WILEY

Notes on the History of Pauperism in England and Wales from 1850, Treated by the Method of Frequency-Curves; with an Introduction on the Method

Author(s): G. Udny Yule

Source: *Journal of the Royal Statistical Society*, Vol. 59, No. 2 (Jun., 1896), pp. 318-357

Published by: [Wiley](#) for the [Royal Statistical Society](#)

Stable URL: <http://www.jstor.org/stable/2979777>

Accessed: 28/06/2014 09:26

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Wiley and Royal Statistical Society are collaborating with JSTOR to digitize, preserve and extend access to *Journal of the Royal Statistical Society*.

<http://www.jstor.org>

NOTES on the HISTORY of PAUPERISM in ENGLAND and WALES from
1850, TREATED by the METHOD of FREQUENCY-CURVES; with an
INTRODUCTION on the METHOD.

By G. UDNY YULE, *Assistant Professor of Applied Mathematics,
University College, London.*

[Read before the Royal Statistical Society, 21st April, 1896.
The Right Hon. the EARL OF VERULAM, Vice-President, in the Chair.]

CONTENTS :

	PAGE
I.—Introduction : on the Methods used.....	318
Sec. 1. Introductory.....	318
„ 2. Artificial Chance; Frequency-Polygons	319
„ 3. Point-Binomial	320
„ 4. Method of Fitting; Moment Coefficients.....	321
„ 5. Continuous and Discontinuous Variation; Frequency- Curves; Normal Curve; Generalised Binomial Curve; Hypergeometrical Series and Derived Curves	323
„ 6. Criterion	326
„ 7. Mode, Median, and Mean	327
„ 8. Standard Deviation	328
„ 9. Skewness	329
„ 10. Probable Errors	330
„ 11. Contributory Causes and Component Cause-Groups.....	331
„ 12. Homogeneity of Material	331
„ 13. Test of Fit	332
II.—On Pauperism in England and Wales, 1850-91	332
Sec. 14. Material; Arrangement of Tables and Plates	332
„ 15. Table I; Plates I—V	334
„ 16. Means and Modes.....	335
„ 17. Standard Deviations	336
„ 18. Criteria	337
„ 19. Ranges	337
„ 20. Indices m , and Ratio m/a	338
„ 21. Skewness	338
„ 22. Point-Binomials ..	339
„ 23. Fit of Curves	340
„ 24. Concluding Remarks.....	341
Supplementary Note on approximate Methods of Determining the Mode	343

I.—*Introduction : on the Methods used.*

Sec. 1. *Introductory.*

CONSIDERING the amount and variety of our English poor law statistics, it seems somewhat strange that so little use has been made of them to give us information on the history of pauperism

in the kingdom as a whole. We have had, more especially of recent years, numerous instances of groupings and sub-divisions of the raw material by "in-door" and "out-door" unions, by north and south, and so on, but in few of these cases do we seem to get much grasp of the broad outlines of the statistical development, the numerical history, of English pauperism. We have had a mass of detail for single unions, hardly anything but a wretched average for the whole country; the wood seems to be hidden behind the trees.

Now surely, to keep to the simplest facts, we might be told a little more. How do the rates of pauperism for the isolated unions group themselves round the average rate? Are they more, or less, scattered round it now than forty years ago? Are rates near the average also the most common and typical or no? If not, what is the most frequent or "fashionable" rate? How has this "fashion" or "mode" in pauperism been varying? Are not these questions that anyone might ask?

But under the guidance of the theory of probability such questions become at once more defined and more numerous. What is the *numerical measure* of the degree of scatter round the average, and how has it altered? What is the *true mode*, as opposed to the one accidentally created by our groupings of the unions? Is the distribution round the mean symmetrical? If not, what is the numerical measure of its skewness, and how has this varied? Finally, can one express immediately by some few constants the whole form of the distribution round the mean?

This question we must stop and consider, for it is the fundamental one.

Sec. 2. *Artificial Chance. Frequency-Polygons.*

To endeavour to clear our ideas, let us revert to a case of so-called artificial chance—to a case in which we can directly compare theoretical conclusions with actual fact. Suppose we take a handful of dice, say fourteen, toss them some hundred or thousand times, and note each time the number of sixes thrown. Let us then record the results graphically. Taking a straight line as base, mark off along it a number of short equal lengths, 01, 12, 23, and so on (Fig. 1); then erect at 0 a vertical representing by its length the number of times that throws with no sixes occurred; if, *e.g.*, there were twenty such throws, make 00' twenty millimetres long. Again at 1 erect a vertical showing on the same scale the number of throws with one six; if there were 100, make 11' 100 millimetres long. Completing the record in this way, and then joining up the tops of the verticals 0', 1', 2', &c., we have a statistical "frequency-polygon." The height

of each vertical, or the elementary strip of area round it, gives the frequency of occurrence of the corresponding number of sixes: the area of the whole polygon must represent on the same scale the number of observations made. The polygon will be in general more or less jagged and rough, but not badly so if the number of observations is large; the figure is drawn from an actual record of 648 throws of fourteen dice (*vide* the table on p. 322).

FIG. 1.

Sec. 3. Point-Binomial.

Now in the case of dice we know *à priori* the data which determine the approximate form of this frequency-polygon. If any event can happen in s ways and fail to happen in r ways, then it is found (in a great variety of cases) that if the event be tried a large number of times, the ratio of successes to failures will be very nearly the ratio of s/r . $\frac{s}{s+r}$ is called the chance of the event succeeding, and $\frac{r}{s+r}$ the chance of its failing; these chances are generally denoted by p and q . We know p and q approximately

for a die (assuming it to be a perfect homogeneous cube); the chance of throwing any given face is $\frac{1}{6}$. Further we know the number of dice, n , and the distance, c , between the verticals we erected. Given p , q , n , and c , we can calculate what frequency to expect for throws with any given number of sixes, *i.e.*, we can draw a theoretical frequency-polygon for our experimental cases and compare it with the actual one. This theoretical polygon is called a point-binomial. The point-binomial is drawn with dotted lines in Fig. 1.

Sec. 4. *General Method of fitting: Moment-Coefficients.*

But now suppose we had to approach the question from the other side. Suppose a friend had given us a record of dice throwings, with the remark that it was a "record of chance distribution" but no further information. Could we make our way back from the statistical polygon to the numbers p , q , n , c , which describe it, if indeed these are sufficient? If we could do so, would the experimental numbers be likely to agree at all reasonably with the theoretical numbers? Both of these questions can fortunately be answered in the affirmative.

The method for passing back from the polygon to its constants, *i.e.*, for fitting a point-binomial to an experimental polygon, is due to Professor Karl Pearson, to whose recent work on "Skew Variation in Homogeneous Material"¹ I shall have frequent occasions to refer. As the procedure is extremely general, and applicable to the fitting of all polygons or curves mentioned in this paper, I shall endeavour to describe it briefly; the description must not be taken too literally, but is correct in principle.

Taking the statistical polygon of Fig. 1, or the numerical results it represents—

(1) Determine the mean or average value of the variable, *i.e.*, the mean number of sixes thrown.

(2) Determine the mean (deviation)² from the mean, μ_2 ; *i.e.*, multiply the length of each observed vertical by the square of its distance from the mean, add up all the figures so obtained, and divide by the total number of observations.

(3) Similarly determine the mean (deviation)³ and mean (deviation)⁴, μ_3 , and μ_4 . These quantities μ may be called the moment-coefficients.

Equating μ_2 , μ_3 , and μ_4 to the corresponding moment-coefficients of the theoretical point-binomial expressed in terms of n , p , q , and c , and remembering that

$$p + q = 1,$$

we have sufficient equations to determine all the unknowns.

¹ "Phil. Trans.," vol. clxxvi (1895), A, pp. 343—414.

Let us revert to experiment to see how close this method will work in practice. The results of five cases are given in full in the following table, treating the experimental figures in three slightly different ways. The moment-coefficients were worked out twice, (1) by treating the observed ordinates as simply loaded proportionally to their lengths, (2) by treating the whole polygon as a system of trapezia, just as it is drawn. These different methods lead to the figures given under "trapezia" and "loaded ordinates" respectively.

Table showing the Results of some Chance Experiments.

