6

REPRINT FROM THE

PROCEEDINGS

OF THE

ROYAL SOCIETY OF EDINBURGH.

SESSION 1910–1911.

VOL. XXXI.—PART II.—(No. 14.)

The Mathematical Theory of Random Migration and Epidemic Distribution.

By John Brownlee, M.D., D.Sc.

EDINBURGH:

Published by ROBERT GRANT & SON, 107 Princes Street, and WILLIAMS & NORGATE, 14 Henrietta Street, Covent Garden, London.

MDCCCCXI.

Price One Shilling and Eightpence.

XIV.—The Mathematical Theory of Random Migration and Epidemic Distribution. By John Brownlee, M.D., D.Sc.

(MS. received May 16, 1910. Read same date.)

The general theory of epidemic disease I have already considered in a communication to this Society.* In that communication I showed that the course of epidemics of all forms of infectious disease obeyed certain very definite laws. In the same paper it was also shown that the distribution of epidemic disease in a uniformly populated area obeyed a law essentially similar. Certain reasons were given why the normal curve of error $y = y_0 e^{-\frac{x^2}{2\sigma^2}}$ might be expected to give an approximate solution in both the cases considered, but why the distribution actually found (type iv.) should be the common form was not at all clear. I think, however, I have now arrived at the solution.

The distribution of an epidemic in space is evidently a problem in chance. If there is an infective group in the middle of a uniformly disposed population, then the distance from which friends come to visit a sick person or the distance a sick person travels while developing the disease determines the subsequent distribution of cases—a distribution, therefore, obeying some law on the average. This problem has since been attacked and solved by Professor Pearson under the title of "The Problem of Random Migration." The case which he considers refers specially to the prevention of malaria, which is now known to be spread through the agency of mosquitoes. The mathematical theory, which is very complex, leads to the determination that the normal surface of error gives a very close representation of this distribution. For epidemiological purposes the result is quite sufficiently close. To make the matter perfectly clear, the conditions of the problem solved are given in Professor Pearson's own words:—

"(1) Breeding grounds and food supply are supposed to have an average uniform distribution over the district under consideration. There is to be no special following of river beds or forest tracks.

"(2) The species scattering from a centre is supposed to distribute itself uniformly in all directions. The average distance through which an individual of the species moves from habitat to habitat

* *Proc. Roy. Soc. Edin.*, June 1906.

will be spoken of as a 'flight,' and there may be n such 'flights' from locus of origin to breeding ground, or again from breeding ground to breeding ground, if the species reproduces more than once. A flight is to be distinguished from a 'flitter,' a mere to-and-fro motion associated with the quest for food or mate in the neighbourhood of the habitat.

"(3) Now, taking a centre, reduced in the idealised system to a point, what would be the distribution after random flights of N individuals departing from this centre? This is the first problem. I will call it the 'Fundamental Problem of Random Migration.'

"(4) Supposing the first problem solved, we have now to distribute such points over an area bounded by any contour, and mark the distribution on both sides of the contour after any number of breeding seasons. The shape of the contour and the number of seasons dealt with will provide a series of problems which may be spoken of as 'Secondary Problems of Migration.'"

The proof of the theory given by Professor Pearson contains also implicitly the proof that if the normal surface of error describes the distribution at any moment, it will at all subsequent times. This can be seen, however, quite easily otherwise. Thus if $y = y_0 e^{-\frac{x^2+y^2}{2\sigma^2}}$ be the distribution of disease where y_0 is the number of cases per unit area at origin and $\sqrt{x^2+y^2}$ is the distance of any point from the origin, then the amount of disease at any point x', y' is $y_0 e^{-\frac{x'^2+y'^2}{2\sigma^2}}$. This element gives rise to a new normal surface $y_0 e^{-\frac{(x-x')^2+(y-y')^2}{2\sigma^2}-\frac{x'^2+y'^2}{2\sigma^2}}$ due to the infection at $x'y'$. If these several surfaces be integrated from $-\infty$ to $+\infty$ with respect to x' and y' successively, we get the new distribution

$$y = y_0 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\frac{-(x-x')^2+(y-y')^2}{2\sigma^2}-\frac{x'^2+y'^2}{2\sigma^2}} dx'dy'$$

$$= y_0 \pi\sigma^2 e^{-\frac{x^2+y^2}{4\sigma^2}},$$

so that the standard deviation σ is multiplied by $\sqrt{2}$—that is, that the slope of the surface is flattened. In other words, the longer the disease is present in a town the more uniformly, other things being equal, it will tend to distribute itself.

At the present point it will perhaps be well to recapitulate the method by which it was considered likely that the normal curve of error should represent the course of an epidemic in time. If there be an amount of infectious disease a: if such element infect pa persons, and if q be the rate of loss of infectivity per unit time, as has already been shown to be probable, the number of each group of infected persons (supposing the supply of susceptible persons large) is $a : ap : ap \times pq : ap^2q \times pq^2 : ap^3q^3 \times pq^3$, etc., or, in general, if x denote the unit time, $y = ap^{x-1}q^{\frac{(x-1)(x-2)}{2}}$, which, as q is necessarily less than unity, is the normal curve of error. If, instead of finite, infinitesimal differences are employed, the result is expressed by the formula

$$y = ap^x q^{\frac{1}{2}x^2}.$$

That the normal curve is to be taken as that from which variation is to be expected both when the space and time distributions of epidemics are examined then seems clear, and it remains to discover in what manner the natural process differs from that so far developed theoretically.

It is, in the first place, to be noted that if the distribution is from a central area instead of a point, a disturbing influence on the shape of the curve comes into play. This can be allowed for at once. For purposes of convenience a two-dimensional solution is given, such a solution being easily extended to three dimensions in any particular case. If we consider the modification of the normal curve produced when the mosquitoes start from an area and not from a point, the moments of the resulting distribution will not be those of the normal curve, that is, not

$$\int_{-\infty}^{\infty} x^n e^{-\frac{x^2}{2\sigma^2}} dx,$$

but

$$\int_{-a}^{a} da \int_{-\infty}^{\infty} (x+a)^n e^{-\frac{x^2}{2\sigma^2}} dx;$$

or, for the even moments, since the odd moments are zero,

$$\mu'_2 = \mu_2 + \frac{a^2}{3}$$

$$\mu'_4 = \mu_4 + 2a^2\mu_2 + \frac{a^4}{5},$$

so that

$$\frac{\mu_4}{\mu_2^2} = \frac{\mu_4 + 2a^2\mu_2 + \frac{a^4}{5}}{\left(\mu_2 + \frac{a^2}{3}\right)^2}$$

$$= \frac{3\mu_2^2 + 2a^2\mu_2 + \frac{a^4}{5}}{\mu_2^2 + \frac{2a^2\mu_2}{3} + \frac{a^4}{9}}$$

since for the normal curve $\mu_4 = 3\mu_2^2$.

The latter is always less than 3, so that if the centre of distribution is extended, that is, if the original mosquitoes are uniformly dispersed over a space bounded by two parallel lines, the subsequent distribution will resemble a curve of type ii. rather than the normal curve. But, as has been said, type iv. almost uniformly occurs. Some other modification is therefore necessary. This may be found in the fact that σ is not constant. In the simplified problem of Professor Pearson $2\sigma^2 = nl^2$ where n is the number of the flights and l the length of the mean flight. The curve is thus given by $y = \frac{\mathrm{N}}{\pi n l^2} e^{-\frac{x^2}{nl^2}}$. If l, however, vary, we do not get the normal curve at all, but a derivative. The frequency of any value of l or σ may be taken as given by $f(\sigma)$ when the limits of σ are α and β. The surface of distribution derived from this $y = \frac{\mathrm{N}}{\pi} \int_\alpha^\beta \frac{f(\sigma)}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}} d\sigma$. Further, if the distribution take place from a definite area as above described, the final form of the distribution of the organism becomes

$$y = y_0 \int_{x-c}^{x+c} \phi x dx \int_\alpha^\beta f(\sigma) e^{-\frac{x^2}{2\sigma^2}} \frac{d\sigma}{\sigma^2} \quad . \quad . \quad . \quad . \quad (1)$$

when ϕx denotes the mode of distribution in the area. This, if the forms of $\phi(x)$ and $(f\sigma)$ are known, I take to be the fundamental epidemic or random migration equation.

To return for a moment to the form found for the time distribution of an epidemic

$$y = ap^x q^{\frac{1}{2}x^2}.$$

This may be put in form

$$y = ae^{x \log p + \frac{1}{2}x^2 \log q}$$

$$= ae^{\frac{1}{2}\log q\left(x + \frac{\log p}{\log q}\right)^2 - \frac{1}{2}\frac{(\log p)^2}{\log q}}$$

$$= ae^{-\frac{(x+c)^2}{\sigma^2} + \frac{c}{\sigma^2}}$$

$$= ae^{-\frac{\xi^2}{\sigma^2} + \frac{c^2}{\sigma^2}} \quad \text{changing the origin of } x \quad . \quad . \quad (2)$$

$$\text{if } \log q = -\frac{2}{\sigma^2} \quad \text{and} \quad \frac{\log p}{\log q} = c.$$

From the symmetry of the epidemic c must in general be a constant, so that we have the relation

$$c \log q = \log p$$

or

$$p = q^c$$

as the relationship between the infectivity and the rate of loss of infectivity. As q is less than unity c will be negative in sign, so that we have as the general rule that as p increases q decreases, or the greater the infectivity the more rapidly it is lost.

