THE CALCULATION OF MOMENTS OF A FREQUENCY-DISTRIBUTION.

By W. F. SHEPPARD.

[*Note.* This paper deals mainly with simplification of method. The results obtained in §§ 8 and 11 are new.]

1. Let the range of observed values of x, the measure whose frequency is under consideration, be from $x_1 - \frac{1}{2}h$ to $x_n + \frac{1}{2}h$, this range being divided into n equal segments h, the values of x at whose middle points are $x_1, x_2, \ldots x_r, \ldots x_n$; and let the areas standing on these segments be $A_1, A_2, \ldots A_r, \ldots A_n$, the sum of these areas, which is the total frequency-area, being 1. Then, if

$$z = f(x) \qquad \ldots\ldots (1)$$

is taken to be the equation to the curve of frequency, we have

$$A_r = \int_{x_r - \frac{1}{2}h}^{x_r + \frac{1}{2}h} f(x)\,dx = \int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x_r + \theta)\,d\theta \qquad \ldots\ldots (2).$$

It is required to find a method for calculating

$$\nu_p \equiv \int_{x_1 - \frac{1}{2}h}^{x_n + \frac{1}{2}h} x^p f(x)\,dx \qquad \ldots\ldots (3),$$

which is the pth moment of the frequency-area about the axis of z; or, more generally, for calculating

$$\int_{x_1 - \frac{1}{2}h}^{x_n + \frac{1}{2}h} \phi(x) f(x)\,dx \qquad \ldots\ldots (4),$$

where $\phi(x)$ is a function of x which either may be given explicitly or may have its values tabulated for a series of values of x.

2. The most general method involves the use of a quadrature-formula. If we denote by I the area of the frequency-curve up to the ordinate z, so that

$$I = \int_{x_1 - \frac{1}{2}h}^{x} f(x)\,dx \qquad \ldots\ldots (5),$$

we have

$$f(x) = \frac{dI}{dx} \qquad \ldots\ldots (6),$$

and therefore

$$\int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h} \phi(x) f(x)\, dx$$

$$= \int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h} \phi(x) \frac{dI}{dx}\, dx$$

$$= \Big[I\phi(x) \Big]_{x=x_1-\frac{1}{2}h}^{x=x_n+\frac{1}{2}h} - \int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h} I\phi'(x)\, dx$$

$$= \phi(x_n+\tfrac{1}{2}h) - \int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h} I\phi'(x)\, dx \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(7).$$

The expression which has here to be subtracted from $\phi(x_n+\frac{1}{2}h)$ is the area, from $x=x_1-\frac{1}{2}h$ to $x=x_n+\frac{1}{2}h$, of a curve whose ordinate at any point is $I\phi'(x)$. The ordinates at successive distances h, commencing and ending with the extreme ordinates, are

$$\left.\begin{array}{l} 0 \\ A_1\phi'(x_1+\frac{1}{2}h) \\ (A_1+A_2)\phi'(x_2+\frac{1}{2}h) \\ \vdots \\ (A_1+A_2+\ldots+A_{n-1})\phi'(x_{n-1}+\frac{1}{2}h) \\ \phi'(x_n+\frac{1}{2}h) \end{array}\right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots(8),$$

and the area can be expressed in terms of these ordinates by a quadrature-formula. If, for instance, we use the ordinary trapezoidal rule for bounding ordinates, we have (using the symbol $\doteq$ to denote approximate equality)

$$\int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h} \phi(x) f(x)\, dx \doteq \phi(x_n+\tfrac{1}{2}h) - h\,\{A_1\phi'(x_1+\tfrac{1}{2}h) + (A_1+A_2)\phi'(x_2+\tfrac{1}{2}h)$$
$$+ \ldots + (A_1+A_2+\ldots+A_{n-1})\phi'(x_{n-1}+\tfrac{1}{2}h) + \tfrac{1}{2}\phi'(x_n+\tfrac{1}{2}h)\} \ldots(9).$$

