Applications of Mathematics to Medical Problems.

By Lieut.-Col. A. G. M'KENDRICK.

(From the Laboratory of the Royal College of Physicians, Edinburgh).

(Read 15th January 1926.   Received 13th August 1926.)

In the majority of the processes with which one is concerned in the study of the medical sciences, one has to deal with assemblages of individuals, be they living or be they dead, which become affected according to some characteristic.   They may meet and exchange ideas, the meeting may result in the transference of some infectious disease, and so forth.   The life of each individual consists of a train of such incidents, one following the other.   From another point of view each member of the human community consists of an assemblage of cells.   These cells react and interact amongst each other, and each individual lives a life which may be again considered as a succession of events, one following the other. If one thinks of these individuals, be they human beings or be they cells, as moving in all sorts of dimensions, reversibly or irreversibly, continuously or discontinuously, by unit stages or *per saltum*, then the method of their movement becomes a study in kinetics, and can be approached by the methods ordinarily adopted in the study of such systems.

It is the object of this communication to approach this field in a systematic manner, to find solutions for some of the variations which may arise, and to illustrate certain of these by examples.

1. *One dimension, irreversible.*

I have been in the habit of employing vector diagrams for the representation of such problems.   They have the advantage that the hypotheses which are adopted are clearly visualized as well by the non-mathematical reader as by the mathematical, and they also aid in helping one to realise the various modifications which may occur, and so to treat the study of the general problem systematically.   To fix ideas let us consider a simple case ; the relation of an assemblage of individuals to common colds.   In the

following series of compartments are classified at any instant the numbers of individuals who have experienced, 0, 1, 2, 3 ... attacks of this complaint. The history of each individual consists of a series of unit steps, originating in the compartment which describes his initial condition. The arrows in the diagram indicate the chance of passage from one compartment to the next—that is to say the chance of experiencing a further attack during the infinitesimal period of time dt.
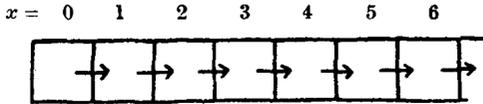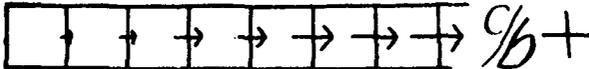

Fig. 1

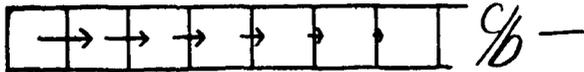Fig. 2

Fig. 3

In fig. 1 these arrows are of equal size, and by this we understand that the successive chances were of constant value ; in fig. 2 the arrows increase in size, denoting an increase of susceptibility with each attack ; in fig. 3 they decrease, which denotes that the individual is becoming decreasingly liable, or in medical parlance he is developing an immunity.

Guided by the diagram, and using the nomenclature $v_x =$ the number of individuals who have experienced x attacks (or shortly "of grade x") ; $f_{t,x} dt =$ the probability that an individual of grade x will pass to grade $x + 1$ in the time dt, and noting that the variation of the number in any grade is the difference between the number of incomers into that grade, and the number who go out from that grade, we have

$$dv_x = (f_{t,x-1} \ v_{x-1} - f_{t,x} \ v_x) \, dt. \quad \dots\dots\dots\dots\dots(1)$$

In this case and in what follows, for the sake of conciseness, the solutions will be given for instantaneous point sources; other initial conditions may be obtained by summation.

In the first place let us assume that $f_{t,x}$ is of the form $\phi_t f_x$, that is to say, that the time function applies generally to the probability of exit from all compartments. (The general case will be considered later in dealing with two dimensional problems.)

Let us adopt the nomenclature $\mu_r \equiv \sum\limits_{x=0}^{\infty} \dfrac{(x-\mu)^r v_x}{N}$ where μ is the mean $\left(\sum\limits_{0}^{\infty} \dfrac{xv_x}{N} \right)$, and N is the total number of individuals.

When $f_x = b + cx$ (a first approximation), we find

$$v_x = N\,\frac{b}{c}\left(\frac{b}{c}+1\right)\dots\left(\frac{b}{c}+x-1\right)\frac{\left(1-\dfrac{\mu}{\mu_2}\right)^x}{x!}\left(\frac{\mu}{\mu_2}\right)^{\frac{b}{c}} \quad \dots\dots(2)$$

(The values for the moments are obtained by differentiating the particular moment and making use of equation (1)).

Thus v_x is the $(x+1)^{\text{th}}$ term of the expansion of the binomial

$$N \left\{ \frac{\mu_2}{\mu} - \left(\frac{\mu_2}{\mu} - 1 \right) \right\}^{-\frac{b}{c}}$$

also $\dfrac{c}{b} = \dfrac{\mu_2 - \mu}{\mu^2}$, and $\mu_3\mu + \mu_2\mu = 2\mu_2^2.$ $\dots\dots\dots\dots\dots\dots\dots\dots(3)$

If we write $\lambda_r \equiv \sum\limits_{0}^{\infty} x^r v_x$, $N = \dfrac{\lambda_1^2(\lambda_2 - \lambda_1)}{2\lambda_2^2 - \lambda_1(\lambda_3 + \lambda_2)}.$ $\dots\dots\dots\dots\dots(4)$

In the case where $\dfrac{c}{b}$ tends to zero, the solution reduces to

$$v_x = Ne^{-\mu}\,\frac{\mu^x}{x!}, \quad \dots\dots\dots\dots\dots\dots\dots\dots(5)$$

i.e. Poisson's limit of the binomial. It is interesting to note that the time function ϕ_t has been eliminated, and does not affect the relative distributions given in (2) and (5). This is of importance in dealing with many problems, for example (a) the effects of seasonal variations which apply generally to individuals of all grades, and which probably operate in all epidemics, (b) variations in the virulence of the organism during the course of the epidemic to which it gives rise, and (c) any variations depending upon the values μ_r or λ_r, which are themselves functions of the time and consequently may be expressed as ϕ_t, are eliminated and do not affect the distribution given by the solution.

Example 1. The following figures denote the number of houses in two suburbs of the town of Luckau in Hanover, in which x cases of cancer had occurred during the period 1875–1898.

				Observed	Calculated
Houses with		0	case	64	65
,,	,,	1	,,	43	40
,,	,,	2	cases	10	12
,,	,,	3	,,	2	2.5
,,	,,	4	,,	1	0.4

The value of $\dfrac{c}{b}$ was negligible (-0.009). The calculated figures were obtained by equation (5). Thus the figures afford no evidence that the occurrence of a late case was influenced by previous cases. Behla, however, from whose communication the table was obtained, writes with regard to those houses which experienced 3 and 4 cases during the period " das kann kein zufall sein."

Example 2. The following figures refer to an epidemic of cholera in a village in India.

				Observed	Calculated
Houses with		0	case	168	37
,,	,,	1	,,	32	34
,,	,,	2	cases	16	16
,,	,,	3	,,	6	5
,,	,,	4	,,	1	1
				223	93

If one assumes that the value of $\dfrac{c}{b}$ was zero, then $N = \lambda_1^2/(\lambda_2 - \lambda_1)$, and the calculated values of v_x are as tabulated. This suggests that the disease was probably water borne, that there were a number of wells, and that the inhabitants of 93 out of 223 houses drank from one well which was infected. On further local investigation it was found that there was one particular infected well from which a certain section of the community drank.

Example 3. A modification of this method depends upon the fact that if conditions be constant over a number of years, the problem of one community over a series of years may be replaced by that of a number of communities over a single year. Bowley in his *Elements of Statistics* gives the following figures for deaths from human anthrax in Great Britain.

				Observed	Calculated
Years with		0–3 deaths		1	0.9
,,	,,	4–7	,,	6	5.6
,,	,,	8–11	,,	7	7.3
,,	,,	12–15	,,	4	4.2
,,	,,	16–23	,,	2	1.9

The closeness of the fit suggests that conditions had been constant, and the value for $\dfrac{c}{b}$ (+ 0.089) suggests that the disease is slightly epidemic in the sense that the chance of occurrence of fresh cases in any year increases with the number which have already occurred in that year.

The method has been extensively employed in various directions epidemiological and other. For further examples the reader is referred to nos. 2–8 of the bibliography.
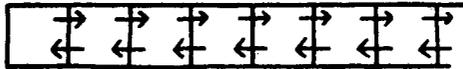

Fig. 4

2. *One dimension, reversible.*

The scheme for reversible cases is given in fig. 4. The variation in this case is the sum of the variations of the forward and backward movements, consequently,

$$dv_x = (f_{x-1} v_{x-1} - f_x v_x)\, dt + (f'_{x+1} v_{x+1} - f'_x v_x)\, dt \ldots\ldots\ldots(6)$$

Where $f = f' = $ constant, $\mu = 0$ and

$$v_x = N e^{-\mu_2} I_x(\mu_2) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(7)$$

Where $I_x(\mu_2)$ is the Bessel function of the xth order for an imaginary argument.