In the case of the epidemic the frequency of each value of σ is quite unknown. As $\log q = -\frac{2}{\sigma^2}$ and as q must from the initial assumption lie between 0 and 1, the limits of σ are evidently $-\infty$ and 0. The situation of the commonest value of σ is therefore between these limits, and must in general lie nearer zero than infinity. In choosing an arbitrary form for σ so as to get an approximation, it must be of such a form that (2) will be integrable and expressible in a form suitable to calculation.

The equation obtained by integrating (2) is $y = \int_0^\infty e^{-\frac{\xi^2}{\sigma^2}+\frac{c^2}{\sigma^2}} f\sigma d\sigma$. It is obvious that if this be finite when $\sigma = 0$ that the term $e^{+\frac{c^2}{\sigma^2}}$ must disappear, so that part of $f(\sigma)$ must be $e^{-\frac{c^2}{\sigma^2}}$: the other part may be taken as $e^{-\frac{\sigma^2}{k^2}}$. The function $e^{-\frac{c^2}{\sigma^2}-\frac{\sigma^2}{k^2}}$ has a maximum when $\sigma = \sqrt{ck}$. The constant c is quite unknown, nor does it seem ascertainable from the method of analysis, but it disappears, and as k is at our disposal it is evident that the maximum value of σ can be placed where the statistics to be examined demand. We thus have as a possible form of the epidemic equation

$$y = \int_0^\infty e^{-\frac{\xi^2}{\sigma^2}-\frac{\sigma^2}{k^2}} dx.$$

To increase the variety of the distribution we might take $\sigma^n e^{-\frac{\xi^2}{\sigma^2}-\frac{\sigma^2}{k^2}}$ as representing the variation of σ: in this case the final integral becomes

$$y = \int_0^\infty \sigma^n e^{-\frac{\xi^2}{\sigma^2}-\frac{\sigma^2}{k^2}} d\sigma.$$

The epidemic due to an organism instantaneously becoming infective and thereafter losing its infectivity at a rate corresponding to the geometrical progression should at least approximately fit the above curve if the distribution of σ has been at all closely guessed. When we turn to Professor Pearson's approximate form of solution of the random migration problem we find that it also has a term with σ in the denominator. The distribution is given by

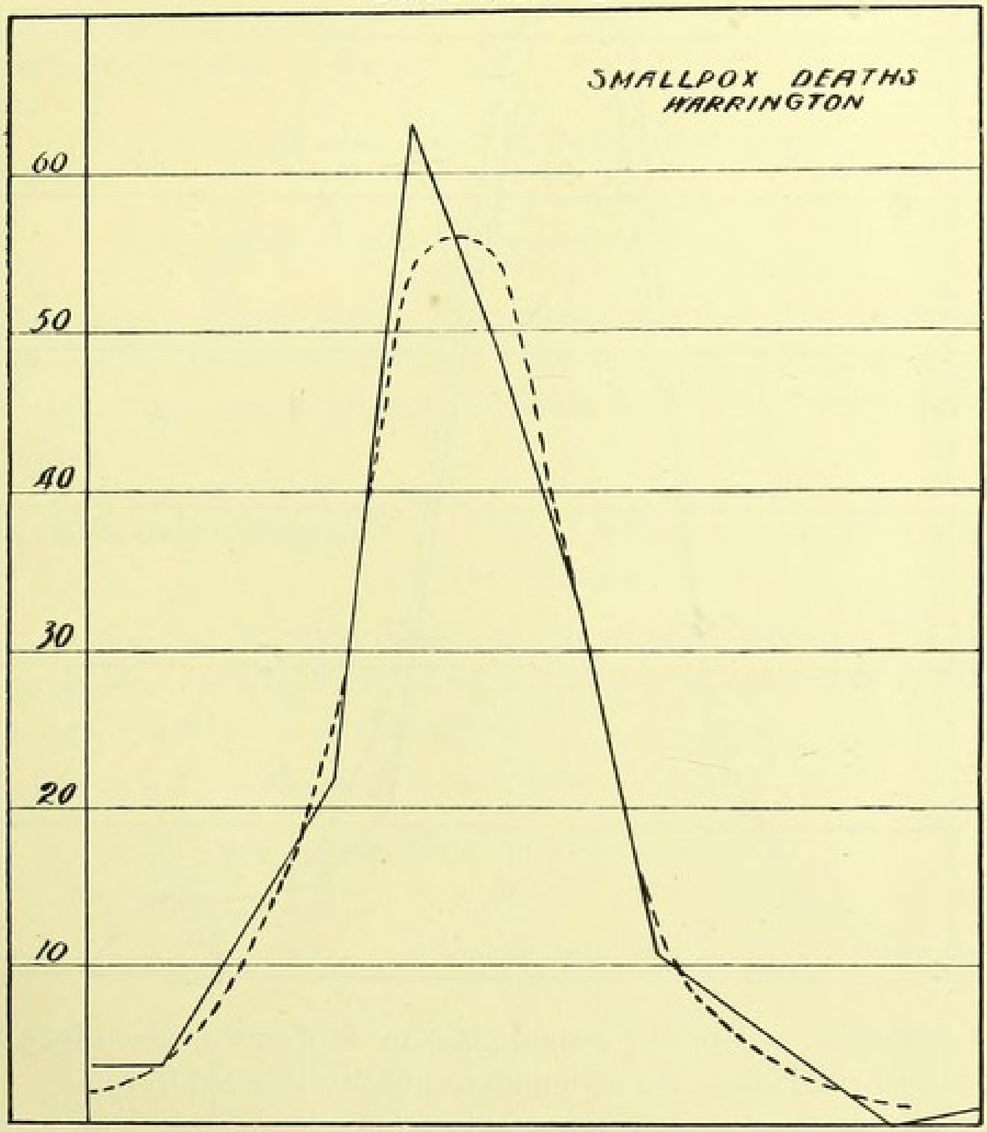
$$y = \frac{N}{2\pi\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$$

where N is the total number of mosquitoes starting from the point of origin and $\sigma = \frac{1}{2}nl^2$, n being equal to the number of flights and l to the length of the average flight. Now, if σ vary from 0 to ∞ for a first approximation, its frequency might be represented by type iii. or, say, $y=\sigma^n e^{-\frac{\sigma}{k}}$, in which case the migration form would be

$$\frac{\mathrm{AN}}{2\pi}\int_0^\infty \sigma^{n-2} e^{-\frac{r^2}{\sigma^2}-\frac{\sigma}{k}}.$$

This form is very intractable as a working integral, and as $\sigma^{2n}e^{-\frac{\sigma^2}{k^2}}$ gives

DIAGRAM I

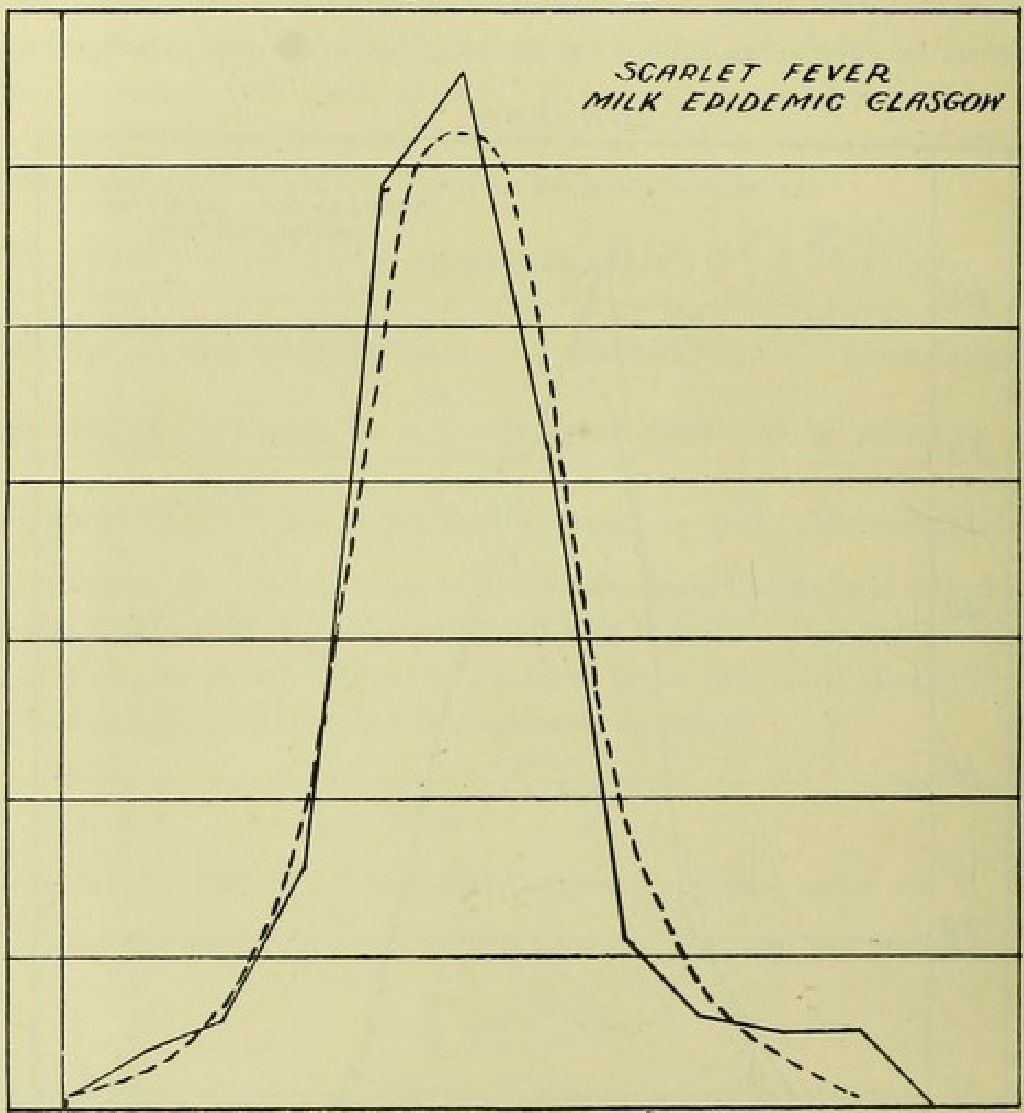


a somewhat similar series of curves, the latter may be substituted with much simplification of the calculation, so that the distribution of a species might be represented by

$$c\int_0^\infty \sigma^{2n+2} e^{-\frac{x^2}{\sigma^2}-\frac{\sigma^2}{k^2}} d\sigma,$$

a form easily reducible for any value of n. This is the same surface as was deduced as a possible one for the epidemic time wave. What has preceded is to a certain extent hypothetical in so far as the assumption of a form for the frequency of σ is connected; otherwise it is directly deduced in the case of the epidemic from the consideration of a large number of

DIAGRAM II.



typical epidemics, and in the second place by Professor Pearson by a rigid mathematical result from the assumptions already referred to.