		<i>n.</i>	<i>p.</i>	<i>q.</i>	<i>c.</i>
2,048 tosses of 10 shillings. (Pearson.)	Theory	10.0	0.50	0.50	1.00
	Trapezia	10.4	0.52	0.48	0.99
	Loaded ordinates.....	10.4	0.53	0.47	0.99
	<i>c</i> as known	11.1	0.53	0.47	—
5's or 6's recorded in 2,363 tosses of 12 dice. (Weldon.)	Theory	12.0	0.67	0.33	1.00
	Trapezia	11.6	0.62	0.38	0.97
	Loaded ordinates.....	11.5	0.65	0.35	0.97
	<i>c</i> as known	10.7	0.61	0.39	—
Drawing balls from an urn 683 times. (Quetelet.)	Theory	6.0	0.50	0.50	1.00
	Trapezia	4.2	0.54	0.46	1.19
	Loaded ordinates.....	4.5	0.54	0.46	1.19
	<i>c</i> as known	6.4	0.52	0.48	—
648 throws of 14 dice (6's recorded).	Theory	14.0	0.83	0.17	1.00
	Trapezia	16.1	0.88	0.12	1.08
	Loaded ordinates.....	14.3	0.83	0.17	1.10
	<i>c</i> as known	22.8	0.90	0.10	—
648 throws of 3 dice (5's or 6's re- corded).	Theory	3.00	0.67	0.33	1.00
	Trapezia	3.68	0.67	0.33	0.97
	Loaded ordinates.....	3.31	0.67	0.33	0.96
	<i>c</i> as known	3.11	0.68	0.32	—

Now in both of these cases we assume that nothing whatever is known of the work to which the figures refer, not even whether it be a case of artificial chance or no. The numbers might have been obtained by measuring the heights of a lot of men, and counting how many there were between 5 ft. and 5 ft. 1 in., how many between 5 ft. 1 in. and 5 ft. 2 in., and so on. The unit *c* is consequently quite unknown, we cannot assume that it corresponds to our arbitrary groups; *c* must be determined like *p*, *q*, and *n*, from the moment-coefficients. But if we know that the figures

refer to a case of real discontinuity, we may simplify the work by putting $c = 1$ immediately, and only reckoning the values of the three remaining constants. The third row of figures headed " c known" refers to this method.

The series given are, I think, sufficient to show that even with a small number of observations (600—700) moderately close results may be obtained. It must be noted that we need not necessarily expect the experimental figures to agree closely with the simple *à priori* theory here applied. If a die be badly cut, the chance of any one face coming uppermost is not $\frac{1}{6}$; if a handful of dice be all badly cut the chance may in fact be different for each one, and the results obtained by tossing them will not be represented by a point-binomial at all. I think this is the reason for the erratic character of the figures in the last two sets, for certainly the dice used for them were very far from true cubes.

It should be remarked that to compare several sets of observations, the same method of reduction should be used for all: different methods lead to slightly divergent results.

Sec. 5. *Continuous and Discontinuous Variation.*

Frequency-Curves.

All these cases of artificial chance have one limitation in common: they can only give rise to discontinuous frequency polygons. In tossing a handful of dice, for instance, we can only throw a whole number of sixes, say four or five, but not any intermediate number. Step-by-step variation like this occurs of course in several cases of "natural" statistics, *e.g.*, in the number of petals on flowers, in the number of children in families, and so on, though there is no *à priori* reason for supposing that the frequency will in this case be represented by a binomial series.

On the other hand, however, we have numerous cases in which there is no obvious unit of variation; the variation is apparently continuous, *e.g.*, in the case of measurements on men or animals. We now want a *curve* such that its area between any two verticals will represent the frequency of values of the variable between corresponding limits for the theoretical representation of our statistics; a discontinuous polygon no longer suffices. In practice of course we have still to go by arbitrary units; we do not count, for example, how many men there are of one exact height, but how many there are with heights between 5 ft. 1 in. and 5 ft. 2 in., how many between 5 ft. 2 in. and 5 ft. 3 in., and so on.

Normal Curve.

Experience shows that such frequency-curves take a very great variety of forms, ranging from approximate symmetry

through all degrees of skewness, to distributions with a maximum at one end of the range, or maxima at both ends and a minimum near the centre. Until the publication of Professor Pearson's memoir the only curve which had been shown to be at all generally applicable was Laplace's or the normal curve

$$y = y_0 e^{-x^2/2\sigma^2} \quad . \quad . \quad . \quad (1)$$

y_0 being the ordinate at the origin and σ a constant. This curve has been considerably employed, notably by Quetelet and by Mr. Francis Galton, for representing various statistics, and its figure is now fairly well known. The curve is symmetrical and tails off more or less rapidly in both directions, according to the value of σ , asymptoting to the base. Only one way of deducing it immediately concerns us: it may be considered as a limit to the form of the symmetrical binomial

$$\left(\frac{1}{2} + \frac{1}{2}\right)^n$$

for all values of n . That is to say, we may consider a continuous variation curve to be generated by rendering possible some form of interpolation between the absolute units of the point-binomial. The question thus suggested itself to Professor Pearson: would a curve having the same relation to the skew binomial as the normal curve has to the symmetrical binomial prove equally useful in statistics? Was it not probable, in fact, that being more general it would be more useful?

Generalised Binomial Curve.

The curve thus obtained (Professor Pearson's Type III) is of the form²

$$y = y_0 \left(1 + \frac{x}{a}\right)^{\gamma a} e^{-\gamma x} \quad . \quad . \quad . \quad (2)$$

It is a curve limited in range at one end (at a distance $(-a)$ from the maximum ordinate) and tailing off to infinity at the other. At the limited end the curve may either be tangential to the base, or may rise sharply at any angle with it. Both of these types occur in actual statistics, the first for example in curves showing the frequency of various heights of the barometer, the second in frequency diagrams of divorce with duration of marriage, of mortality with age, at certain periods, and so on.³ For special values of the constants the maximum ordinate of the curve may be at the limit of the range. Evidently this is a great advance in generality on the symmetrical normal curve.

² This curve was obtained earlier by De Forrest, *vide* "Nature," vol. lii, p. 317 (1895). ("The Analyst," vols. vi, ix, and x.)

³ Various examples will be found in Professor Pearson's paper *loc. cit.*

Hypergeometrical Series and Derived Curves.

But we need not stop at this point in the analogues with artificial chance. Let us return to one of our examples of the case of a point-binomial, Queletet's drawings of balls from a bag. To get this point-binomial each single ball had to be returned when drawn and recorded, then another ball drawn and so on; the records only being grouped into clusters of six afterwards. But suppose all six balls had been drawn at once, what then? Evidently a point-binomial would no longer serve to describe the results. There are now four constants instead of three: the chance of drawing the first single ball p , the number N in the bag, the number n drawn out at once, and the distance c between the ordinates. These data lead to a more general form of polygon, that may be called, after the series to which it corresponds, a "point-hypergeometrical." Using the same geometrical relation that holds between the normal curve and the symmetrical-binomial, Professor Pearson derived from this two new curves, the first (his Type IV):—

$$y = y_0 \left(1 + \frac{x^2}{a^2} \right)^{-m} e^{-r \tan^{-1} \frac{x}{a}} \quad . \quad . \quad . \quad (3)$$

the second (his Type I)—

$$y = y_0 \left(1 + \frac{x}{a_1} \right)^{m_1} \left(1 - \frac{x}{a_2} \right)^{m_2} \quad . \quad . \quad . \quad (4)$$

where $m_1/a_1 = m_2/a_2$.

The first or second form results from a given hypergeometrical according as the number of balls withdrawn lies within certain limits or no, both are of great generality, and include Form (2) and the normal curve as special cases. Form (3) is of unlimited range in both directions, Form (4) limited in both directions; both curves are, in general, skew. The justification of all these curves, derived on the hypothesis of possible fractionising of the discontinuous steps of the polygon, lies in experience; all are found to give very close representations of frequency polygons in many classes of statistics.

Curve (4) is the form which we shall find occurs uniformly in the distributions dealt with in the sequel, so it is as well to note the chief relations of its shape to the values of its constants.

The value $x = 0$ corresponds to the maximum ordinate of the curve. The curve extends to a distance a_1 to the left, and a_2 to the right of the maximum ordinate.

If the curve is to be tangential to the base at both extremities, m_1 and m_2 must both be positive and greater than unity.

If $a_1 = a_2$ or $m_1 = m_2$ the curve becomes symmetrical (but not normal).