When $f \neq f'$, but both are constant, $\mu = \mu_3$ and

$$v_x = N e^{-\mu_2} \left(\frac{\mu_2 + \mu_3}{\mu_2 - \mu_3} \right)^{\frac{x}{2}} I_x \left(\sqrt{\mu_2^2 - \mu_3^2} \right). \quad \ldots\ldots\ldots\ldots(8)$$

(The value μ_3 is written here in place of μ, as the position for $x = 0$ may not be known.)

Where $f_x = f_x' = b + cx$,

$$v_x = N b \, (b + c) \ldots (b + \overline{cx - 1}) \, (1 + ct) \quad \left(2 \frac{b}{c} + x \right) \frac{t^x}{x!}$$

$$F \left(-\frac{b}{c}, \ -\frac{b}{c} + 1, \ x + 1, \ c^2 t^2 \right) \quad \ldots\ldots(9)$$

3. *Two dimensional cases and correlation.* The schema for these, if the variables x and y are independent, is obviously as shown in figure 5.
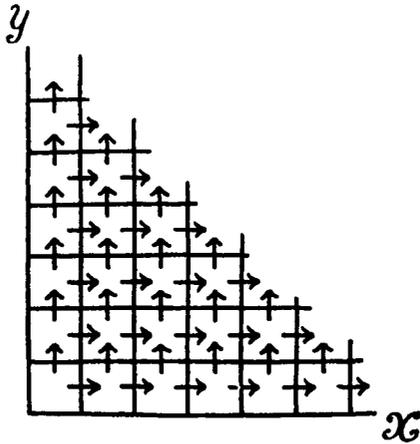

Fig. 5

There are now two movements, one in the x dimension, and one in the y dimension. Consequently

$$dv_{x,\,y} = (f_{x-1,\,y} \, v_{x-1,\,y} - f_{xy} v_{xy}) \, dt + (g_{x,\,y-1} \, v_{x,\,y-1} - g_{x,\,y} \, v_{x,\,y}) dt \quad (10)$$

When $f_{x,\,y} = b + cx$, and $g_{x,\,y} = d + ey$,

$$v_{x,\,y} = v_{0,\,0} \, \frac{b}{c} \left(\frac{b}{c} + 1 \right) \ldots \left(\frac{b}{c} + x - 1 \right)$$

$$\frac{\left(1 - \frac{\mu_{x1}}{\mu_{x2}} \right)^x}{x!} \, \frac{d}{e} \left(\frac{d}{e} + 1 \right) \ldots \left(\frac{d}{e} + y - 1 \right) \frac{\left(1 - \frac{\mu_{y1}}{\mu_{y2}} \right)^y}{y!} \quad (11)$$

where $\mu_{x1} \equiv \sum\limits_{0}^{\infty} \sum\limits_{0}^{\infty}{}_x v_{xy} \div N$

$\mu_{xr} \equiv \sum\limits_{0}^{\infty} \sum\limits_{0}^{\infty} (x - \mu_{x1})^r v_{x,r} \div N$

also $\dfrac{c}{b} = \dfrac{\mu_{x2} - \mu_{x1}}{\mu_{x1}^2}$ and $\dfrac{e}{d} = \dfrac{\mu_{y2} - \mu_{y1}}{\mu_{y1}^2}$ ....(12)

In the case where $\dfrac{c}{b}$ and $\dfrac{e}{d}$ tend to zero,

$$v_{x,y} = N e^{-(\mu_{x1} + \mu_{y1})} \frac{(\mu_{x1})^x}{x!} \frac{(\mu_{y1})^y}{y!} \quad ....(13)$$
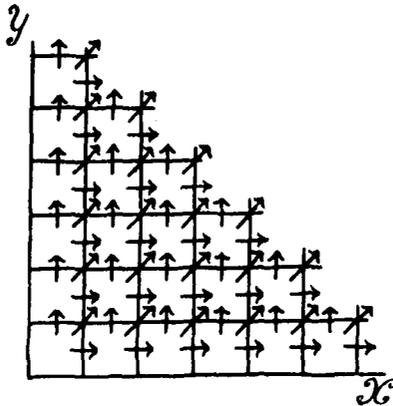

Fig. 6

There is, however, a third direction of movement and this is illustrated in figure 6: individuals may move from the compartment x, y into the compartment $x + 1$, $y + 1$. In figure 5 where there is no oblique movement of this sort, we assumed that the probabilities of movement $f_{xy} dt$ and $g_{xy} dt$ were so small that in comparison with them their product is negligible. This is true of events which are independent. But if when an event of the one sort is likely to happen, an event of the other sort is also likely to happen, then oblique movement is no longer negligible. This relation between the two types of events is the logician's definition of correlation. I have been in the habit of calling it "oblique"

correlation for the following reason. It is obvious that when no correlation exists between x and y the numbers v_{xy} will, when a sufficient period of time has elapsed, be so arranged that equal values of $v_{x,y}$ will lie on contours which have their major and minor axes parallel, the one to the x coordinate and the other to the y coordinate. When oblique correlation exists (as denoted by the oblique arrows) there will be a tendency for these *ovoid* forms to be dilated in the $x = y$ direction and so oblique *ovoid* forms will result. But this type of oblique contour may be also brought about by mathematical considerations of another nature. Thus if $f_{x,y}$ is an increasing function of y, a " shear " movement will tend to push the upper portion of a non-correlated distribution in the x direction, and similarly if $g_{x,y}$ is an increasing function of x there will be a shear movement pushing the portion on the right in an upward direction. This type of movement is shown in the third diagram of figure 7. I have been in the habit of denoting this variety by the name "shear" correlation. Statistically it means that the chance of an occurrence of one sort depends upon the number of previous occurrences of the other sort.

No correlation.   Oblique.   Shear.

Fig. 7
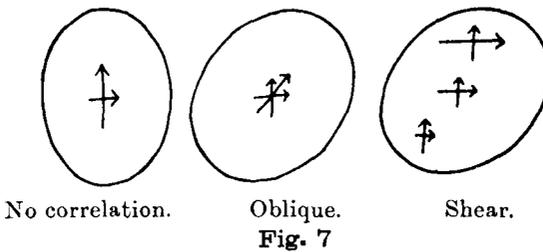
" Shear " correlation may be illustrated in the following example from vital statistics. If x be the number of males, y the number of females, and $v_{x,y}$ denote the number of communities containing x males and y females, then the probability of a child (male or female) being born will depend (subject to the type of marriage which is in operation) upon the number of parents of both sexes, that is to say it will be a function of xy. In the resulting distribution the largest values of $v_{x,y}$ will lie along the diagonal $x = y$, or in other words there will tend to be an equality of sexes in the communities

The general equation, which includes both types of correlation, is

$$
\left.
\begin{aligned}
\frac{dv_{x,\,y}}{dt} &= f_{x-1,\,y}\, v_{x-1,\,y} \; - f_{x,\,y}\, v_{xy} \\
&+ g_{x,\,y-1}\, v_{x,\,y-1} - g_{x,\,y}\, v_{xy} \\
&+ h_{x-1,\,y-1}\, v_{x-1,\,y-1,} \; - h_{x,\,y}\, v_{z,\,y}
\end{aligned}
\right\} \quad \dots\dots\dots\dots(14)
$$

In the case where f, g and h are all constants, that is to say a simple case of oblique correlation, with no shear, the solution is

$$
v_{x,\,y} = N e^{-(\mu_{x2} + \mu_{y2} - \mu_{yx})} \sum_{s\,=\,0} \frac{(\mu_{x2} - \mu_{xy})^{x-s}}{(x-s)!} \frac{(\mu_{y2} - \mu_{xy})^{y-s}}{(y-s\,!} \frac{\mu_{xy}{}^{s}}{s!} \quad \dots\dots(15)
$$

where μ_{xy} is written for $\sum\limits_{0}^{\infty}\sum\limits_{0}^{\infty} (xy - \mu_{x1}\,\mu_{y1})\, v_{xy} \div N$,

and the upper limit of the summation in the expression for v_{xy} is $s = x$ or y whichever is least.

