Example :—Determine the intrinsic accuracy of an error curve of Type IV. and the efficiency of the method of moments in location and scaling.

Since

$$\overline{\phi''} = -\frac{\overline{r+1}\,\overline{r+2}\,\overline{r+4}}{\overline{r+4}^2+\nu^2},$$

$$\sigma_{\hat{m}}^2 = \frac{a^2}{n}\cdot\frac{\overline{r+4}^2+\nu^2}{\overline{r+1}\,\overline{r+2}\,\overline{r+4}};$$

and the intrinsic accuracy of the curve is

$$\frac{1}{a^2}\,\frac{\overline{r+1}\,\overline{r+2}\,\overline{r+4}}{\overline{r+4}^2+\nu^2};$$

but

$$\sigma^2_{m_\mu} = \frac{a^2}{n}\cdot\frac{r^2+\nu^2}{r^2\,\overline{r-1}},$$

therefore the efficiency of the method of moments in location is

$$\frac{r^2\,\overline{r-1}\,(\overline{r+4}^2+\nu^2)}{\overline{r+1}\,\overline{r+2}\,\overline{r+4}\,(r^2+\nu^2)}. \qquad (3)$$

When $\nu = 0$, we have for curves of Type VII. an efficiency of location

$$1 - \frac{6}{\overline{r+1}\,\overline{r+2}}.$$

The efficiency of location of these curves vanishes at $r = 1$, at which value the standard deviation becomes infinite. Although values down to -1 give admissible frequency curves, the conventional limit at which curves are reckoned as heterotypic is at $r = 7$. For this value the efficiency is

$$\frac{49}{132}\cdot\frac{121+\nu^2}{49+\nu^2},$$

which varies from 91·67 per cent. for the symmetrical Type VII. curve, to 37·12 per cent. when $\nu \to \infty$ and the curve to Type V.

Turning to the question of scaling, we find

$$\overline{\xi^2\phi''}-1 = -\frac{\overline{r+1}\,(2\,\overline{r+4}+\nu^2)}{\overline{r+4}^2+\nu^2},$$

whence

$$\overline{\xi^2\phi''}-1-\frac{\overline{\xi\phi''}^2}{\overline{\phi''}} = -\frac{2\,\overline{r+1}}{r+4},$$

and

$$\sigma_a^2 = \frac{a^2}{n}\cdot\frac{r+4}{2\,\overline{r+1}};$$

Click here to view part 1 of paper

the intrinsic accuracy of scaling is therefore independent of ν. Now for these curves

$$\beta_2 = \frac{3\,\overline{r-1}}{\overline{r-2}\;\overline{r-3}}\left(r+6-\frac{8r^2}{r^2+\nu^2}\right),$$

so that

$$\frac{\beta_2-1}{4} = \frac{r^3\,\overline{r-2}+\nu^2\left(r^2+10r-12\right)}{2\,\overline{r-2}\;\overline{r-3}\;\left(r^2+\nu^2\right)},$$

and

$$\sigma^2_{a_\mu} = \frac{a^2}{n}\cdot\frac{r^3\,\overline{r-2}+\nu^2\left(r^2+10r-12\right)}{2\,\overline{r-2}\;\overline{r-3}\;\left(r^2+\nu^2\right)}.$$

The efficiency of the method of moments for scaling is thus

$$\frac{\overline{r-2}\;\overline{r-3}\;\overline{r+4}\;\left(r^2+\nu^2\right)}{\overline{r+1}\;\left\{r^3\,\overline{r-2}+\nu^2\left(r^2+10r-12\right)\right\}}; \quad . \quad . \quad . \quad . \qquad (4)$$

when $\nu = 0$, we have for curves of Type VII. an efficiency of scaling

$$1-\frac{12}{r\,\overline{r+1}}.$$

The efficiency of the method of moments in scaling these curves vanishes at $r = 3$, where β_2 becomes infinite; for $r = 7$, the efficiency of scaling is

$$\frac{55}{2}\cdot\frac{49+\nu^2}{1715+107\nu^2},$$

varying in value from 78·57 per cent. for the symmetrical Type VII. curve, to 25·70 per cent. when $\nu \rightarrow \infty$ and the curve to Type V.

10. The Efficiency of the Method of Moments in fitting the Pearsonian Curves.

The Pearsonian group of skew curves are obtained as solutions of the equation

$$\frac{1}{y}\frac{dy}{dx} = \frac{-(x-m)}{a+bx+cx^2}; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (5)$$

algebraically these fall into two main classes,

$$df \propto \left(1+\frac{x}{a_1}\right)^{m_1}\left(1-\frac{x}{a_2}\right)^{m_2} dx$$

and

$$df \propto \left(1+\frac{x^2}{a^2}\right)^{-\frac{r+2}{2}} e^{-\nu\tan^{-1}\frac{x}{a}}\,dx,$$

according as the roots of the quadratic expression in (5) are real or imaginary.

The first of these forms may be rewritten

$$df \propto \left(1-\frac{x^2}{a^2}\right)^{-\frac{r+2}{2}} e^{-\nu \tanh^{-1}\frac{x}{a}}\, dx,$$

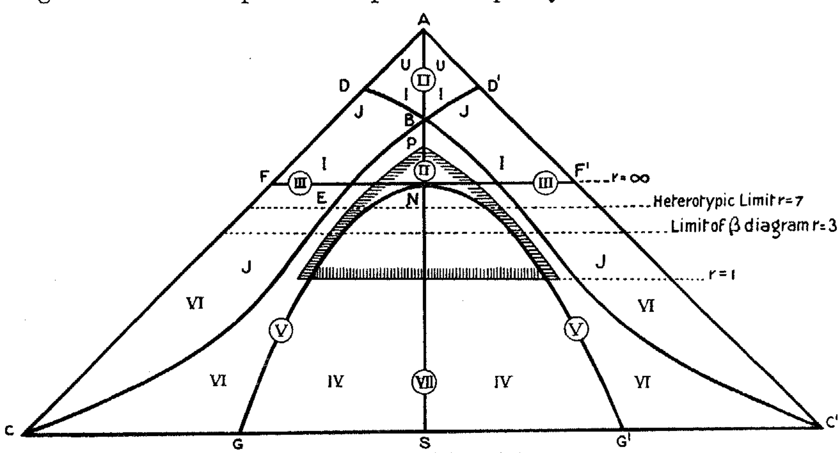
r being negative, showing its affinity with the second class.

In order that these expressions may represent frequency curves, it is necessary that the integral over the whole range of the curve should be finite ; this restriction acts in two ways :—

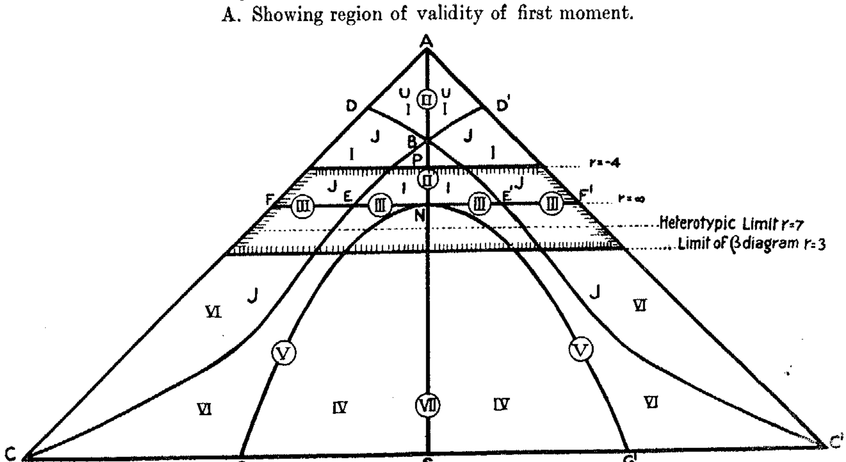
(1) When the curve terminates at a finite value of x, say $x = a_2$, the power to which $a_2 - x$ is raised must be greater than -1.

(2) When the curve extends to infinity, the ordinate, when x is large, must diminish more rapidly than $\frac{1}{x}$;

In Fig. 2 is shown a conspectus of all possible frequency curves of the Pearsonian type ;



A. Showing region of validity of first moment.



B. Showing region of validity of second moment.

Fig. 2. Conspectus of Pearsonian system of frequency curves.

the lines AC and AC′ represent the limits along which the area between the curve and a vertical ordinate tends to infinity, and on which m_1, or m_2, takes the value -1 ; the line CC′ represents the limit at which unbounded curves enclose an infinite area with the horizontal axis ; at this limit $r = -1$.