The ordinates will generally be vanishingly small for some distance before the curve comes to the limit of its range in either direction. The smaller the ratio of m_1/a_1 the further do the ordinates remain sensible towards either limit.

It may be said that too much weight cannot be attached to the value of the range, for all these curves, normal included, are only limits to polygons which themselves never extend to infinity, although the curves do. We can, however, often check the range in practice. There may be some obvious physical limit to our frequency-curve in one direction, and we can see how close the range gives this. The range may thus justify itself: if we find that it gives one known limit reliably, why not accept the other limit as equally valid? There may be really a limited range in types (1) and (3), but there is no means of getting it; this does not seem to be any reason for refusing to accept the limits where they are determinate.

Sec. 6. *Criterion.*

As I stated in briefly describing the method of fitting a point-binomial to a series of observations (sec. 4), the same method remains available for fitting any type of frequency-curve. But before we can proceed to actual fitting we require some criterion to tell us which of the forms (1)–(4) to use. This criterion is afforded very simply by the moment-coefficients, μ_2 , μ_3 , &c. Let

$$\beta_1 = \mu_3^2 / \mu_2^3$$

$$\beta_2 = \mu_4 / \mu_2^2$$

β_1 and β_2 are then pure numerics. Also let

$$\kappa = 6 + 3\beta_1 - 2\beta_2.$$

Then if

$$(1) \quad \left. \begin{array}{l} \kappa = 0 \\ \beta_1 = 0 \end{array} \right\}$$

(approximately) for the observations, a normal curve may be used. If

$$(2) \quad \left. \begin{array}{l} \kappa = 0 \\ \text{but } \beta_1 \text{ is not zero} \end{array} \right\}$$

the distribution is of the form given in equation (2)—(generalised binomial type).

If

$$(3) \quad \left. \begin{array}{l} \kappa \text{ is negative} \\ \text{whatever the value of } \beta_1 \end{array} \right\}$$

the distribution is of the trigonometrical type, equation (3); and finally, if

$$(4) \quad \left. \begin{array}{l} \kappa \text{ is positive} \\ \text{whatever the value of } \beta_1 \end{array} \right\}$$

the distribution is of the limited range type equation (4).

The number κ we will call "the criterion," as it tells us to which type of distribution any frequency polygon belongs. In the first and second cases the relation $\kappa = 0$ will never, of course, be exactly satisfied; it may be read " κ must be small compared with its own probable error."

It should further be noticed that for the normal curve $\beta_2 = 3$; for the symmetrical form of the trigonometrical curve β_2 is greater than 3; while for the symmetrical form of the limited range curve β_2 is less than 3. These relations come at once by putting $\beta_1 = 0$ in the criterion.

In fitting any statistical frequency polygon to a theoretical curve, the moment-coefficients must be reckoned, the β 's and the criterion determined, before any further steps can be taken. One of the most curious points that has presented itself in handling statistics by these generalised curves is the apparent constancy of type, *i.e.*, of the sign of the criterion, in certain classes of material.

Sec. 7 *Mode, Median, and Mean.*

Let us now pass on and consider certain terms that can be defined generally without reference to any special type of distribution.

Let C (Fig. 2) be the centroid or centre of gravity of any frequency-curve. The vertical through C meets the axis of

FIG. 2.

measurement in a point termed the mean or average value of the variable. We are using this term in its ordinary sense; it scarcely requires further explanation.

Let dd be a vertical dividing the area between the curve and axis into two equal portions. The point in which dd meets the axis is termed the median. Values of the variable greater and less than the median occur with equal frequency.

Finally let the point in which the maximum ordinate cuts the

base correspond to a value M of the variable. Professor Pearson has proposed to call this value the "mode," *i.e.*, it is the "fashion-able" value.

If the distribution be symmetrical, mode, median, and mean will coincide, but not otherwise. If the distribution be skew, the mode becomes more important than the mean; it becomes, in fact, the type, the value that strikes the eye by its constant repetition when one looks down a column of values of the variable.

The median value has been largely used by Mr. Francis Galton on account of its practical simplicity. Thus, to get the median height of, say, 100 men, they have simply to be ranged in order of their heights, and the fiftieth man from the end measured.

An approximate method of obtaining the mode when the mean and the median are given, based on an empirical formula of Professor Pearson's, is described at the end of this paper.⁴

Sec. 8. *Standard Deviation of a Distribution.*

Consider the second moment-coefficient, the quantity denoted in sec. 4 by μ_2 . μ_2 is the mean square deviation [*i.e.*, mean (deviation)²] of the curve round its centroid vertical; it is therefore of the order of the square of a length. Let

$$\sigma = \sqrt{\mu_2}$$

σ is then a length, and a very important one, for it measures the mean distance of the curve from the centroid vertical, *i.e.*, the degree of "scatter" of the distribution from the mean. This quantity has been variously called "standard deviation," "mean error," "error of mean square," and so on; in dynamics it occurs as "swing radius" and "radius of gyration." The first term is the one we shall use.

In the case of a skew-curve it might seem best to refer the curve to the mode instead of the mean, and state separately the standard deviations in excess and defect of the mode; unluckily however these quantities do not lend themselves to easy calculation. In any case a rough and handy rule holds good, that a total range of six times the standard deviation contains nearly the whole frequency; it is a crude but convenient measure of the practical extension of the curve.

The *range* that occurs in the curve of type (4) is not, it must be noted, in any way a measure of scatter or variation. In a curve of that type you may hold the range fixed, and then proceed to vary the standard deviation as you please. Thus the two curves in fig. 3 have been drawn with the same range; it would evidently

⁴ Supplementary note p. 343.

FIG. 3.

be useless to take as a measure of variation any quantity which would be the same for both. The equations of the two curves are—

$$(1.) \quad y = 10 \left(1 - \frac{x^2}{100}\right)^5$$

$$(2.) \quad y = 14.91 \left(1 - \frac{x^2}{100}\right)^{12}$$

They have equal areas and the same range, viz., 20 centimetres.⁵ The standard deviations are—

$$\sigma_1 = \frac{10}{\sqrt{13}} = 2.77$$

$$\sigma_2 = \frac{10}{\sqrt{27}} = 1.92$$

It should be noticed that in the normal curve the constant σ in the exponent is the standard deviation.

Sec. 9. Skewness.

The two curves of Fig. 3 are examples of the symmetrical form of the limited range distribution; this case is however exceptional. As a rule distributions will be more or less skew, and it will be convenient to have some numerical measure of their departure from

⁵ Each of the large squares represents 5 cms. on the original scale.

symmetry. Whatever be taken as this measure, it must be a pure number, and must be easily calculable. Professor Pearson defines skewness in his paper as the ratio of the distance between the mode and mean to the standard deviation.

This ratio is zero for a symmetrical curve, where mode and mean coincide, and is never greater than unity. In all the skew-curves (2) to (4) the skewness may be expressed directly in terms of the constants.

Sec. 10. *Probable Errors.*

Now these are all numerical measures descriptive of the original distribution: but to what extent can we rely on the numbers when obtained?

Thus, suppose we make a series of physical observations of any kind, group the results into a frequency-curve and determine its mean, standard deviation, and so on. However steady the methods and material may be in reality, we shall never, on repeating the observations, arrive at exactly the same values of the constants. To take a concrete case, return to dice. Repeat a series like that of example (4) in the table r times, and work out the r resulting values of the chance p . Now form a new frequency-curve of the values of p . This derived curve will have some standard deviation σ' , and, evidently, the smaller σ' the more can we rely on the determination of p from a single set of observations. The standard deviation σ' can in general be estimated *à priori* for any constant of a chance distribution, and gives us a most valuable measure of the accuracy with which that constant can be determined. In general not σ' but a number

$$t = .674489 \dots \sigma'$$

is used as the actual measure, and called the Probable Error of the constant; it might be better called, in Mr. Galton's nomenclature, the Quartile Deviation (or error) as it is not probable in any sense. Deviations greater and less than the quartile occur with equal frequency in the long run; *i.e.*, the quartile may be defined as that point on the axis of measurement the vertical through which cuts off a quarter of the area of the frequency polygon on one side.

The distributions of skew curve constants may in general be taken as very closely normal, otherwise some more complex measure would have to be adopted, or the probable errors in excess and defect given separately.

The "probable errors" are quoted after the constants in several of the tables of the appendix.⁶

⁶ I am indebted to Professor Pearson for permission to use his unpublished results for the calculation of these "probable errors" of skew curve constants.

