1 = Y_{k+1} = \ldots = Y_{k+i} = 0 \), so that as many individuals are found below \( \xi_k + \frac{c}{2} = x_k \) as below \( \xi_{k+1} + \frac{c}{2} = x_{k+1} \), and as below \( \xi_{k+i} + \frac{c}{2} = x_{k+i} \), from which follows \( I_k = I_{k+1} = \ldots = I_{k+i} \).

So the probability a posteriori \( p \), which is also the most probable value of the probability a priori, remains constant, whence also \( z \) assumes \( i + 1 \) times the same value \( (z_k = z_{k+1} = \ldots = z_{k+i}) \). Consequently of all the points \( P_h(x_k, z_k), (k = 1, \ldots, n - 1) \), the set \( P_h, P_{h+1}, \ldots, P_{h+i} \) is

situated in the defined area.

The occurrence of such reactions is not infrequent, espec. quite from the great number of domains is rather exceptional, since here a small abnormal variation of \( p \), and stagnation of individuals grow smaller, which
situated on a horizontal line \((\varepsilon = \text{const.})\), so that in the domain \(x_0 \ldots x_{h+i}\) the derived function \(\frac{dz}{dx} = f'(x)\) would be zero, were it not that this would correspond to an infinite reaction \(\eta = \frac{1}{f'(x)}\).

A slight deviation from the given points is therefore necessary. On account of the uncertainty in the correspondence \((x, \varepsilon)\) we may depart from the horizontal line and give the curve a slope, however little it may be. This, to be sure, enables us to get rid of the infinite value of the reaction \(\eta\), but we are obliged anyhow to assign to \(\eta\) a considerable value, especially near the centre \((\varepsilon = 0)\), because the correspondence \((\varepsilon, z)\) is quite certain there, so that the value of \(f'(x)\) may differ but very little from zero.

So a scantiness of individuals in a certain domain indicates a powerful reaction at this spot, which makes the corresponding value of \(x\) hypersensitive, so that the least occasion suffices to move \(x\) from that value.

When however this scantiness occurs near the limits of the whole domain \((x_0 \ldots x_n)\) this inference is not nearly so sure in consequence of the greater uncertainty in the correspondence \((x, \varepsilon)\). But if the gap \(\varepsilon_h \ldots \varepsilon_{h+i}\) is rather large, we are, even near the limits, obliged to admit that \(f'(x)\) is very small within this interval and that accordingly the reaction-function here assumes a strikingly high value.

If the frequencies \(Y_h \ldots Y_{h+i}\) are not exactly zero but yet abnormally small, the above considerations still hold. We may illustrate them by Example V: Length of Wheat-ears under scanty feeding given by Dr. C. de Bruycker (see p. 56).

II. Excessive frequencies within the limits of the domain.

If for a set of successive class-intervals (with centres \(\varepsilon_h \ldots \varepsilon_{h+i}\)) abnormally large frequency-numbers \(Y_h \ldots Y_{h+i}\) are found, the fraction \(p_k = \frac{Y_1 + \ldots + Y_h}{N}\) increases rapidly in the interval between \(x_{h-1}\) and \(x_{h+i}\), and so does \(z\).

Hence the function \(f'(x)\) reaches very large values in that interval and the reaction-function very small ones, so that the growth almost stagnates. So individuals with small \(x\) may eventually overtake the individuals whose \(x\) lies in the domain in question, and — if also negative growth is admitted — individuals with large \(x\) may pass into such with smaller \(x\).

The consequence is an accumulation of individuals in this domain, which explains the high frequency-numbers.
If the uncertainty in the correspondence \((x, z)\) is of such a kind that the slope of \(z\) (be it in a single point) may be assumed infinite (so that the tangent-line at this points becomes parallel to the axis of \(z\)), then the reaction may be considered to be zero there.

Another explanation of such an accumulation with the aid of a many-valued function \(z = f(x)\) will be given in a continuing article to be published in the Proceedings of the Kon. Akad. v. Wet. te Amsterdam, referred to by C. A.

Example IV: Diameter of spores of Mucor Mucedo, measured by Mr. G. Postma (see p. 50) may serve for an illustration of the preceding case.

III. Excessively great frequencies at the limits of the frequency-domain.

We next consider the case that the frequencies \(Y_1, Y_2, \ldots Y_i\) at the lower limit, or the frequencies \(Y_{n-i}, Y_{n-i+1}, \ldots Y_n\) at the upper limit, or those at both limits, are abnormally large. This takes place, for instance, when the frequencies form a series ascending to either or both limits.

If the first frequencies \(Y_1, \ldots Y_i\) are large, the quantity \(z\) must rise in the first interval \(x_0 \ldots x_i\) from \(-\infty\) to either a small negative or a positive value. If the last frequencies \(Y_{n-i}, \ldots Y_n\) are large, then \(z\) must rise in the last intervals \(x_{n-j} \ldots x_n\) from either a negative or a small positive value to \(+\infty\).

Now if we stick to the above developed theory, we arrive at a very peculiar and hence improbable form of the function \(f(x)\), as will be shown later on (appendix to Ch. II, II p. 66).

In order to make the theory applicable also to this case without having to operate with less acceptable functions it must be generalised by dropping the suppositions incidentally introduced for the sake of simplification. Such a generalisation will be expounded in the C. A. (see above).

6. Proportional reaction.

Two sets of individuals of the same kind may be subject to the same causes of growth, with only this difference, that the reactions of one set, characterised by \(x\) are \(\lambda\) times as strong as those of the other, represented by \(x_1\).

The elementary increments are therefore resp.

\[ \alpha = \beta \psi(x) \quad \text{and} \quad \alpha_1 = \lambda \beta \psi(x_1). \]

So we have, according to (1) (p. 32) and (3) (p. 34)

\[ \Delta Z = m \beta = \frac{m \alpha}{\psi(x)} = \frac{\Delta x}{\psi(x)}, \]

\[ = \frac{m \alpha_1}{\lambda \psi(x_1)} = \frac{\Delta x_1}{\lambda \psi(x_1)}. \]
The individuals, being supposed entirely homogeneous, are likely to have the same initial value $X$.