The symmetrical curves of Type II.

$$df \propto \left(1-\frac{x^2}{a^2}\right)^{-\frac{r+2}{2}}$$

extend from the point N, representing the normal curve, at which r is infinite, through the point P at which $r = -4$, and the curve is a *parabola*, to the point B ($r = -2$), where the curve takes the form of a *rectangle* ; from this point the curves are U-shaped, and at A, when the arms of U are hyperbolic, we have the limiting curve of this type, which is the discontinuous distribution of *equal or unequal dichotomy* ($r = 0$).

The unsymmetrical curves of Type I. are divided by PEARSON into three classes according as the terminal ordinate is infinite at neither end, at one end (J curves), or at both ends (U curves) ; the dividing lines are C′BD and CBD′, along which one of the terminal ordinates are finite (m_1, or m_2, $= 0$) ; at the point B, as we have seen, both terminal ordinates are finite.

The same line of division divides the curves of Type III.,

$$df \propto x^p e^{-x}\, dx,$$

at the point E ($p = 0$), representing a simple exponential curve ; the J curves of Type III. extend to F ($p = -1$), at which point the integral ceases to converge. In curves of Type III., r is infinite ; ν is also infinite, but one of the quantities m_1 and m_2 is finite, or zero ($= p$) ; as p tends to infinity we approach the normal curve

$$df \propto e^{-\frac{1}{2}x^2}\, dx.$$

Type VI., like Type III., consists of curves bounded only at one end ; here r is positive, and both m_1 and m_2 are finite or zero. For the J curves of Type VI. both m_1 and m_2 are negative, but for the remainder of these curves they are of opposite sign, the negative index being the greater by at least unity in order that the representative point may fall above CC′ ($r = -1$).

Type V. is here represented by a parabola separating the regions of Types IV. and VI. ; the typical equation of this type of curve is

$$df \propto x^{-\frac{r+3}{2}} e^{-\frac{1}{x}} dx.$$

As r tends to infinity the curve tends to the normal form ; the integral does not become divergent until $\frac{r+3}{2} = 1$, or $r = -1$. On curves of Type V., then, r is finite or zero, but ν is infinite.

In Type IV.

$$df \propto \left(1+\frac{x^2}{a^2}\right)^{-\frac{r+2}{2}} e^{-\nu \tan^{-1}\frac{x}{a}};$$

we have written ν, not as previously for the difference between m_1 and m_2, for these quantities are now complex, and their difference is a pure imaginary, but for the difference divided by $\sqrt{-1}$; ν is then real and finite throughout Type IV., and it vanishes along the line NS, representing the symmetrical curves of Type VII.

$$df \propto \left(1+\frac{x^2}{a^2}\right)^{-\frac{r+2}{2}}$$

from $r = \infty$ to $r = -1$.

The Pearsonian system of frequency curves has hitherto been represented by the diagram (13, p. 66), in which the co-ordinates are β_1 and β_2. This is an unsymmetrical diagram which, since β_1 is necessarily positive, places the symmetrical curves on a boundary, whereas they are the central types from which the unsymmetrical curves diverge on either hand; further, neither of the limiting conditions of these curves can be shown on the β diagram; the limit of the U curves is left obscure,* and the other limits are either projected to infinity, or, what is still more troublesome, the line at infinity cuts across the diagram, as occurs along the line $r = 3$, for there β_2 becomes infinite. This diagram thus excludes all curves of Types VII., IV., V., and VI., for which $r < 3$.

In the β diagram the condition r = constant yields a system of concurrent straight lines. The basis of the representation in fig. 2 lies in making these lines parallel and horizontal, so that the ordinate is a function of r only. We have chosen $r = y - \frac{1}{y}$, and have represented the limiting types by the simplest geometrical forms, straight lines and parabolas, by taking

$$4\epsilon = r^2 + \nu^2 = \frac{(\overline{1+y}^2 + x^2)(\overline{1-y}^2 - x^2)}{y(x^2+y)}.$$

It might have been thought that use could have been made of the criterion,

$$\kappa_2 = \frac{\beta_1(\beta_2+3)^2}{4(4\beta_2-3\beta_1)(2\beta_2-3\beta_1-6)} = 1 - \frac{r^2}{4\epsilon},$$

by which PEARSON distinguishes these curves; but this criterion is only valid in the region treated by PEARSON. For when $r = 0$, $\kappa_2 = 1$, and we should have to place a variety of curves of Types VII., IV., V., and VI., all in Type V. in order to adhere to the criterion.

This diagram gives, I believe, the simplest possible *conspectus* of the whole of the Pearsonian system of curves; the inclusion of the curves beyond $r = 3$ becomes neces-

* The true limit is the line $\beta_2 = \beta_1 + 1$, along which the curves degenerate into simple dichotomies.

sary as soon as we take a view unrestricted by the method of moments; of the so-called heterotypic curves between $r = 3$ and $r = 7$ it should be noticed that they not only fall into the ordinary Pearsonian types, but have finite values for the moment coefficients β_1 and β_2; they differ from those in which r exceeds 7, merely in the fact that the value of β_2, *calculated from the fourth moment of a sample*, has an infinite probable error. It is therefore evident that this is not the right method to treat the sample, but this does not constitute, as it has been called, "the failure of Type IV.," but merely the failure of the method of moments to make a valid estimate of the form of these curves. As we shall see in more detail, the method of moments, when its efficiency is tested, fails equally in other parts of the diagram.

In expression (3) we have found that the efficiency of the method of moments for location of a curve of Type IV. is

$$E = \frac{r^2\,\overline{r-1}\,(\overline{r+4}^2+\nu^2)}{\overline{r+1}\;\overline{r+2}\;\overline{r+4}\,(r^2+\nu^2)},$$

whence if we substitute for r and ν in terms of the co-ordinates of our diagram, we obtain a general formula for the efficiency of the method of moments in locating Pearsonian curves, which is applicable within the boundary of the zero contour (fig. 3). This may

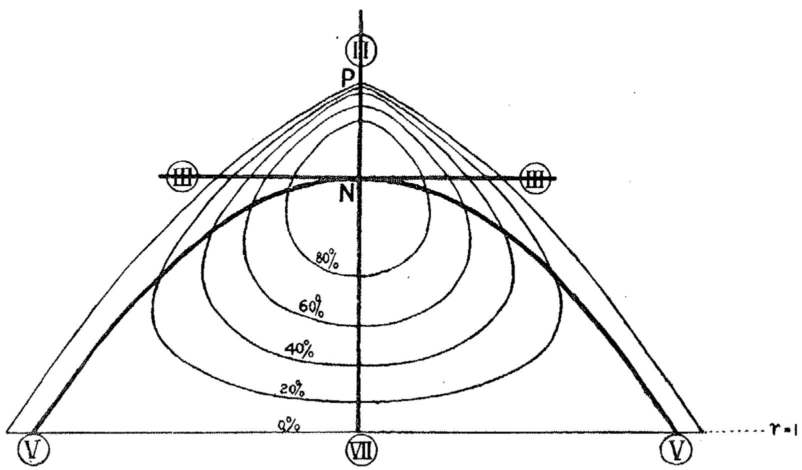


Fig. 3. Region of validity of the first moment (the mean) applied in the location of Pearsonian curves showing contours of efficiency.

be called the region of validity of the first moment; it is bounded at the base by the line $r = 1$, so that the first moment is valid far beyond the heterotypic limit; its other boundary, however, represents those curves which make a finite angle with the axis at the end of their range (m_1, or m_2, $= 1$); all J curves (m_1, or m_2, < 0) are thus excluded. This boundary has a double point at P, which thus forms the apex of the region of validity.

In fig. 3 are shown the contours along which the efficiency is 20, 40, 60, and 80 per cent. For high efficiencies these contours tend to the system of ellipses,

$$8x^2+6y^2 = 1-\mathrm{E}.$$

In a similar manner, we have obtained in expression (4) the efficiency of the second moment in fitting Pearsonian curves. The region of validity in this case is shown in fig. 4; this region is bounded by the lines $r = 3$, $r = -4$, and by the limits

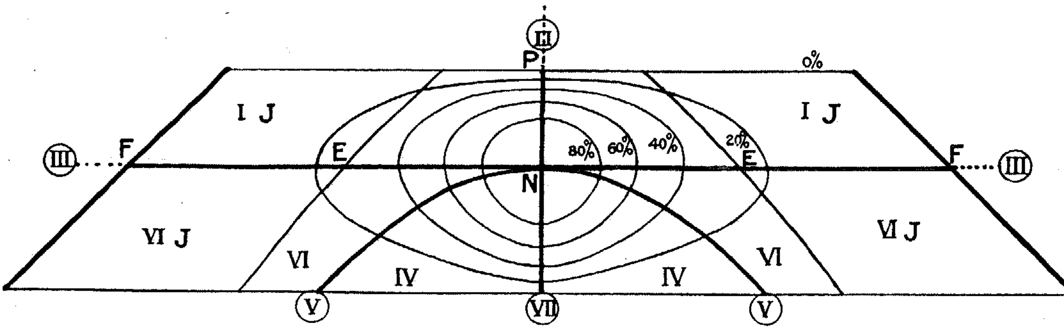


Fig. 4. Region of validity of the second moment (standard deviation) applied in scaling of Pearsonian curves, showing contours of efficiency.