Sec. 11. *Contributory Causes and Component Cause-Groups.*

Besides definitions of these various measures, there are one or two points of nomenclature on which I wish to be clear.

Consider again the case of generation of a frequency polygon by dice throwing. Each die as it falls and rolls has its final position determined by various causes, which cannot be precisely specified, and may (somewhat loosely) be described as "infinite" in number. Each of these causes might be termed a "contributory cause" in the formation of the chance distribution. Professor Pearson, however, uses the phrase in a different sense,⁷ in the sense, namely, of each die being a "contributory cause." The distinction is important, for while the number of contributory causes in the first sense is very large, the number of component cause-groups (considered as one group to each die) is not only finite, but often very small.

Now when we pass from fitting a point-binomial to a known discontinuous number of observations to fitting it to a series known to be really continuous, our numbers n , p , and q no longer admit of anything but a metaphorical interpretation, if indeed they can be considered as anything further than constants describing the form of the distribution. To avoid the slight confusion of terms as above, we shall refer to n in every case as the number of "component cause-groups." Just as n in the case of dice throwing is, at it were, the number of distinct channels through which the multitude of small contributory causes affect the form of the polygon, so we assume as a tentative working hypothesis that it bears the same significance in the case of continuous variation; there is no need for n to be large, however great we may conceive the number of contributory causes.⁸

Sec. 12. *Homogeneity of Material.*

To pass on to another point of terminology, what is to be understood when we speak of statistical material as "homogeneous" or otherwise?

All that we shall imply by the term is this, that the distribution of frequency arising from measurements on the given material is capable of adequate representation by one single theoretical frequency-curve. This definition is very loose, but it scarcely seems possible to go further. Two points about it should be noticed; first, that material may be homogeneous with regard to one measurement while non-homogeneous with regard to another: the measurement should always be specified if not evident from

⁷ *E.g., loc. cit.*, p. 358.

⁸ These remarks and some of the preceding were suggested by some of Professor Edgeworth's criticisms in the *Journal* for September, 1895, p. 510.

the context; secondly, that the material is not necessarily non-homogeneous because it can, by some method of sorting, be split up into two or more superposed distributions. Common sense or experience may have to be called in to tell us whether the representation of the material is adequate or no, or the theory of probability may have to be again employed to find whether the odds against divergences of fit are greater than seems permissible.

The definition of homogeneity might appear more naturally based on the condition that an increase in the number of observations should tend to bring the statistical polygon closer and closer to some theoretical form. I have purposely excluded this condition. Useful though it might be in theory, or in a physical case where the number of observations could be easily increased, in economics it would frequently be, from the very nature of the case, simply impossible to apply. The material we are handling is often limited, and the very introduction of fresh data may create the non-homogeneity for which we are seeking a test.

Sec. 13. *Test of Fit.*

Professor Pearson has proposed in his paper a test for the closeness of fit of a theoretical curve to a statistical polygon, namely, the ratio

$$\frac{\text{area between curve and polygon}}{\text{area of whole curve or polygon}}$$

the area between curve and polygon being all counted positive. The better the fit the smaller this ratio. The test is of course very rough; it lays the same stress on very probable divergences near the mode, and very improbable divergences in the tail of the curve, but it is convenient to have some such ready means of comparison.

In a large series of normal curve fits containing some 1,000 observations each, the ratio ranged between 6 per cent. and 9 per cent. With skew-curves the results are generally better, averaging about 5 per cent. "misfit."

We may now pass on to consider the special case of "pauperism" with the aid of such frequency-distributions and numerical measures as we have here very briefly described.

II.—*On Pauperism in England and Wales.*

Sec. 14. *Material used.*

The raw material on which the present study of pauperism is based is contained in the series of returns that generally go by the name of "Pell's Returns," viz., No. 214, 1876; No. 339, 1882; and No. 266, 1892. The return published in 1876 is retrospective, containing the figures for 1850, 1860, and 1870. The returns give,

inter alia, the percentage of the population in receipt of relief for every union in England and Wales on the 1st January of each census year, but there are irregularities. First, the return for 1850 is very defective, no information being available for many unions: secondly, the earlier return does not use as it should the figures taken from the "B returns" for 1st January, 1851, 1861, and 1871, but gives the percentage of paupers for 1st January, 1850, on the population for 1851, and so on. To keep this in mind I have labelled the plates, tables, &c., for these years 1850, 1860, and 1870; the last two returns are correct for 1881 and 1891. It scarcely seemed worth while to go through the considerable labour of forming correct returns for the earlier census years; for 1851, in fact, there are no published data. The returns given are sufficiently close to the decades to afford a clear picture of the advance or the reverse that has taken place in each successive period. The average pauperisms for the earlier years would, however, all be a little too low, owing to the populations being taken too high.

The figures in these returns include paupers of all classes; lunatics and vagrants have not been struck out. It seems not unfair to exclude these two classes when comparing two individual unions, but in a case like the present, where we shall never mention any union by name but only consider the country as a whole, it appears better to retain them. The number of lunatics and vagrants in a union may not be due to any local conditions over which that union has control, but they form an important part of the pauperism of the kingdom, a part that there seems no reason to exclude when we are simply comparing the kingdom with itself at different stages of its progress.

Owing to my having originally worked out the data for 1881 and 1891 not from the original returns, but from the figures given in the appendix to Mr. Charles Booth's work on the "Aged Poor," these data refer to registration districts and not unions. In most cases the two are coterminous, but in a few instances two or more unions go to a district. There are now 647 unions in England and Wales, grouped into 632 registration districts; the constants, &c., of the distributions would be practically unaffected by such a small reduction by grouping.

A frequency polygon showing the number of unions (in England and Wales) having any assigned rate of pauperism was formed for each of the years mentioned above, the unit of grouping chosen being 0.5 per cent. (see the explanation of Table I below). The first four moment-coefficients of these five frequency polygons were then reckoned, and each of them was fitted first to a skew frequency-curve and then to a binomial-polygon, so as to

condense the whole history into a few constants. The data referring to the statistical and to the theoretical distributions are given in the tables of the appendix as follows:—

Table I. The figures corresponding to the above five frequency polygons.

Table II. Equations of the curves fitted to the polygons.

Table III. Moment-coefficients; criterion; standard deviation.

Table IV. Modes and means.

Table V. Probable errors of skew-curve constants; range; skewness.

Table VI. Constants of the point-binomials fitted to the polygons.

The plates contain the following diagrams:—

Plate I. Curve of mean pauperism (rates of total pauperism to total population) for each year from 1850-91. This curve is based on the weekly returns, and gives the average for the year, not the figures for 1st January. It has been given for the sake of general comparison, and will not be much noticed in the sequel.

Plates II—VI. The frequency-polygons, with the fitted curves and point-binomials superposed.

Sec. 15.

Table I is read thus: taking 1850 for example, there were in England and Wales:—

Only one union in which the proportion of the population in receipt of relief on 1st January lay between 0.25 and 0.75 per cent.

Four in which the proportion lay between 0.75 and 1.25 per cent.

Two in which the proportion lay between 1.25 and 1.75 per cent.

* * * * *

Thirty-three in which the proportion lay between 3.75 and 4.25 per cent.

Forty-six in which the proportion lay between 4.25 and 4.75 per cent.,

and so on. The whole trend of the figures is shown by the black-lined polygons of Plates II—VI. Just glancing over these diagrams, we see at once that other changes, quite as important as changes in the average pauperism, have taken place. The standard deviation has decreased; the unions are more closely grouped round the mode. The period of stagnation, of positive reaction, from 1860-70 is characterised, as we see, not only by an increase in

the mean and the mode, but by an almost complete absence of change in the form of the distribution. The contrast of this with the alteration effected during the following decade is almost startling, an alteration of which a mere average leaves one perforce completely ignorant. Let us proceed to consider the numerical measures of these changes.

Sec. 16. *Means and Modes.*

It might be the more orderly and logical arrangement to commence *ab initio* with the criterion, and then to pass on to a discussion of the constants of skew-curves and point-binomials. We will find it convenient, however, to give the first place to means, modes, and standard deviations, and clear up a few points with regard to these fundamental quantities which have an immediate and obvious physical meaning.