The moments may be obtained as follows   We have for v_{00} the relation $\dfrac{dv_{00}}{dt} = -(f_{00} + g_{00} + h_{00})\, v_{00}$

or if m be written for $\log \dfrac{N}{v_{00}}$

$$
\frac{dm}{dt} = f_{00} + g_{00} + h_{00}.
$$

In the case where

$$
\begin{aligned}
f_{x,\,y} &= \alpha_1 + \alpha_2 x + \alpha_3 y \\
g_{x,\,y} &= \beta_1 + \beta_2 x + \beta_3 y \\
h_{x,\,y} &= \gamma_1 + \gamma_2 x + \gamma_3 y,
\end{aligned}
$$

if we also write

$$
\begin{aligned}
P_r &\equiv \frac{\alpha_r}{\alpha_1 + \beta_1 + \gamma_1} \\[4pt]
Q_r &\equiv \frac{\beta_r}{\alpha_1 + \beta_1 + \gamma_1} \\[4pt]
R_r &\equiv \frac{\gamma_r}{\alpha_1 + \beta_1 + \gamma_1}
\end{aligned}
$$

equation (14) takes the form

$$
\begin{aligned}
\frac{dv_{x,\,y}}{dm} &= (P_1 + P_2(x-1) + P_3 y)v_{x-1,\,y} \; - (P_1 + P_2 x + P_3 y)\, v_{x,\,y} \\
&+ (Q_1 + Q_2 x + Q_3(y-1))\, v_{x,\,y-1} - (Q_1 + Q_2 x + Q_3 y)\, v_{x,\,y} \\
&+ (R_1 + R_2(x-1) + R_3(y-1))\, v_{x-1,\,y-1} - (R_1 + R_2 x + R_3 y)\, v_{x,\,y}.
\end{aligned}
$$

Hence by differentiating the expressions for the moments we have

$$
\left.
\begin{aligned}
\frac{d\mu_{x1}}{dm} &= (P_1 + R_1) + (P_2 + R_2)\,\mu_{x1} + (P_3 + R_3)\,\mu_{y1} \\
\frac{d\mu_{y1}}{dm} &= (Q_1 + R_1) + (Q_2 + R_2)\,\mu_{x1} + (Q_3 + R_3)\,\mu_{y1} \\
\frac{d\mu_{x2}}{dm} &= (P_1 + R_1) + (P_2 + R_2)\,\mu_{x1} + (P_3 + Q_3)\,\mu_{y1} + 2\,(P_2 + R_2)\,\mu_{x2} \\
&\quad + 2\,(P_3 + R_3)\,\mu_{xy} \\
\frac{d\mu_{y2}}{dm} &= (Q_1 + R_1) + (Q_2 + R_2)\,\mu_{x1} + (Q_3 + R_3)\,\mu_{y1} + 2\,(Q_2 + R_2)\,\mu_{xy} \\
&\quad + 2\,(Q_3 + R_3)\,\mu_{y2} \\
\frac{d\mu_{xy}}{dm} &= R_1 + R_2\,\mu_{x1} + R_3\,\mu_{y1} + (Q_2 + R_2)\,\mu_{x2} + (P_3 + R_3)\,\mu_{y2} \\
&\quad + (P_2 + R_2 + Q_3 + R_3)\,\mu_{xy}
\end{aligned}
\right\} \quad \ldots\ldots(16)
$$

(We may remind the reader at this point that oblique correlation depends upon the values R_1, R_2 and R_3, whereas shear correlation depends upon P_3 and Q_2.) These equations may be solved by the usual methods.

In the particular case where f, g and h are all constants, we have from the above,

$$
\left.
\begin{aligned}
\mu_{x1} &= \mu_{x2} = (P_1 + R_1)\,m \\
\mu_{y1} &= \mu_{y2} = (Q_1 + R_1)\,m \\
\mu_{xy} &= R_1\,m
\end{aligned}
\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots(17)
$$

whence
$$
\frac{\mu_{xy}}{\sqrt{\mu_{x2}\,\mu_{y2}}} = \frac{R_1}{\sqrt{(P_1 + R_1)(Q_1 + R_1)}} \quad \ldots\ldots\ldots\ldots(18)
$$

an expression from which m and consequently t has been eliminated. This is of course the ordinary Pearsonian correlation coefficient r.

It is easy to show, by summations of the values of $v_{x,y}$, that if the statistics are arranged in the four fold table

	not A	A
not B	a	b
B	c	d

$$
r = \frac{\log \dfrac{a\,(a + b + c + d)}{(a + b)(a + c)}}{\sqrt{\log \dfrac{a + b + c + d}{a + b} \cdot \log \dfrac{a + b + c + d}{a + c}}} \quad \ldots\ldots\ldots(19)
$$

and it is interesting to note that when $\dfrac{c + d}{a + b}$, $\dfrac{b + d}{a + c}$, and $\dfrac{ad - bc}{(a + b)(a + c)}$ are so small that their squares may be neglected, this value of r reduces to $\dfrac{ad - bc}{\sqrt{(a + b)(a + c)(b + d)(c + d)}}$ which is the Pearsonian coefficient ϕ.

4. *Summations.*

Before leaving this part of the subject attention may be drawn to the following schema fig. (8) which shows the effect of collecting the values of v_{xy} in either diagonal direction, *i.e.* according to the
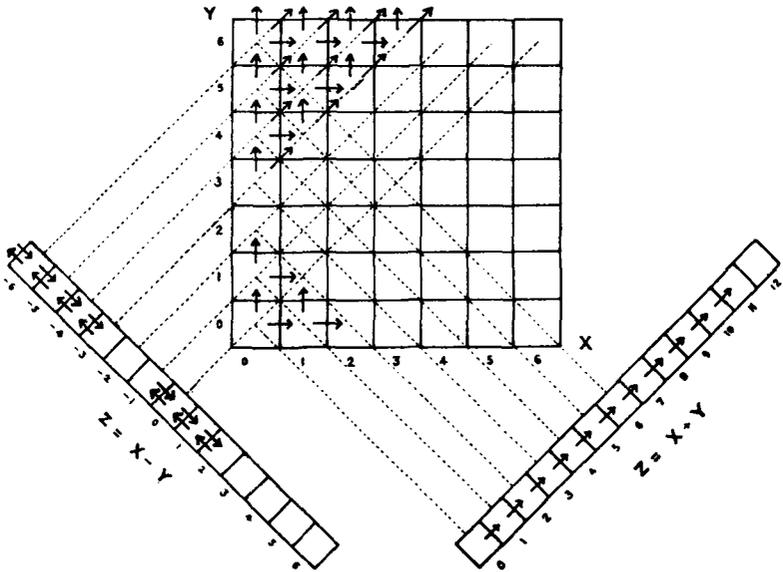

Fig. 8

relation $n = x + y$ or $n = x - y$. Such summations may be written in the forms

$$S_n = \sum_{s=0}^{n} v_{n-s,\,s}$$

$$\text{and } T_n = \sum_{s=0}^{\infty} v_{n+s,\,s}$$

$$T_{-n} = \sum_{s=0}^{\infty} v_{s,\,n+s}.$$

These are important in medical statistical problems. For example if x denote the number of fresh infections cf a disease, and y denote the number of relapses, then $x + y = n$ denotes the number of attacks of the disease. The functions $f_{xy}\,dt$, and $g_{xy}\,dt$, for fresh infections and relapses respectively, are certainly not identical, and as it is impossible to differentiate clinically between fresh infection and relapse, statistics are only available in the form S_n.

It is at once apparent from the figure, that where f, g and h are constants the result of the summation $n = x + y$, is to convert an irreversible two dimensional schema into a reversible schema in one dimension (the effect of oblique correlation does not appear), and the solution is the Bessel form which has already been dealt with.

Summation according to $n = x + y$ gives the equation

$$\frac{dS_n}{dm} = (P + Q)(S_{n-1} - S_n) + R(S_{n-2} - S_n) \quad \dots\dots\dots(20)$$

where for convenience the suffixes of P_1, Q_1 and R_1 are dropped, whence for the moments according to n we have

$$\left.\begin{aligned}
\mu &= (P + Q + 2R)\,m = (1 + R)\,m \\
\mu_2 &= (P + Q + 4R)\,m = (1 + 3R)\,m \\
\mu_3 &= (P + Q + 8R)\,m = (1 + 7R)\,m
\end{aligned}\right\} \quad \dots\dots\dots\dots(21)$$

Consequently $R = \dfrac{\mu - \mu_2}{\mu_2 - 3\mu}$ $\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(22)$

and $\mu_3 = 3\mu_2 - 2\mu$ $\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(23)$

From the latter expression we also obtain the relation

$$N = \frac{3\lambda_1(\lambda_2 - \lambda) \pm \lambda_1 \sqrt{(9\lambda_2^2 - 7\lambda_1^2 - 8\lambda_1\lambda_3 + 6\lambda_1\lambda_2)}}{2(\lambda_3 - 3\lambda_2 + 2\lambda_1)} \quad \dots\dots(24)$$

Also after some reduction we find

$$r^2 = \frac{R^2}{(P + R)(Q + R)} = \frac{\left(2 + \dfrac{P}{Q} + \dfrac{Q}{P}\right)R^2}{1 + \left(\dfrac{P}{Q} + \dfrac{Q}{P}\right)R + R^2} \quad \dots\dots\dots\dots(25)$$

$\left(\text{and thus when } P = Q,\ r = \dfrac{2R}{1 + R} = \dfrac{\mu_2}{\mu} - 1.\right)$ $\dots\dots\dots\dots\dots(26)$

Finally we have

$$S_n = Ne^{-\frac{3\mu - \mu_2}{2}} \sum_{s=0}^{n/2} \frac{(2\mu - \mu_2)^{n-2s}}{(n - 2s)!} \frac{\left(\dfrac{\mu_2 - \mu_1}{2}\right)^s}{s!}. \quad \dots\dots\dots\dots\dots\dots(27)$$

Example 4. In a phagocytic experiment it seemed likely that bacteria which were being ingested were not all discrete, some of them were united into pairs. If one considers for the moment that they were of two types, and that a pair consisted of one of each type, then the above analysis is applicable.