(m_1, or m_2, $= -1$) on which $r^2+\nu^2$ vanishes. This statistic is therefore valid for certain J curves, though the maximum efficiency among the J curves is about 30 per cent. As before, the contours are centred about the normal curve (N) and for high efficiencies tend to the system of concentric circles,

$$12x^2+12y^2 = 1-\mathrm{E},$$

showing that the region of high efficiency is somewhat more restricted for the second moment, as compared to the first.

The lower boundary to the efficiencies of these statistics is due merely to their probable errors becoming infinite, a weakness of the method of moments which has been partially recognised by the exclusion of the so-called heterotypic curves ($r < 7$). The stringency of the upper boundary is much more unexpected; the probable errors of the moments do not here become infinite; only the ratio of the probable errors of the moments to the probable error of the corresponding optimum statistics is great and tends to infinity as the size of the sample is increased.

That this failure as regards location occurs when the curve makes a finite angle with the axis may be seen by considering the occurrence of observations near the terminus of the curve.

Let

$$df = kx^a\,dx$$

in the neighbourhood of the terminus, then the chance of an observation falling within a distance x of the terminus is

$$f = \frac{k}{\alpha+1} x^{\alpha+1} = k' x^{\alpha+1},$$

and the chance of n observations all failing to fall in this region is

$$(1-f)^n$$

or, when n is great, and f correspondingly small,

$$e^{-fn}.$$

Equating this to any finite probability, e^{-a}, we have

$$k' x^{\alpha+1} = \frac{a}{n},$$

or, in other words, if we use the extreme observation as a means of locating the terminus, the error, x, is proportional to

$$n^{-\frac{1}{\alpha+1}};$$

when $\alpha < 1$, this quantity diminishes more rapidly than $n^{-\frac{1}{2}}$, and consequently for large samples it is much more accurate to locate the curve by the extreme observation than by the mean.

Since it might be doubted whether such a simple method could really be more accurate than the process of finding the actual mean, we will take as example the location of the curve (B) in the form of a rectangle,

$$df = \frac{dx}{a}, \qquad m - \frac{a}{2} < x < m + \frac{a}{2},$$

and

$$df = 0,$$

outside these limits.

This is one of the simplest types of distribution, and we may readily obtain examples of it from mathematical tables. The mean of the distribution is m, and the standard deviation $\frac{a}{\sqrt{12}}$, the error $m_\mu - m$, of the mean obtained from n observations, when n is reasonably large, is therefore distributed according to the formula

$$\frac{1}{a}\sqrt{\frac{6n}{\pi}}\, e^{-\frac{6nx^2}{a^2}}\, dx.$$

The difference of the extreme observation from the end of the range is distributed according to the formula

$$\frac{n}{a} e^{-\frac{n\xi}{a}}\, d\xi;$$

if ξ is the difference at one end of the range and η the difference at the other end, the joint distribution (since, when n is considerable, these two quantities may be regarded as independent) is

$$\frac{n^2}{a^2} e^{-\frac{n}{a}\overline{\xi+\eta}}\, d\xi\, d\eta.$$

Now if we take the mean of the extreme observations of the sample, our error is

$$\tfrac{1}{2}\,\overline{\xi-\eta},$$

for which we write x; writing also y for $\xi+\eta$, we have the joint distribution of x and y,

$$\frac{n^2}{a^2} e^{-\frac{n}{a}y}\, dx\, dy.$$

For a given value of x the values of y range from $2|x|$ to ∞, whence, integrating with respect to y, we find the distribution of x to be

$$df = \frac{n}{a} e^{-\frac{2n}{a}|x|}\, dx,$$

the double exponential curve shown in fig. 5.

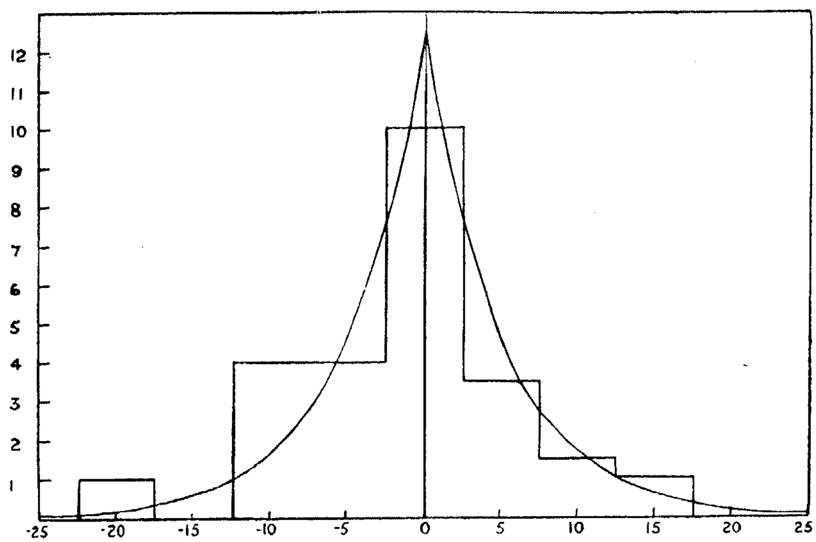


Fig. 5. Double exponential frequency curve, showing distribution of 25 deviations.

The two error curves are thus of a radically different form, and strictly no value for the efficiency can be calculated; if, however, we consider the ratio of the two standard deviations, then

$$\frac{\sigma_{\hat{m}}^2}{\sigma^2_{m_\mu}} = \frac{a^2}{2n^2} \div \frac{a^2}{12n} = \frac{6}{n}$$

when n is large, a quantity which diminishes indefinitely as the sample is increased.

For example, we have taken from VEGA (14) sets of digits from the table of Natural Logarithms to 48 places of decimals. The last block of four digits was taken from the logarithms of 100 consecutive numbers from 101 to 200, giving a sample of 100 numbers distributed evenly over a limited range. It is sufficient to take the three first digits to the nearest integer; then each number has an equal chance of all values between 0 and 1000. The true mean of the population is 500, and the standard deviation 289. The standard error of the mean of a sample of 100 is therefore 28·9.

Twenty-five such samples were taken, using the last five blocks of digits, for the logarithms of numbers from 101 to 600, and the mean determined merely from the highest and lowest number occurring, the following values were obtained :—

Digits.	1st hundred.			2nd hundred.			3rd hundred.			4th hundred.			5th hundred.		
	Lowest.	Highest.	$\hat{m}-m$.	Lowest.	Highest.	$\hat{m}-m$.	Lowest.	Highest.	$\hat{m}-m$.	Lowest.	Highest.	$\hat{m}-m$.	Lowest.	Highest.	$\hat{m}-m$.
45–48	24	978	+ 1·0	39	980	+ 9·5	1	999	0	16	983	− 0·5	18	994	+6·0
41–44	35·5	993	+14·0	3	960	−18·5	6	997	+1·5	1	978	−10·5	4	979	−8·5
37–40	9	988	− 1·5	11	999	+ 5·0	31	984	+7·5	4	978	− 9·0	2	986	−6·0
33–36	7	995	+ 1·0	13	997	+ 5·0	4	998	+1·0	0	994	− 3·0	3	981	−8·0
29–32	1	988	− 5·5	3	988	− 4·5	4	992	−2·0	1	996	− 1·5	21	977	−1·0

It will be seen that these errors rarely exceed one-half of the standard error of the mean of the sample. The actual mean square error of these 25 values is 6·86, while the calculated value, $\sqrt{50}$, is 7·07. It will therefore be seen that, with samples of only 100, there is no exaggeration in placing the efficiency of the method of moments as low as 6 per cent. in comparison with the more accurate method, which in this case happens to be far less laborious.