For the pth moment, $\phi(x) \equiv x^p$, and therefore

$$\nu_p = \int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h} x^p f(x)\, dx = (x_n+\tfrac{1}{2}h)^p - p\mathfrak{A} \quad \ldots\ldots\ldots\ldots\ldots\ldots(10),$$

where $\mathfrak{A}$ is the area of a curve whose ordinates at successive distances h, commencing and ending with the extreme ordinates, are

$$0,\ A_1(x_1+\tfrac{1}{2}h)^{p-1},\ (A_1+A_2)(x_2+\tfrac{1}{2}h)^{p-1},\ \ldots$$
$$\ldots (A_1+A_2+\ldots+A_{n-1})(x_{n-1}+\tfrac{1}{2}h)^{p-1},\ (x_n+\tfrac{1}{2}h)^{p-1}.$$

This area can be calculated by a quadrature-formula.

3. There is an alternative method, known as "correction of raw moments," which is applicable only to a "quasi-normal" curve, i.e., to a curve which at the extremities of the range is so close to the base $z=0$ that the ordinates from $f(x_1-h)$ to $f(x_1)$ and from $f(x_n)$ to $f(x_n+h)$ may be regarded as negligible. This method will be considered in the following paragraphs; exact and approximate formulae being distinguished by the use of the symbols $=$ and $\doteq$.

4. The *raw moments* are obtained by massing each of the areas $A_1, A_2, \dots A_n$ along its central ordinate, so that, if ρ_p denotes the raw pth moment,

$$\rho_p \equiv A_1 x_1^p + A_2 x_2^p + \dots + A_n x_n^p \quad\text{.....................(11)}.$$

Now the expression on the right-hand side of (11) is an approximate expression for the area from $x = x_1 - \frac{1}{2}h$ to $x = x_n + \frac{1}{2}h$ of a continuous curve whose ordinates at $x = x_1$, $x = x_2$, ... $x = x_n$ are $A_1 x_1^p/h$, $A_2 x_2^p/h$, ... $A_n x_n^p/h$, the ordinate at $x = x_r$ being

$$A_r x_r^p/h = x_r^p \cdot \frac{1}{h}\int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x_r + \theta)\, d\theta.$$

The condition of continuity obviously requires that this should be the expression for the ordinate, whatever the value of r may be. If therefore we write

$$R_p \equiv \int_{x_1 - \frac{1}{2}h}^{x_n + \frac{1}{2}h} x^p \cdot \frac{1}{h}\int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x + \theta)\, d\theta \,.\, dx \quad\text{..................(12)},$$

we have, approximately,

$$\rho_p \doteq R_p \quad\text{...................................(13)}.$$

Now A_r/h is equal to the average value of the ordinate $f(x)$ of the original curve for values of x between $x_r - \frac{1}{2}h$ and $x_r + \frac{1}{2}h$. We therefore take a new curve

$$z = F(x) \quad\text{...................................(14)},$$

related to the original curve in such a way that the ordinate of the new curve corresponding to $x = \xi$ is equal to the average value of the ordinate of the original curve for values of x between $\xi - \frac{1}{2}h$ and $\xi + \frac{1}{2}h$, so that

$$F(x) \equiv \frac{1}{h}\int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x + \theta)\, d\theta \quad\text{.......................(15)}.$$

This curve may be called the spurious curve of frequency, since its ordinates at successive distances h are proportional to the frequencies found by classification with intervals h. Also, by (12),

$$R_p = \int_{x_1 - \frac{1}{2}h}^{x_n + \frac{1}{2}h} x^p F(x)\, dx \quad\text{..........................(16)},$$

and therefore R_p is equal to the pth moment of the spurious curve, taken between the same limits as the original curve.