The figures were as follows—

				Observed	Calculated
Leucocytes containing 0 bacteria				269	268
,,	,,	1	,,	4	7
,,	,,	2	,,	26	23
,,	,,	3	,,	0	0.6
,,	,,	4	,,	1	1.1

and, since in this case $P_i = Q$, $r = 0.86$.

This example rests upon incomplete assumptions, and upon insufficient data. It is introduced only to show how, when $P = Q$. the correlation coefficient may be calculated from one dimensional data.

5. *Restricted Cases.*

So far we have been dealing with the complete case in which entry into all compartments in the schema is possible. Let us now turn to particular cases in which certain limitations are introduced.
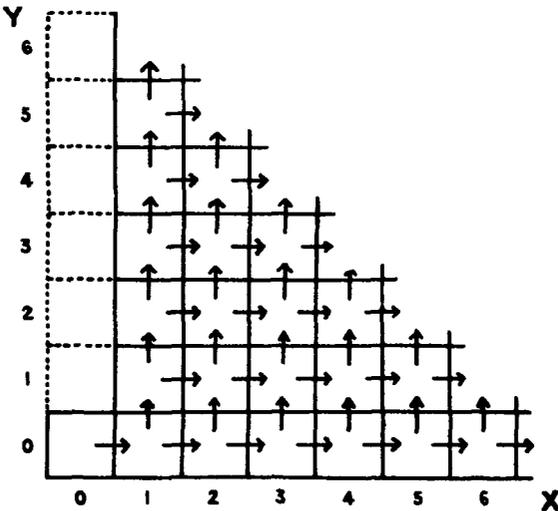

Fig. 9

Example 5. The problem of house infection may be approached in a manner rather different from that suggested in 1. In the accompanying figure (9) the coordinate x denotes infections arising from without the house, and the coordinate y denotes infections originating within the house itself. There can be no entrance into the upper compartments of the zero column, for an internal infection is only possible after an infection from the exterior has taken place. The chance of an external infection occurring is proportional to the number of cases in all houses except the house with which one is dealing, and the probability of occurrence of an internal infection depends upon the number of cases which have already occurred within the house itself.

Thus, in its simplest form, we have

$$\frac{dv_{x,\,y}}{dt} = k\left\{ \left(\Sigma\Sigma \left(x + y \right) v_{x,\,y} - n + 1 \right) v_{x-1,\,y} - \left(\Sigma\Sigma \left(x + y \right) v_{x,\,y} - n \right) v_{x,\,y} \right\}$$
$$+ \, l\left\{ \left(n - 1 \right) v_{x,\,y-1} - n\,v_{x,\,y} \right\} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(28)$$

where kdt and ldt denote the chances of external and internal infections respectively.

Hence for the value S_n we have

$$\frac{dS_n}{dt} = kN\mu\left(S_{n-1} - S_n\right) + \left(l - k\right)\left\{\left(n - 1\right)S_{n-1} - nS_n\right\} \quad \dots\dots(29)$$

where the moment μ is taken with reference to n.

Consequently $\quad \dfrac{dN\mu}{dt} = \left\{k\left(N - 1\right) + l\right\}N\mu$

and $\quad \dfrac{l}{k} = \dfrac{N\left(\mu - \mu_0\right)}{\log\dfrac{N}{S_0}} - N + 1, \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(30)$

where μ is the initial value of μ.

That is to say from the total number of cases $(N\mu)$, the total number of houses (N), and the number of infected houses S_0, we can find the ratio of the probability of external infection to that of internal infection, as in most cases μ_0 is small as compared with μ.*

---

* If the whole distribution is given we can use the expressions

$$\frac{l}{k} = 1 + N\,\frac{\mu_2 - \mu}{\mu_2 + \mu} \; (31), \text{ and } \frac{3\mu}{\mu_2} - \frac{2\mu}{\mu_3} = 1 \; (32).$$

In the case of the cancer statistics one thus obtains $\dfrac{l}{k} = 0\cdot966$ and for the left side of the second expression (32) the value $1\cdot23$.

In the case of bubonic plague, from four epidemics in consecutive years in a certain village, $\frac{l}{k} = 199$. For a similar number of epidemics in a neighbouring village the value was 231. Allowing for a large error, the probability of internal infection would appear to be about 200 times as great as that of external infection. In the case of cancer, from one group of statistics a value of 10 was obtained, and from another a value of -9. Each of these numbers is the difference between two large numbers and has a large error. The high value of the ratio $\frac{l}{k}$ in the case of plague may be easily understood when one remembers that the disease is transmitted by fleas, a species of animal which does not as a rule travel far from its own neighbourhood. In the case of cancer the figures afford no evidence either of insect infection or of infection by contagion. Present theories regarding the transmission of *kala azar* point to transmission through either the bedbug or the sand fly. The former is a very local insect, the latter is supposed to be distinctly localised, though not as strictly so as either the flea or the bed-bug. From figures which have been placed at my disposal the value of $\frac{l}{k}$ was of the order 70, suggesting that the sand fly is the more likely carrier.

Example 6. The problem of infection and relapse in malaria is similar to that dealt with in the last example. The figures which follow relate to sizes of spleen of young children, and those who have studied the disease consider that the size of the spleen bears a direct relation to the number of attacks which the patient has experienced. As in this case the disease is endemic (that is to say it is in a more or less steady state) the number of sources may be taken as constant (a).

The equations are consequently

$$\frac{dv_{00}}{dt} = -kav_{00}, \quad \frac{dv_{10}}{dt} = kav_{00} - (ka + l)v_{10}$$

and for values of y greater than zero,

$$\frac{dv_{x,y}}{dt} = ka(v_{x-1,y} - v_{xy}) + l(v_{x,y-1} - v_{x,y}) \quad \left.\right\} \quad \ldots(33)$$

where kdt is the probability of external infection, and ldt is the probability of relapse.

From this we deduce the series of equations.

$$\frac{dS_0}{dt} = - kaS_0$$

$$\frac{dS_1}{dt} = kaS_0 - (ka + l)\,S_1$$

and for values of n greater than unity,

$$\frac{dS_n}{dt} = (ka + l)\,(S_{n-1} - S_n).$$

$$\left. \right\} \quad \dots\dots\dots(34)$$

Thus we find $\quad \dfrac{l}{ka} = \dfrac{m - \mu}{1 - m - e^{-m}}$ $\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(35)$

and $S_n = S_0 \dfrac{ka}{l}\left(\dfrac{ka}{l} + 1\right)^{n-1}\left[1 - \sum_{0}^{n-1}\dfrac{\left(\dfrac{lm}{ka}\right)^r}{r!}\ e^{-\frac{lm}{ka}}\ \right]$ $\dots\dots\dots\dots(36)$

The figures are as follows

Values of $n =$	0	1	2	3	4	5	6	7	l/ka
A. Aden Malarious Villages									
obs.	89	38	36	26	10	9	1	1	
calc.	(89)	39	35	24	13	6	2	1	1.78
B. Aden Sheikh Othman									
obs.	1006	38	23	13	7	0	4		
calc.	(1006)	34	25	16	7	3	1		25.6
C. Punjab (1909) Spleen rates 30–40%									
obs.	264	67	44	15	3	2			
calc.	(264)	70	39	16	5	1			2.23
D. Punjab (1909) Spleen rates 50–60%									
obs.	81	38	26	16	6	3	2		
calc.	(81)	35	28	17	8	3	1		1.56