The next assumption is the rate at which infective organisms are evolved. This may be taken as represented by $\phi(x)$, so that at any point x the ordinate will be represented by the integral of the curve taken between $x-c$ and $x+c$. We then, finally, obtain as the form of the epidemic

$$y = y_0 \int_{x-c}^{x+c} \phi(x) \int_0^\infty \sigma^{2n} e^{-\frac{x^2}{\sigma^2} - \frac{\sigma^2}{k^2}}.$$

It will be shown later that good agreement of theory with fact is obtained if $n=0$ and $\phi(x)=c$ (a constant), so that, in the first instance, the form

$$y=a\int_{x-c}^{x+c}dx\int_0^\infty e^{-\frac{x^2}{\sigma^2}-\frac{\sigma^2}{k^2}}d\sigma$$

will be considered.

As is well known

$$\int_0^\infty e^{-\frac{x^2}{\sigma^2}-\frac{\sigma^2}{k^2}}d\sigma=\sqrt{\pi}ke^{\pm\frac{2x}{k}},$$

the positive sign referring to the negative branch of x and the negative sign to the positive branch.

Hence the equation of the curve is

$$y=ak-ake^{-\frac{2c}{k}}\cosh\frac{2x}{k}\text{ from }x=0\text{ to }x=c,$$

$$y=ake^{-\frac{2x}{k}}\sinh\frac{2c}{k}\text{ from }x=c\text{ to }x=\infty.$$

The curve is symmetrical. Examples of this are given in Diagrams I. and II. When these are compared with Diagrams VII. and XXIII. respectively of my former paper,* it is seen how much better the fit now obtained is.

Mathematical Formulæ of Curves which represent possible Epidemic Forms.

Equations and formulæ which might describe more or less approximately epidemic or migration forms will now be considered in detail. The following symbols are used throughout:—

(1) Let $y=f(x)$ be the curve of distribution

$$\text{A}=\int_{-\infty}^{\infty}ydx,\qquad\text{area of the curve}$$

$$\mu_2=\frac{1}{\text{A}}\int_{-\infty}^{\infty}x^2ydx,\qquad\text{where the axis of } y \text{ is taken through the centre of gravity}$$

$$\mu_3=\frac{1}{\text{A}}\int_{-\infty}^{\infty}x^3ydx,\text{ etc.};$$

and

$$\nu_1=\frac{1}{\text{A}}\int_0^{\infty}xydx.$$

When the origin is not in a line through the centre of gravity the moments are denoted by

$$\mu'_1,\ \mu'_2,\ \mu'_3,\text{ etc.}$$

* *Proc. Roy. Soc. Edin.*, vol. xxvi. pp. 491 and 507.

To complete the symbols

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3},\ \beta_2 = \frac{\mu_4}{\mu_2^2}$$

The fundamental forms which have been chosen for investigation are

$$y = a\int_0^\infty \sigma^{2n} e^{-\frac{x^2}{\sigma^2} - \frac{\sigma^2}{k^2}} d\sigma \quad \text{and} \quad y = a\int_0^\infty \sigma^n e^{-\frac{x^2}{\sigma^2} - \sigma\gamma}.$$

These are symmetrical round the origin. On the hypothesis before mentioned they are integrated for each value of x from $x-c$ to $x+c$, multiplying each term by a suitable function $\phi(x)$, so that for the working equations we have,

$$y = a\int_{x-c}^{x+c} \phi x dx \int_0^\infty \sigma^{2n} e^{-\frac{x^2}{\sigma^2} - \frac{\sigma^2}{k^2}} d\sigma \qquad . \quad . \quad . \quad . \quad . \text{ (M)}$$

$$y = a\int_{x-c}^{x+c} \phi x dx \int_0^\infty \sigma^n e^{-\frac{x^2}{\sigma^2} - y\sigma} \qquad . \quad . \quad . \quad . \quad . \text{ (N)}$$

(2) The first of these forms is much easier to evaluate, and in addition gives the closest representation of the facts. Considering it in the first instance and assuming that $\phi(x) = a$ (a constant), we have the curve of distribution in time or space given by

$$y = a\int_{x-c}^{x+c} dx \int_0^\infty \sigma^{2n} e^{-\frac{x^2}{\sigma^2} - \frac{\sigma^2}{k^2}} d\sigma.$$

This form implies either that the rate at which infective organisms are given off is constant or that the distribution of organisms before migration begins is uniform.

In the application required n is in general zero. The areas and moments of the first three curves, however, are given in the following table:—

	$n=0.$	$n=1.$	$n=2.$
A	$ck^2\sqrt{\pi}$	$ck^4\sqrt{\pi}$	$2ck^6\sqrt{\pi}$
ν_1	$\frac{k}{2}$	$\frac{3k}{4}$	$\frac{15k}{16}$
μ_2	$\frac{k^2}{2} + \frac{c^2}{3}$	$k^2 + \frac{c^2}{3}$	$\frac{3}{2}k^2 + \frac{c^2}{3}$
μ_4	$\frac{3}{2}k^4 + c^2k^2 + \frac{c^4}{5}$	$\frac{9}{2}k^4 + 2c^2k^2 + \frac{c^4}{5}$	$9k^4 + 6c^2k^2 + \frac{c^4}{5}$

From the value of $\frac{\mu_4}{\mu_2^2}$ obtained from the statistics $\frac{k^2}{c^2}$ can be easily calculated, and thence k. The latter value may be compared with value of k deduced from the value of ν_1, and the nearest value of n ascertained. The equation of the curve corresponding to $n=0$ is as follows:—

$$y = ak - ake^{\frac{-2c}{k}} \cosh\frac{2x}{k} \text{ from } x=0 \text{ to } x=c$$

$$y = ake^{\frac{-2x}{k}} \sinh\frac{2c}{k} \text{ from } x=c \text{ to } x=\infty.$$

Those for higher values of n may be obtained by differentiating the above values with respect to $\frac{1}{k^2}$.

(3) To find k and c from the moments is easy, but for convenience a table is given by which the values of $\frac{k}{c}$ may be obtained when the value of $\frac{1}{6}\beta_2$, *i.e.* $\frac{1}{6}\frac{\mu_4}{\mu_2^2}$, has been calculated.

Table showing the values of $\frac{k}{c}$ for different values of

$$\frac{\frac{k^4}{c^4}+\frac{2}{3}\frac{k^2}{c^2}+\frac{2}{15}}{\frac{k^4}{c^4}+\frac{4}{3}\frac{k^2}{c^2}+\frac{4}{9}}=\frac{1}{6}\frac{\mu_4}{\mu_2^2}=\frac{1}{6}\beta_2 .$$

$\frac{k}{c}$.	$\frac{1}{6}\beta_2$.	Differences.
·50	·4314	
		227
·55	·4541	
		230
·60	·4771	
		232
·65	·5003	
		230
·70	·5233	
		226
·75	·5459	
		222
·80	·5679	
		213
·85	·5892	
		205
·90	·6097	
		196
·95	·6293	
		185
1·00	·6478	
		346
1·10	·6826	
		310
1·20	·7136	
		275
1·30	·7411	
		244
1·40	·7655	
		216
1·50	·7871	

(4) A special case arises in random migration if the area to be invaded is bounded by a straight boundary on the one side of which there is an infinite field uniformly stocked with the organism which is migrating. In the case of animals actually migrating this form has been found by Pro-

fessor Pearson, but it is much simpler in the case of plant life. We have to deal with the integral

$$y = a\int_0^\infty e^{\frac{-2(x+\zeta)}{k}}\,d\zeta$$

$$y = \frac{ka}{2}e^{\frac{-2x}{k}}.$$

Application of this will be given.