Such a value for the efficiency of the mean in this case is, however, purely conventional, since the curve of distribution is outside the region of its valid application, and the two curves of sampling do not tend to assume the same form. It is, however, convenient to have an estimate of the effectiveness of statistics for small samples, and in such cases we should prefer to treat the curve of distribution of the statistic as an error curve, and to judge the effectiveness of the statistic by the *intrinsic accuracy* of the curve as defined in Section 9. Thus the intrinsic accuracy of the curve of distribution of the mean of all the observations is

$$\frac{12n}{a^2},$$

while that of the mean of the extreme values is

$$\frac{4n^2}{a^2},$$

so yielding a ratio $3/n$. It is probable that this quantity may prove a suitable substitute for the efficiency of a statistic for curves beyond its region of validity.

To determine the efficiency of the moment coefficients β_1 and β_2 in determining the form of a Pearsonian curve, we must in general apply the method of Section 8 to the calculation of the simultaneous distribution of the four parameters of those curves when estimated by the method of maximum likelihood. Expressing the curve by the formula appropriate to Type IV., we are led to the determinant

$$\begin{vmatrix} \frac{\overline{r+1}\,\overline{r+2}\,\overline{r+4}}{a^2(\overline{r+4}^2+\nu^2)} & -\frac{\overline{r+1}\,\overline{r+2}\,\nu}{a^2(\overline{r+4}^2+\nu^2)} & \frac{\overline{r+1}\,\overline{r+2}}{a(\overline{r+2}^2+\nu^2)} & -\frac{\overline{r+1}\,\nu}{a(\overline{r+2}^2+\nu^2)} \\[2ex] -\frac{\overline{r+1}\,\overline{r+2}\,\nu}{a^2(\overline{r+4}^2+\nu^2)} & \frac{\overline{r+1}\,(\overline{2r+4}+\nu^2)}{a^2(\overline{r+4}^2+\nu^2)} & -\frac{\overline{r+1}\,\nu}{a(\overline{r+2}^2+\nu^2)} & \frac{r+2+\nu^2}{a(\overline{r+2}^2+\nu^2)} \\[2ex] \frac{\overline{r+1}\,\overline{r+2}}{a(\overline{r+2}^2+\nu^2)} & -\frac{\overline{r+1}\,\nu}{a(\overline{r+2}^2+\nu^2)} & \frac{\partial^2}{\partial\nu^2}\log F & \frac{\partial^2}{\partial\nu\,\partial r}\log F \\[2ex] -\frac{\overline{r+1}\,\nu}{a(\overline{r+2}^2+\nu^2)} & \frac{r+2+\nu^2}{a(\overline{r+2}^2+\nu^2)} & \frac{\partial^2}{\partial\nu\,\partial r}\log F & \frac{\partial^2}{\partial r^2}\log F \end{vmatrix}$$

as the Hessian of $-L$, when

$$F = e^{-\frac{1}{2}\nu\pi}\int_0^\pi e^{\nu\theta}\sin^r\theta\,d\theta.$$

The ratios of the minors of this determinant to the value of the determinant give the standard deviations and correlations of the optimum values of the four parameters obtained from a number of large samples.

In discussing the efficiency of the method of moments in respect of the *form* of the curve, it is doubtful if it be possible to isolate in a unique and natural manner, as we have done in respect of *location* and *scaling*, a series of parameters which shall successively represent different aspects of the process of curve fitting. Thus we might find the efficiencies with which r and ν are determined by the method of moments, or those of the parametric functions corresponding to β_1 and β_2, or we might use m_1 and m_2 as independent parameters of form; but in all these cases we should be employing an arbitrary pair of measures to indicate the relative magnitude of corresponding contour ellipses of the two frequency surfaces.

For the symmetrical series of curves, the Types II. and VII., the two systems of ellipses are coaxial, the deviations of r and ν being uncorrelated; in the case of Type VII. we put $\nu = 0$, in the determinant given above, which then becomes

$$\begin{vmatrix} \frac{\overline{r+1}\,\overline{r+2}}{r+4} & 0 & \frac{r+1}{r+2} & 0 \\ 0 & \frac{2\overline{r+1}}{r+4} & 0 & \frac{1}{r+2} \\ \frac{r+1}{r+2} & 0 & \frac{1}{2}F\left(\frac{r}{2}\right) & 0 \\ 0 & \frac{1}{r+2} & 0 & \frac{1}{4}\left\{F\left(\frac{r-1}{2}\right)-F\left(\frac{r}{2}\right)\right\} \end{vmatrix}$$

and falls in the two factors

$$\left[\frac{r+1}{2\,\overline{r+4}}\left\{F\left(\frac{r-1}{2}\right)-F\left(\frac{r}{2}\right)\right\}-\frac{1}{\overline{r-2}^2}\right]\left[\frac{\overline{r+1}\,\overline{r+2}}{2\,\overline{r+4}}F\left(\frac{r}{2}\right)-\frac{\overline{r+1}^2}{\overline{r+2}^2}\right],$$

so that

$$n\sigma_\nu^{\,2} = \frac{2\,\overline{r+2}^3}{\overline{r+2}^3\,F\left(\frac{r}{2}\right)-2\,\overline{r+1}\,\overline{r+4}},$$

and

$$n\sigma_r^{\,2} = \frac{4\,\overline{r+1}\,\overline{r+2}^2}{\overline{r+1}\,\overline{r+2}^2\left\{F\left(\frac{r-1}{2}\right)-F\left(\frac{r}{2}\right)\right\}-2\,\overline{r+4}}.$$

The corresponding expressions for the method of moments are

$$n\sigma_{\nu_\mu}^{\,2} = \frac{3}{8}\cdot\frac{r^2\,\overline{r-2}^2\,(r^2+r+10)}{\overline{r-1}\,\overline{r-3}\,\overline{r-5}},$$

and

$$n\sigma_{r_\mu}^{\,2} = \frac{2}{3}\cdot\frac{r\,\overline{r-1}^2\,\overline{r-3}\,(r^2-r+18)}{\overline{r-5}\,\overline{r-7}}.$$

Since for moderately large values of r, we have, approximately,

$$\overline{r+2}^3\,F\left(\frac{r}{2}\right)-2\,\overline{r+1}\,\overline{r+4} = \frac{16}{3}\left(1-\frac{1}{5\,\overline{r+2}^2}\right),$$

and

$$\overline{r+1}\;\overline{r+2}^2\left\{F\left(\frac{r-1}{2}\right)-F\left(\frac{r}{2}\right)\right\}-2\,\overline{r+1}\,\overline{r+4} = 6-\frac{4}{\overline{r+2}^2};$$

we have, approximately, for the efficiency of ν_μ,

$$\frac{\left(\overline{r+2}^3+\tfrac{1}{5}\,\overline{r+2}\ldots\right)\overline{r-1}\,\overline{r-3}\,\overline{r-5}}{(r^2+r+10)\,r^2\,\overline{r-2}^2},$$

or, when r is great,

$$1-\frac{28\cdot 8}{r^2};$$

and for the efficiency of r_μ,

$$\frac{\left(\overline{r+2}^2+\tfrac{2}{3}\ldots\right)\overline{r+1}^2\,\overline{r-5}\,\overline{r-7}}{(r^2-r+18)\,r\,\overline{r-1}^2\,\overline{r-3}};$$

or, when r is great,

$$1-\frac{53\cdot\dot{3}}{r^2}.$$

The following table gives the values of the transcendental quantities required, and the efficiency of the method of moments in estimating the value of ν and r from samples drawn from Type VII. distribution.

r.	$\overline{r+2}^2\, F\left(\frac{r}{2}\right) - 2\overline{r+1}\,\overline{r+4}$.	Efficiency of ν_μ.	$\overline{r+1}^2\,\overline{r+2}^2 \times\left\{F\left(\frac{r-1}{2}\right)-F\left(\frac{r}{2}\right)\right\} - 2\overline{r+1}\,\overline{r+4}$.	Efficiency of r_μ.
5	5·31271	0		
6	5·31736	0·2572		
7	5·32060	0·4338	5·9473	0
8	5·32296	0·5569	5·9574	0·1687
9	5·32472	0·6449	5·9649	0·3130
10	5·32607	0·7097	5·9706	0·4403
11	5·32713	0·7586	5·9750	0·5207
12	5·32797	0·7963	5·9787	0·5935
13	5·32866	0·8259	5·9810	0·6519
14	5·32919	0·8497	5·9839	0·6990
15			5·9853	0·7376
16			5·9870	0·7694
17			5·9883	0·7959
18			5·9895	0·8182

It will be seen that we do not attain to 80 per cent. efficiency in estimating the form of the curve until r is about 17·2, which corresponds to $\beta_2 = 3\cdot 42$. Even for symmetrical curves higher values of β_2 imply that the method of moments makes use of less than four-fifths of the information supplied by the sample.