5. In building up the value of R_p as given by (12), any particular value of $f(x)/h$ is multiplied by the value of x^p corresponding to the central ordinate of every strip of breadth h in which $f(x)$ is included; i.e. it is multiplied by the pth power of the distance of $f(x)$ from every ordinate in the strip of breadth h whose central ordinate is the axis of z. Hence, subject to certain corrections in respect of the strips at the two ends,

$$\int x^p \cdot \frac{1}{h}\int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x + \theta)\, d\theta \,.\, dx = \int f(x) \cdot \frac{1}{h}\int_{-\frac{1}{2}h}^{\frac{1}{2}h} (x + \theta)^p\, d\theta \,.\, dx.$$

To obtain this result by more exact methods, suppose that the range of values of x is extended by at least $\frac{1}{2}h$ in each direction, and that a solid is generated by moving the frequency-area at right angles to itself through a distance $\frac{1}{2}h$ on each side. The section of this solid by a plane through the ordinate $f(x)$, at right angles to the central section, is a rectangle of height $f(x)$ and base h; and the volume of the strip comprised between two such planes, through ordinates $f(x)$ and $f(x+dx)$ of the central section, is $hf(x)\,dx$. If, however, we turn these planes round, so as to be inclined at an angle of 45° to the central section, the volume comprised between them will be $\int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x+\theta)\,d\theta\,.\,dx = hF(x)\,dx$. Hence, if we write

$$S_p \equiv \int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h} f(x)\,.\,\frac{1}{h}\int_{-\frac{1}{2}h}^{\frac{1}{2}h} (x+\theta)^p\,d\theta\,.\,dx \quad \ldots\ldots\ldots\ldots(17),$$

and compare it with

$$R_p \equiv \int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h} x^p\,.\,\frac{1}{h}\int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x+\theta)\,d\theta\,.\,dx \ldots\ldots\ldots\ldots(12),$$

we see that each of the expressions hS_p and hR_p represents the sum of all the values of $\xi^p dV$, where dV is an element of a certain portion of the solid, and ξ is its distance, measured parallel to the axis of x, from a plane through the axis of z inclined at an angle of 45° to the frequency-area, and that in each case the portion of the solid for which the summation is made is bounded by parallel planes through the ordinates $f(x_1-\frac{1}{2}h)$ and $f(x_n+\frac{1}{2}h)$ of the central section; but that in the case of hS_p these planes are at right angles to the central section, while in the case of hR_p they are inclined at an angle of 45° to it. The difference between hS_p and hR_p is therefore the difference between the values of $\Sigma\xi^p dV$ for certain portions of the solid standing on triangular bases at its extremities. In the case of a quasi-normal curve these values are negligible, provided p is not too great, and therefore hR_p and hS_p are approximately equal, i.e.

$$R_p \doteq S_p \ldots\ldots\ldots\ldots(18).$$

Hence, by (13),

$$\rho_p \doteq S_p \ldots\ldots\ldots\ldots(19).$$

6. Now expand $(x+\theta)^p$ in powers of θ, substitute in (17), and integrate. Then, substituting $\nu_1, \nu_2, \nu_3, \ldots$ for their values as given by (3),

$$S_p = \nu_p + \frac{p(p-1)}{2\,.\,3}(\tfrac{1}{2}h)^2\nu_{p-2} + \frac{p(p-1)(p-2)(p-3)}{2\,.\,3\,.\,4\,.\,5}(\tfrac{1}{2}h)^4\nu_{p-4} + \ldots \quad \ldots(20),$$

the series continuing till we reach ν_1 or ν_0; and therefore, by (19),

$$\rho_p \doteq \nu_p + \frac{p(p-1)}{2\,.\,3}(\tfrac{1}{2}h)^2\nu_{p-2} + \frac{p(p-1)(p-2)(p-3)}{2\,.\,3\,.\,4\,.\,5}(\tfrac{1}{2}h)^4\nu_{p-4} + \ldots \quad \ldots(21).$$

Writing $p = 1, 2, 3, \ldots$ successively in (21), and remembering that $\rho_0 = \nu_0 = 1$, we get a series of equations for determining $\nu_1, \nu_2, \nu_3, \ldots$ in terms of $\rho_1, \rho_2, \rho_3, \ldots$; and it will be found that these equations give

$$\left.\begin{aligned} &\nu_1 \doteq \rho_1 \\ &\nu_2 \doteq \rho_2 - \tfrac{1}{12}h^2 \\ &\nu_3 \doteq \rho_3 - \tfrac{1}{4}h^2\rho_1 \\ &\nu_4 \doteq \rho_4 - \tfrac{1}{2}h^2\rho_2 + \tfrac{7}{240}h^4 \\ &\nu_5 \doteq \rho_5 - \tfrac{5}{6}h^2\rho_3 + \tfrac{7}{48}h^4\rho_1 \\ &\;\vdots \qquad \vdots \end{aligned}\right\} \ldots\ldots\ldots\ldots(22).$$

The first of these results is of special importance, as showing that in the case of a quasi-normal curve the mean as given by the raw first moment is approximately the correct mean.