The mode is the value of pauperism corresponding to the maximum ordinate of the fitted curve; it is obtained, by somewhat lengthy arithmetic, from the values of the three moment-coefficients μ_2 , μ_3 , and μ_4 . The mean is the average pauperism deduced at once from the figures of Table I.

This average, the "union average," is quite a different thing from the mean pauperism usually given in returns and quoted by the Registrar-General: the former is based on the union as unit; it takes a single union as a single experiment, and gives you the mean result of these experiments; the latter is the ratio of the total number of paupers in the country to the total population. This ratio for 1st January of each year is given in the last column of Table IV for the sake of comparison, and it will be seen that it is uniformly considerably lower than the union average. We can only conclude that the mean population of unions with a high pauperism is low, so that the effect of the bad unions on the average is minimised when all are loaded proportionally to their population. The ratio of paupers to population (*cf.* Table IV) decreased from 5·8 per cent. in 1850 to 2·69 in 1891: the union mean from 6·51 per cent. in 1850, to 3·29 in 1891. In both cases almost the entire drop took place between 1870 and 1880.

Now the union average would be a very useful quantity if the frequency polygon were symmetrical, but when this is skew it can scarcely be said to be anything but a mere analytical convenience. It is not in any way typical. The mode is the type. The mode corresponds to the greatest frequency of occurrence. It is the modal value of the pauperism, not the mean value, that strikes one on looking at a map of the pauperism of English unions, that catches the eye on glancing through a return. It is the mode

consequently which should be used in comparing the state of the country at different periods, or in comparing a section of the country with the whole—provided materials are sufficient to warrant the extra labour. Nothing indeed can replace the data for the entire distribution, but the mode is at least better than the mean.

The values of the modes are given in the first column of Table IV, and their probable errors in the second column. We see that the mode is only slightly less reliable than the mean, its probable error averaging about $1\frac{1}{2}$ per cent. of its value.

The mode reflects the reaction of 1870 more clearly than the mean: we drop from a mode of 5.82 per cent. of pauperism in 1850 to 4.66 per cent. in 1860, rising again to 5.04 per cent. during the following decade. As with the means, almost the entire subsequent reduction was effected between 1870 and 1880.

Sec. 17. *Standard Deviation.*

The standard deviations are given in Table III. They range, practically speaking, from $2\frac{1}{2}$ per cent. in 1850 to $1\frac{1}{4}$ per cent. in 1891, *i.e.*, the standard deviation, or the divergence of the unions from the mean, has been halved. The decrease between 1861 and 1871 is very slight, but there is no actual rise such as takes place in the mean and mode. The change has been a little more continuous.

Surely this decrease of divergence is as great an improvement as the decrease of the mode. If the standard deviation had remained constant, we would have had in 1891 a considerable proportion of unions with 10 to 15 per cent. of their populations in receipt of relief (on one day), notwithstanding all the decrease in the mean. The change must point to greater uniformity either in economic conditions or in the administration of the law. Whether or to what extent the first factor has come into play I do not know, but the second most certainly has.⁹ The value of the work that has been done by all who have striven, officially or unofficially, to create a single ideal for the scattered unions is thus doubly exhibited, directly in the standard deviation, indirectly in the mode; neither characteristic seems to me complete without the other.

It was mentioned in sec. 8 that a range of six times the standard deviation formed a rough measure of the actual extension of the frequency-polygon. This rule may easily be verified in the

⁹ For one form of evidence I may be permitted to refer to my paper on "The Correlation of Total Pauperism, with Proportion of Out-relief."—"Economic Journal," p. 603, December, 1895; *vide* particularly the values of the standard deviations, pp. 607 and 608.

present case; ranges of 15 per cent. in 1850 and $7\frac{1}{2}$ per cent. in 1891 contain nearly all the actual deviations.

Sec. 18. *Criterion.*

Let us now pass on to consider the fitting of theoretical frequency-curves to these distributions. The three moment-coefficients, the numbers we called β_1 and β_2 , and the criterion κ , are given for each year in Table III.

The latter quantity is of the highest importance, as it shows by its sign, we remember (*vide* sec. 6), the type of curve to which the polygon must be fitted. The criterion in the present case remains positive throughout; all the curves from 1850 to 1890, notwithstanding their great diversity of form, belong to the "*limited-range*" type (Equation 4, sec. 5). The variation in the magnitude of the criterion is however somewhat curious; it starts in 1850 (when pauperism was rapidly decreasing, *cf.* Plate I) with the value 0.6; in 1860 and 1870, or, roughly speaking, at the beginning and end of the retrograde period, it rose to 0.9; in 1881 it had dropped to 0.7, and in 1891 to 0.4. The distributions appear to be slowly passing over to the simple binomial type (Equation 2), for which $\kappa = 0$.

The actual equations to the fitted curves are given in Table II. The unit of x in these equations, as in the moment-coefficients, is 0.5 per cent., *i.e.*, the unit of grouping. If

$$y = y_0 \left(1 + \frac{x}{a_1}\right)^{m_1} \left(1 - \frac{x}{a_2}\right)^{m_2}$$

be taken as before as the general form, y_0 is the maximum ordinate, the zero value of x corresponding to the mode (Table IV). The curve extends for a_1 units to the left of the mode, and a_2 units to the right, so that the total theoretical range is $a_1 + a_2$ units.

The probable errors of the indices m are tabulated separately in the following table (V), and also the limits of the range with its probable error.

Sec. 19. *Range.*

The present curves form admirable examples of our previous remark, that the range is in no way a measure of scatter or divergence from the mean. The range in 1860 runs up to 14.63 per cent.; in 1891 up to 15.10 per cent. To call the divergences in these years nearly equal would evidently be absurd. The standard deviations of the same years are 2.08 and 1.25 per cent. respectively.

The lower limit of the range comes in every case very close to the absolute origin, the zero value of pauperism. In 1850 and 1891 it lies slightly to the negative side, but in neither year is

there any sensible negative frequency. We might probably have considerably simplified the arithmetic by assuming zero to be always the lower limit of the range, but it seemed best on the whole to use the more general if more laborious method. It is of considerable interest to find the curve giving so reasonably closely the limit that is known *à priori*, and warrants us I think in accepting the other.

The only considerable change that has taken place in the range is the drop from 23.7 per cent. in 1850 to 14.6 per cent. in 1860; from 1860 onwards the alterations are smaller. 1891 shows a rise; this is in fact inevitable, if, as we saw in the last section, the distribution is going to pass over to the binomial type. The theoretical range in the binomial type is infinite.

Sec. 20. Indices m and Ratio m/a .

The indices m , and the ratios of the m 's to the quantities¹⁰ a (the ranges from the mode) are important characteristics of the distribution, though one cannot, as in the case of the range, give them any immediate geometrical interpretation. Their values will be found in Table V.

One of the most curious points to be noted is the way in which the indices are affected during the 1860-70 period. The changes are most noticeable in m_2 , the values of which are, from 1850 onwards, 11.22, 5.58, 5.80, 10.67, 20.17. It is not quite clear here what the variations indicate: if the general tendency has been to a rise of m_2 one would have expected a high, not a low, value in 1860. The average pauperism in 1860 (*vide* Plate I) was lower than it had ever been before; the rise did not commence till the winter of 1860-61, and the present figures refer to January, 1860. I must nevertheless confess with the figures for 1870, 1881, 1891 before me, that if m_2 shows a fall in the next available return, I shall expect a rise in the mean to follow, whether it has then risen or no.

The ratio of m/a also shows an irregular rise from 1870 to 1891, but 1860 is in this case no exception. As remarked in sec. 5, the larger the ratio of m/a the smaller the standard deviation of the curve compared with its range.

Sec. 21. Skewness.

The skewness of the distributions is another function of considerable interest (Table V). As in the case of m_2 , the skewness is less in 1860 than in 1850. From 1860-70 there is a considerable fall, then a rise to 1881, and a slight further rise in 1891.

¹⁰ Remember that $m_1/a_1 = m_2/a_2$.

Now Professor Pearson has noted, in the paper to which we have been frequently referring, that in a variety of cases in which the statistical material is varying, and the direction of variation is known *à priori*, the mode is shifted away from the mean in the direction in which such variation is taking place: he then suggests that this may give us a criterion for the direction of variation in all cases, though it will not be safe to lay stress on it until more evidence has been accumulated.