On the other side of the normal point, among the Type II. curves, very similar formulæ apply. The fundamental Hessian is

$$\begin{vmatrix} \frac{\overline{r-1}\,\overline{r-2}}{r-4} & 0 & -\frac{r-1}{r-2} & 0 \\ 0 & \frac{2\,\overline{r-1}}{r-4} & 0 & -\frac{1}{r-2} \\ -\frac{r-1}{r-2} & 0 & \frac{1}{2}\digamma\left(\frac{r-2}{2}\right) & 0 \\ 0 & -\frac{1}{r-2} & 0 & \frac{1}{4}\left\{\digamma\left(\frac{r-2}{2}\right)-\digamma\left(\frac{r-1}{2}\right)\right\} \end{vmatrix}$$

where r is written for the positive quantity, $-r$, whence

$$n\sigma_p^{\,2} = \frac{2\,\overline{r-2}^3}{\overline{r-2}^3\,\digamma\left(\frac{r-2}{2}\right)-2\,\overline{r-1}\,\overline{r-4}}$$

and

$$n\sigma_r^{\,2} = \frac{4\,\overline{r-1}\,\overline{r-2}^2}{\overline{r-1}\,\overline{r-2}^2\left\{\digamma\left(\frac{r-2}{2}\right)-\digamma\left(\frac{r-1}{2}\right)\right\}-2\,\overline{r-4}}$$

Now since

$$\digamma\left(\frac{r-2}{2}\right) = \digamma\left(\frac{r-4}{2}\right)-\frac{4}{\overline{r-2}^2},$$

it follows that

$$\overline{r-2}^3\,\digamma\left(\frac{r-2}{2}\right)-2\,\overline{r-1}\,\overline{r-4} = \overline{r-2}^3\,\digamma\left(\frac{r-4}{2}\right)-2r\,\overline{r-3},$$

which is the same function of $r-4$ as

$$\overline{r+2}^3\,\digamma\left(\frac{r}{2}\right)-2\,\overline{r+1}\,\overline{r+4}$$

is of r.

In a similar manner

$$\overline{r+1}^2\,\overline{r-2}^2\left\{\digamma\left(\frac{r-2}{2}\right)-\digamma\left(\frac{r-1}{2}\right)\right\}-2\,\overline{r-1}\,\overline{r-4}$$
$$= \overline{r-1}^2\,\overline{r-2}^2\left\{\digamma\,\frac{r-4}{2}-\digamma\,\frac{r-3}{2}\right\}-2\,\overline{r-2}\,\overline{r+1},$$

which is the same function of $r-3$ as

$$\overline{r+1}^2\,\overline{r+2}^2\left\{\digamma\left(\frac{r-1}{2}\right)-\digamma\left(\frac{r}{2}\right)\right\}-2\,\overline{r+1}\,\overline{r+4}$$

is of r.

In all these functions and those of the following table, r must be substituted as a positive quantity, although it must not be forgotten that r changes sign as we pass from Type VII. to Type II., and we have hitherto adhered to the convention that r is to be taken positive for Type VII. and negative for Type II.

r.	$\overline{r-2}^3 \digamma\left(\frac{r-2}{2}\right) - \overline{2r-1}\,\overline{r-4}$.	Efficiency of ν_μ.	$\overline{r-1}^2\,\overline{r-2}^2 \times \left\{\digamma\left(\frac{r-2}{2}\right) - \digamma\left(\frac{r-1}{2}\right)\right\} - \overline{2r-1}\,\overline{r-4}$.	Efficiency of r_μ.
2	4	0	4	0
3	4·93480	0·0576	5·1595	0·0431
4	5·15947	0·2056	5·5648	0·1445
5	5·23966	0·3590	5·7410	0·2613
6	5·27578	0·4865	5·8305	0·3708
7	5·29472	0·5857	5·8813	0·4653
8	5·30576	0·6615	5·9126	0·5441
9	5·31271	0·7198	5·9331	0·6090
10	5·31736	0·7650	5·9473	0·6624
11	5·32060	0·8005	5·9574	0·7063
12	5·32296	0·8287	5·9649	0·7427
13	5·32472	0·8516	5·9706	0·7731
14	5·32607	0·8702	5·9750	0·7986
15			5·9787	0·8202

In both cases the region of validity is bounded by the rectangle, at the point B (fig. 2, p. 343). Efficiency of 80 per cent. is reached when r is about 14·1 ($\beta_2 = 2·65$). Thus for symmetrical curves of the Pearsonian type we may say that the method of moments has an efficiency of 80 per cent. or more, when β_2 lies between 2·65 and 3·42. The limits within which the values of the parameters obtained by moments cannot be greatly improved are thus much narrower than has been imagined.

11. The Reason for the Efficiency of the Method of Moments in a Small Region surrounding the Normal Curve.

We have seen that the method of moments applied in fitting Pearsonian curves has an efficiency exceeding 80 per cent. only in the restricted region for which β_2 lies between the limits 2·65 and 3·42, and as we have seen in Section 8, for which β_1 does not exceed 0·1. The contours of equal efficiency are nearly circular or elliptical within these limits, if the curves are represented as in fig. 2, p. 343, and are ultimately centred round the normal point, at which point the efficiencies of all parameters tend to 100 per cent. It was, of course, to be expected that the first two moments would have 100 per cent. efficiencies at this point, for they happen to be the optimum statistics for fitting the normal curve. That the moment coefficients β_1 and β_2 also tend to 100 per cent. efficiency in this region suggests that in the immediate neighbourhood of the normal curve the departures from normality specified by the Pearsonian formula agree with those of that system of curves for which the method of moments gives the solution of maximum likelihood.

The system of curves for which the method of moments is the best method of fitting may easily be deduced, for if the frequency in the range dx be

$$y\,(x, \theta_1, \theta_2, \theta_3, \theta_4)\,dx,$$

then

$$\frac{\partial}{\partial\theta}\log y$$

must involve x only as polynomials up to the fourth degree; consequently

$$y = e^{-a^2(x^4+p_1x^3+p_2x^2+p_3x+p_4)},$$

the convergence of the probability integral requiring that the coefficient of x^4 should be negative, and the five quantities a, p_1, p_2, p_3, p_4 being connected by a single relation representing the fact that the total probability is unity.

Typically these curves are bimodal, and except in the neighbourhood of the normal point are of a very different character from the Pearsonian curves. Near this point, however, they may be shown to agree with the Pearsonian type; for let

$$y = Ce^{-\frac{x^2}{2\sigma^2}+k_1\frac{x^3}{\sigma^3}+k_2\frac{x^4}{\sigma^4}}$$

represent a curve of the quartic exponent, sufficiently near to the normal curve for the squares of k_1 and k_2 to be neglected, then

$$\frac{d}{dx}\log y = -\frac{x}{\sigma^2}\left(1-3k_1\frac{x}{\sigma}-4k_2\frac{x^2}{\sigma^2}\right)$$
$$= -\frac{x}{\sigma^2\left(1+3k_1\dfrac{x}{\sigma}+4k_2\dfrac{x^2}{\sigma^2}\right)},$$

neglecting powers of k_1 and k_2. Since the only terms in the denominator constitute a quadratic in x, the curve satisfies the fundamental equation of the Pearsonian type of curves. In the neighbourhood of the normal point, therefore, the Pearsonian curves are equivalent to curves of the quartic exponent; it is to this that the efficiency of μ_3 and μ_4, in the neighbourhood of the normal curve, is to be ascribed.

12. Discontinuous Distributions.

The applications hitherto made of the optimum statistics have been problems in which the data are ungrouped, or at least in which the grouping intervals are so small as not to disturb the values of the derived statistics. By grouping, these continuous distributions are reduced to discontinuous distributions, and in an exact discussion must be treated as such.

If p_s be the probability of an observation falling in the cell (s), p_s being a function of the required parameters $\theta_1, \theta_2 \ldots$; and in a sample of N, if n_s are found to fall into that cell, then

$$\mathrm{S}(\log f) = \mathrm{S}(n_s \log p_s).$$

If now we write $\bar{n}_s = p_s \mathrm{N}$, we may conveniently put

$$\mathrm{L} = \mathrm{S}\left(n_s \log \frac{n_s}{\bar{n}_s}\right),$$

where L differs by a constant only from the logarithm of the likelihood, with sign reversed, and therefore the method of the optimum will consist in finding the *minimum* value of L. The equations so found are of the form

$$\frac{\partial \mathrm{L}}{\partial \theta} = -\mathrm{S}\left(\frac{n_s}{\bar{n}_s}\frac{\partial \bar{n}_s}{\partial \theta}\right) = 0. \qquad (6)$$

It is of interest to compare these formulæ with those obtained by making the Pearsonian χ^2 a minimum.