7. If we want the moments about the mean it is usually simplest to calculate them about some convenient axis and then transform to the mean. If N_1, N_2, N_3, ... are the moments, about any ordinate, of a number of frequencies (whether continuous or discontinuous) whose sum is unity, the pth moment about the mean is

$$M_p = N_p - pN_1N_{p-1} + \frac{p(p-1)}{1.2}N_1^2N_{p-2} - \ldots + (-)^{p-1}pN_1^{p-1}N_1 + (-)^pN_1^p \quad \ldots(23).$$

For calculating N_1, N_2, N_3, ... by the method of § 2 it is most convenient to take as axis an ordinate bounding one of the given areas A_r; but for the method of § 6 it is most convenient to take the mid-ordinate of one of these areas. When we have found the raw moments $\rho_1, \rho_2, \rho_3, \ldots$, we may either find the corrected moments $\nu_1, \nu_2, \nu_3, \ldots$ by (22), and then transform to the mean by (23); or we may find the raw moments about the mean by (23), and then correct these by (22). The latter is perhaps the simpler method in most cases. If we denote the raw moments about the mean, as found from the raw first moment, by $\pi_1, \pi_2, \pi_3, \ldots$ we have

$$\left.\begin{array}{l} \pi_1 = 0 \\ \pi_2 = \rho_2 - \rho_1^2 \\ \pi_3 = \rho_3 - 3\rho_1\rho_2 + 2\rho_1^3 \\ \pi_4 = \rho_4 - 4\rho_1\rho_3 + 6\rho_1^2\rho_2 - 3\rho_1^4 \\ \pi_5 = \rho_5 - 5\rho_1\rho_4 + 10\rho_1^2\rho_3 - 10\rho_1^3\rho_2 + 4\rho_1^5 \\ \vdots \quad \vdots \end{array}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots(24).$$

and the corrected moments about the true mean are

$$\left.\begin{array}{l} \mu_1 = 0 \\ \mu_2 \doteq \pi_2 - \frac{1}{12}h^2 \\ \mu_3 \doteq \pi_3 \\ \mu_4 \doteq \pi_4 - \frac{1}{2}h^2\pi_2 + \frac{7}{240}h^4 \\ \mu_5 \doteq \pi_5 - \frac{5}{6}h^2\pi_3 \\ \vdots \quad \vdots \end{array}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(25).$$

It is easy to verify that we get the same results by expressing $\nu_1, \nu_2, \nu_3, \ldots$ in terms of $\rho_1, \rho_2, \rho_3, \ldots$ by (22), and $\mu_2, \mu_3, \ldots$ in terms of $\nu_1, \nu_2, \nu_3, \ldots$ by (23).

8. In applying these results to the calculation of probable errors we must make allowance for the error introduced by the grouping of measurements into classes. Suppose, for instance, that there are n individuals, and that each value of x is measured to the nearest inch, so that $h = 1$ inch. If we knew the exact values of x, we could calculate the mean $\nu_1' \equiv \Sigma x/n$ of these values, and take this to be equal to ν_1. There would then be an error $\nu_1' - \nu_1$, and the mean square of error would be $\mu_2/n = (\nu_2 - \nu_1^2)/n$. Since, however, the measurements are only taken to the nearest multiple of h, the value of x as tabulated for each individual differs

from the true value of x for this individual by an error θ, which may have any value between $-\frac{1}{2}h$ and $+\frac{1}{2}h$; and the mean as calculated will not be ν_1' but $\nu_1' + \Sigma\theta/n$. The mean square of error of the mean as calculated will therefore be $(\nu_2 - \nu_1^2)/n + \frac{1}{12}h^2/n \doteq (\rho_2 - \rho_1^2)/n$; and the probable error will therefore be, not $\cdot 67449\sqrt{\mu_2/n}$, but $\cdot 67449\sqrt{\pi_2/n}$, where π_2 is the mean square of deviation from the mean for the spurious curve. This indeed is obvious from the fact that we are deducing the mean from the raw first moment ρ_1', so that the probable error is the probable error of ρ_1, which is $\cdot 67449\sqrt{\pi_2/n}$.