Example 7. The problem of the behaviour of epidemics may be considered from various points of view, one of which is as follows. Let us trace what is likely to be the distribution of a

number of epidemics in similar communities each starting from a single case, and let us adopt as variables the relation

$$\text{case } (n) = \text{infections} (x) - \text{recoveries} (y).$$
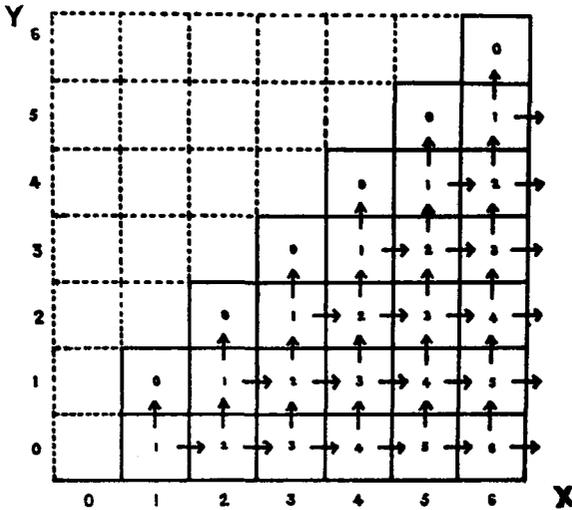
The course of affairs is illustrated in figure 10

Fig. 10

We notice that each epidemic starts initially in the compartment $(1, 0)$, *i.e.* one case which has not yet recovered, and also that when a compartment $n = x - y = 0$ is arrived at, the epidemic has come to an end for there are no more cases. Thus when the probability of recovery exceeds the probability of infection all epidemics must come to rest in one or other of the compartments $n = 0$. If we assume that kdt is the chance that an individual may convey infection, then the probability that an epidemic of n cases will receive an additional case is $kndt$. Similarly the chance that it will lose a case is $lndt$. We assume also, in the first instance, that the population concerned in each epidemic is unlimited. Consequently from the schema

$$\frac{dv_{x,\,y}}{dt} = k\{(\overline{x-1} - y)\,v_{x-1,\,y} - (x - y)\,v_{x,\,y}\}$$
$$+ l\,\{(x - \overline{y-1})\,v_{x,\,y-1} - (x - y)\,v_{x,\,y}\}\ldots\ldots(37)$$

I have obtained solutions of this equation for serial values of x and y up to $v_{4,4}$, but the expressions are not concise, and do not suggest a general form. If, however, one is content with finding the values $v_{1,1}, v_{2,2}, v_{3,3} \ldots$ after an infinite time— that is to say if the problem is to find the distribution of a group of epidemics, which have reached finality, classified according to the number of cases which they have experienced, then the problem is less refractory, and the solution is

$$v_{n,n} = a_n \left(\frac{k}{k+l}\right)^{n-1} \left(\frac{l}{k+l}\right)^n N \ldots\ldots\ldots\ldots\ldots(38)$$

where
$$a_1 = 1$$
$$a_2 = a_1 a_1 = 1$$
$$a_3 = a_1 a_2 + a_2 a_1 = 2$$
$$\vdots$$
$$a_n = a_1 a_{n-1} + a_2 a_{n-2} + a_3 a_{n-3} + \ldots a_{n-1} a_1.$$

These values may be obtained more concisely by means of the formula $a_n = \dfrac{2n-3}{n} 2a_{n-1}$, for which I am indebted to Dr W. O. Kermack. It follows also from equation (37) as well as from equation (38) that $\dfrac{l}{k} = \dfrac{\mu}{\mu-1}$. It will be seen that the coefficients a_n are the number of different paths by which an individual moving either east or north may pass from the compartment $(1, 0)$ to the compartment (n, n). The resulting curves for $v_{n,n}$ are of the J type, that is to say they are like exponential curves with a negative index, in which the tail is unduly prolonged.

During the great epidemic of influenza in 1918, very accurate statistics were collected by the Australian Government,[9] of epidemics which had occurred in incoming ships during their voyage to Australia. The figures are as follows

Number of epidemics with 1 case						-	34
,,	,,	,,	,,	2	cases	-	15
,,	,,	,,	,,	3	,,	-	5
,,	,,	,,	,,	4	,,	-	3
,,	,,	,,	,,	5	,,	-	8
,,	,,	,,	,,	6	,,	-	1
,,	,,	,,	,,	7	,,	-	3
,,	,,	,,	,,	8	,,	-	4
,,	,,	,,	,,	9	,,	-	2
			more than 9 cases			-	17

I have been unable to fit these figures by means of equation (38); on the one hand the approximations are probably insufficient, and on the other the crews of the ships varied very considerably. But the J character of the distribution is evident. It is also interesting to note that in 82 per cent. of the ship epidemics the number of cases was less than ten. Now influenza is a peculiarly difficult disease to diagnose in the individual case; consequently epidemics of less than 10 cases are in ordinary circumstances seldom recognised and reported. The conclusion is suggested that as our limited experience of epidemic influenza is based upon statistics which may relate only to a small selected minority of the total number of epidemics, it may be in no sense representative, and may even be misleading.

If in the second place we wish to investigate the distribution of epidemics according to their duration, we have to find the value of $S_0 = \overset{\infty}{\underset{0}{\Sigma}} v_{n,\,n}$ as a function of the time. The problem then reduces to one in a single dimension n, in which there is an instantaneous point source at $n = 1$, and in which (for $l > k$) all epidemics finally enter the compartment S_0.

Let us now turn to the case where the population affected by each epidemic is limited, and of constant value p.

The equation is

$$\frac{dv_{x,\,y}}{dt} = k \left\{ (p - \overline{x - 1})(\overline{x - 1} - y)\, v_{x-1,\,y} - (p - x)(x - y)\, v_{x,\,y} \right\}$$
$$+ l \left\{ (x - \overline{y - 1})\, v_{x,\,y-1} - (x - y)\, v_{x,\,y} \right\}. \quad \ldots\ldots\ldots\ldots\ldots(39)$$

For magnitude of epidemics we have as before to find the value of $v_{n,\,n}$, when t tends to infinity.

These are found to be

$$\left.\begin{array}{l}
v_{11} = \alpha_{p-1}\, N \\
v_{22} = \beta_{p-1}\, \alpha^2_{p-2}\, N \\
v_{33} = \beta_{p-1}\, \beta_{p-2}\, \alpha^2_{p-3}\, N S_3 \\
v_{44} = \beta_{p-1}\, \beta_{p-2}\, \beta_{p-3}\, \alpha^2_{p-4}\, N S_4 \\
\text{etc.}
\end{array}\right\} \quad \ldots\ldots\ldots\ldots(40)$$

where α_{p-r} is written for $\dfrac{l}{k\,(p\,-\,r)+l}$ and β_{p-r} for $\dfrac{k\,(p\,-\,r)}{k\,(p\,-\,r)+l}$

The values of S_3, S_4 ... are related as follows

$$S_3 = \alpha_{p-3} + \alpha_{p-2}$$
$$S_4 = \alpha_{p-4}\,(\alpha_{p-4} + p_{p-3} + \alpha_{p-2}) + \alpha_{p-3}(\alpha_{p-3} + \alpha_{p-2}),\ \text{etc.}$$

These may be written symbolically in the form

$$S_3 = (3 + 2)$$
$$S_4 = 4\,(4 + 3 + 2) + 3\,(3 + 2)\ \text{etc.}$$

Let us further, taking the suffix 5 as an example, use the notation

$$A_5 = 5 + 4 + 3 + 2$$
$$B_5 = 5\,(5 + 4 + 3 + 2) + 4\,(4 + 3 + 2) + 3\,(3 + 2)$$
$$C_5 = 5\,\{5\,(5 + 4 + 3 + 2) + 4\,(4 + 3 + 2) + 3\,(3 + 2)\}$$
$$\qquad\qquad + 4\,\{4\,(4 + 3 + 2) + 3\,(3 + 2)\}$$

whence $B_5 = 5A_5 + 4A_4 + 3A_3$
$$\qquad\ C_5 = 5B_5 + 4B_4.$$

Similarly $B_6 = 6A_6 + 5A_5 + 4A_4 + 3A_3$
$$\qquad\ C_6 = 6B_6 + 5B_5 + 4B_4$$
$$\qquad\ D_6 = 6C_6 + 5C_5.$$

Then $S_3 = A_3$
$$\qquad S_4 = 4A_4 + 3A_3 = B_4$$
$$\qquad S_5 = 5B_5 + 4B_4 = C_5$$
$$\qquad S_6 = 6C_6 + 5C_5 = D_6$$
$$\qquad S_7 = 7D_7 + 6D_6 = E_7,\ \text{etc.}$$

In the first instance the values A_3, A_4, ... , etc., are calculated, and from these are obtained successively the B's, then the C's and so on. The character of the curves for the values of $v_{n,\,n}$ may be seen from the following numerical examples in the first of which $N = 1$, $\dfrac{l}{k} = 2$, and $p = 5$: and in the second $N = 1$, $\dfrac{l}{k} = 2$, $p = 6$.

$v_{1,\,1}$	0.33	0.29
$v_{2,\,2}$	0.11	0.08
$v_{3,\,3}$	0.09	0.06
$v_{4,\,4}$	0.13	0.08
$v_{5,\,5}$	0.34	0.13
$v_{6,\,6}$		0.37
	1.00	1.01

The distributions are of the U type.