Hence

$$Z = \int_{\chi}^{x} \frac{dx}{\psi(x)} = \int_{\chi}^{\chi_1} \frac{dx_1}{\lambda \psi(x_1)},$$

or

$$M + \zeta = F(x) - F(X) = \frac{1}{\lambda} [F(x_1) - F(X)].$$

The same value $\zeta$ corresponds in one distribution to $x$, in the other to $x_1$, which is generally different from $x$. Inversely, a same result of observation $x = x_1 = \xi$, which is in one distribution joined to the value $\zeta$, is in the other connected with a different value, say $\zeta_1$. These values $\zeta$ and $\zeta_1$ are found from the equation

$$M + \zeta = F(\xi) - F(X),$$

whence

$$M + \zeta_1 = \frac{1}{\lambda} [F(\xi) - F(X)],$$

The difference $\zeta_1 - \zeta = \lambda (z_1 - z)$ does not contain $M$, the mean growth of $Z$. The undisturbed or initial value $X$, which otherwise is inseparably bound to $M$ by the relation $F(X) + M$ (see § 2, form. (4), p. 35) can now be determined by itself.

The above formula shows that for $x = X$ we have $\zeta_1 = \zeta$ and $z_1 = z$.

By tracing both the curves $z = f(x)$ and $z_1 = f_1(x)$ in the same system of coordinates we evidently find the initial (undisturbed) value $X$ as the abscissa of the point of intersection of the curves.

From

$$z + hM = h \{F(x) - F(X)\} \text{ and } z_1 + hM = \frac{h}{\lambda} \{F(x_1) - F(X)\}$$

it follows that the curve $z_1 = f_1(x_1)$ may be obtained by enlarging all ordinates reckoned from a certain line $z = -hM$ in a certain ratio $\frac{1}{\lambda}$.

It is by this property of the two curves that proportional reaction may be recognised.

The reaction-curves also have proportional ordinates (with the ratio $\lambda$) as was to be expected on account of our starting point.

See Example VI: Summer and winter barometric heights at den Helder (p. 57).
EXAMPLES.

CHAPTER III.

General remarks.

The frequency-numbers \( Y \) have been divided by the whole number \( N \)
the individuals; the quotients \( y \) are the ordinates of the points — marked
by a cross (\( \times \)) — of the frequency-curve (\( \cdots \)).

The points \( (x, z) \) — marked by a dot (\( \cdot \)) — are joined by the smooth
curve (\( \cdots \)) representing the normal function.

The reaction-curve (\( \cdots \)) is obtained by graphical differentiation of
the normal function.

The scale of \( x \) is given at the bottom of the figure.

The unit of \( y \) is different in the different examples. It is so chosen
that the frequency-curve is always of a convenient size.

The scale of \( y \) is marked at the right margin of the figure.

The unit of \( z \) in the original figure is 5 cm; it is 2.5 cm in the
(reduced) reproduction.

The scale of \( z \) is marked at the left margin of the figure.

As the values of \( \eta \) contain an arbitrary constant factor, the unit of \( \eta \)
is chosen according to circumstances.

Example I.
Circumference of the chest of recruits, given by A. Quetelet (Anthro-
pomérie, Bruxelles, 1871, p. 289).

Unit of \( x \): 1 inch; class-range = 1 unit = 1 inch; \( N = 1516 \).

Normal distribution: \( z = 0.334 (x - 35.0) \).

<table>
<thead>
<tr>
<th>( x )</th>
<th>( Y )</th>
<th>( z )</th>
<th>( x )</th>
<th>( Y )</th>
<th>( z )</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>2</td>
<td>-2.13</td>
<td>35</td>
<td>310</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>4</td>
<td>-1.88</td>
<td>36</td>
<td>251</td>
<td>+0.18</td>
</tr>
<tr>
<td>30</td>
<td>17</td>
<td>-1.53</td>
<td>37</td>
<td>181</td>
<td>+0.52</td>
</tr>
<tr>
<td>31</td>
<td>55</td>
<td>-1.15</td>
<td>38</td>
<td>103</td>
<td>+0.85</td>
</tr>
<tr>
<td>32</td>
<td>102</td>
<td>-0.84</td>
<td>39</td>
<td>42</td>
<td>+1.19</td>
</tr>
<tr>
<td>33</td>
<td>180</td>
<td>-0.50</td>
<td>40</td>
<td>19</td>
<td>+1.49</td>
</tr>
<tr>
<td>34</td>
<td>242</td>
<td>-0.18</td>
<td>41</td>
<td>6</td>
<td>+1.81</td>
</tr>
<tr>
<td>35</td>
<td>310</td>
<td>+0.18</td>
<td>42</td>
<td>2</td>
<td>+2.13</td>
</tr>
</tbody>
</table>
Example II.
Threshold of sensation, measured by Prof. G. Heymans (J. C. Kapteyn, Skew Frequency Curves, p. 25).
Unit of $x$: 1 decigramme; class-range = 1 unit = 1 decigr.; $N = 120$.

Logarithmic distribution: $z = 3.49 \times 10 \log \frac{x}{4.78}$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$Y$</th>
<th>$z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>-1.69</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>-1.11</td>
</tr>
<tr>
<td>3</td>
<td>23</td>
<td>-0.48</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
<td>-0.13</td>
</tr>
<tr>
<td>5</td>
<td>21</td>
<td>+0.18</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>+0.42</td>
</tr>
<tr>
<td>7</td>
<td>15</td>
<td>+0.73</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>+0.81</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Reproduction on logarithmic paper.

Example III.
Unit of $x$: 10 £; class-range from 10 £ to 500 £; $N = 5829.9$ thousand.