Similarly if we take μ_2 to be equal to $\pi_2' - \frac{1}{12}h^2$, where π_2' is the value of π_2 as calculated from the actual measurements, the error in μ_2, due to the limitation of number of observations, is equal to the error in π_2; and the mean square of this latter error is $(\pi_4 - \pi_2^2)/n$. Hence the probable error in the standard deviation will be, not $\cdot 67449\sqrt{\dfrac{\mu_4 - \mu_2^2}{4\mu_2}\Big/n}$, but $\cdot 67449\sqrt{\dfrac{\pi_4 - \pi_2^2}{4\mu_2}\Big/n}$.

It may be observed that, if we only want the mean and the standard deviation and their respective probable errors, we can conveniently calculate $\frac{1}{4}(\pi_4 - \pi_2^2)$ from the formula (obtained from (24))

$$\tfrac{1}{4}(\pi_4 - \pi_2^2) = \tfrac{1}{4}(\rho_4 - 4\rho_1\rho_3 + 3\rho_2^2) - \pi_2^2 \quad\text{.................(26).}$$

9. The formulae (22), giving the corrected moments in terms of the raw moments, have been obtained indirectly, by first expressing the raw moments in terms of the corrected moments. It would be desirable to obtain them directly.

The argument of § 5 is quite general, and we might replace x^p in (18) by $\psi(x)$, provided that the values of $\psi(x) f(x)$ are negligible between $x = x_1 - h$ and $x = x_1$, and between $x = x_n$ and $x = x_n + h$. We then have

$$\int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h} f(x)\,.\,\frac{1}{h}\int_{-\frac{1}{2}h}^{\frac{1}{2}h} \psi(x+\theta)\,d\theta\,.\,dx \doteq \int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h} \psi(x)\,.\,\frac{1}{h}\int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x+\theta)\,d\theta\,.\,dx \text{...(27).}$$

If therefore we can find a function $\psi_p(x)$ such that

$$\frac{1}{h}\int_{-\frac{1}{2}h}^{\frac{1}{2}h} \psi_p(x+\theta)\,d\theta \equiv x^p \quad\text{.........................(28),}$$

we shall have

$$\begin{aligned}\nu_p &= \int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h} x^p f(x)\,dx \\ &\doteq \int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h} \psi_p(x) F(x)\,dx\,. \\ &\doteq A_1\psi_p(x_1) + A_2\psi_p(x_2) + \dots + A_n\psi_p(x_n) \quad\text{.............. (29).}\end{aligned}$$

It is clear from (28) that $\psi_0(x) = 1$ and $\psi_1(x) = x$. Also, if we write $\psi_p(x) \equiv x^p + \chi_p(x)$, we find from (28) that $\frac{1}{h}\int_{-\frac{1}{2}h}^{\frac{1}{2}h} \chi_p(x+\theta)\,d\theta$ is of degree $p - 2$ in x. Hence, repeating the process, we find that $\psi_p(x)$ is a rational integral function of x of degree p, containing only terms in x^p, x^{p-2}, x^{p-4}, ... We may therefore write

$$\psi_p(x) \equiv x^p + k_1 p(p-1)(\tfrac{1}{2}h)^2 x^{p-2} + k_2 p(p-1)(p-2)(p-3)(\tfrac{1}{2}h)^4 x^{p-4} + \dots \quad \text{(30)}$$

$$= \left\{1 + k_1(\tfrac{1}{2}h)^2 \frac{d^2}{dx^2} + k_2(\tfrac{1}{2}h)^4 \frac{d^4}{dx^4} + \dots\right\} x^p \quad \text{(30 A)},$$

where k_1, k_2, ... are coefficients to be determined. Now substitute from (30 A) in (28), and integrate. Then since