6. *Generalisation.*

The equations of all the examples with which we have been engaged may be written generally in the form

$$
\left.
\begin{aligned}
\frac{dv_{x,\,y}}{dt} &= f_{x-1,\,y}\,v_{x-1,\,y} \;-\; f_{x,\,y}\,v_{x,\,y} \;+\; f''_{x+1,\,y}\,v_{x+1,\,y} \;-\; f''_{x,\,y}\,v_{x,\,y} \\
&\;+\; g_{x,\,y-1}\,v_{x,\,y-1} \;-\; g_{x,\,y}\,v_{x,\,y} \;+\; g'_{x,\,y+1}\,v_{x,\,y+1} \;-\; g'_{x,\,y}\,v_{x,\,y} \\
&\;+\; h_{x-1,\,y-1}\,v_{x-1,\,y-1} \;-\; h_{x,\,y}\,v_{x,\,y} \;+\; h'_{x+1,\,y+1}\,v_{x+1,\,y+1} \\
&\hspace{8cm} -\; h'_{x,\,y}\,v_{x,\,y} \\
&\;+\; i_{x-1,\,y+1}\,v_{x-1,\,y+1} \;-\; i_{x,\,y}\,v_{x,\,y} \;+\; i'_{x+1,\,y-1}\,v_{x+1,\,y-1} \;-\; i'_{x,\,y}\,v_{x,\,y}
\end{aligned}
\right\} \quad (41)
$$

in which f, g, h and i refer to forward translations, and f', g', h' and i' refer to backward translations, and this generalisation may obviously be extended to any number of dimensions. If in place of a continuous variation through time we are dealing with a succession of events, then for $-\dfrac{dv}{dt}$ we write $v_{t-1} - v_t$.

Thus generally for any two variables x and y, a general equation may be built up by equating the right hand side of the preceding equation to zero. This may be represented schematically in the form
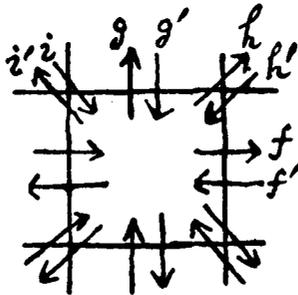

Fig. 11

The arrows no longer represent translations, but are now differences between the numbers in the compartments which they connect. The schema, and its corresponding equation, now indicate the relationship which exists between the numbers in a particular

compartment and those in the compartments with which it is contiguous. In 2 dimensions the number of nearest neighbours is 8, in 3 it is 26, and in n dimensions it is $3^n - 1$. So long as the variables are discontinuous, and proper units are employed, the above schema is generally descriptive, and the problem becomes one of a matrix of pure numbers. In certain cases which I have examined, more distant compartments also have an influence, but these cannot be dealt with here.

Example 8. For a purely discontinuous example we may turn to ordinary probabilities. Let p be the probability of r successes in n events. The variables in this case are $y = n$ and $x = r$. Also $h = p$, and $g = 1 - p$. In an event an individual may either have experienced a failure in which case he moves up one compartment in the n direction, or he may have been successful in which case he moves obliquely in a north-easterly direction (fig. 12). There can obviously be no success without an event, so no horizontal arrows are drawn. This then is a problem of correlation, and its equation is
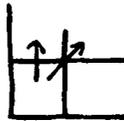

Fig. 12

$$q\,(v_{n-1,\,r} - v_{n,\,r}) + p\,(v_{n-1,\,r-1} - v_{n,\,r}) = 0 \ \ldots\ldots\ldots\ldots(42)$$

The solution of this equation, in the case in which initially no events, and no successes have occurred, is the well known expression,

$$v_{n,\,r} = N \frac{n\,!}{(n-r)\,!\,r\,!}\,(1 - p)^{n-r}\,p^r \ \ldots\ldots\ldots\ldots(43)$$

For the coefficient of correlation r we find

$$r^2 = \frac{p}{1 - (1 - p)\dfrac{n-4}{n+2}}, \quad \ldots\ldots\ldots\ldots(44)$$

which shows that r is not in this case a satisfactory coefficient. Again if a success entails a gain, and a failure a loss, then if y denotes the number of events, and x the number of gains, $h = p$ and $i' = i - p$. The schema is as in fig. 13, and the well known result follows.
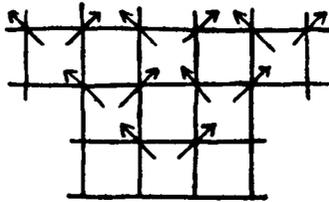

Fig. 13

7. *Continuous Variables.*

A consideration of continuous variables leads us from the foregoing general equation, on the one hand into the domain of mathematical physics, and on the other into wider statistical fields. Each uncorrelated term of equation may in this case be written as

$$\frac{f_{x \pm k} \, v_{x \pm k} - f_{x, \, y} \, v_{x, \, y}}{k}$$

which after expansion becomes

$$\pm \frac{\partial f v}{\partial x} + \frac{k}{2 \, !} \frac{\partial^2 f v}{\partial x^2} \pm \frac{k^2}{3 \, !} \frac{\partial^3 f v}{\partial x^3} + \dots$$

Similarly for the correlated terms we have

$$\frac{h_{x \pm k, \, y \pm k} \, v_{x \pm k, \, y \pm k} - h_{x, \, y} \, v_{x, \, y}}{k}$$
$$= \left(\pm \frac{\partial h v}{\partial x} \pm \frac{\partial h v}{\partial y} \right) + \frac{k}{2 \, !} \left(\frac{\partial^2 h v}{\partial x^2} + 2 \frac{\partial^2 h v}{\partial x \, \partial y} + \frac{\partial^2 h v}{\partial y^2} \right) \pm \dots$$

It is to be noted that in considering correlation a movement in the oblique direction consists of a movement of one unit in each dimension; hence if a unit is denoted by k it is the same for both x and y.

By making use of the above forms of expansion, the general equation may be built up. Thus for example in a reversible case

in the three dimensions of space, when k is so small that its square may be neglected, and the functions of f, g, etc., are equal constants

$$\frac{\partial v}{\partial t} = f \frac{k}{2!} \left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} \right), \quad \dots\dots\dots\dots(45)$$

which is the ordinary equation for diffusion (of matter or of heat). In this case k is the length of the mean freepath of molecule or electron as the case may be. If k tended to zero there would be no diffusion, and if k were large it would be necessary to include higher differentials. Thus Fourier's equation contains in its essence the idea of heterogeneity or discreteness.

Similarly in a reversible case in two dimensions in which the variables x and y are correlated we have

$$\frac{\partial v}{\partial t} = \frac{k}{2!} \left(\frac{\partial^2 fv}{\partial x^2} + \frac{\partial^2 gv}{\partial y^2} + \frac{\partial^2 hv}{\partial x^2} + 2 \frac{\partial^2 hv}{\partial x \partial y} + \frac{\partial^2 hv}{\partial y} \right) \quad \dots\dots(46)$$

which when f, g and h are constant is the equation of the correlated frequency surface.

Again in the general case in three dimensions without oblique correlation, if h tends to zero we have

$$\frac{\partial v}{\partial t} + \frac{\partial (f - f') v}{\partial x} + \frac{\partial (g - g') v}{\partial y} + \frac{\partial (e - e') v}{\partial z} = 0, \text{ or } = \frac{dv}{dt} \dots(47)$$

The left-hand side of the equation is equal to zero if there is no variation of material (*i.e.* of v), and if there be a variation it is equal to $\frac{dv}{dt}$, where this differential expresses the rate of variation of substance in following the element in its movement. If the functions f, g, e, f', g', e', are constants this is the ordinary equation of hydrodynamics. In actual fact k does not tend to zero. Diffusion must always take place as an element of the fluid flows onwards, but the place left empty is filled up from neighbouring elements, and the diffusion is so to speak neutralised.

Example 9. Let us now turn to a problem in vital statistics. Let the variables be the time (t) and the age (θ)—thus $v_{t, \theta}$ denotes the number of individuals aged θ at the time t. Time and age are continuous ($k \to 0$), are absolutely correlated ($r = 1$) and the rate

of movement through both time and age are the same and are constant. The translation is irreversible (fig. 14). Thus if there is no death rate
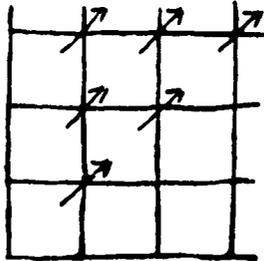

Fig. 14

$$\frac{v_{t-k,\ \theta-k}\ -\ v_{t,\ \theta}}{k} = 0 \quad \text{or} \quad \frac{\partial v}{\partial t} + \frac{\partial v}{\partial \theta} = 0.$$

If there be a death rate then

$$\frac{\partial v}{\partial t} + \frac{\partial v}{\partial \theta} = \frac{dv}{dt} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots(48)$$

where $\dfrac{dv}{dt}$ has the same significance as in the hydrodynamical equation, viz., it is the variation in the element as one follows it in its movement.