(5) Secondly, the form

$$y = a\int_{x-c}^{x+c}dx\int_0^\infty \sigma^n e^{-\frac{x^2}{\sigma^2}-\gamma\sigma}\,d\sigma$$

may be considered. In this the distribution of frequency of σ is taken as $\sigma^{n+2}e^{-\gamma\sigma}$. The moments are easily calculated, and are given by

$$A = 2c\sqrt{\pi}\,\frac{\Gamma(n+2)}{\gamma^{n+2}}$$

$$\nu_1 = \frac{1}{2\sqrt{\pi}}\,\frac{n+2}{\gamma}$$

$$\mu_2 = \frac{c^2}{3} + \frac{(n+3)(n+2)}{2\gamma^2}$$

$$\mu_4 = \frac{c^4}{5} + \frac{c^2(n+3)(n+2)}{\gamma^2} + \frac{3(n+5)(n+4)(n+3)(n+2)}{\gamma^4}.$$

These equations are solved most easily if c and n be eliminated. This gives, if $2\sqrt{\pi}\nu_1$ is denoted by ζ,

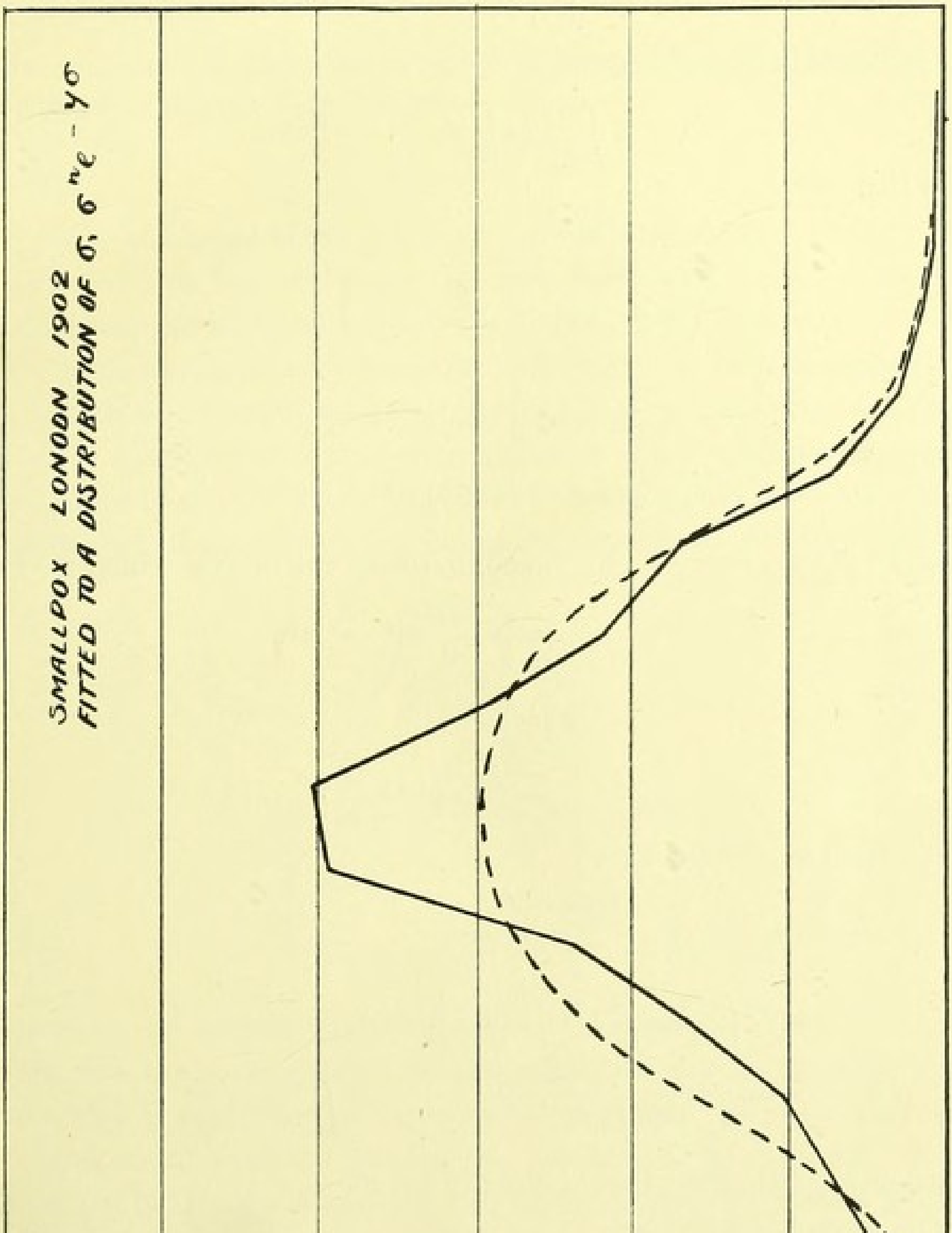
$$\frac{9}{2}\zeta\frac{1}{\gamma^3} + \frac{36}{5}\zeta^2\frac{1}{\gamma^2} + \left(\frac{12}{5}\zeta^3 + \frac{6}{5}\mu_2\zeta\right)\frac{1}{\gamma} + \left(\frac{9}{5}\mu_2{}^2 + \frac{6}{5}\mu_2\zeta^2 - \frac{3}{10}\zeta^4 - \mu_4\right) = 0.$$

The rest of the solution is easy. I have not been able to arrive at the curve without mechanical quadrature. In the instance in which it has been fitted (Diagram III.) it proves exceedingly unsuitable, and it is therefore evident that the form discussed cannot represent the distribution of σ even approximately.

Asymmetry.

(6) Asymmetry in the epidemic curve deserves some notice. As we found before, $\frac{\log p}{\log q}$ must in general be equal to a constant, as the epidemic is symmetrical. If, however, $\frac{\log p}{\log q}$ is a function of σ it may, on account of the near symmetry of the epidemic, be assumed that it can be expanded in terms of σ of a rapidly convergent nature so that all the terms in the expansion except that of σ may be neglected.

DIAGRAM III.



The expression

$$a\int_{x-c}^{x+c}dx\int_0^\infty e^{-\frac{x^2}{\sigma^2}-\frac{\sigma^2}{k^2}}d\sigma$$

becomes on this hypothesis

$$a\int_{x-c}^{x+c}dx\int_0^\infty e^{\frac{-(x-a\sigma)^2}{\sigma^2}-\frac{\sigma^2}{k^2}}d\sigma.$$

In this case

$$\begin{aligned}
\mathrm{A} &= c\sqrt{\pi}k^2 \\
\mu'_1 &= \sqrt{\pi}ak \\
\mu'_2 &= \frac{c^2}{3}+\frac{k^2}{2}+a^2k^2 \\
\mu'_3 &= \frac{9}{4}\sqrt{\pi}ak+\frac{3}{2}\sqrt{\pi}a^3k^3+a\sqrt{\pi}c^2k \\
\mu'_4 &= \frac{k^4}{2}(4a^4+12a^2+3)+k^2c^2(1+2a^2)+\frac{c^4}{5}
\end{aligned}$$

when μ'_1, μ'_2, μ'_3, etc., are the moments round the origin, which gives

$$\mu_2=\frac{c_2}{3}+\frac{k^2}{2} \quad \text{if} \quad a^2 \doteqdot 0$$

$$\mu_3=\frac{9}{4}\sqrt{\pi}ak^3$$

$$\mu_4=\frac{3k^4}{2}+k^2c^2+\frac{c^4}{5}.$$

For the form of the curve

$$\begin{aligned}
y &= a\int_{x-c}^{x+c}dx\int_0^\infty e^{-\frac{(x-a\sigma)^2}{\sigma^2}-\frac{\sigma^2}{k^2}}d\sigma \\
&= ae^{-a^2}\int_{x-c}^{x+c}dx\int_0^\infty e^{-\frac{x^2}{\sigma^2}-\frac{\sigma^2}{k^2}}\left(1-\frac{a}{\sigma}\right)d\sigma \\
&= ae^{-a^2}\int_{x-c}^{x+c}dx\int_0^\infty e^{-\frac{x^2}{\sigma^2}-\frac{\sigma^2}{k^2}}d\sigma - aae^{-a^2}\int_{x-c}^{x+c}dx\int\frac{1}{\sigma}e^{-\frac{x^2}{\sigma^2}-\frac{\sigma^2}{k^2}}d\sigma
\end{aligned}$$

the solution of the first part is already given.

The second is equal to

$$-aae^{-a^2}\frac{2\pi}{2}\int_{x-c}^{x+c}\mathrm{H}_0^{(1)}(ix)dx.*$$

This method for accounting for symmetry has likewise not been successful in representing the statistics.

(7) When the distribution is symmetrical around a centre it is evident that we have to deal with a similar integration of Professor Pearson's form and that a distribution of $y=ae^{-\frac{r}{k}}$ results.

* A table of this function is given in Jahnke und Emde, *Funktionentafeln*, p. 135.

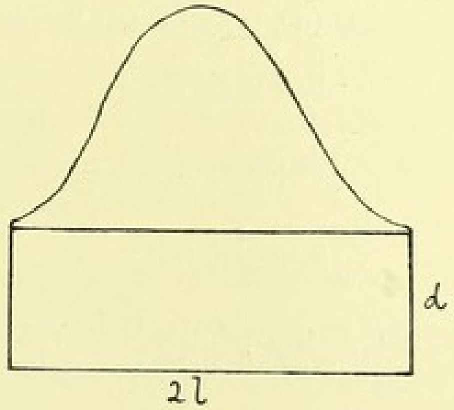
This can be fitted by radial moments, or, in general, more easily by first reducing the statistics so that the numbers per unit zone is taken as a basis of calculation.

The distribution may also be summed parallel to one axis and compared with a table of the integral

$$a\int_0^\infty e^{-\frac{\sqrt{x^2+y^2}}{k}}dx.$$

In some cases, especially when an epidemic occurs in a locality where a disease is more or less uniformly endemic, it is useful to have the means of separating the epidemic portion of the disease from that which is endemic. Such a combination is very common in the case of such diseases as scarlet fever, enteric fever, diarrhœa, etc. Successive approximations are necessary to obtain a solution, but a first approximation can be obtained by using the normal curve to represent the statistics of the course of the epidemic.