$$\frac{1}{h}\int_{-\frac{1}{2}h}^{\frac{1}{2}h} (x+\theta)^q \, d\theta = x^q + \frac{q(q-1)}{2\,.\,3}(\tfrac{1}{2}h)^2 x^{q-2} + \frac{q(q-1)(q-2)(q-3)}{2\,.\,3\,.\,4\,.\,5}(\tfrac{1}{2}h)^4 x^{q-4} + \dots$$

$$= \left\{1 + \frac{1}{3\,!}(\tfrac{1}{2}h)^2 \frac{d^2}{dx^2} + \frac{1}{5\,!}(\tfrac{1}{2}h)^4 \frac{d^2}{dx^2} + \dots\right\} x^q,$$

we have

$$\left\{1 + \frac{1}{3\,!}(\tfrac{1}{2}h)^2 \frac{d^2}{dx^2} + \frac{1}{5!}(\tfrac{1}{2}h)^4 \frac{d^4}{dx^4} + \dots\right\}\left\{1 + k_1(\tfrac{1}{2}h)^2 \frac{d^2}{dx^2} + k_2(\tfrac{1}{2}h)^4 \frac{d^4}{dx^4} + \dots\right\} x^p = x^p,$$

and therefore the coefficients k_1, k_2, ... are such as to make

$$\left(1 + \frac{\phi^2}{3!} + \frac{\phi^4}{5!} + \dots\right)(1 + k_1\phi^2 + k_2\phi^4 + \dots) = 1 \quad \text{(31)}$$

identically. This gives

$$1 + k_1\phi^2 + k_2\phi^4 + \dots = 1 \div \left(1 + \frac{\phi^2}{3!} + \frac{\phi^4}{5!} + \dots\right)$$

$$= 1 \div \frac{\sinh\phi}{\phi}$$

$$= \frac{\phi}{\sinh\phi}$$

$$= 1 - \frac{P_1}{2!}\phi^2 + \frac{P_2}{4!}\phi^4 - \dots \quad \text{(32)},$$

where, if B_1, B_2, B_3, ... are Bernoulli's numbers,

$$P_r \equiv (2^{2r} - 2)B_r \quad \text{(33)},$$

the first few values being

$$P_1 = \tfrac{1}{3}, \quad P_2 = \tfrac{7}{15}, \quad P_3 = \tfrac{31}{21}, \quad P_4 = \tfrac{127}{15}, \dots \quad \text{(34)}.$$

Equating coefficients in (32), and substituting in (30 A) and (30),

$$\psi_p(x) = \left\{1 - \frac{P_1}{2!}(\tfrac{1}{2}h)^2 \frac{d^2}{dx^2} + \frac{P_2}{4!}(\tfrac{1}{2}h)^4 \frac{d^4}{dx^4} - \dots\right\} x^p \quad \text{(35)}$$

$$= x^p - \tfrac{1}{12}\frac{p(p-1)}{1\,.\,2} h^2 x^{p-2} + \tfrac{7}{240}\frac{p(p-1)(p-2)(p-3)}{1\,.\,2\,.\,3\,.\,4} h^4 x^{p-4} - \dots \quad \text{(36)}.$$

This result might have been obtained at once from (28) by a more liberal use of symbolic methods, since, if we write $D \equiv d/dx$, this gives

$$\frac{\sinh \frac{1}{2}hD}{\frac{1}{2}hD}\psi_p(x) \equiv x^p,$$

and therefore

$$\psi_p(x) \equiv \frac{\frac{1}{2}hD}{\sinh \frac{1}{2}hD} x^p$$

$$\equiv \left\{1 - \frac{P_1}{2!}(\tfrac{1}{2}hD)^2 + \frac{P_2}{4!}(\tfrac{1}{2}hD)^4 - \dots\right\} x^p,$$

which is identical with (35).