For

$$\chi^2 = \mathrm{S}\frac{(n_s - \bar{n}_s)^2}{\bar{n}_s},$$

and therefore

$$1 + \chi^2 = \mathrm{S}\left(\frac{n_s^2}{\bar{n}_s}\right),$$

so that on differentiating by $d\theta$, the condition that χ^2 should be a minimum for variations of θ is

$$-\mathrm{S}\left(\frac{n_s^2}{\bar{n}_s^2}\frac{\partial \bar{n}_s}{\partial \theta}\right) = 0. \qquad (7)$$

Equation (7) has actually been used (12) to "improve" the values obtained by the method of moments, even in cases of normal distribution, and the Poisson series, where the method of moments gives a strictly sufficient solution. The discrepancy between these two methods arises from the fact that χ^2 is itself an approximation, applicable only when $\bar{n}_s$ and n_s are large, and the difference between them of a lower order of magnitude. In such cases

$$\mathrm{L} = \mathrm{S}\left(n_s \log \frac{n_s}{\bar{n}_s}\right) = \mathrm{S}\left(\overline{m+x} \log \frac{m+x}{m}\right) = \mathrm{S}\left\{x + \frac{x^2}{2m} - \frac{x^3}{6m^2} \ldots\right\},$$

and since

$$\mathrm{S}(x) = 0,$$

we have, when x is in all cases small compared to m,

$$L = \tfrac{1}{2}S\left(\frac{x^2}{m}\right) = \tfrac{1}{2}\chi^2$$

as a first approximation. In those cases, therefore, when χ^2 is a valid measure of the departure of the sample from expectation, it is equal to 2L; in other cases the approximation fails and L itself must be used.

The failure of equation (7) in the general problem of finding the best values for the parameters may also be seen by considering cases of fine grouping, in which the majority of observations are separated into units. For the formula in equation (6) is equivalent to

$$S\left(\frac{1}{\bar{n}_s}\frac{\partial \bar{n}_s}{\partial \theta}\right)$$

where the summation is taken over all the observations, while the formula of equation (7), since it involves n_s^2, changes its value discontinuously, when one observation is gradually increased, at the point where it happens to coincide with a second observation.

Logically it would seem to be a necessity that that population which is chosen in fitting a hypothetical population to data should also appear the best when tested for its goodness of fit. The method of the optimum secures this agreement, and at the same time provides an extension of the process of testing goodness of fit, to those cases for which the χ^2 test is invalid.

The practical value of χ^2 lies in the fact that when the conditions are satisfied in order that it shall closely approximate to 2L, it is possible to give a general formula for its distribution, so that it is possible to calculate the probability, P, that in a random sample from the population considered, a worse fit should be obtained; in such cases χ^2 is distributed in a curve of the Pearsonian Type III.,

$$df \propto \left(\frac{\chi^2}{2}\right)^{\frac{n'-3}{2}} e^{-\frac{\chi^2}{2}} d\left(\frac{\chi^2}{2}\right)$$

or

$$df \propto L^{\frac{n'-3}{2}} e^{-L}\, dL,$$

where n' is one more than the number of degrees of freedom in which the sample may differ from expectation (17).

In other cases we are at present faced with the difficulty that the distribution L requires a special investigation. This distribution will in general be discontinuous (as is that of χ^2), but it is not impossible that mathematical research will reveal the existence of effective graduations for the most important groups of cases to which χ^2 cannot be applied.

We shall conclude with a few illustrations of important types of discontinuous distribution.

1. *The Poisson Series.*

$$e^{-m}\left(1,\ m,\ \frac{m^2}{2!},\ \ldots,\ \frac{m^x}{x!},\ \ldots\right)$$

involves only the single parameter, and is of great importance in modern statistics. For the optimum value of m,

$$\mathrm{S}\left\{\frac{\partial}{\partial m}(-m+x\log m)\right\} = 0,$$

whence

$$\mathrm{S}\left(\frac{x}{\hat{m}}-1\right) = 0,$$

or

$$\hat{m} = \bar{x}.$$

The most likely value of m is therefore found by taking the first moment of the series.

Differentiating a second time,

$$-\frac{1}{\sigma_{\hat{m}}^2} = \mathrm{S}\left(-\frac{x}{m^2}\right) = -\frac{n}{m},$$

so that

$$\sigma_{\hat{m}}^2 = \frac{m}{n},$$

as is well known.

2. *Grouped Normal Data.*

In the case of the normal curve of distribution it is evident that the second moment is a sufficient statistic for estimating the standard deviation ; in investigating a sufficient solution for grouped normal data, we are therefore in reality finding the optimum correction for grouping ; the SHEPPARD correction having been proved only to satisfy the criterion of consistency.

For grouped normal data we have

$$p_s = \frac{1}{\sigma\sqrt{2\pi}}\int_{x_s}^{x_{s+1}} e^{-\frac{\overline{x-m}^2}{2\sigma^2}}\,dx,$$

and the optimum values of m and σ are obtained from the equations,

$$\frac{\partial \mathrm{L}}{\partial m} = \mathrm{S}\left(\frac{n_s}{p_s}\frac{\partial p_s}{\partial m}\right) = 0,$$

$$\frac{\partial \mathrm{L}}{\partial \sigma} = \mathrm{S}\left(\frac{n_s}{p_s}\frac{\partial p_s}{\partial \sigma}\right) = 0\ ;$$

or, if we write,

$$z = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\overline{x-m}^2}{2\sigma^2}},$$

we have the two conditions,

$$S\left(\frac{n_s}{p_s}\overline{z_s - z_{s+1}}\right) = 0$$

and

$$S\left\{\frac{n_s}{p_s}\left(\frac{x_s}{\sigma}z_s - \frac{x_{s+1}}{\sigma}z_{s+1}\right)\right\} = 0.$$

As a simple example we shall take the case chosen by K. SMITH in her investigation of the variation of χ^2 in the neighbourhood of the moment solution (12).

Three hundred errors in right ascension are grouped in nine classes, positive and negative errors being thrown together as shown in the following table :—

0″·1 arc	0–1	1–2	2–3	3–4	4–5	5–6	6–7	7–8	8–9
Frequency	114	84	53	24	14	6	3	1	1

The second moment, without correction, yields the value

$$\sigma_\nu = 2\cdot282542.$$

Using SHEPPARD'S correction, we have

$$\sigma_\mu = 2\cdot264214,$$

while the value obtained by making χ^2 a minimum is

$$\sigma_{\chi^2} = 2\cdot355860.$$

If the latter value were accepted we should have to conclude that SHEPPARD'S correction, even when it is small, and applied to normal data, might be altogether of the wrong magnitude, and even in the wrong direction. In order to obtain the optimum value of σ, we tabulate the values of $\frac{\partial L}{\partial \sigma}$ in the region under consideration; this may be done without great labour if values of σ be chosen suitable for the direct application of the table of the probability integral (13, Table II.). We then have the following values :—

$\frac{1}{\sigma}$	0·43	0·44	0·45	0·46
$\frac{\partial L}{\partial \sigma}$	+15·135	+2·149	−11·098	−24·605
$\Delta^2 \frac{\partial L}{\partial \sigma}$		−0·261	−0·260	

By interpolation,

$$\frac{1}{\hat{\sigma}} = 0{\cdot}441624$$

$$\hat{\sigma} = 2{\cdot}26437.$$

We may therefore summarise these results as follows :—

Uncorrected estimate of σ	$2{\cdot}28254$
SHEPPARD'S correction	$-0{\cdot}01833$
Correction for maximum likelihood	$-0{\cdot}01817$
" Correction " for minimum χ^2	$+0{\cdot}07332$

Far from shaking our faith, therefore, in the adequacy of SHEPPARD'S correction, when small, for normal data, this example provides a striking instance of its effectiveness, while the approximate nature of the χ^2 test renders it unsuitable for improving a method which is already very accurate.

It will be useful before leaving the subject of grouped normal data to calculate the actual loss of efficiency caused by grouping, and the additional loss due to the small discrepancy between moments with SHEPPARD'S correction and the optimum solution.