Writing $\dfrac{dv}{dt} = -f_{t,\ \theta} v_{t,\ \theta}$, where $f_{t,\ \theta} dt$ is the probability of dying at age θ and at time t, we have

$$\frac{\partial v}{\partial t} + \frac{\partial v}{\partial \theta} = -f_{t,\ \theta}\, v_{t,\ \theta}\,.$$

Whence

$$v_{t,\ \theta} = v_{t-\theta,\ 0}\, e^{-\int_0^\theta f(t\,-\,\theta\,+\,t,\ t)dt} \quad \dots\dots\dots\dots\dots(49)$$

which may be translated as follows :—the number of persons aged θ at the time t, is equal to the number born at the time $t - \theta$ reduced exponentially by the successive death rates (taken for convenience discontinuously)

for age 0 at the time $t - \theta$
for age 1 at the time $t - \theta + 1$
for age 2 at the time $t - \theta + 2$ and so on up to $\xi = \theta$.

If we neglect variations in the death rate with the time, we know that $f(\theta)$ has roughly the form drawn in fig. 15.

Fig. 15

Thus for a number of pairs of values θ_1, and θ_2, $f(\theta_1) = f(\theta_2)$. Now if a population is decreasing according to the relation $\dfrac{\partial v}{\partial t} = -cv$, where c is constant, the equation takes the form

$$\frac{\partial v_\theta}{\partial \theta} = -(f_\theta - c)\, v_\theta.$$

If c is less than the minimal value of $f(\theta)$, then v_θ will always decrease with increasing θ.

If c is equal to the minimal value of $f(\theta)$ there will be one level point, and if c is greater than the minimal value but less than $f(0)$ there will be two level points.

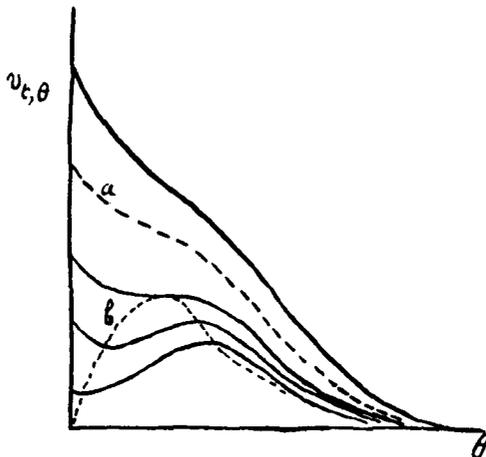Thus age distribution curves will have forms roughly as in fig. 16.

Fig. 16

In this diagram the dotted curve a is intended to represent the form when the population is steady, and the dotted line b is the locus of the level points. The curves above a are those of increasing populations, and those below it represent the age distributions of populations which are on the decrease. The hump on the curve occurs between the ages of 20 and 40, the child-bearing age, consequently the more a population is on the decrease, the greater is the tendency for the earlier part to be pushed up by the birth of children. This is an interesting example of a natural "governor" mechanism.

Example 10 (with Dr W. O. Kermack). The problem of the course of an epidemic of a contagious disease is similar to the last. The variables are as before, the time (t) and the age (θ), and these are absolutely correlated (fig. 14). We have to deal with an infective rate $\phi(t, \theta)$, and a rate of removal (including death and recovery) $\psi(t. \theta)$. We shall suppose that these are independent of the time. Let us treat the problem in the first instance as if t and θ were discontinuous variables.

The equation is then

$$v_{t-1, \theta-1} - v_{t, \theta} = \psi(\theta) v_{t, \theta} \quad \dots\dots\dots\dots\dots(50)$$

whence

$$v_{t, \theta} = \frac{v_{t-1, \theta-1}}{1 + \psi(\theta)} = \frac{v_{t-2, \theta-2}}{(1 + \psi(\theta))(1 + \psi(\theta-1))} = v_{t-\theta, 0} B_\theta, \quad (51)$$

where B_θ is written for the reciprocal of

$$(1 + \psi(\theta))(1 + \psi(\theta-1)) \dots (1 + \psi(1)).$$

Now the number $v_{t, 0}$ denotes the number of persons who became infected at the time t, and this by hypothesis is equal to $\overset{t}{\underset{1}{\Sigma}}\phi_\theta v_{t, \theta}$, whence

$$v_{t, 0} = \overset{t}{\underset{1}{\Sigma}} \phi_\theta B_\theta v_{t-\theta, 0} = \overset{t}{\underset{1}{\Sigma}} A_\theta v_{t-\theta, 0} \dots\dots\dots\dots\dots(52)$$

where for conciseness A_θ is written for $\phi_\theta B_\theta$.

Hence $\sum\limits_{0}^{\infty} x^t v_{t,\,0} = \sum\limits_{0}^{\infty} x^t \sum\limits_{1}^{t} A_\theta\, v_{t-\theta,\,0} + N_0$ where $N_0 \equiv v_{00}$

$$= \sum\limits_{0}^{\infty} x^t\, v_{t,\,0} \sum\limits_{1}^{\infty} x^\theta\, A_\theta + N_0$$

x being chosen so that the series are convergent,

$$\therefore \quad \sum\limits_{0}^{\infty} x^t\, v_{t,\,0} = \frac{N_0}{1 - \sum\limits_{1}^{\infty} x^\theta\, A_\theta} \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(53)$$

Let N_t denote the number of persons infected at the time t,

then $\quad N_t = \sum\limits_{0}^{t} v_{t,\,\theta} = \sum\limits_{0}^{t} B_\theta\, v_{t-\theta,\,0}$

and $\quad \sum\limits_{0}^{\infty} x^t N_t = \sum\limits_{0}^{\infty} x^t \sum\limits_{0}^{t} B_\theta\, v_{t-\theta,\,0}$

$$= \sum\limits_{0}^{\infty} x^t v_{t,\,0} \sum\limits_{0}^{\infty} x^\theta\, B_\theta,$$

$$= \frac{N_0 \sum\limits_{0} x^\theta\, B_\theta}{1 - \sum\limits_{1}^{\infty} x^\theta\, A_\theta} \qquad \dots\dots\dots\dots\dots\dots\,\dots\dots\dots\dots\,\dots\dots\dots\dots(54)$$

$v_{t,\,0}$ and N_t are the coefficients of x^t in the expansions of the above expressions (53) and (54).

By cross multiplication, expansion and equating like powers of x it is easy to show also that

$$N_t = B_t N_0 + \sum\limits_{1}^{t} A_\theta\, N_{t-\theta} \qquad \dots\dots\dots\dots\dots\dots\dots(55)$$

Passing to the continuous form; from equation (53) we have

$$\int_0^\infty x^t v_{t,\,0}\, dt = \frac{N_0}{1 - \int_0^\infty x^\theta\, A_\theta\, d\theta}$$

$$\left(\text{where} \quad A_\theta = \phi_\theta\, B_\theta = \phi_\theta\, e^{-\int_0^\theta \psi(t)\,dt}\right)$$

or $\quad \int_0^\infty e^{-zt} v_{t,\,0}\, dt = \frac{N_0}{1 - \int_0^\infty e^{-z\theta}\, A_\theta\, d\theta} \qquad \dots\dots\dots\dots\dots\dots\dots(56)$

whence using Laplace's transformation

$$v_{t,\,0} = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} e^{zt}\int_0^\infty e^{-zt} v_{t,\,0}\,dt\,dz = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} \frac{N_0\,e^{tz}}{1-\int_0^\infty e^{-\theta z}A_\theta\,d\theta}\,dz\ \Bigg\}\ \ \ldots\ldots(.$$

or shortly $\quad v_{t,\,0} = L\left(\dfrac{N_0}{1-\int_0^\infty e^{-\theta z}A_\theta\,d\theta}\right)$

This then must under the conditions of the case be also a solution by equation (52) in its integral form,

$$v_{t,\,0} = \int_0^t A_\theta\,v_{t-\theta,\,0}\,d\theta\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots(58)$$

This solution appears to hold only when singularities exist in $v_{t,\,\theta}$. In the case which we are discussing there is a singularity at $t = 0$, since one assumes that there is an instantaneous infinite infection rate at the origin. If a singularity of similar type exists at a point $t = t_0$ then it can be shown that the solution is

$$v_{t,\,0} = L\left(\frac{N_0 e^{-zt_0}}{1-\int_0^\infty e^{-\theta z}A_\theta\,d\theta}\right)\ \ldots\ldots\ldots\ldots\ldots(59)$$