Substituting from (36) in (29), we find

$$\nu_p \doteq \Sigma A_r \psi_p(x_r)$$

$$\doteq \Sigma A_r \left\{x_r^p - \tfrac{1}{12}\frac{p(p-1)}{1.2} h^2 x_r^{p-2} + \tfrac{7}{240}\frac{p(p-1)(p-2)(p-3)}{1.2.3.4} h^4 x_r^{p-4} - \dots\right\}$$

$$\doteq \rho_p - \tfrac{1}{12}\frac{p(p-1)}{1.2} h^2 \rho_{p-2} + \tfrac{7}{240}\frac{p(p-1)(p-2)(p-3)}{1.2.3.4} h^4 \rho_{p-4} - \dots \quad \text{......(37)},$$

as the general formula of which those in (22) are particular cases.

10. The preceding results may be extended to any function $\phi(x)$ which can be expressed as a series in positive integral powers of x. Writing

$$\chi(x) \equiv \left\{1 - \frac{P_1}{2!}(\tfrac{1}{2}h)^2 \frac{d^2}{dx^2} + \frac{P_2}{4!}(\tfrac{1}{2}h)^4 \frac{d^4}{dx^4} - \dots\right\} \phi(x) \quad \text{............(38)}$$

$$\equiv \phi(x) - \frac{P_1}{2!}(\tfrac{1}{2}h)^2 \phi''(x) + \frac{P_2}{4!}(\tfrac{1}{2}h)^4 \phi^{\text{iv}}(x) - \dots \quad \text{......(38 A)},$$

we have

$$\frac{1}{h}\int_{-\frac{1}{2}h}^{\frac{1}{2}h} \chi(x+\theta)\, d\theta = \phi(x) \quad \text{..........................(39)},$$

and therefore, by (27),

$$\int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h} \phi(x) f(x)\, dx \doteq \int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h} \chi(x) F(x)\, dx \quad \text{..................(40)}$$

$$\doteq \Sigma\chi(x_r) A_r \quad \text{..........................(40 A)},$$

provided that the values of $\phi(x)f(x)$, $\phi''(x)f(x)$, ... are negligible between $x = x_1 - h$ and $x = x_1$ and between $x = x_n$ and $x = x_n + h$.

To replace the approximate formula (40) by an exact formula, we must take account of the difference between the values of $\Sigma\chi(\xi)\, dVh$ as calculated for the two solids considered in § 5. It is easy to show that

$$\int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h} \phi(x) f(x)\, dx = \int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h} \chi(x) F(x)\, dx + T \quad \text{..............(41)},$$

where

$$T = \frac{1}{h}\int_{-\frac{1}{2}h}^{\frac{1}{2}h} \int_0^{\theta} \{\chi(x_n + \tfrac{1}{2}h + \phi) f(x_n + \tfrac{1}{2}h + \phi - \theta)$$

$$- \chi(x_1 - \tfrac{1}{2}h + \phi) f(x_1 - \tfrac{1}{2}h + \phi - \theta)\}\, d\phi \,.\, d\theta \quad \text{......(42)}.$$

Also to replace (40 A) by an exact formula, we have by the Euler-Maclaurin formula, adapted for summation of mid-ordinates,

$$\int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h} \chi(x) F(x)\, dx = \Sigma\chi(x_r) A_r + R \quad \text{..................(43)},$$

where

$$R = \left[\left\{\frac{P_1}{2!}(\tfrac{1}{2}h)^2\frac{d}{dx} - \frac{P_2}{4!}(\tfrac{1}{2}h)^4\frac{d^3}{dx^3} + \dots\right\}\chi(x)F(x)\right]_{x=x_1-\frac{1}{2}h}^{x=x_n+\frac{1}{2}h} \quad \text{......(44)},$$

and therefore

$$\int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h}\phi(x)f(x)\,dx = \Sigma\chi(x_r)A_r + R + T \quad \text{..................(45)}.$$

11. The cases in which we require the value of a quantity which can be expressed in the form $\int\phi(x)f(x)\,dx$, where $\phi(x)$ is not of the form x^p, are usually cases in which $\phi(x)$ is not given explicitly in terms of x, but its values are tabulated for a series of values of x. If, for instance, we wished to find the rate of mortality of a certain section of the population, whose age-distribution was given, we should take $\phi(x)$ to denote the rate of mortality at age x, and the above integral would then give the rate of mortality of a population the proportion of which between ages x and $x + dx$ was $f(x)\,dx$.