In the present set of diagrams we have unluckily no single curve referring to a period during which the mean was markedly rising. Let us see, however, in what way the skewness actually did vary. In 1850, during a period of improvement, it was 0.240; in 1860, after the very rapid decrease during the preceding decade, it reached 0.318. After the stagnation of 1860-70 it fell to 0.204, and then rose again during the improvement of the next twenty years to 0.253 and 0.275. This seems to me to rather remarkably bear out the spirit of Professor Pearson's suggestion: the mode is on that side of the mean to which the secular variation is tending, and periods of stagnation shift it back. It would require, however, a reaction more severe than any we have experienced to actually reverse the sign of the skewness.

Sec. 22. *Point-Binomials*.¹¹

These form some of the chief points to be noted with regard to the skew-curve constants and general characteristics of the distributions. A little remains to be said with regard to the binomials, the constants of which are given in Table VI. It must be remembered here that the skew-curves are complete in themselves; they give the whole form of each distribution. Further work would in general be needless. I have simply worked through the binomials to see if they would throw any further light on the subject, and to exemplify the method.

We have in the point-binomial three independent constants, say p , n , and c (*vide* sec. 5).

The chance p falls from 1850 to 1870, rises very largely during the following decade, and very slightly between 1881-91. The greater p the larger the proportion of unions below the average.

The number of contributory cause groups n varies quite differently. There is a considerable fall between 1850-60, and a continual rise, most markedly during the last decade, from 1870 onwards. The variation in n is in fact most curiously like that of the skew-curve exponent m_2 , as the following table shows:—

¹¹ In the binomial constants for 1891 which were given as an example in Professor Pearson's paper (p. 405), there are errors for which I am responsible; the figures now given are correct.

Year.	1850.	1860.	1870.	1881.	1891.
n	6.31	3.03	3.34	5.86	13.54
m_2	11.22	5.58	5.80	10.67	20.17
m_2/n	1.8	1.8	1.7	1.8	1.5

If we may give n anything more than a purely formal significance (compare sec. 10), the rise from 3.3 in 1870 to 13.5 in 1890 is of considerable importance. It would indicate a greater complexity in the economic conditions of pauperism, a change probably beneficial, as tending to minimise variations due to any one factor. But whether such interpretation is valid or no can hardly be decided till we have a much wider field of examples on which to base conclusions. The growth of n and decrease of c again points to the passage of the distribution towards the binomial and ultimately the normal form.

Sec. 23. *Fit of Curves.*

One curious change in the statistical distributions we have left hitherto unnoticed, though it strikes the eye immediately; I refer to the smoothness of the polygon for 1891 compared to the jagged character of all the preceding ones. I am wholly unable to account for this. The change is partly indicated in the following table, showing the fits of the skew-curves to the statistical polygons as measured by Professor Pearson's method:—

Year.	1850.	1860.	1870.	1881.	1891.
Misfit per cent.....	11.4	10.0	10.0	9.3	4.5

These numbers may I think be considered fairly satisfactory, though in the earlier years the errors are large. In a series of normal curves based mostly on a larger number of observations (1,000—2,000) the "misfits" ranged roughly between 7 and 9 per cent.

The greater homogeneity (sec. 11) in 1891 might be conceived to be due to the last twenty years' central direction and public discussion in tending to unify, more or less, the conceptions and actions of separate unions; we have noted such an effect as indicated by the standard deviation. But it is not clear why such a "smoothing" should not have made itself more strongly felt before 1850, when the central authority had been for long acting with a firm hand.

Sec. 24. *Concluding Remarks.*

It is almost impossible to summarise the points we have now been taking *seriatim*. The best and briefest "abstract," especially for those who do not care for symbols, is given by the plates themselves; they tell far more than a series of bald statements to the effect that the mean has been halved, the mode halved, the standard deviation halved, and so on, since 1850.

There are, however, a few general characteristics of the symbols still remaining that seem worthy of attention.

Variations in the rate of progress are reflected quite as clearly in the numerical characteristics of the form of the distribution as in the mean or mode, the standard deviation and skewness serving particularly well for this purpose.

Whatever constant we choose to take (mode, mean, standard deviation, skewness, index m , number of contributory cause-groups, or any other), we find a considerable change between 1850 and 1860, practically no change from 1860-70, a considerable change again from 1870-81, and a small further change in the same direction from 1881-91.

The constants may, however, be sorted into two classes: in the first the 1850-60 alteration is in the same direction as that effected between 1870 and 1891; in the second it is in the opposite direction. The first class includes mode, mean, standard deviation, and ratio of m/a (in the limited range curves). The second includes index m (in the limited range curves), chances and number of contributory cause-groups (in the point-binomials), the criterion, and the skewness.

It is not always easy to see exactly what a change in some numerical characteristic will connote, but I think it will be found that the above differences point to this: While in both periods of progress 1850-60 and 1870-80 there was a general shift of the whole distribution to the left (bringing about the lessening of mean, mode, and standard deviation), in the earlier period the change in the bad unions was disproportionately large (lessening the skewness and ratio of p/q); in the latter period the change has been more general and uniform, the curve for 1881 is in fact very approximately simply the 1850 curve squeezed up shorter (*cf.* Table II), so that the mean, or modal, unions must have undergone the greatest proportional change between 1870 and 1880 in order to bring the distribution back to its original form.

I am only too sensible of several faults of omission and commission from which this paper suffers.

As we have already noticed (sec. 13), the material is not quite uniform. The first three returns refer to the decennial years 1850, 1860, and 1870; the last two to the census years 1881 and 1891; the first three also refer to unions, the last two to registration districts.¹² Except in the case of 1850, both faults might have been corrected, but the necessary labour seemed wholly out of proportion to the end in view. No stress has to be laid on any small differences in the numbers, it is large and fundamental alterations with which we are concerned. Why the earlier Pell returns should have been made up in such an odd form is, however, quite another question, and one to which the answer is not obvious.

It must further be admitted that five frequency-curves do not form an adequate statistical record of forty years of poor law administration as regards even the one point with which we have dealt. Here, however, the fault is not mine. I attempted an interpolation for the year 1886. The population of each union was taken as the mean of the populations in 1881 and 1891, a procedure of course crude in the extreme and absolutely unreliable in the single instance; I only hoped that errors might be sometimes in excess, sometimes in defect, and so not tell seriously on the whole distribution. The number of paupers was obtained from the B return for the 1st January of that year. But the attempt proved a failure. After working out the moment-coefficients, the criterion was found to be negative, namely:—

$$= - \cdot 2634.$$

In view of the really remarkable uniformity exhibited by the previous series this was considered sufficient to justify the entire rejection of the results. The errors produced of simple interpolation must have been not random, but in some way systematic; any such systematic errors might easily bring about a change of type.

It is of little use at this hour to bewail the long series of statistics that have been rendered mere exasperating futilities by the lack of sufficiently frequent population returns, but surely they may well serve as warnings for the future. This country expends, I suppose, a considerable sum in procuring detailed returns of births, deaths, marriages, crimes, paupers, and so forth, yet without a knowledge of the population none of these returns are adequate, none convey any certain information, or lead to any accurate conclusions.

The chief objection to a more frequent census is, we are told, its cost. There may be some more practical reason. To the

¹² In my figures of Table I.

theoretical statistician it cannot but seem strange to simultaneously grant the cost of half useless returns, and withhold the cost of making them useful. We hope sincerely that the efforts of this Society towards a more frequent census may soon be crowned with success, and that we may not be condemned to exist for at least half of each decade in a condition of groping ignorance as to some of the simplest and most important questions of fact.

Even the small amount of matter I have been able to give illuminates, I hope, to some extent the purely statistical side of poor law history, and may perhaps, while showing where we stand under the present law, serve as some basis of comparison for the effects of any future changes.

The method employed is, I feel convinced, a most valuable one, not simply from the *picturesque* point of view (if I may be permitted to use the term, not in the æsthetic sense, but in the sense simply of giving a clear general picture of events), but from the point of view of numerical measurement. A diagram is good but lacks "quantification." You may have, we will suppose, two frequency diagrams for different years or different subjects before you; you may see that they are characteristically different; you may be able to do a little statistical word-painting and describe those characteristic differences; but such methods are lengthy and vague. Two or three curve-constants will express the whole thing briefly, neatly, and *exactly*. There is good reason to believe that, from a few hundred observations, they give you an approximation to the form of the diagram obtainable from an indefinitely extended series.

Supplementary Note on the Determination of the Mode.