Logarithmic distribution: $z = 1.11 \times 10 \log \frac{x - 0.50}{0.94 - 0.50}$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$Y: 1000$</th>
<th>$z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3175.</td>
<td>+0.08</td>
</tr>
<tr>
<td>1</td>
<td>1451.</td>
<td>+0.58</td>
</tr>
<tr>
<td>2</td>
<td>441.6</td>
<td>+0.79</td>
</tr>
<tr>
<td>3</td>
<td>259.8</td>
<td>+0.97</td>
</tr>
<tr>
<td>4</td>
<td>151.0</td>
<td>+1.10</td>
</tr>
<tr>
<td>5</td>
<td>90.4</td>
<td>+1.20</td>
</tr>
<tr>
<td>6</td>
<td>104.1</td>
<td>+1.36</td>
</tr>
<tr>
<td>7</td>
<td>150.0</td>
<td></td>
</tr>
</tbody>
</table>

Reproduction on logarithmic paper.
Example IV.
Diameter of Spores of Mucor Mucedo, measured by Mr. G. Postma (unpublished).
Unit of $x$: 3.27 $\mu$; class-range = 1 unit = 3.27 $\mu$; $N = 330$.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$Y$</th>
<th>$z$</th>
<th>$x$</th>
<th>$Y$</th>
<th>$z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>3</td>
<td></td>
<td>18</td>
<td>26</td>
<td>-0.44</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>-1.67</td>
<td>19</td>
<td>50</td>
<td>-0.28</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>-1.48</td>
<td>20</td>
<td>106</td>
<td>0.00</td>
</tr>
<tr>
<td>13</td>
<td>7</td>
<td>-1.40</td>
<td>21</td>
<td>33</td>
<td>+0.65</td>
</tr>
<tr>
<td>14</td>
<td>11</td>
<td>-1.20</td>
<td>22</td>
<td>10</td>
<td>+1.00</td>
</tr>
<tr>
<td>15</td>
<td>12</td>
<td>-1.00</td>
<td>23</td>
<td>6</td>
<td>+1.18</td>
</tr>
<tr>
<td>16</td>
<td>25</td>
<td>-0.85</td>
<td>24</td>
<td>7</td>
<td>+1.33</td>
</tr>
<tr>
<td>17</td>
<td>26</td>
<td>-0.62</td>
<td>25</td>
<td>3</td>
<td>+1.67</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>-0.44</td>
<td>26</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Example V.
Length of wheat-ears, grown under unfavorable circumstances (closely sown in poor soil), measured by Dr. C. De Bruyker. (Handelingen van het 13e Vlaamsche Natuur- en Geneeskundig Congres, p. 172).
Unit of $x$: 1 mm; class-range = 10 units = 1 cm; $N = 372$.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$Y$</th>
<th>$z$</th>
<th>$x$</th>
<th>$Y$</th>
<th>$z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.5</td>
<td>28</td>
<td>-1.02</td>
<td>80.5</td>
<td>37</td>
<td>+0.18</td>
</tr>
<tr>
<td>30.5</td>
<td>101</td>
<td>-0.28</td>
<td>90.5</td>
<td>58</td>
<td>+0.37</td>
</tr>
<tr>
<td>40.5</td>
<td>30</td>
<td>-0.13</td>
<td>100.5</td>
<td>35</td>
<td>+0.76</td>
</tr>
<tr>
<td>50.5</td>
<td>15</td>
<td>-0.06</td>
<td>110.5</td>
<td>15</td>
<td>+1.18</td>
</tr>
<tr>
<td>60.5</td>
<td>26</td>
<td>+0.07</td>
<td>120.5</td>
<td>2</td>
<td>+1.70</td>
</tr>
<tr>
<td>70.5</td>
<td>24</td>
<td>+0.18</td>
<td>130.5</td>
<td>1</td>
<td>+1.97</td>
</tr>
<tr>
<td>80.5</td>
<td></td>
<td></td>
<td>140.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example VI.
Summer and Winter barometric heights at Den Helder.

Summer months: frequency-curve: _______; normal function: _______; \( N = 22602 \).

Winter months: frequency-curve: _______; normal function: _______; \( N_1 = 22188 \).

Unit of \( x \): 1 mm. mercury; class-range = 1 unit = 1 mm. mercury.
Proportional reaction: \( \lambda = \eta_i : \eta = 1.98 \).
Undisturbed value \( X = 761.2 \).

<table>
<thead>
<tr>
<th>( x )</th>
<th>( Y )</th>
<th>( z )</th>
<th>( Y_1 )</th>
<th>( z_1 )</th>
<th>( x )</th>
<th>( Y )</th>
<th>( z )</th>
<th>( Y_1 )</th>
<th>( z_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>718</td>
<td>1</td>
<td>-2.77</td>
<td>718</td>
<td>1</td>
<td>-2.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>719</td>
<td>0</td>
<td>-2.77</td>
<td>719</td>
<td>0</td>
<td>-2.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>720</td>
<td>0</td>
<td>-2.77</td>
<td>720</td>
<td>0</td>
<td>-2.77</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>721</td>
<td>2</td>
<td>-2.58</td>
<td>721</td>
<td>2</td>
<td>-2.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>722</td>
<td>2</td>
<td>-2.48</td>
<td>722</td>
<td>2</td>
<td>-2.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>723</td>
<td>0</td>
<td>-2.48</td>
<td>723</td>
<td>0</td>
<td>-2.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>724</td>
<td>4</td>
<td>-2.37</td>
<td>724</td>
<td>4</td>
<td>-2.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>725</td>
<td>3</td>
<td>-2.31</td>
<td>725</td>
<td>3</td>
<td>-2.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>726</td>
<td>1</td>
<td>-2.30</td>
<td>726</td>
<td>1</td>
<td>-2.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>727</td>
<td>5</td>
<td>-2.23</td>
<td>727</td>
<td>5</td>
<td>-2.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>728</td>
<td>9</td>
<td>-2.14</td>
<td>728</td>
<td>9</td>
<td>-2.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>729</td>
<td>11</td>
<td>-2.07</td>
<td>729</td>
<td>11</td>
<td>-2.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>730</td>
<td>23</td>
<td>-1.96</td>
<td>730</td>
<td>23</td>
<td>-1.96</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>731</td>
<td>20</td>
<td>-1.90</td>
<td>731</td>
<td>20</td>
<td>-1.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>732</td>
<td>28</td>
<td>-1.83</td>
<td>732</td>
<td>28</td>
<td>-1.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>733</td>
<td>39</td>
<td>-1.75</td>
<td>733</td>
<td>39</td>
<td>-1.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>734</td>
<td>47</td>
<td>-1.68</td>
<td>734</td>
<td>47</td>
<td>-1.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>735</td>
<td>52</td>
<td>-1.62</td>
<td>735</td>
<td>52</td>
<td>-1.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>736</td>
<td>58</td>
<td>-1.58</td>
<td>736</td>
<td>58</td>
<td>-1.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