Let the whole amount of the disease be as represented in the diagram



The endemic prevalence of the disease is represented by the rectangular base and the epidemic portion by the curve above. In general the parts of the curve beyond the limits may be neglected as only affecting the result to a small extent. Taking the middle of the rectangle as origin, denoting its height by d and its length $2l$, where l is known and d unknown, we have for the area and for the moments of the rectangle round an axis through the middle,

$$A = 2ld$$

$$A\mu_2 = \frac{2l^3d}{3}$$

$$A\mu_4 = \frac{2l^5d}{5}.$$

For the normal curve the equation is

$$y = y_0 e^{\frac{-(x-c)^2}{\sigma^2}},$$

and its area and moments become

$$A' = y_0 \sqrt{\pi}\sigma$$
$$A\mu'_1 = y_0 \sqrt{\pi}\sigma c$$
$$A\mu'_2 = y_0 \sqrt{\pi}\frac{\sigma^3}{2} + y_0 \sqrt{\pi}\sigma c^2$$
$$A\mu'_3 = y_0 \sqrt{\pi}\frac{\sigma^3 c}{2} + y_0 \sqrt{\pi}\frac{\sigma c^3}{3}.$$

Adding the corresponding portions of the moments, we have the following equations for determining the four unknowns, d, σ, c, and y_0,

$$A + A_1 = 2ld + y \;= B_1 \qquad \ldots \qquad (a)$$
$$A\mu'_1 = y_0 \sqrt{\pi}\sigma c = B_2 \qquad \ldots \qquad (b)$$
$$A'\mu'_2 + A\mu_2 = \frac{2l^3 d}{3} + y_0 \sqrt{\pi}\frac{\sigma^3}{2} + y_0 \sqrt{\pi}\sigma c^2 = B_3 \qquad \ldots \qquad (c)$$
$$A\mu'_3 = y_0 \sqrt{\pi}\frac{\sigma^3 c}{2} + y_0 \sqrt{\pi}\frac{\sigma c^3}{3} = B_4 \qquad \ldots \qquad (d)$$

where B_1, B_2, etc., denote the area and moments obtained from the statistics.

By using (b) the equations (a), (c), and (d) become

$$\frac{B_2}{c} + 2ld = B_1 \qquad \ldots \qquad (\alpha)$$
$$\frac{\sigma^2 B_2}{2c} + B_2 c + \frac{2ld^3}{3} = B_3 \qquad \ldots \qquad (\beta)$$
$$\frac{\sigma^2}{2}B_2 + \frac{c^2}{3}B_2 = B_4 \qquad \ldots \qquad (\gamma)$$

From α and β

$$\frac{\sigma^2}{2c}B_2 - \frac{B_2 l^2}{3c} + B_2 c = B_3 - \frac{l^2}{3}B_1,$$

or

$$2B_2 c^2 - (3B_3 - l^2 B_1)c + 3B_4 - B_2 l^2 = 0,$$

which gives c.

Hence immediately we have

$$\sigma^2 = 2\frac{B_4}{B_2} - \frac{2}{3}c^2$$
$$y_0 = \frac{B_2}{\sqrt{\pi}\sigma c}$$
$$2ld = B_1 - \frac{B_2}{c},$$

which complete the solution.

When the theory just described is applied to the elucidation of epidemic processes it is found to be fairly satisfactory. It gives a very good fit to many epidemics, as will be shown. It is also capable of expressing the moment relationships of epidemics satisfactorily. Thus it is generally found that $\mu_4 > 3\mu_2$: this is explained, but the form also allows of $\mu_4 < 3\mu_2$ without a change of theory.

Where this theory fails, however, is in its application to seasonable epidemics such as enteric fever when averaged over a number of years. This might be expected. Though the hypothesis that the infectious organism is given out for a definite time at a constant rate may be sufficiently accurate when applied to a single epidemic, it can hardly be expected to hold when an average of years is considered. An early or a late epidemic of the disease is the product of special weather conditions which will occur very seldom. So that though the epidemics occurring at the usual time will tend to have their irregularities smoothed, the beginning and the end of the compound curve will be framed on a different law from the middle. An endeavour has been made to frame an approximate formula to meet this, but none has been found fitting the facts.

To test the likelihood of a curve fitting, Professor Pearson's method is used throughout. To do this the actual and theoretical numbers are differenced, the difference squared and divided by the corresponding theoretical number. These are summed. The sum is denoted by χ^2. The table for testing curve fitting * is then consulted, and the value of P obtained.

Thus if $P = \cdot 8$ it signifies that in testing the matter over again a worse fit might be expected eight times out of ten, so the fit is good. The opposite interpretation is given when P is small, say $= \cdot 2$, when a worse fit might be expected only in two out of ten trials.

Plague in Hong-Kong.

The epidemics of plague in Hong-Kong are typical of the great majority in all the East. They are nearly bilaterally symmetrical; they spring from

Table showing the Number of Plague Cases, Actual and Theoretical, in the Epidemics of Plague in Hong-Kong in 1902 and 1904.

	1902. Number of Cases.		1904. Number of Cases.		
	Actual.	Theoretical.	Actual.	Actual Figures corrected so as to correspond with the Mean.	Theoretical.
March .	2	3·8	4	3·6	7·0
April .	27	38·5	40	34·4	35
May .	157	147·0	135	122·0	126·4
June .	194	178·7	194	186·0	180·5
July .	131	147·0	96	109·2	126·4
August .	50	38·5	19	29·3	35·1
September .	2	3·8	9	11·3	7·0

* *Biometrika*, vol. i. p. 155.

a zero line where the disease is nearly completely absent. The only point of difficulty in treating them mathematically is in settling how far the calculation of the moments is to be carried into the interepidemic period. In general good fits are obtained if the interepidemic cases are neglected.

Two epidemics are shown in the figures given in the table.

The fit of the first of these is not good as it stands; but when it is noted that the symmetry with regard to the first moment is produced from an unequal distribution with respect to the mean of a kind that the sums of the equidistant terms on both sides of the mean are nearly equal, we may put the distribution as follows:—

	Actual.	Theoretical.
2+2	4	7·6
27+50	77	77·0
157+131	288	294·0
194	194	178·7

This gives $\chi^2 = 3{\cdot}83$ or $P = {\cdot}28$.

Thus the variations in the rise cancel those of the fall. This method is subject to criticism, but the fit of any individual epidemic can hardly be expected to be good. With regard to the epidemic of 1904 the fit may be said to be good. Here $\chi^2 = 5{\cdot}96$ or $P = {\cdot}42$.

It is not necessary to give a large number of examples of the way in which the present theory fits the facts. In many instances better fits are obtained than those shown in my previous paper, where type iv. was found to closely represent the epidemic form. Two examples, however, may be given, that for the smallpox deaths in Warrington, and that for the milk epidemic of scarlet fever in Glasgow. These are shown in Diagrams I. and II. In both the correspondence of the facts with theory is very close, much closer than with type iv. Before leaving this part of the subject an example of the use of the distribution of $\sigma^n e^{-\gamma\sigma}$ for σ may be given. This is the only example I have thoroughly worked out, but in a number of others I have obtained the medium value. In none is there any evidence that the curve obtained has any resemblance to the facts. It is therefore very improbable that $\sigma^n e^{-\gamma\sigma}$ can represent the variation of σ. The example illustrated is that of smallpox in London in 1902. As will be seen by referring to my previous paper, type iv. gives a quite different representation from the curve shown in this case (Diagram III.).

Examples of the application to random migration will now be given; those with animal forms will be considered, and those with plant forms thereafter. With the former it is difficult to secure suitable examples. *Daphnia pulex* was used in many experiments. This crustacean does not move so consistently in one direction as others. It moves by jerks, and, moving more or less vertically, fulfils more nearly than any other Professor Pearson's criterion that motion in any one direction is as likely as in any other. It has also the advantages of being not specially attracted by light and of being more or less opaque, so that it is easily photographed. *Cyclops*, *Littorina rudis*, etc., were also used as is described. Of plants only two species were investigated, an Oscillatorian and *Aspargia hispida*.

Daphnia pulex.

Experiments were made in two ways with *Daphnia pulex*. In some experiments a large number of the crustaceans were placed inside a cylindrical tube in a large flat dish of white porcelain and then liberated; in others the water flea was allowed to take up its position as it liked in the dish. It usually chose to distribute itself from one corner along one side of the dish. Examples of the manner in which this happened are given below. The corner being an impenetrable boundary may be taken as representing a centre of diffusion, so that the simple fundamental integral should apply and the grouping should conform to the exponential. This is what takes place (Diagram IV.).

Table showing the Number of Daphnia per Unit of Length along the Margin of the Plate from one Corner.