The *mode* or most frequent value of the variable is a quantity of so much importance, that simple, if only approximate, methods for determining it are of considerable interest. The complete and exact method, requiring the determination of all moment-coefficients up to the fourth, is exceedingly laborious. Of the two following approximations, the first requires only μ_2 and μ_3 , the second only the median, in addition to the mean which is necessary in both cases.

Method I.

Assume the distribution to belong to the binomial type (equation 2, sec. 5), then

$$\text{distance from mode to mean} = \frac{1}{2}\mu_3/\mu_2.$$

Let us see how close this formula will work for the distributions

given in the preceding paper. The arithmetic is very brief; thus, in 1850

$$\begin{aligned}\text{Distance} &= \frac{1}{2} \frac{\mu_3}{\mu_2} = \frac{1}{2} \frac{55 \cdot 2199}{25 \cdot 2688} = 1 \cdot 092 \text{ units.} \\ &= \cdot 546 \text{ per cent.} \\ \text{Mean} &= 6 \cdot 508 \quad ,, \\ \therefore \text{Mode} &= 5 \cdot 962 \quad ,, \\ \text{True mode, from limited range curve} &= 5 \cdot 815 \quad ,,\end{aligned}$$

The table below gives the results for each year, obtained in the same way, with the true modes (Table IV of appendix) for comparison. These figures were worked out by slide rule, so only three places are given.

Year.	1850.	1860.	1870.	1881.	1891.
Mode { approximate..	5.96	4.84	5.17	3.34	3.03
true	5.82	4.66	5.04	3.24	2.99

The binomial-curve mode is in every case nearer the mean than the true mode, but the difference is not great. The approximation would serve for many practical purposes.

Method II.

By determining the mean and the median.

This approximation again depends on the binomial curve, for which the relation

$$\begin{aligned}\text{distance between mean and median} \\ = \frac{1}{3} \text{ of distance between mean and mode}\end{aligned}$$

is *very nearly* true.¹³

The only point here is as to the calculation of the median, which we remember (sec. 7) is given us by the vertical dividing the frequency-polygon into two portions of equal area.

Let $y_0, y_1, y_2 \dots y_n$ be successive ordinates of the frequency-polygon. Then the area of the polygon up to y_n , treated as a series of trapezia, is

$$\frac{1}{2} y_0 c + \frac{y_0 + y_1}{2} c + \frac{y_1 + y_2}{2} c + \dots + \frac{y_{n-1} + y_n}{2} c,$$

or as we may write it for more convenient calculation

$$c(y_0 + y_1 + y_2 + \dots + \frac{1}{2} y_n) \quad \dots \quad (1)$$

This seems on the whole a better formula to use than that

¹³ "Skew Variation," &c., p. 375. The relation is only approximate for the binomial type, and may hold closely for the others. It is an empirical result; there is no theoretical proof.

based on the assumption that the frequency-polygon is built up of rectangles, namely—

$$\begin{aligned} & \text{area up to } c/2 \text{ beyond } y_n \\ &= c(y_0 + y_1 + y_2 + \dots + y_n) \quad . \quad . \quad . \quad (2) \end{aligned}$$

The first of these expressions corresponds to the frequency-polygon as we draw it; the second is the ordinary straightforward procedure.

Neither method has any serious objection, and the results obtained by the two will in general only differ slightly. The first seems the better (especially when we have to interpolate between the ordinates of the polygon), as it follows more closely the contour of the fitted frequency curve.

We can exemplify the method by the figures for 1850:—

Total number of unions = 588		
∴ number up to median = 294		
y_0	=	1
y_1	=	4
y_2	=	2
y_3	=	7
y_4	=	11
y_5	=	21
y_6	=	28
y_7	=	23
y_8	=	46
y_9	=	55
y_{10}	=	40
<hr/>		
$\frac{1}{2}y_{11}$	=	22·5
		<hr/>
Number up to y_{11}		248
or 6 per cent.... }		45 = y_{11}
		22 = $\frac{1}{2}y_{12}$
<hr/>		
315 = number up to y_{12} , or 6·5 per cent.		
270·5 corresponds to 6·00 per cent.		
315·0	„	6·50 „

Therefore interpolating

294 corresponds to.....	6·264 per cent. = median.
Mean	= 6·508 „
Difference	= 244 „
× 3	= 732
Mean	= 6·508
∴ mode	= 5·776

The table below gives the values of the mean and mode

similarly calculated for each year, with the true modes again inserted for comparison :—

Year.	1850.	1860.	1870.	1881.	1891.
Median	6·26	5·00	5·38	3·52	3·21
Approximate mode	5·78	4·61	5·25	3·22	3·04
True mode	5·82	4·66	5·04	3·24	2·99

On the whole this method seems to give better results than the first; the figures lie sometimes on one side and sometimes on the other side of the true mode, and are not uniformly biased in one direction. The approximation is surprisingly close for every year except 1870.

For any set of observations then which give a frequency-distribution *not very far removed from the binomial type*,¹⁴ we may calculate the mode simply and easily, thus—

- (1). Calculate the mean.
- (2). Calculate the median.
- (3). Then the mode = the mean — 3 (mean — median).

The condition italicised is however rather puzzling; we can only decide whether a distribution is approximately binomial or no by calculating the four moment-coefficients and the criterion! But judging from the present case the condition is fairly lax, and will frequently not obtrude itself. I should not myself be inclined to call some of these distributions (*e.g.*, 1860 and 1870) “approximately binomial” at all. The approximation may possibly hold more generally than has yet been proved.

If in a long series of curves, referring to the same material, the approximation had been shown at length to hold for two or three, it would probably be quite safe to employ it for determining the mode in the remainder. It is obvious that, at all events in such flat-topped distributions as those occurring in 1860 and 1870, it will be difficult to get anything even remotely accurate by interpolating between the greatest ordinates actually occurring.

¹⁴ Excluding the particular form of binomial curve in which the ordinate is infinite at one end.

APPENDIX.

In several of the following tables the constants have been calculated to more places than are probably necessary.

The "probable errors" were only reckoned some months after the completion of the rest of the work.

TABLE I, giving, for each Year, the Number of Unions with the Pauperism specified in the left hand column. These figures are shown graphically by the thick lined Polygons in Plates II—VI.

Pauperism per Cent.	Year.				
	1850.	1860.	1870.	1881.	1891.
0·5	1	1	1	—	—
1·0	4	4	2	4	18
1·5	2	15	5	30	48
2·0	7	21	18	58	72
2·5	11	29	24	88	89
3·0	21	48	47	83	00
3·5	28	65	51	97	90
4·0	33	60	52	73	175
4·5	46	50	59	52	60
5·0	55	58	52	58	40
5·5	40	46	48	41	21
6·0	45	45	63	26	11
6·5	44	54	52	15	5
7·0	35	37	44	3	1
7·5	44	38	40	—	1
8·0	31	24	39	3	—
8·5	27	14	16	—	1
9·0	34	15	12	—	—
9·5	21	8	8	1	—
10·0	11	5	8	—	—
10·5	12	2	1	—	—
11·0	11	3	2	—	—
11·5	7	1	2	—	—
12·0	7	1	—	—	—
12·5	3	—	1	—	—
13·0	1	—	—	—	—
13·5	3	—	—	—	—
14·0	4	—	—	—	—
Total....	588*	644	647	632†	632†

* No returns from many unions.

† Registration districts. In a few cases two or more unions go to a registration district, so that the total of 647 is reduced to 632.

TABLE II, giving the Equations to the Skew-Curves for each Year.

1850.

$$y = 45.952 \left(1 + \frac{x}{11.9545} \right)^{3.7808} \left(1 - \frac{x}{35.4900} \right)^{11.2242}$$

1860.

$$y = 58.568 \left(1 + \frac{x}{9.2128} \right)^{2.5780} \left(1 - \frac{x}{19.9521} \right)^{5.5832}$$

1870.

$$y = 61.614 \left(1 + \frac{x}{9.9291} \right)^{3.0164} \left(1 - \frac{x}{18.8950} \right)^{5.7972}$$

1881.

$$y = 89.653 \left(1 + \frac{x}{5.8594} \right)^{3.1778} \left(1 - \frac{x}{19.6719} \right)^{10.6687}$$

1891.

$$y = 99.9065 \left(1 + \frac{x}{7.1993} \right)^{5.9953} \left(1 - \frac{x}{24.2203} \right)^{20.1697}$$

The unit of x in all these equations is 0.5 per cent. The origin of each equation is the value of the mode for the same year (*vide* Table IV).