Similarly

$$N_t = L\left(\frac{N_0\int_0^\infty e^{-\theta z}B_\theta\,d\theta}{1-\int_0^\infty e^{-\theta z}A_\theta\,d\theta}\right)\ \ldots\ldots\ldots\ldots\ldots(60)$$

and this is by equation (55) a solution of the integral equation

$$N_t = B_t N_0 + \int_0^t A_\theta\,N_{t-\theta}\,d\theta,\ \ldots\ldots\ldots\ldots\ldots\ldots(61)$$

a result which has been obtained by Fock. [10]

Thus if the infection and removal rates are known and can be expressed as continuous functions, both $v_{t,\,0}$ and N_t can be obtained. In actual experience $\phi(\theta)$ and $\psi(\theta)$ are not available, but statistics regarding $v_{t,\,0}$ are readily accessible, and values of N_t could also be obtained. The above operations must therefore be reversed. Equation (56) may be written in the form

$$\int_0^\infty e^{-z\theta}A_\theta\,d\theta = 1 - \frac{N_0}{\int_0^\infty e^{-zt}v_{t,\,0}\,dt}$$

whence

$$\phi_\theta B_\theta = A_\theta = L \left(1 - \frac{N_0}{\displaystyle\int_0^\infty e^{-zt} v_{t,0}\, dt} \right) \quad \dots\dots\dots\dots(62)$$

where the exponential within the operator L is in this case $e^{z\theta}$.
Also from equation (54) in its continuous form, and from (56)

$$\int_0^\infty e^{-z\theta} B_\theta\, d\theta = \frac{\displaystyle\int_0^\infty e^{-zt} N_t\, dt}{\displaystyle\int_0^\infty e^{-zt} v_{t,0}\, dt} \quad \dots\dots\dots\dots\dots(63)$$

whence

$$B_\theta = L \left(\frac{\displaystyle\int_0^\infty e^{-zt} N_t\, dt}{\displaystyle\int_0^\infty e^{-zt} v_{t,0}\, dt} \right) \quad \dots\dots\dots\dots\dots(64)$$

Thus if N_t and $v_{t,0}$ can be expressed as known functions, the values of $\phi(\theta)$ and $\psi(\theta)$ may be deduced. For the numerical solution of the problem the reader is referred to Whittaker (*Proc. R.S.*, 94, p. 367, 1918). From which it will be seen that if a series of numerical values of $v_{t,0}$ are known the corresponding values of A_θ may be determined from equation (58) in the form

$$v_{t,0} = \int_0^t A_{t-\theta}\, v_{\theta,0}\, d\theta.$$

The values of B_θ may then be determined directly by means of equation (61)

Thus the values of the functions $\phi(\theta)$ and $\psi(\theta)$ may be calculated either formally or numerically from statistics giving the values of $v_{t,0}$ and N_t.

In the above we have considered the case in which there is no seasonal effect on the ϕs and the ψs. If that be not the case then ϕ and ψ are both functions of t and θ. If $\phi(t, \theta) = \alpha_t \beta_\theta$, and $\psi(t, \theta) = \alpha'_t \beta'_\theta$, then the above method of treatment can be applied, but not otherwise. In many instances the product form may reasonably be used. For example, if α_t depends upon the number of mosquitos, and the disease is malaria, then it would be

reasonable to suppose that the effect of a multiplication of the number of mosquitos would be to multiply the chances of infection at the various case ages θ.

These results may obviously be applied to the previous example if the birth rate can be expressed in the form $\phi(\theta)$.

8. *Vorticity in statistical problems.*

Returning to equations (16) which also hold when x and y are continuous, and also if the constants be suitably modified when the progression is reversible, and omitting the R coefficients we have for the equations of the coordinates of the centre of gravity of the moving system

$$\frac{d\mu_x}{dm} = P_1 + P_2 \mu_x + P_3 \mu_y$$

$$\frac{d\mu_y}{dm} = Q_1 + Q_2 \mu_x + Q_3 \mu_y$$

whence

$$\frac{d^2 \mu_x}{dm^2} - (P_2 + Q_3) \frac{d\mu_x}{dm} + (P_2 Q_3 - P_3 Q_2) \mu_x = P_3 Q_1 - P_1 Q_3$$

and

$$\frac{d^2 \mu_y}{dm^2} - (P_2 + Q_3) \frac{d\mu_y}{dm} + (P_2 Q_3 - P_3 Q_2) \mu_y = Q_2 P_1 - Q_1 P_2$$

The motion is periodic if

$$- 4 P_3 Q_2 > (P_2 - Q_3)^2$$

therefore periodicity depends primarily upon the condition that one, but not both, of P_3 and Q_2 must be negative.

Again if $(P_2 + Q_3)$ is negative the motion will be damped, whilst if $(P_2 + Q_3)$ is positive there will be augmentation of amplitude.

Example 11. During the course of a microbic disease a battle is raging between microbes on the one hand and protective "antibodies" on the other. These antibodies are themselves called into being by the action of the microbes. The process is reversible for both variables ($x =$ microbes and $y =$ antibodies).

Now microbes multiply by compound interest, hence P_2 is positive ($= a_2$), and are destroyed at a rate proportional to the

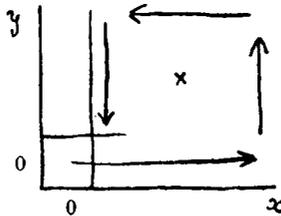number of antibodies present, hence P_3 is negative $(= -a_3)$. The state of affairs is as shown in figure 17.

Fig. 17

Antibodies are produced at a rate proportional to the number of microbes hence Q_2 is positive $(= b_2)$, and are destroyed in the general metabolism, in such a manner that the more the antibody the more the destruction, hence Q_3 is negative $(= -b_3)$.

Hence if $4a_3 b_2 > (a_2 + b_3)^2$ a vortex is formed, and if $a_2 - b_3 > 0$ there will be augmentation of amplitudes, whilst if $b_3 - a_2 > 0$ there will be damping.

Also when $t = \infty$

$$\mu_x = \frac{P_3 Q_1 - P_1 Q_3}{P_2 Q_3 - P_3 Q_2} = \frac{P_1 b_3 - Q_1 a_3}{a_3 b_2 - a_2 b_3}$$

and

$$\mu_y = \frac{P_1 Q_2 - P_2 Q_1}{P_2 Q_3 - P_3 Q_2} = \frac{P_1 b_2 - Q_1 a_2}{a_3 b_2 - a_2 b_3}.$$

From these considerations arrived at from the approximate equations, we can deduce at once—

(1) Augmentation of amplitude $(a_2 > b_3)$ will lead to total and relatively sudden extinction of organisms at the end of an attack (or series of attacks) for the motion will enter the column $x = 0$.

(2) Damping $(b_3 > a_2)$ will lead to a final state at which both organism and antibodies will continue in a steady state, their relative proportions depending upon the above relations.

(3) Relapses will occur when $4a_3 b_2 > (a_2 + b_3)^2$. They will seldom, if ever, occur where amplitudes augment $(a_2 > b_3)$ they will be a prominent feature as a_2 and b_3 tend to equality, and they will be absent when b_3 is very much greater than a_2, as in this case the motion becomes a-periodic.

This is only a rough approximation but it places in evidence the main features of infectious diseases, viz, (a) termination by "crisis" and complete extinction of the disease, or (b) gradual decline by "lysis," with continued "carriage" of the disease. It draws attention also to (c) the occurrence of relapses in the intermediate types. This latter consideration is exemplified in Malta Fever, in which there may occur a series of relapses, and in which both types ($a_2 > b_3$) and ($b_3 > a_2$) are found to occur.

A better approximation is obtained by introducing a third variable, the temperature. The differential equations for the coordinates of the centre of gravity of the system are then of the third order, and more in accordance with experience (see also reference 11).

In conclusion the author desires to thank Dr W. O. Kermack for continued help and criticism, and Mr E T. Copson for encouragement and advice in many directions.

---

1. BEHLA.   Cent. f. Bact. (orig.) 24, Abt. 1.   1898.

2. M'KENDRICK, *Science Progress.*   Jan. 1914.

3. ,, *Proc. Lond. Math. Soc.*, Ser. 2, Vol. 13, p. 401.   1914.

4. ,, *Indian Journal of Medical Research*, Vol. 2, p. 882.   1915.

5. ,,   ,,   ,,   ,,   ,,   ,,   Vol. 3, p. 271.   1915.

6. ,,   ,,   ,,   ,,   ,,   ,,   Vol. 3, p. 266.   1915.

7. ,,   ,,   ,,   ,,   ,,   ,,   Vol. 3, p. 667.   1916.

8. ,,   *Nature*, Vol. 104, p. 660.   1920.

9. *Commonwealth of Australia Quarantine Service, Publication* 18.   1919

10. FOCK, *Math. Zeit*, 21, 161.   1924.

11. M'KENDRICK, *Indian Journal of Medical Research*, Vol. 6, p. 614. 1919,