To calculate the loss of efficiency involved in the process of grouping normal data, let

$$v = \frac{1}{a}\int_{\xi-\frac{1}{2}a}^{\xi+\frac{1}{2}a} f(\xi)\,d\xi,$$

when $a\sigma$ is the group interval, then

$$v = f(\xi) + \frac{a^2}{24}f''(\xi) + \frac{a^4}{1920}f^{\text{iv}}(\xi) + \frac{a^6}{322{,}560}f^{\text{vi}}(\xi) + \dots$$

$$= f(\xi)\left\{1 + \frac{a^2}{24}(\xi^2-1) + \frac{a^4}{1920}(\xi^4-6\xi^2+3) + \frac{a^6}{322{,}560}(\xi^6-15\xi^4+45\xi^2-15) + \dots\right\},$$

whence

$$\log v = \log f + \frac{a^2}{24}(\xi^2-1) - \frac{a^4}{2880}(\xi^4+4\xi^2-2) + \frac{a^6}{181{,}440}(\xi^6+6\xi^4+3\xi^2-1) - \dots,$$

and

$$\frac{\partial^2}{\partial m^2}\log v = -\frac{1}{\sigma^2} + \frac{1}{\sigma^2}\left\{\frac{a^2}{12} - \frac{a^4}{720}(3\xi^2+2) + \frac{a^6}{30{,}240}(5\xi^4+12\xi^2+1) - \dots\right\},$$

of which the mean value is

* $$-\frac{1}{\sigma^2}\left\{1 - \frac{a^2}{12} + \frac{a^4}{144} - \frac{a^6}{4320}\dots\right\},$$

* Read $-\dfrac{1}{\sigma^2}\left[1 - \dfrac{a^2}{12} + \dfrac{a^4}{144} - \dfrac{a^6}{1728} + \dfrac{31a^8}{25.12^4} - \dfrac{313a^{10}}{175.12^5}\right]$

neglecting the periodic terms; and consequently

$$\sigma_{\hat{m}}^2 = \frac{\sigma^2}{n}\left(1+\frac{a^2}{12}-\frac{a^6}{2880}\ldots\right). \quad *$$

Now for the mean of ungrouped data

$$\sigma_{\hat{m}}^2 = \frac{\sigma^2}{n},$$

so that the loss of efficiency due to grouping is nearly $\frac{a^2}{12}$.

The further loss caused by using the mean of the grouped data is very small, for

$$\sigma_{\nu_1}^2 = \frac{\nu_2}{n} = \frac{\sigma^2}{n}\left(1+\frac{a^2}{12}\right),$$

neglecting the periodic terms; the loss of efficiency by using ν_1 therefore is only

$$\frac{a^6}{2880}. \quad *$$

Similarly for the efficiency for scaling,

$$\frac{\partial^2}{\partial\sigma^2}\log v = \frac{1}{\sigma^2}-\frac{3\xi^2}{\sigma^2}+\frac{1}{\sigma^2}\left\{\frac{a^2}{12}(10\xi^2-3)-\frac{a^4}{360}(9\xi^4+21\xi^2-5)\right.$$
$$\left.+\frac{a^6}{30,240}(26\xi^6+110\xi^4+36\xi^2-7)-\frac{a^8}{1,814,400}(51\xi^8+315\xi^6+351\xi^4-55\xi^2+9)+\ldots\right\},$$

of which the mean value is

$$-\frac{2}{\sigma^2}\left\{1-\frac{a^2}{6}+\frac{a^4}{40}-\frac{a^6}{270}+\frac{83a^8}{129,600}\ldots\right\};$$

neglecting the periodic terms; and consequently

$$\sigma_{\hat{\sigma}}^2 = \frac{\sigma^2}{2n}\left\{1+\frac{a^2}{6}+\frac{a^4}{360}-\frac{a^8}{10,800}\ldots\right\}.$$

For ungrouped data

$$\sigma_{\hat{\sigma}}^2 = \frac{\sigma^2}{2n},$$

so that the loss of efficiency in scaling due to grouping is nearly $\frac{a^2}{6}$. This may be made as low as 1 per cent. by keeping a less than $\frac{1}{4}$.

The further loss of efficiency produced by using the grouped second moment with SHEPPARD's correction is again very small, for

$$\sigma_{\nu_2}^2 = \frac{\nu_4-\nu_2^2}{n} = \frac{2\sigma^4}{n}\left(1+\frac{a^2}{6}+\frac{a^4}{360}\right)$$

neglecting the periodic terms.

* For $\frac{a^6}{2880}$, read $\frac{a^8}{86400}$.

Whence it appears that the further loss of efficiency is only

$$\frac{a^8}{10,800}.$$

We may conclude, therefore, that the high agreement between the optimum value of σ and that obtained by SHEPPARD's correction in the above example is characteristic of grouped normal data. The method of moments with SHEPPARD's correction is highly efficient in treating such material, the gain in efficiency obtainable by increasing the likelihood to its maximum value is trifling, and far less than can usually be gained by using finer groups. The loss of efficiency involved in grouping may be kept below 1 per cent. by making the group interval less than one-quarter of the standard deviation.

Although for the normal curve the loss of efficiency due to moderate grouping is very small, such is not the case with curves making a finite angle with the axis, or having at an extreme a finite or infinitely great ordinate. In such cases even moderate grouping may result in throwing away the greater part of the information which the sample provides.

3. *Distribution of Observations in a Dilution Series.*

An important type of discontinuous distribution occurs in the application of the dilution method to the estimation of the number of micro-organisms in a sample of water or of soil. The method here presented was originally developed in connection with Mr. CUTLER's extensive counts of soil protozoa carried out in the protozoological laboratory at Rothamsted, and although the method is of very wide application, this particular investigation affords an admirable example of the statistical principles involved.

In principle the method consists in making a series of dilutions of the soil sample, and determining the presence or absence of each type of protozoa in a cubic centimetre of the dilution, after incubation in a nutrient medium.

The series in use proceeds by powers of 2, so that the frequency of protozoa in each dilution is one-half that in the last.

The frequency at any stage of the process may then be represented by

$$m = \frac{n}{2^x},$$

when x indicates the number of dilutions.

Under conditions of random sampling, the chance of any plate receiving 0, 1, 2, 3, protozoa of a given species is given by the Poisson series

$$e^{-m}\left(1, m, \frac{m^2}{2!}, \frac{m^3}{3!}, \dots\right),$$

and in consequence the proportion of sterile plates is

$$p = e^{-m},$$

and of fertile plates

$$q = 1 - e^{-m}.$$

In general we may consider a dilution series with dilution factor a so that

$$\log p = -\frac{n}{a^x},$$

and assume that s plates are poured from each dilution.

The object of the method being to estimate the number n from a record of the sterile and fertile plates, we have

$$L = S_1 (\log p) + S_2 (\log q)$$

when S_1 stands for summation over the sterile plates, and S_2 for summation over those which are fertile.

Now

$$\frac{\partial p}{\partial \log n} = -\frac{\partial q}{\partial \log n} = p \log p,$$

so that the optimum value of n is obtained from the equation,

$$\frac{\partial L}{\partial \log n} = S_1 (\log p) - S_2 \left(\frac{p}{q} \log p\right) = 0.$$

Differentiating a second time,

$$\frac{\partial^2 L}{\partial (\log n)^2} = S_1 (\log p) - S_2 \left\{\frac{p \log p}{q}\left(\log p + 1 + \frac{p \log p}{q}\right)\right\};$$

now the mean number of sterile plates is ps, and of fertile plates qs, so that the mean value of $\frac{\partial^2 L}{\partial (\log n)^2}$ is

$$-\frac{1}{\sigma^2_{\log \hat{n}}} = sS\left\{p \log p - p \log p \left(\log p + 1 + \frac{p}{q} \log p\right)\right\} = -sS\left\{\frac{p}{q} (\log p)^2\right\},$$

the summation, S, being extended over all the dilutions.

It thus appears that each plate observed adds to the weight of the determination of $\log n$ a quantity

$$w = \frac{p}{q} (\log p)^2.$$

We give below a table of the values of p, and of w, for the dilution series $\log p = 2^{-x}$ from $x = -4$ to $x = 11$.

x.	p.	w.	S (w) (per cent.).
−4	0·00000011254	0·000029	0·001
−3	0·0003354626	0·021477	0·906
−2	0·01831564	0·298518	13·485
−1	0·1353353	0·626071	39·865
0	0·3678794	0·581977	64·387
1	0·6065307	0·385374	80·625
2	0·7788008	0·220051	89·897
3	0·8824969	0·117350	94·842
4	0·9394131	0·060567	97·394
5*	0·9692332	0·030764	98·690
6	0·9844964	0·015503	99·343
7	0·9922179	0·007782	99·671
8	0·9961014	0·003899	99·836
9	0·9980488	0·001951	99·918
10	0·9990239	0·000976	99·959
11	0·9995118	0·000488	99·979
Remainder		0·000488	
Total		2·373265	

For the same dilution constant the total S (w) is nearly independent of the particular series chosen. Its average value being $\frac{\pi^2}{6 \log a}$, or in this case 2·373138. The fourth column shows the total weight attained at any stage, expressed as a percentage of that obtained from an infinite series of dilutions. It will be seen that a set of eight dilutions comprise all but about 2 per cent. of the weight. With a loss of efficiency of only 2 to $2\frac{1}{2}$ per cent., therefore, the number of dilutions which give information as to a particular species may be confined to eight. To this number must be added a number depending on the range which it is desired to explore. Thus to explore a range from 100 to 100,000 per gramme (about 10 octaves) we should require 10 more dilutions, making 18 in all, while to explore a range of a millionfold, or about 20 octaves, 28 dilutions would be needed.