The formulae of § 10 apply to such cases, if the age-distribution is quasi-normal, provided that $\frac{1}{2}h$ is so small that, for every value of x_r from x_1 to x_n, $\phi(x_r + \theta)$ can be expanded in a series of powers of θ which is convergent between the limits $\theta = \pm\frac{1}{2}h$. This is practically equivalent to a condition that when the values of $\phi(x)$ are tabulated at intervals of h in x the successive differences converge with sufficient rapidity to enable interpolation-formulae to be used. When this condition is satisfied, we have from (40 A)

$$\int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h}\phi(x)f(x)\,dx \doteq \chi(x_1)A_1 + \chi(x_2)A_2 + \dots + \chi(x_n)A_n,$$

where $\chi(x)$ is given by (38 A). To adapt this for practical calculation, we must express $\chi(x_r)$ in terms of the tabulated values of $\phi(x)$. It is only necessary to consider two cases.

(i) Suppose that the series of values for which $\phi(x)$ is tabulated comprises the mid-values ... x_1, x_2, ... x_n, ... Then, confining ourselves to central-difference formulae, with the notation I have adopted elsewhere*, and using symbolical methods, we have

$$\delta = 2\sinh\tfrac{1}{2}hD,$$

and†

$$\chi(x) = \frac{\frac{1}{2}hD}{\sinh\frac{1}{2}hD}\phi(x) = (1 - \tfrac{1}{24}\delta^2 + \tfrac{3}{640}\delta^4 - \tfrac{5}{7168}\delta^6 + \dots)\,\phi(x) \quad \text{......(46)},$$

so that, denoting $\phi(x)$ by u,

$$\int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h}uf(x)\,dx \doteq \sum_{r=1}^{r=n}(u_r - \tfrac{1}{24}\delta^2u_r + \tfrac{3}{640}\delta^4u_r - \dots)A_r \quad \text{............(47)}.$$

* *Proc. Lond. Math. Soc.* Vol. XXXI. p. 449.

† *Ibid.* p. 465, formulae (75).

(ii) Suppose that the series of values for which $\phi(x)$ is tabulated comprises the determining values ... $x_1 - \frac{1}{2}h$, $x_1 + \frac{1}{2}h$, ... $x_n + \frac{1}{2}h$, ... Then, if

$$\mu = \cosh \tfrac{1}{2}hD,$$

we have*

$$\chi(x) = \frac{\frac{1}{2}hD}{\sinh \frac{1}{2}hD}\,\phi(x) = \mu\,(1 - \tfrac{1}{6}\delta^2 + \tfrac{1}{30}\delta^4 - \tfrac{1}{140}\delta^6 + \ldots)\,\phi(x) \quad \ldots\ldots(48),$$

so that

$$\int_{x-\frac{1}{2}h}^{x_n+\frac{1}{2}h} uf(x)\,dx \triangleq \sum_{r=1}^{r=n} \tfrac{1}{2}\{(u_{r-\frac{1}{2}} - \tfrac{1}{6}\delta^2 u_{r-\frac{1}{2}} + \tfrac{1}{30}\delta^4 u_{r-\frac{1}{2}} - \ldots)$$
$$+ (u_{r+\frac{1}{2}} - \tfrac{1}{6}\delta^2 u_{r+\frac{1}{2}} + \tfrac{1}{30}\delta^4 u_{r+\frac{1}{2}} - \ldots)\}\,A_r \ldots(49).$$

This last result may also be written

$$\int_{x_1-\frac{1}{2}h}^{x_n+\frac{1}{2}h} uf(x)\,dx \triangleq \sum_{r=0}^{r=n} \{u_{r+\frac{1}{2}} - \tfrac{1}{6}\delta^2 u_{r+\frac{1}{2}} + \tfrac{1}{30}\delta^4 u_{r+\frac{1}{2}} - \ldots\}\,\tfrac{1}{2}(A_r + A_{r+1}) \ldots(50),$$

each of the extreme areas A_0 and A_{n+1} being zero.

* *Ibid.* formulae (74).