G. Postma
<table>
<thead>
<tr>
<th>Summer</th>
<th>Winter</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>$Y$</td>
</tr>
<tr>
<td>754</td>
<td>722</td>
</tr>
<tr>
<td>755</td>
<td>873</td>
</tr>
<tr>
<td>753</td>
<td>1046</td>
</tr>
<tr>
<td>757</td>
<td>1172</td>
</tr>
<tr>
<td>753</td>
<td>1324</td>
</tr>
<tr>
<td>759</td>
<td>1438</td>
</tr>
<tr>
<td>760</td>
<td>1549</td>
</tr>
<tr>
<td>761</td>
<td>1567</td>
</tr>
<tr>
<td>762</td>
<td>1601</td>
</tr>
<tr>
<td>763</td>
<td>1671</td>
</tr>
<tr>
<td>764</td>
<td>1425</td>
</tr>
<tr>
<td>765</td>
<td>1310</td>
</tr>
<tr>
<td>766</td>
<td>1179</td>
</tr>
<tr>
<td>767</td>
<td>969</td>
</tr>
<tr>
<td>763</td>
<td>778</td>
</tr>
<tr>
<td>769</td>
<td>613</td>
</tr>
<tr>
<td>770</td>
<td>395</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Summer</th>
<th>Winter</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td>$Y$</td>
</tr>
<tr>
<td>770</td>
<td>395</td>
</tr>
<tr>
<td>771</td>
<td>278</td>
</tr>
<tr>
<td>772</td>
<td>191</td>
</tr>
<tr>
<td>773</td>
<td>110</td>
</tr>
<tr>
<td>774</td>
<td>57</td>
</tr>
<tr>
<td>775</td>
<td>20</td>
</tr>
<tr>
<td>776</td>
<td>13</td>
</tr>
<tr>
<td>777</td>
<td>6</td>
</tr>
<tr>
<td>778</td>
<td>1</td>
</tr>
<tr>
<td>779</td>
<td>1</td>
</tr>
<tr>
<td>780</td>
<td>85</td>
</tr>
<tr>
<td>781</td>
<td>41</td>
</tr>
<tr>
<td>782</td>
<td>31</td>
</tr>
<tr>
<td>783</td>
<td>13</td>
</tr>
<tr>
<td>784</td>
<td>9</td>
</tr>
<tr>
<td>785</td>
<td>6</td>
</tr>
<tr>
<td>786</td>
<td>2</td>
</tr>
</tbody>
</table>
APPENDIX TO CHAPTER 1.

Note to art. 3, remark 2.

According to formula (10) of the text, the deviations executed by any one individual $K$, under the influence of any one cause $C_h$ is:

$$
\Delta x = \frac{A_{h,K}}{F'(x)} \frac{1}{2} \frac{F''(x)}{(F'(x))^2} A_{h,K}^2
$$

Required are the cases in which these deviations can be symmetrical for the individuals of any size $x$. With some generalisation of the term, we call symmetrical the deviations $\Delta x$ if

$$
\Sigma \Delta x = 0.
$$

It thus is required to find the cases in which this sum, taken over all the individuals of an arbitrary size $x$, is zero for any cause $C_h$. Substituting (a) in (b) we have

$$
\Sigma \frac{A_{h,K}}{F'(x)} = \frac{1}{2} \frac{F''(x)}{(F'(x))^2} \Sigma A_{h,K}^2
$$

the sums being taken over the indices $K$.

A first solution will evidently be

$$
F''(x) = 0.
$$

Consequently

$$
F'(x) = a + bx
$$

if, at the same time the deviations are such that

$$
\Delta A_{h,K} = 0 \text{ (summation over } K) \nonumber
$$

In the case that $F''(x)$ is not zero, let

$$
\frac{\Sigma A_{h,K}}{\Sigma A_{h,K}^2} = B \text{ (summation over } K). \nonumber
$$

If we suppose $B$ the same for each of the causes $C_h$, we get another solution, for then

$$
\frac{F''(x)}{(F'(x))^2} = 2B,
$$

which integrated gives

$$
\frac{1}{F'(x)} = -(C + 2Bx) \text{ or } F'(x) = -\frac{1}{C + 2Bx}
$$
and integrating again:

\( (g) \quad F(x) = -\frac{1}{2B} \log(C + 2Bx) + C' \)

which is easily brought under the forms of the text.

The forms \((d)\) and \((g)\) are thus seen to be the only ones for \(F(x)\), in which symmetrical deviations of the \(x\) are possible.

We can easily verify that in these cases — provided the deviations are indeed symmetrical — the arithmetical mean \(\bar{x}\) is indeed \(\bar{x}_0\).