Unit of Length.	Actual (*a*).	Theoretical.	Actual (*b*).	Theoretical.
0 – 1	18	18·1	23	20·7
1 – 2	11	10·1	11	13·6
2 – 3	3	5·6	7	9·1
3 – 4	4	3·9	6	6·0
4 – 5	2	2·5	5	3·9
5 – 6	1	1·5	1	2·6
6 – 7	...	1·1	4	1·7
7 – 8	1	·7	2	1·1
8 – 9	...	...	1	·75

The curve is given,

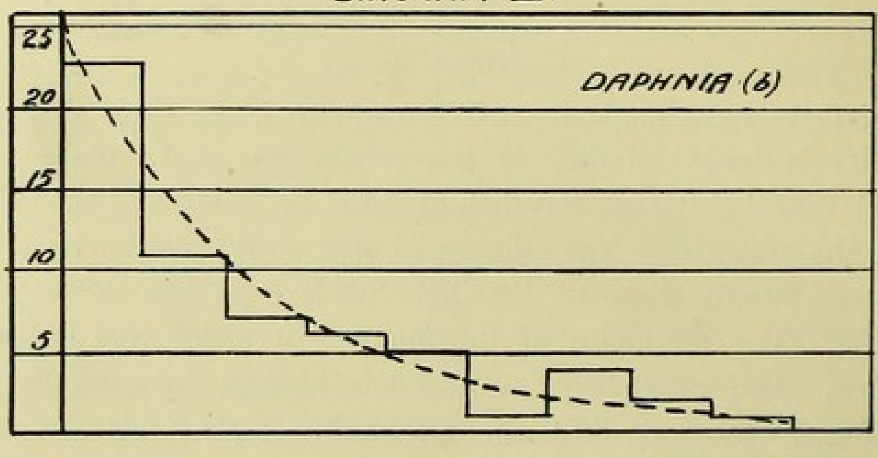
$$\text{for} \quad (a) \text{ by } y = 18 \cdot e^{-\frac{x}{2\cdot 225}}$$

$$\text{for} \quad (b) \text{ by } y = 25 \cdot e^{-\frac{x}{2\cdot 4}}.$$

In the first example the fit is excellent ($\chi^2 = 2{\cdot}8$ P $= {\cdot}9$), in the second not so good ($\chi^2 = 6{\cdot}34$ P $= {\cdot}6$). The want of fit, however, is largely due to the group of four near the tail end of the distribution, which contributes half of the divergence. Except for this group a further divergence might be expected nine times in each ten trials.

In several instances the Daphniæ, instead of grouping themselves along one edge of the dish, arranged themselves as if one corner of the dish were the centre of attraction. From this point outwards in all directions of the quadrant their numbers diminished. This, however, was observed only once, when the light was good enough for instantaneous photography. To

DIAGRAM IV.



count the numbers of the organism, concentric circles were drawn on the plate and the organisms between each pair of concentric circles in the quadrant counted. Some difficulty was experienced, however, at the greater distances, as in places it seemed that the outlyers of other groupings were even invading the periphery of this. All the organisms were counted, however, and none rejected. The figures are given in the following table :—

	i. Total Numbers.	ii. Number per Unit Area. Actual.	iii. Number per Unit Area. Calculated.	iv. Calculated Total.
Centre quadrant . .	19	19	18·4	18·4
1st ,, zone .	33	11	11·1	33·3
2nd ,, ,, .	37	7·4	6·78	33·9
3rd ,, ,, .	18	2·67	4·16	29·12
4th ,, ,, .	24	2·67	2·55	22·95
5th ,, ,, .	19	1·72	1·57	16·27
6th ,, ,, .	17	1·31	·96	12·5
7th ,, ,, .	11	·73	·59	8·9
	178	46·50	46·11	177·14

It is easier to examine this grouping when the numbers are reduced to the population per unit area. On the hypothesis the grouping should be given by $y = ae^{-\frac{r}{k}}$ (par. 7) and the section of this by the plane of $y=0$ gives $y = ae^{\pm\frac{x}{k}}$. The numbers reduced to the population per unit area are given in column ii. of the table. If we take the area and first moment and integrate from $x=0$ to $x=8$, the limit to which a count was made, we get $y = 23{\cdot}04e^{-\frac{x}{2{\cdot}056}}$. This gives $\chi^2 = 6{\cdot}13$ or P = ·53, a fit as good as could be expected, when the uncertainty regarding the numbers in the outer zone is remembered. A large part of the value of χ^2 is, as before noted, due to one zone alone. In all these groupings there seem to be secondary centres which interfere with results when the numbers are small.

The methods of dispersal from a centre were also investigated. For this many Daphniæ (from 100 to 200) were placed in the centre of the dish with a depth of water of about $\frac{3}{16}$ of an inch. They were contained in a cylindrical tube about $\frac{1}{2}$ inch in diameter. When the level of the water was the same on both sides of the tube and when the light was good and the camera ready, the tube was removed, the dispersal watched, and at a suitable moment instantaneously photographed. The photographs of course show no detail, the organisms being simply marked by a paler spot on the negative. In all cases a few Daphniæ were found greatly more energetic than the rest. These were generally above the mean size and probably represented an older generation. They were so exceptional that they possibly should be rejected from the statistics; all, however, have been included.

The experiments took much time. It was very difficult to manipulate the organism without damaging a number and thus introducing a new factor.

In the case of the negative from which the following table is made the centre of the group was found by counting the number of organisms in each half-inch square of the plate and calculating the mean. This being found, circles were drawn round this of diameter $\frac{1}{2}$ inch, 1 inch, $1\frac{1}{2}$ inches, etc., and the organisms in each zone counted.

The numbers in each zone are given in the table, column i.

The curve is given by $y = 39{\cdot}2e^{-\frac{x}{1{\cdot}309}}$, hence $\chi^2 = 9{\cdot}7$, which gives P = ·4.

In this case, again, one zone gives a large part of the value of χ^2, and it is also to be noted that again it is the third zone. Another fact of interest in all the experiments of this class is that the centre of the migration is not the original centre of dispersal; the whole mass has moved towards the light.

	i.	ii.	iii.	iv.
	Total Numbers.	Number per Unit Area.		Total Numbers calculated.
		Actual.	Calculated.	
Centre circle . . .	28	28	27·8	27·8
1st zone . . .	42	14	12·5	37·5
2nd " . . .	28	5·6	5·8	29·0
3rd " . . .	11	1·57	2·71	19·0
4th " . . .	7	·78	1·26	11·3
5th " . . .	7	·63	·58	6·3
6th " . . .	5	·40	·274	3·6
7th " . . .	4	·26	·128	1·9
8th " . . .	2	·12	·06	1·0
9th " . . .	1	·05	·028	·53

The dispersion of Daphnia may also be compared with the formula $y = ae^{-\frac{r}{k}}$ by summing parallel to one axis. A comparison is given below when the Daphniæ were so counted. The numbers are as follows:—

Actual.	Theoretical.
2	1·9
7	9·4
46	39·5
35	39·5
11	9·4
1	1·9

giving $\chi^2 = 2{\cdot}34$ or $P = {\cdot}75$.

Cyclops.

Cyclops were used in a few experiments but found very unsatisfactory. Not only did they photograph with great difficulty on account of their transparency, but they moved on dispersal strongly towards the light.

A sectional count of one instance is given (Diagram V.). It is not treated as a whole, but it is to be noticed that the advancing edge is very closely given by the exponential curve which has been fitted to it.

The values are:—

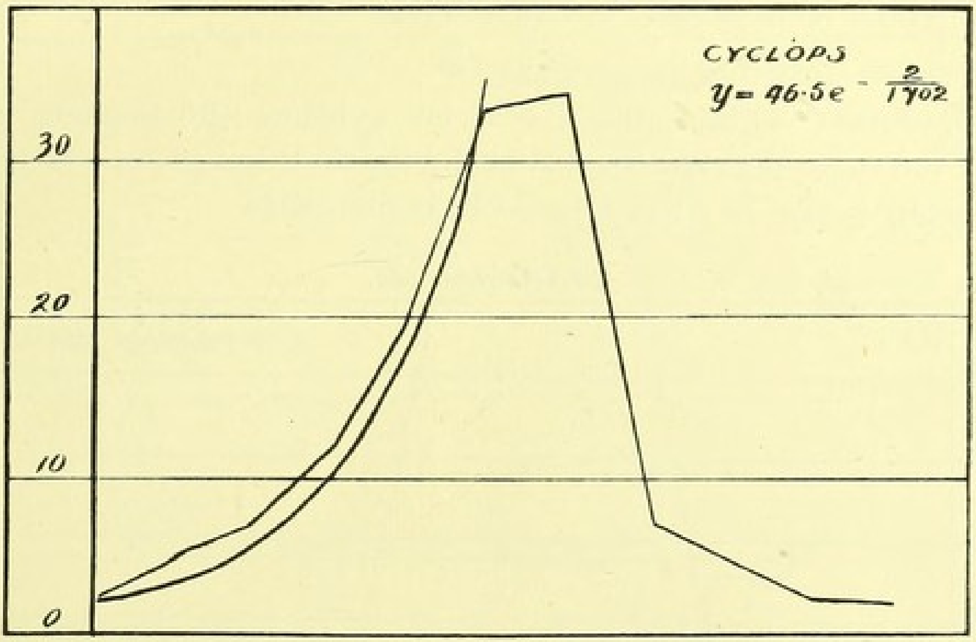
Actual.	Calculated.
33	35·3
20	19·2
12	10·7
7	5·9
5	3·3
2	1·8

which gives $\chi^2 = 1{\cdot}61$ or $P = {\cdot}9$, so that the fit is very close.

Other Crustacea.

Some copepods were experimented with, but their transparency and rate of motion unfitted them for this purpose; they were also too strongly attracted to light. In addition an attempt was made with a fresh-water

DIAGRAM V



isopod, but a sufficiently large dish could not be obtained to make the experiment of any value.

Littorina rudis.