In practice it would be exceedingly laborious to calculate the optimum value of n for each series observed (of which 38 are made daily). On the advice of the statistical department, therefore, Mr. Cutler adopted the plan of counting the total number of sterile plates, and taking the value of n which on the average would give that number. When a sufficient number of dilutions are made, $\log n$ is diminished by $\frac{1}{s} \log a$ for each additional sterile plate, and even near the ends of the series the appropriate values of n may easily be tabulated. Since this method of estimation is of wide application, and appears at first sight to be a very rough one, it is important to calculate its efficiency.

* Several figures in this table have been corrected.

For any dilution the variance in the number of sterile plates is

$$spq,$$

and as the several dilutions represent independent samples, the total variance is

$$s\mathrm{S}(pq),$$

hence

$$\sigma^2_{\log n} = \frac{(\log a)^2}{s}\mathrm{S}(pq).$$

Now $\mathrm{S}(pq)$ has an average value $\frac{\log 2}{\log a}$, therefore taking $a = 2$,

$$(\log a)^2 = \cdot 480453,$$

and

$$\mathrm{S}(pq) = 1$$

being very nearly constant and within a small fraction of unity; whence the efficiency of the method of counting the sterile plates is

$$\frac{6}{\pi^2 \log 2} = 87{\cdot}71 \text{ per cent.},$$

a remarkably high efficiency, considering the simplicity of the method, the efficiency being independent of the dilution ratio.

13. SUMMARY.

During the rapid development of practical statistics in the past few decades, the theoretical foundations of the subject have been involved in great obscurity. Adequate distinction has seldom been drawn between the sample recorded and the hypothetical population from which it is regarded as drawn. This obscurity is centred in the so-called " inverse " methods.

On the bases that the purpose of the statistical reduction of data is to obtain statistics which shall contain as much as possible, ideally the whole, of the relevant information contained in the sample, and that the function of Theoretical Statistics is to show how such adequate statistics may be calculated, and how much and of what kind is the information contained in them, an attempt is made to formulate distinctly the types of problems which arise in statistical practice.

Of these, problems of Specification are found to be dominated by considerations which may change rapidly during the progress of Statistical Science. In problems of Distribution relatively little progress has hitherto been made, these problems still affording a field for valuable enquiry by highly trained mathematicians. The principal purpose of this paper is to put forward a general solution of problems of Estimation.

Of the criteria used in problems of Estimation only the criterion of Consistency has hitherto been widely applied; in Section 5 are given examples of the adequate and inadequate application of this criterion. The criterion of Efficiency is shown to be a special but important case of the criterion of Sufficiency, which latter requires that the whole of the relevant information supplied by a sample shall be contained in the statistics calculated.

In order to make clear the nature of the general method of satisfying the criterion of Sufficiency, which is here put forward, it has been thought necessary to reconsider BAYES' problem in the light of the more recent criticisms to which the idea of "inverse probability" has been exposed. The conclusion is drawn that two radically distinct concepts, both of importance in influencing our judgment, have been confused under the single name of *probability*. It is proposed to use the term *likelihood* to designate the state of our information with respect to the parameters of hypothetical populations, and it is shown that the quantitative measure of likelihood does not obey the mathematical laws of probability.

A proof is given in Section 7 that the criterion of Sufficiency is satisfied by that set of values for the parameters of which the likelihood is a maximum, and that the same function may be used to calculate the efficiency of any other statistics, or, in other words, the percentage of the total available information which is made use of by such statistics.

This quantitative treatment of the information supplied by a sample is illustrated by an investigation of the efficiency of the method of moments in fitting the Pearsonian curves of Type III.

Section 9 treats of the location and scaling of Error Curves in general, and contains definitions and illustrations of the *intrinsic accuracy*, and of the *centre of location* of such curves.

In Section 10 the efficiency of the method of moments in fitting the general Pearsonian curves is tested and discussed. High efficiency is only found in the neighbourhood of the normal point. The two causes of failure of the method of moments in locating these curves are discussed and illustrated. The special cause is discovered for the high efficiency of the third and fourth moments in the neighbourhood of the normal point.

It is to be understood that the low efficiency of the moments of a sample in estimating the form of these curves does not at all diminish the value of the notation of moments as a means of the comparative specification of the form of such curves as have finite moment coefficients.

Section 12 illustrates the application of the method of maximum likelihood to discontinuous distributions. The POISSON series is shown to be sufficiently fitted by the mean. In the case of grouped normal data, the SHEPPARD correction of the crude moments is shown to have a very high efficiency, as compared to recent attempts to improve such fits by making χ^2 a minimum; the reason being that χ^2 is an expression only approximate to a true value derivable from likelihood. As a final illustration of the scope of the new process, the theory of the estimation of micro-organisms by the dilution method is investigated.

Finally it is a pleasure to thank Miss W. A. MACKENZIE, for her valuable assistance in the preparation of the diagrams.

REFERENCES.

(1) K. PEARSON (1920). "The Fundamental Problem of Practical Statistics," 'Biom.,' xiii., pp. 1–16.

(2) F. Y. EDGEWORTH (1921). "Molecular Statistics," 'J.R.S.S.," lxxxiv., p. 83.

(3) G. U. YULE (1912). "On the Methods of Measuring Association between two Attributes." 'J.R.S.S.,' lxxv., p. 587.

(4) STUDENT (1908). "The Probable Error of a Mean," 'Biom.," vi., p. 1.

(5) R. A. FISHER (1915). "Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population," 'Biom.," x., 507.

(6) R. A. FISHER (1921). "On the 'Probable Error' of a Coefficient of Correlation deduced from a Small Sample," 'Metron.," i., pt. iv., p. 82.

(7) R. A. FISHER (1920). "A Mathematical Examination of the Methods of Determining the Accuracy of an Observation by the Mean Error and by the Mean Square Error," 'Monthly Notices of R.A.S., lxxx., 758.

(8) E. PAIRMAN and K. PEARSON (1919). "On Corrections for the Moment Coefficients of Limited Range Frequency Distributions when there are finite or infinite Ordinates and any Slopes at the Terminals of the Range," 'Biom.,' xii., p. 231.

(9) R. A. FISHER (1912). "On an Absolute Criterion for Fitting Frequency Curves," 'Messenger of Mathematics,' xli., p. 155.

(10) BAYES (1763). "An Essay towards Solving a Problem in the Doctrine of Chances," 'Phil. Trans.,' liii., p. 370.

(11) K. PEARSON (1919). "Tracts for Computers. No. 1: Tables of the Digamma and Trigamma Functions," By E. PAIRMAN, Camb. Univ. Press.

(12) K. SMITH (1916). "On the 'best' Values of the Constants in Frequency Distributions," 'Biom., xi., p. 262.

(13) K. PEARSON (1914). "Tables for Statisticians and Biometricians," Camb. Univ. Press.

(14) G. VEGA (1764). "Thesaurus Logarithmorum Completus," p. 643.

(15) K. PEARSON (1900). "On the Criterion that a given System of Deviations from the Probable in the case of a Correlated System of Variables is such that it can be reasonably supposed to have arisen from Random Sampling," 'Phil. Mag.,' l., p. 157.

(16) K. PEARSON and L. N. G. FILON (1898). "Mathematical Contributions to the Theory of Evolution. IV.—On the Probable Errors of Frequency Constants, and on the influence of Random Selection on Variation and Correlation," 'Phil. Trans.," cxci., p. 229.

(17) R. A. FISHER (1922). "The Interpretation of χ^2 from Contingency Tables, and the Calculation of P," 'J.R.S.S.,' lxxxv., pp. 87–94.

(18) K. PEARSON (1915). "On the General Theory of Multiple Contingency, with special reference to Partial Contingency," 'Biom.," xi., p. 145.

(19) K. PEARSON (1903). "On the Probable Errors of Frequency Constants," 'Biom.,' ii., p. 273, Editorial.

(20) W. F. SHEPPARD (1898). "On the Application of the Theory of Error to Cases of Normal Distribution and Correlations," 'Phil. Trans.,' A., cxcii., p. 101.

(21) J. M. KEYNES (1921). "A Treatise on Probability," Macmillan & Co., London.