In the case \((d)\), the equation of the frequency curve, according to formula (15) of the text is:

\[
\Omega(x) = \frac{b}{\sqrt{2\pi} e} e^{-\frac{1}{2\sigma^2} (x - \bar{x})^2 - M^2}.
\]

The condition of symmetry, according to \((e)\), being \(\Sigma A_k = 0\) (for every cause \(\gamma_k\)) we have by formula (13) of the text \(M = 0\), therefore

\[
\Omega(x) = \frac{b}{\sqrt{2\pi} e} e^{-\frac{1}{2\sigma^2} (x - \bar{x})^2}.
\]

This is a normal curve having really its centre of gravity \((x = \bar{x})\) at \(x = \bar{x}_0\).

In the case \((g)\), the frequency curve, according to formula (15) of the text, will be

\[
\Omega(x) = \frac{1}{\sqrt{2\pi} e} C + 2Bx
\]

The limits of this curve lie at the points

\[ \begin{align*}
    x &= \infty \quad \text{and} \quad x = -\frac{C}{2B}
\end{align*} \]

for which \(F(x)\) becomes respectively \(-\infty\) and \(+\infty\).

The arithmetical mean therefore is

\[
\bar{x} = \frac{1}{e\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2} (x - \bar{x})^2} d\bar{x} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2} (x - \bar{x})^2} - M^2 d\bar{x}.
\]

Put

\[
\frac{1}{e\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2} (x - \bar{x})^2} - M^2 d\bar{x} = z.
\]

We get

\[
\bar{x} = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2} x} + \frac{C + 2Bx_0}{2B} e^{-\frac{1}{2\sigma^2} x} - \frac{1}{2B} \log(C + 2Bx) - M d\bar{x}.
\]

Putting in the last

\[ \bar{x} = \frac{Be\sqrt{2}}{2} = t. \]
We get, because \( \int_{-\infty}^{\infty} e^{-y^2} dy = V_\pi \),
\[
\bar{x} = - \frac{C}{2B} + \frac{2B}{C + 2Bx_0^*} e^{-2B(M - Bx_0^*)}.
\]

According to (f) the deviations will be symmetrical in this case, if, for every cause \( C_h \)
\[
\frac{\Sigma A_h, K}{\Sigma A^2_h, K} = B
\]
or with the notation of art. 1
\[
B = \frac{\Sigma (A_h + \alpha_{h,K})}{\Sigma (A_h + \alpha_{h,K})^2} \quad \text{(summation over } K). \]

It was shown in art. 2 that, in the case of an infinity of causes — which is here assumed — the \( A_h \) must be of higher order of smallness than the \( \alpha_{h,K} \), so that, \( \Sigma \alpha_{h,K} \) being = 0, we may put
\[
B = \frac{\Sigma A_h}{\Sigma \alpha_{h,K}^2} = \frac{A_h}{\alpha_h^2} \quad \text{or} \quad A_h = B\alpha_h^2.
\]

Taking the sum of the similar equations for the whole of all the causes we have, according to formulae (13) and (14) of the text
\[
M = B\alpha^2
\]
which, in (h) gives
\[
\bar{x} = x_0^*.
\]
APPENDIX TO CHAPTER II.

In the appendix to chapter II we propose to give firstly some demonstrations and explanatory notes referring to the theory worked out above, secondly a short discussion of the equation of the frequency-curve.

The further development of this theory will be given in the C. A. (see p. 52). In this latter extension we partly apply this theory — facilitated as it is by the two simplifications introduced on pp. 37, 38 — to a few new cases, which are — though a little more intricate — closely related to those already treated.

In the C. A. we shall also expand the sphere of action of our theory by dropping the simplifications mentioned, particularly in investigating high frequency-numbers at the limits of the whole domain.

I. Explanatory notes.

A. The probable error $\varrho_z$ of $z$.

From

$$p = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt$$

ensues

$$\frac{dp}{dz} = e^{-\frac{z^2}{2}} \sqrt{\pi}.$$

A small deviation $\Delta z$ from a certain value $z$ is accompanied by a usually also small deviation $\Delta p$ from the corresponding value of $p$; this deviation may be approximated by

$$\Delta p = \frac{e^{-\frac{z^2}{2}}}{\sqrt{\pi}} \Delta z.$$

So to the probable error $\varrho_z$ of $z$ corresponds the probable error $\varrho_p$ of $p$, according to the (approximative) relation

$$\varrho_p = \frac{e^{-\frac{z^2}{2}}}{\sqrt{\pi}} \varrho_z,$$

whence

$$\varrho_z = \sqrt{\frac{2p(1-p)}{N}} \varrho_p = e^{\frac{z^2}{2}} \varrho_p \sqrt{\frac{2p(1-p)}{N}}.$$

[$\varrho = 0.47694, N = \text{whole number of individuals, } p = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt.$]
Now, for \( z = 0 \) we have \( e^z = 1 \) and \( p = \frac{1}{2} \), so that

\[
|q_2|_{z=0} = 1, e^{2,1,1,1,\pi} = e^{\sqrt{\pi}/2N} = 0.598.
\]

In the neighbourhood of \( z = 0 \) we may, without serious error, still use the same value for \( q_2 \).

An expression can also be given for \( q_2 \), which holds near the limits \( z = \pm \infty \). Here we make use of the so-called „Error-function” introduced by Glaisher and defined by

\[
\text{Erf} \ z = \int_{-\infty}^{+\infty} e^{-t^2} dt,
\]

which, for large values of \( z \), can be expanded in powers of \( \frac{1}{z} \) in this way:

\[
\text{Erf} \ z = \frac{e^{-z^2}}{2z} \left( 1 - \frac{1}{2z^3} + \frac{1.3.5}{(2^3z^5)} - \frac{1.3.5.7}{(2^5z^7)} + \ldots \right) \)