Several experiments were made with this mollusc. Numbers were placed in various vessels and their movements observed. It was found best to put them at the bottom of a deep glass jar with a little salt water

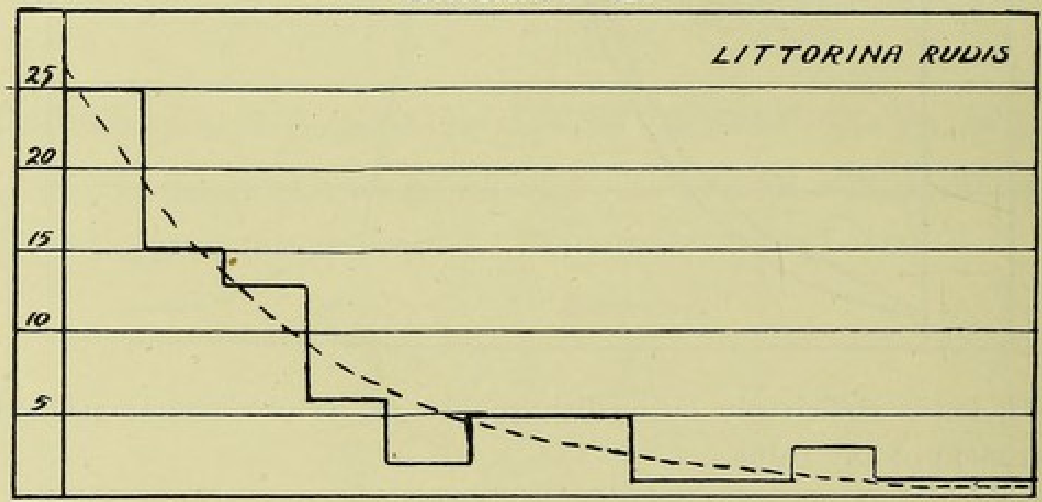
Distance in inches.	Numbers actual.	Numbers theoretical.
0 – ·5	25	23·2
·5 – 1	15	16·3
1 – 1·5	13	11·5
1·5 – 2·0	6	8·2
2·0 – 2·5	2	5·8
2·5 – 3·0	5	4·1
3·0 – 3·5	5	3·0
3·5 – 4·0	1	2·2
4·0 – 4·5	1	1·4
4·5 – 5·0	3	1·0
5·0 – 5·5	1	·7
5·5 – 6·0	1	·5

in the bottom and leave them to climb out; this they did, passing the edge and falling on the floor. Latterly a large number were placed in the jar one day and next day those on the side of the jar counted. It seems evident that we are dealing with the tail of a moving mass, and that this tail should obey the law of the exponential. In one experiment the number of Littorina in each half inch of the jar, measuring from the top, were as shown in table on page 283.

This gives a distribution $y = 2{\cdot}72e^{-\frac{x}{2{\cdot}572}}$.

As it stands $\chi^2 = 10{\cdot}2$, so that $P = {\cdot}53$, but, as before with Daphnia, 40 per cent. of the value of χ^2 is accounted for by the third last group. If this be subtracted $P = {\cdot}87$. In either case the fit is admissible.

DIAGRAM VI.



Other Marine Animals.

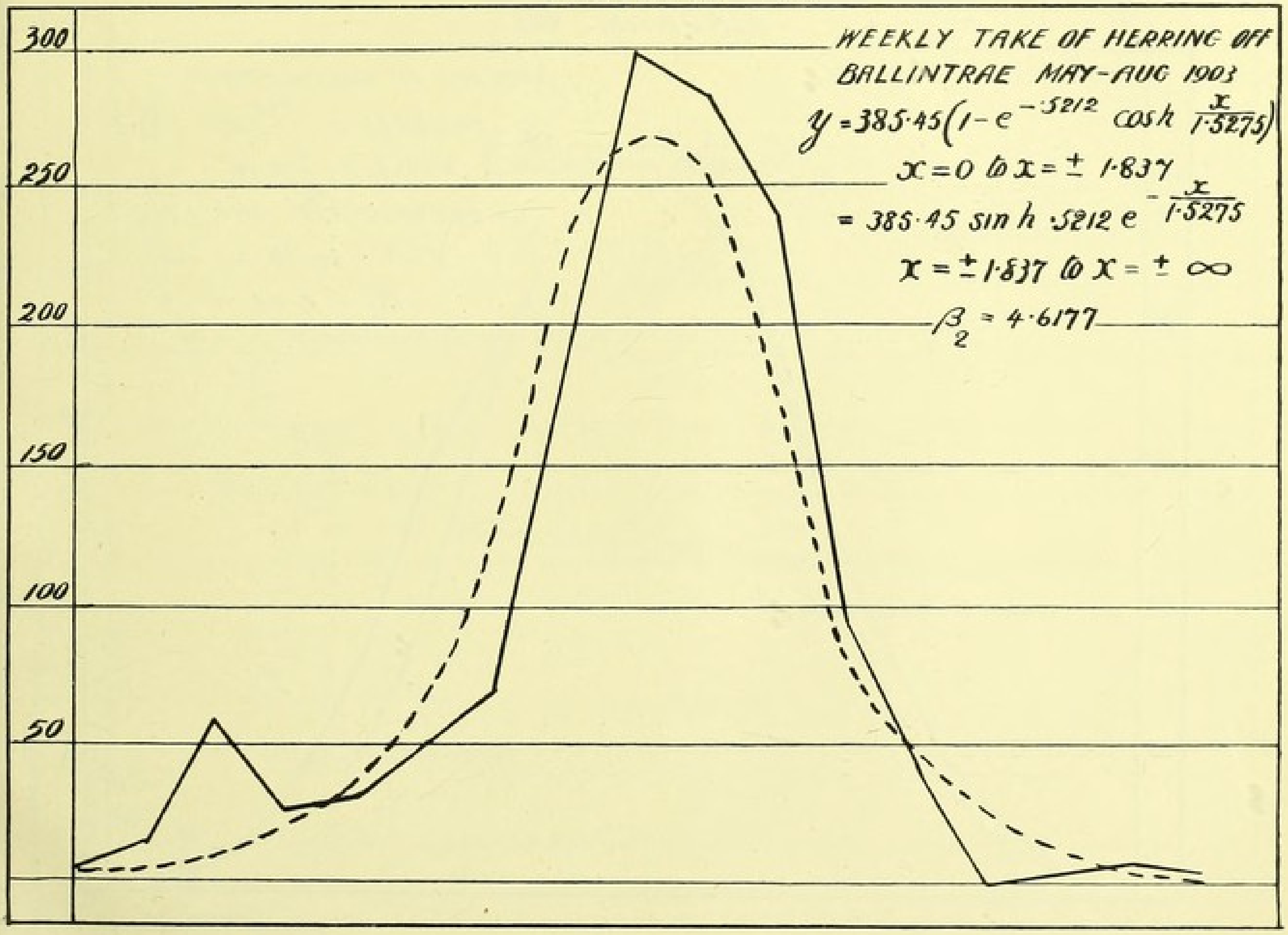
With regard to large animals few experiments can be recorded. Some observations were made on fish. In fish migration the form of the shoal should be capable of measurement. In this case the tanks in the Millport Biological Station seemed to afford some chance of success. One of the large tanks contained a small shoal of saithe (*Gadus virens*). These were photographed several times, but to get sufficient light for an instantaneous plate was difficult. In one such photograph the shoal is on the point of turning. The symmetry is remarkable, the numbers from the left to the right in each unit of length being as follows:—

1, 2, 2·5, 6, 6, 2·5, 1, 1. Total 22.

The numbers are too small to allow of differentiation of the type of curve, as they can be fitted either to the exponential curve described or to the normal curve of error.

Data for the other fish are lacking with the exception of herring. The weekly takes at given stations afford some guide, but not much. One selected from the Fishery Board Report of the take of herring off Ballantrae is illustrated in the diagram (Diagram VII.). Here the elevation at the beginning of the ascent prevents accurate curve fitting, but the general

DIAGRAM VII.