For large negative values \( -z \) the probability \( p \) is very small, so that \( 1 - p \) is nearly 1. So the probable error \( q_p \) of \( p \) is approximated by

\[
q_p = e^{\sqrt{2p}}
\]

and the corresponding probable error \( q_\zeta \) of \(-z\) by

\[
q_\zeta = \sqrt{\frac{2\pi}{N}} \cdot e^{\sqrt{2p}} = \sqrt{\frac{2\pi}{N}} \cdot e^{\sqrt{2p}} V_p,
\]

or, since

\[
p = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2} dt = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2} dt = \frac{1}{\sqrt{\pi}} \cdot \text{Erf} \ z = \frac{1}{\sqrt{\pi}} \cdot \frac{e^{-z^2}}{2z} \left| 1 - \frac{1}{2z^2} + \ldots \right|
\]

\[
q_\zeta = \sqrt{\frac{2\pi}{N}} \cdot e^{\sqrt{2p}} \cdot \frac{1}{\sqrt{\pi}} \cdot \frac{e^{-z^2}}{2z} \left| 1 - \frac{1}{2z^2} + \ldots \right|
\]

or

\[
q_\zeta = \frac{\sqrt{\pi}}{\sqrt{N}} \cdot \frac{e^{-z^2}}{\sqrt{\zeta}} \left| 1 - \frac{1}{2z^2} + \ldots \right|
\]

A very tolerable approximation is still given by

\[
q_\zeta = \frac{\sqrt{\pi}}{\sqrt{N}} \cdot \frac{e^{-z^2}}{\sqrt{\zeta}}
\]

The smallest a-posterioric probability \( p \) which can occur in a set of \( N \) individuals is obviously
\[
p' = \frac{1}{N};
\]
of course this value is found at the lower limit of the frequency-domain.
At the upper limit the value for \( p \) is at the utmost
\[
p'' = \frac{N - 1}{N}.
\]
Hence the expression \( \frac{p(1 - p)}{N} \) has for its lower limit
\[
\frac{p'(1 - p')}{N} = \frac{p''(1 - p'')}{N} = \frac{1}{N} \left( \frac{1}{N} - \frac{1}{N} \right) = \frac{1}{N^2},
\]
and approximately
\[
\frac{p'(1 - p')}{N} = \frac{1}{N^2} = p'^2,
\]
so that an approximative value for the probable error \( \varepsilon_p \), of the corresponding \( p' \) is found in
\[
\varepsilon_p = \varepsilon \sqrt{2 \cdot p'},
\]
or, expressed in terms of the corresponding value \( z' \) of \( z \):
\[
\varepsilon_p = \varepsilon \sqrt{2 \cdot \frac{e^{-z'^2}}{\pi \cdot 2z'}} \left( 1 - \frac{1}{2z^2} + \ldots \right).
\]
Hence the approximative value for the probable error \( \varepsilon_z \) of \( z' \) is
\[
\varepsilon_z = \sqrt{\pi \cdot e^{\sigma^2}} \varepsilon_p = \varepsilon \sqrt{2 \pi \cdot e^{\sigma^2} \cdot \frac{e^{-z'^2}}{\pi \cdot 2z'}} \left( 1 - \frac{1}{2z^2} + \ldots \right)
\]
or, more roughly,
\[
\varepsilon_z = \varepsilon \sqrt{\frac{1}{2} \cdot \frac{1}{z'}}.
\]
So we conclude that the probable error in the largest \( z \) which may be found near the limits of the domain, or the probable error in the maximum values of \( |z_1| \) and \( z_{n-1} \), is nearly inversely proportional to these maximum values themselves; hence this error becomes less with increasing values of the number \( N \) of the individuals.

The following table shows the numerical relations between the values of \( z \) and those of \( p, \ v, \sqrt{pN}, \) and \( \varepsilon_z \):

<table>
<thead>
<tr>
<th>( \pm z )</th>
<th>0</th>
<th>0.1</th>
<th>0.5</th>
<th>1.0</th>
<th>1.5</th>
<th>2.0</th>
<th>2.5</th>
<th>3.0</th>
<th>3.5</th>
<th>4.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p ) or ( 1 - p )</td>
<td>0.500</td>
<td>0.444</td>
<td>0.340</td>
<td>0.0786</td>
<td>0.0169</td>
<td>0.00234</td>
<td>203 \times 10^{-6}</td>
<td>110 \times 10^{-5}</td>
<td>372 \times 10^{-5}</td>
<td>771 \times 10^{-5}</td>
</tr>
<tr>
<td>( \varepsilon_z \sqrt{N} )</td>
<td>0.598</td>
<td>0.630</td>
<td>0.653</td>
<td>0.677</td>
<td>0.684</td>
<td>3.153</td>
<td>8.838</td>
<td>32.20</td>
<td>152.3</td>
<td>932.8</td>
</tr>
<tr>
<td>( \varepsilon_z )</td>
<td>0.400</td>
<td>0.321</td>
<td>0.246</td>
<td>0.191</td>
<td>0.152</td>
<td>0.126</td>
<td>0.107</td>
<td>0.0928</td>
<td>0.0819</td>
<td></td>
</tr>
</tbody>
</table>
For instance \( N = 10000 \) furnishes a minimum value \( p' = 0.0001 \) for \( p \), and accordingly \( z' = 2.630 \). The probable error of \( z \) takes the values

\[
\begin{array}{cccccccc}
\pm z & 0 & 0.1 & 0.5 & 1.0 & 1.5 & 2.0 & 2.5 & 2.630 \\
\theta_z & 0.00593 & 0.00600 & 0.00655 & 0.00877 & 0.01464 & 0.03153 & 0.08838 & 0.12065 \\
\end{array}
\]

B. Rigorous discussion of the shape of the curves \( V \) and \( W \) (pp. 46–50).

We shall now prove rigorously that the curve \( V \) is concave downward and \( W \) convex downward.