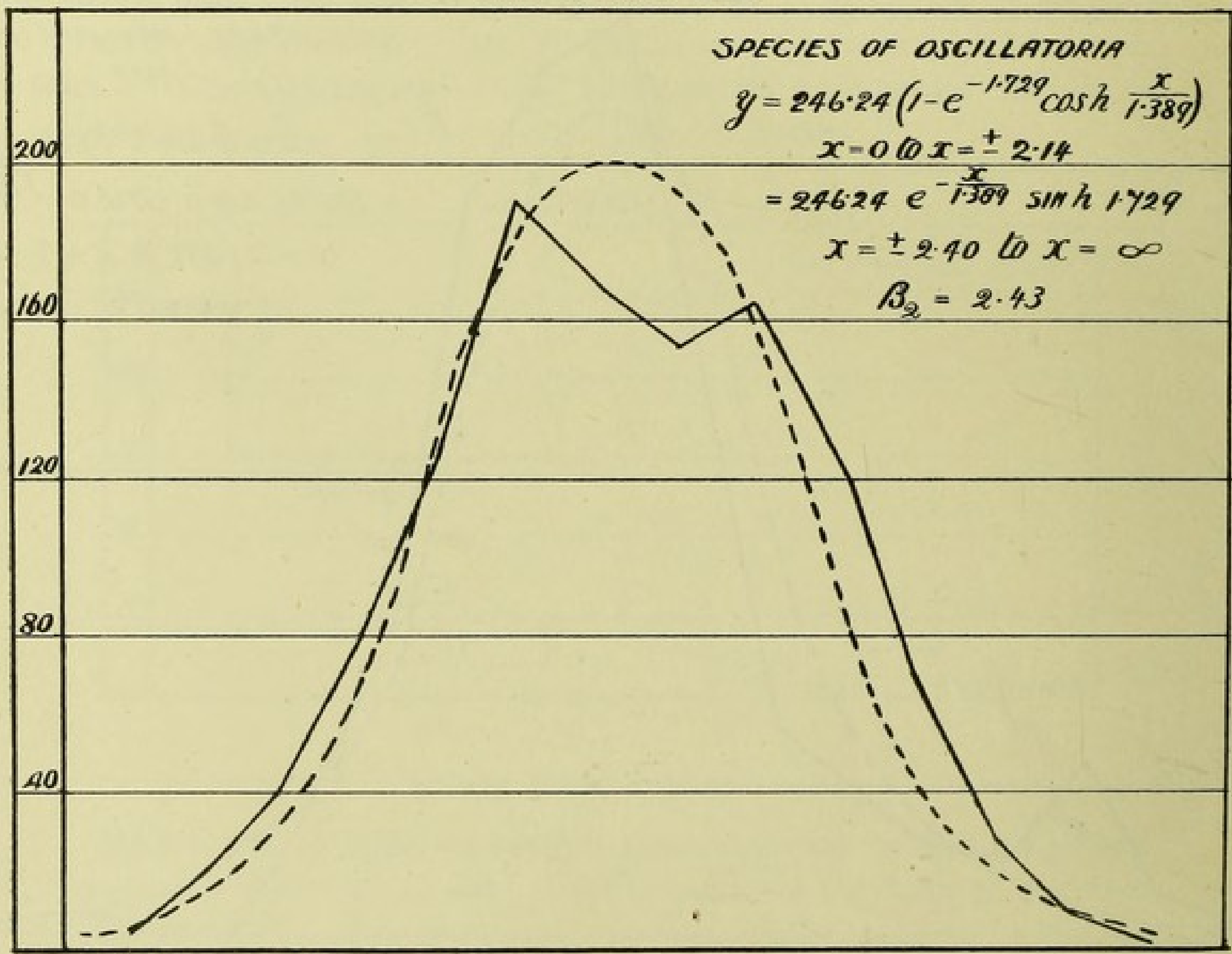
correspondence is good. It was hoped that with the figures at a fixed point some better results might be obtained, and the Fishery Board very kindly furnished me with the figures of the weekly takes in the trammel nets at Ballantrae. But these figures proved insufficient to determine the shoal form; in no case were the curves continuous. Factors such as the end of the open time, storms, etc., interfered with the returns to such an extent that it was not possible to use the figures.*

* Herring catch, Ballantrae, 1903, May–August, *Twenty-fifth Ann. Rep. Fish. Board Sc.*, p. 174.

Oscillatoria.

A species of Oscillatorian (not identified) was also experimented with. This, when put in a mass in the bottom of a test tube, started to climb the wall to the surface. After the migration had taken place so that the beginning and end of the mass was fairly clear, the test tube was carefully

DIAGRAM VIII.



emptied and filled with melted gelatine. When this was set the number of filaments per millimetre of length on the test tube was counted under the microscope. The numbers in each space of four millimetres were as follows :—

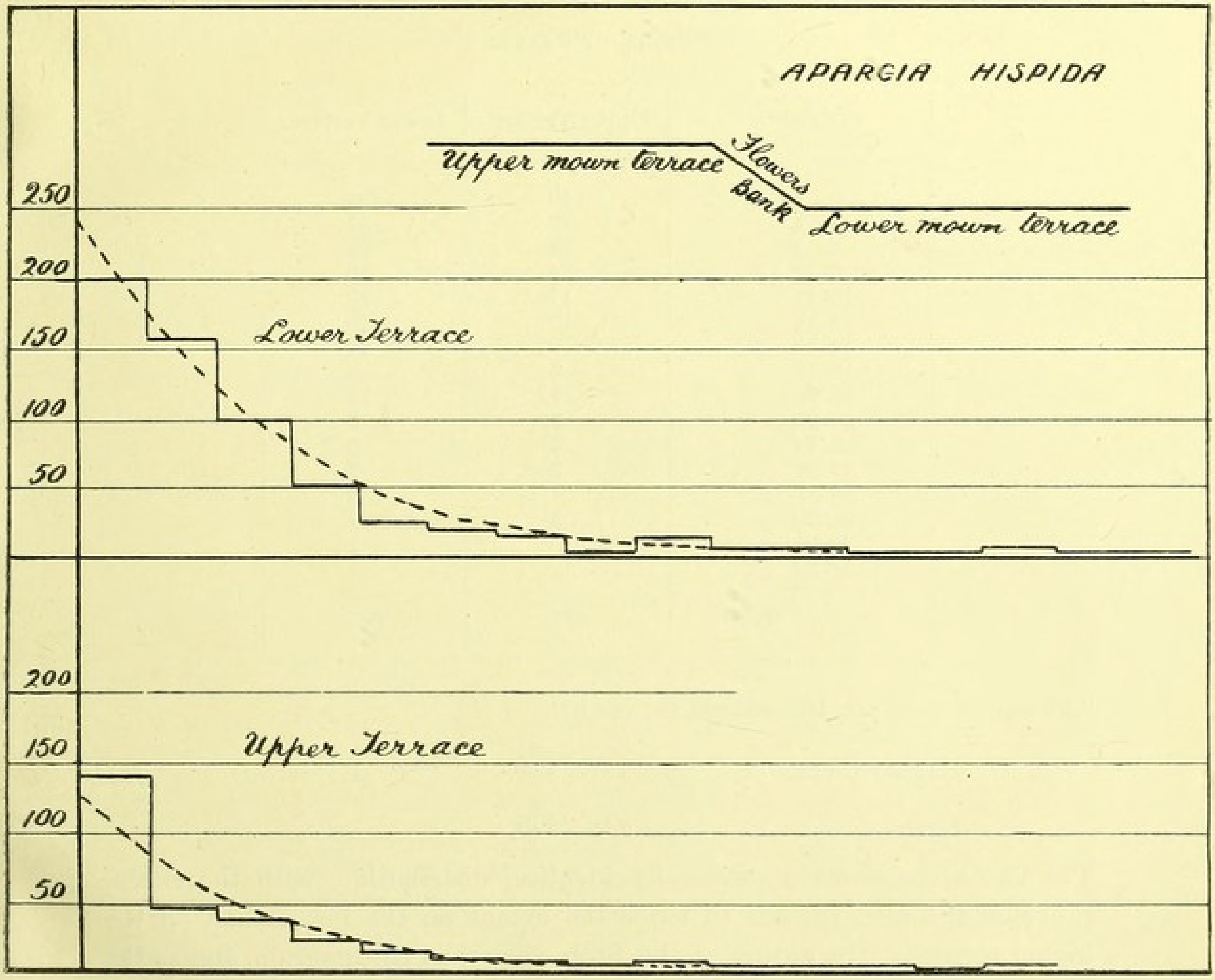
2, 20, 42, 80, 125, 190, 168, 154, 164, 128, 70, 28, 10, 2.

When fitted to a migration curve the correspondence is as illustrated in the accompanying diagram (Diagram VIII.). The fit is not good, but the shape of the curve suggests that it is made up of two and not one migration system.

Apargia hispida.

In the Belvidere Hospital there are long straight terraces adjacent to the Clyde constructed as in the diagram (Diagram IX.). It is the custom to mow the level parts of the terraces with a lawn mower.

DIAGRAM IX.



No plant, therefore, of any high habit can ever seed there. It can only grow herbaceously. On the bank between the terraces, however, such plants can develop, and a considerable stretch of the bank is thickly overgrown with Apargia. The seeds of this scatter over the lower and higher terraces, when they germinate and form plants in the grass.

This case can easily be considered as one where there is an infinite uniform source on one side of a straight line, and therefore may be represented by $y = ae^{-\frac{x}{k}}$.

To obtain the form of distribution the ground on both terraces was lined with parallel lengths of string two feet apart for a distance of fifteen yards, and the number of plants in each rectangular space counted. The numbers are given in this table:—

Number of Plants.

Distance.	Upper Terrace.	Lower Terrace.
0– 2 feet	140	201
2– 4 „	47	157
4– 8 „	39	99
8–10 „	24	50
10–12 „	17	22
12–14 „	10	20
14–16 „	9	16
16–18 „	5	4
18–20 „	7	12
20–22 „	3	6
22–24 „	4	3
24–26 „	3	3
26–28 „	1	3
28–30 „	3	4
30–32 „	...	1
32–34 „	...	1
Total	312	605

The equation of the theoretical curves are:—

Higher terrace $y = 127e^{-\frac{x}{2 \cdot 458}}$.

Lower „ $y = 248e^{-\frac{x}{2 \cdot 445}}$.

The two areas show a practically identical distribution, with the exception that the seed spreads to twice the extent on the lower than on the higher terrace. The nature of the fit is shown on the diagram, and as the soil of the locality is all forced, clay coming to the surface in patches, and drainage being very irregular, it is as good as might be expected.

Conclusions.

(1) The general principles which underlie both epidemic distribution in space and time and random migration are identical.

(2) Both can be deduced almost directly from the laws of chance through assumptions which have considerable *a priori* probability.

(3) It is found in both cases that the exponential curve can be taken as giving a fair representation of most of the facts. The manner in which this curve, however, appears is somewhat different in the two instances. It is doubtful why this formula should express both results.

NOTE.—The experimental work recorded in this paper was done chiefly at the laboratory of the Millport Marine Biological Station, and I desire to thank the Superintendent for much assistance.

(*Issued separately February* 25, 1911.)