Introducing the number \( e \), base of the Neperian logarithms, and the modulus \( \mu = M - 10\log e = 0.434295 \ldots \), we find from

\[
u' = 10\log (x - x_0') \quad \text{or} \quad x = x_0' + 10^\nu
\]

the relations

\[
x = x_0' + \frac{u'}{M}
\]

and

\[
u = 10\log (x - x_0') = M \log (x - x_0) = M \log (\frac{u'}{M} + x_0' - x_0),
\]

or, putting

\[
x_0' - x_0 = \sigma,
\]

\[
u = M \log (\frac{u'}{M} + \sigma)
\]

and

\[
z = \lambda (u - u_m) = \lambda u + \text{const.} = \lambda M \log (\frac{u'}{M} + \sigma) + \text{const.}
\]

Hence

\[
\frac{dz}{d\nu'} = \frac{\lambda}{\frac{u'}{M} + \sigma} \times \frac{u'}{M} = \frac{\lambda}{1 + \sigma e^{-\frac{u'}{M}}}
\]

and

\[
\frac{d^2z}{d\nu'^2} = \frac{-\lambda}{\frac{u'}{M} + \sigma} \times \frac{-\sigma - \frac{u'}{M}}{M e^{-\frac{u'}{M}}} = \frac{\lambda \sigma e^{-\frac{u'}{M}}}{M(1 + \sigma e^{-\frac{u'}{M}})^2}
\]

Since \( \lambda > 0 \), \( \frac{d^2z}{d\nu'^2} \) has the same sign as \( \sigma \).

Two small a value \( x_0' \) makes \( \sigma < 0 \), consequently \( \frac{d^2z}{d\nu'^2} < 0 \) (concave downward).

Too large a value \( x_0' \) makes \( \sigma > 0 \), consequently \( \frac{d^2z}{d\nu'^2} > 0 \) (convex downward).

With increasing \( u' \) the absolute value of \( \frac{d^2z}{d\nu'^2} \) becomes smaller.

As the slope varies but little, also the curvature decreases for increasing \( u' \), viz. left above.
In the same manner the second logarithmic curve is rigorously treated as follows.

From
\[ u' = 10 \log (x_n' - x) \]

we derive
\[ x = x_n' - 10^u' = x_n' - e^{x'} \]

and
\[ u = 10 \log (x_n - x) = M \log (x_n - x) = M \log (x_n - x_n' + e^{x'}) \]

or, putting
\[ x_n' - x_n = \tau \]

and
\[ u = M \log (e^{x'} - \tau) \]

and
\[ z = \lambda (u_m - u) = \text{const.} - \lambda u = \text{const.} - \lambda M \log (e^{x'} - \tau) \]

whence
\[ \frac{dz}{du'} = \frac{-\lambda M \times e^{x'}}{M} = \frac{-\lambda}{1 - \tau e^{-x'}} \]

and
\[ \frac{d^2z}{du'^2} = \frac{+\lambda \times \tau e^{-x'}}{M(1 - \tau e^{-x'})^2} = \frac{\lambda \tau e^{-x'}}{M(1 - \tau e^{-x'})^2} \]

On account of \( \lambda > 0 \), \( \frac{d^2z}{du'^2} \) and \( \lambda \) have the same sign.

Too small a value of \( x_n' \) makes \( \tau < 0 \), consequently \( \frac{d^2z}{du'^2} < 0 \) (concave downward). Too large a value of \( x_n' \) makes \( \tau > 0 \), consequently \( \frac{d^2z}{du'^2} > 0 \) (convex downward).

With increasing \( u' \), that is right below, the absolute value of \( \frac{d^2z}{du'^2} \) becomes less, and (on account of the small variation of the slope \( \frac{dz}{du'} \)) the curvature decreases also.

II. The equation of the frequency-curve.

The area of the frequency-curve bounded by the ordinate-line \( x \), amounts to
\[ W^x_{x_n} = \int_{x_n}^x \frac{dW}{dx} \, dx \]
so that for the ordinate \( y \) of the frequency-curve

\[ y = \frac{dW}{dx}. \]

Now

\[ \frac{dW}{dz} = \frac{1}{\sqrt{\pi}} e^{-z^2}, \]

hence

\[ \frac{dW}{dx} = \frac{dW}{dz} \cdot \frac{dz}{dx} = \frac{1}{\sqrt{\pi}} e^{-\left[f(x)^2\right]} \cdot f'(x). \]

So the equation of the frequency-curve is found to be

\[ y = \frac{f'(x)}{\sqrt{\pi}} e^{-\left[f(x)^2\right]}. \]

If \( z = f(x) = \infty \) for \( x = \xi \), the factor \( e^{-z^2} = e^{-\left[f(x)^2\right]} \) becomes infinitesimal of an excessively high order. If \( y \) shall be finite, we must make \( f'(x) \) in \( x = \xi \) infinite of the same order as \( e^{-z^2} \). This can only be done with very peculiar forms of the function \( f(x) \), so that we are usually inclined to attribute a finite value of \( y \) to a likewise finite value of \( z = f(x) \).
APPENDIX TO CHAPTER III.

As this paper is going through the press, Miss. Dr. T. TAMMIS, the well known botanist, sends us kindly a curious frequency curve showing strong accumulation at the lower limit. The curve is given, further below, under the head \( Y \). As this might be a good test case, we requested that no particulars should be communicated before we had derived the normal function \( z \) and the reaction curve \( (\eta) \) in the ordinary way.

Though we can give no figure it must be easy to follow the course of the latter curve from the numbers \( (\eta) \) in our table. The value of the ordinates show that it starts from zero and then rises extremely abruptly. A maximum however is soon reached at about \( z = 25 \), after which it steadily decreases, so that the reaction for \( z = 100 \) is already below half what it is at maximum.

The meaning of this is of course, that the individuals evidently have great difficulty in starting their growth. There seems to be an almost insuperable impediment against beginning growth. Those individuals, however, who succeed in overcoming the first difficulty then begin to grow very rapidly indeed, the rapidity increasing till the size 25 is reached. After that the growth begins to diminish; it gradually decreases, to only one tenth of the maximum growth for the individuals of size 100 and below.

All this proves to be in good agreement with what has been really observed. Dr. TAMMIS writes: "The case I sent you is as follows: the quantities communicated are Stalk-lengths of Linum crepitans, a variety of the ordinary flax. They were measured, at a moment in which the growth had not yet ceased, by Miss A. HAGA. The seeds were sown in a great deep flower-pot. Their number was purposely taken very high, so that they were extremely crowded. At starting, therefore, the difficulty for each seed was to get a root into the soil. It seems allowable to assume that all the seeds germinated. This has necessarily entailed an intense struggle and many individuals must not have succeeded or not sufficiently succeeded. For those who really got their root in the soil there

\[
\begin{align*}
\begin{array}{c|c}
 x & \\
 0 & \\
 5 & \\
 10 & \\
 15 & \\
 20 & \\
 25 & \\
 30 & \\
 35 & \\
 40 & \\
 45 & \\
 50 & \\
 55 & \\
 60 & \\
 65 & \\
 70 & \\
 75 & \\
 80 & \\
 85 & \\
 90 & \\
 95 & \\
 100 & \\
 105 & \\
 110 & \\
 115 & \\
\end{array}
\end{align*}
\]

1 In finding the reaction curve the normal function \( z \) was first smoothed.
now came a good time. There was plenty of food for a good many of
very small plants. The case however changed when the plants, becoming
greater, required more room. Then a second struggle ensued, viz the
struggle for the available food in the too narrow room. The plants now
became more and more impeded in their growth.

"It seems to me that the conclusions from your curve are well in
accordance with the facts."

<table>
<thead>
<tr>
<th>x</th>
<th>Y</th>
<th>z</th>
<th>η</th>
<th>x</th>
<th>Y</th>
<th>z</th>
<th>η</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>148</td>
<td>0</td>
<td>0</td>
<td>115</td>
<td>34</td>
<td>-0.181</td>
<td>93</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>-0.852</td>
<td>185</td>
<td>120</td>
<td>46</td>
<td>-0.139</td>
<td>79</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>-0.824</td>
<td>215</td>
<td>125</td>
<td>56</td>
<td>-0.073</td>
<td>69</td>
</tr>
<tr>
<td>15</td>
<td>8</td>
<td>-0.813</td>
<td>230</td>
<td>130</td>
<td>63</td>
<td>-0.000</td>
<td>47</td>
</tr>
<tr>
<td>20</td>
<td>7</td>
<td>-0.783</td>
<td>240</td>
<td>135</td>
<td>65</td>
<td>0.084</td>
<td>39</td>
</tr>
<tr>
<td>25</td>
<td>10</td>
<td>-0.764</td>
<td>243</td>
<td>140</td>
<td>94</td>
<td>0.170</td>
<td>33</td>
</tr>
<tr>
<td>30</td>
<td>9</td>
<td>-0.739</td>
<td>241</td>
<td>145</td>
<td>65</td>
<td>0.301</td>
<td>30</td>
</tr>
<tr>
<td>35</td>
<td>9</td>
<td>-0.721</td>
<td>236</td>
<td>150</td>
<td>83</td>
<td>0.400</td>
<td>23</td>
</tr>
<tr>
<td>40</td>
<td>10</td>
<td>-0.700</td>
<td>227</td>
<td>155</td>
<td>55</td>
<td>0.537</td>
<td>23</td>
</tr>
<tr>
<td>45</td>
<td>14</td>
<td>-0.680</td>
<td>214</td>
<td>160</td>
<td>71</td>
<td>0.642</td>
<td>24</td>
</tr>
<tr>
<td>50</td>
<td>13</td>
<td>-0.653</td>
<td>196</td>
<td>165</td>
<td>43</td>
<td>0.796</td>
<td>22.5</td>
</tr>
<tr>
<td>55</td>
<td>13</td>
<td>-0.625</td>
<td>179</td>
<td>170</td>
<td>46</td>
<td>0.919</td>
<td>21</td>
</tr>
<tr>
<td>60</td>
<td>21</td>
<td>-0.600</td>
<td>166</td>
<td>175</td>
<td>26</td>
<td>1.081</td>
<td>19.5</td>
</tr>
<tr>
<td>65</td>
<td>16</td>
<td>-0.560</td>
<td>156</td>
<td>180</td>
<td>24</td>
<td>1.211</td>
<td>13.5</td>
</tr>
<tr>
<td>70</td>
<td>22</td>
<td>-0.534</td>
<td>148</td>
<td>185</td>
<td>16</td>
<td>1.381</td>
<td>17.5</td>
</tr>
<tr>
<td>75</td>
<td>19</td>
<td>-0.495</td>
<td>142</td>
<td>190</td>
<td>11</td>
<td>1.564</td>
<td>18.5</td>
</tr>
<tr>
<td>80</td>
<td>21</td>
<td>-0.462</td>
<td>136</td>
<td>195</td>
<td>2</td>
<td>1.810</td>
<td>15.5</td>
</tr>
<tr>
<td>85</td>
<td>19</td>
<td>-0.429</td>
<td>131</td>
<td>200</td>
<td>2</td>
<td>1.89</td>
<td>15</td>
</tr>
<tr>
<td>90</td>
<td>38</td>
<td>-0.388</td>
<td>127</td>
<td>205</td>
<td>1</td>
<td>2.01</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>21</td>
<td>-0.341</td>
<td>123</td>
<td>210</td>
<td>0</td>
<td>2.10</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>-0.311</td>
<td>119</td>
<td>215</td>
<td>0</td>
<td>2.10</td>
<td></td>
</tr>
<tr>
<td>105</td>
<td>23</td>
<td>-0.272</td>
<td>113</td>
<td>220</td>
<td>2</td>
<td>2.10</td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>43</td>
<td>-0.240</td>
<td>104</td>
<td>225</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>115</td>
<td></td>
<td>-0.181</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
EXAMPLE I
Circumference of the chest of recruits.

EXAMPLE XVI.
Summer and Winter barometric heights.