TABLE III.—Moment-Coefficients, Criterion, and Standard Deviation.

	1. μ_2 .	2. μ_3 .	3. μ_4 .	4. β_1 .	5. β_2 .	6. Criterion κ .	7. Probable Error of Criterion.	8. Standard Deviation per Cent.
1850	25.2668	55.2199	1895.729	.189035	2.96945	.628198	.014	2.51
1860	17.3859	24.6627	817.4970	.115742	2.70454	.938154	.015	2.08
1870	16.4083	18.6879	720.6289	.079055	2.67660	.883959	.017	2.02
1881	7.51163	10.1596	170.8800	.243529	3.02847	.673653	.012	1.37
1891	6.31889	6.62465	122.1815	.173942	3.06002	.401790	.010	1.25

Note.—The unit of length in μ_2 , μ_3 , and μ_4 is the unit of grouping, 0.50 per cent. The standard deviation is expressed in per cent. β_1 , β_2 , and κ are numerics, and unaffected by changes of scale. For the mode of reckoning the moment-coefficients, *vide* Karl Pearson, "Skew Variation, &c.," "Philosophical Transactions," vol. 186 A (1895), pp. 348–50 (trapezia-method).

TABLE IV.—*Modes and Means, with their Probable Errors.*

Year.	Modal Pauperism.	Probable Error.	Mean Pauperism.	Probable Error.	Ratio per Cent. of Total Paupers to Population.
1850	5·815	·083	6·508	·063	5·8*
1860	4·657	·058	5·195	·049	4·26
1870	5·038	·064	5·451	·049	4·77
1881	3·240	·041	3·676	·032	3·12
1891	2·987	·036	3·289	·032	2·69

* Approximate only.

TABLE V.—*Probable Errors of m_1 and m_2 ; Details of Range; Skewness.*

Year.	m_1 .	Probable Error.	m_2 .	Probable Error.	Ratio m/α .	Range.				Skewness.
						From	To	Total.	Probable Error.	
						Per cent.	Per cent.	Per cent.		
1850	3·78	·196	11·22	·320	·316	—·16	23·56	23·72	·40	·240
1860	2·58	·098	5·58	·156	·380	+·05	14·63	14·58	·22	·318
1870	3·05	·136	5·80	·180	·307	+·07	14·48	14·41	·22	·204
1881	3·18	·129	10·67	·269	·543	+·31	13·07	12·76	·21	·253
1891	6·00	·351	20·17	·631	·833	—·61	15·10	15·71	·28	·275

TABLE VI.—*Point-Binomial Constants.*

Year.	Chances.		Number of Component Cause Groups. n .	Distance between Ordinates. c .	Origin at
	p .	q .			
				Per cent.	Per cent.
1850	·78	·22	6·31	2·25	1·14
1860	·69	·31	3·03	2·32	0·73
1870	·66	·34	3·34	2·12	0·95
1881	·80	·20	5·86	1·31	0·86
1891	·83	·17	13·54	0·87	0·41

DISCUSSION *on* MR. YULE'S PAPER.

MR. FRANCIS GALTON¹ said that statisticians owed a large debt of gratitude to University College for the variety of the statistical investigations carried on there, both mathematical and experimental. Those pursued by Professor Karl Pearson, assisted by the author of the present paper, and those by Professor Weldon, had greatly raised the level of statistical science; they had enlarged its powers and widely extended its applications. Ordinary practical statisticians need not be alarmed at the technical terms, which would become familiar in time, nor at the high mathematics of the original investigations, with whose results they would be chiefly concerned.

The arithmetical operations were laborious; on the other hand, shorter methods sufficiently close for most purposes were almost sure to be discovered, such as that curious but as yet wholly empirical method of deducing the position of the mode from those of the median and the mean, which was found to hold good in a large class of cases. Mr. Yule's memoir was very welcome as an illustration of the process by which the new method was applied, and it enabled statisticians to form a useful idea of its value as compared to those of other methods. He (Mr. Galton) would venture to criticise one of the conclusions, namely, as to the trustworthiness of the calculated position of the mode, which was reckoned in the present cases to be nearly equal to that of either the mean or median. The values assigned to the mode for the years 1850 and 1860 differed however so largely from the judgment of his eye in respect to Plates II and III, that he could not accept them, and was compelled to suspect some insecurity in the foundation upon which the calculations were based. He should in both cases locate the mode much more to the left.

His next criticism was of a different kind, namely, an expression of doubt whether the new method *when applied to the particular cases* before them gave results that were better or even as good as the method of *percentiles* or, on this occasion perhaps, of *deciles*. Professor K. Pearson's formulæ were admirably applicable whenever large series of observations of the same variable always yielded the same smoothly running curve; but the present case was very different. The data ran roughly, and were unchecked by others of a similar kind, so that they did not seem to justify elaborate work. It might reasonably be maintained that such data as these admitted of being interpreted almost equally well in a variety of ways. Quite possibly the interpretation here given might be the most probable of all, and yet other interpretations might differ little from it in their respective probabilities. After

¹ In response to a request made at the meeting, Mr. Galton has written the memorandum on p. 392, giving fuller information as to the method of percentiles.

all, the conclusions derived by this method seemed to be little if at all more quantitative or more complete for ordinary statistical purposes than those derived from the method of deciles, while the diagrams which expressed them were certainly less graphic and less suitable for inter-comparison.

Mr. A. H. BAILEY said that without meaning to deny the great value of the paper, he thought that it was impossible to follow it sufficiently upon the reading for its discussion. He certainly had not had time to study it thoroughly, but his own work had familiarised him with the method suggested in the paper. The law of mortality had been represented by a curve in which the abscissæ represented the years of life, and the ordinates the corresponding mortality at each age. That was of course a simpler matter, showing clearly the law of mortality from the observations. Now there was no law of pauperism, and therefore the cases were not parallel.

Mr. A. W. FLUX felt very strongly how great an advantage it was to be able to describe a complicated set of figures by a curve, or by two or three standard measurements which should define that curve, as had been done in the paper before them. But the facts represented in the curves were not absolutely the facts recorded in the statistical returns, although the smooth-running curve did keep very close to the polygon formed by the actual figures. He felt very strongly, too, how far that simple diagram, which Mr. Galton had given them, went towards expressing nearly, if not quite, all that there was in this valuable paper, and yet the paper undoubtedly pointed out a way to develop statistical records in a more effective fashion than had hitherto been done. It seemed to him that the three measurements which practically characterised the curve, viz., the mode, the standard deviation, and the range, left out a great deal which was very valuable. With regard to pauperism in particular, they wanted to know what was the effect of a given policy; but it was absolutely impossible to represent that in these curves. Those parts of the country where various causes had reduced pauperism most rapidly were not separated out distinctly as places, and any connection between the cause and the result in particular districts was necessarily hidden from the representation here given. The figures varied, but the distribution of the figures only was shown. If there had been a definite change in the policy of the whole kingdom, then they might be able to trace the results on the whole kingdom to that definite change, but in some places there had been very little change, and in others great change of policy, and they were not able by this method to connect these changes with the results. It would have been interesting if they could have seen how the mode and the standard deviation varied in the course of a year, but there would be a difficulty very similar to that which Mr. Yule found in dealing with the figures of 1886. In the course of a year the population in many unions changed a great deal and, as they did not know what these changes were, it was

not possible to get the percentages of pauperism to population. The great difficulty in dealing with the figures for 1886 was due to the fact that, although they knew the number of paupers in each union for that year, they did not know the population sufficiently accurately to be able to get the percentages. Was it not possible that the method which had been adopted of using the figures of population for 1851, combined with the figures of pauperism for 1850, and similarly for 1860 and 1870, might give results for those three periods not quite comparable with the results for 1881 and 1891, and that the changes of population in the course of that one year might have caused these results to be misleading? Following the suggestion of Mr. Yule in the paper, he had very hastily compared one hundred unions in Yorkshire, Lancashire, and Cheshire with those for the whole country, and the results he obtained were very curious. He found that the standard deviation for this area for 1891 was less than any of the standard deviations that Mr. Yule had worked out. Another very curious thing was that, in the case of these one hundred unions in 1881, the median was not on the same side of the mean as in all the five cases which Mr. Yule had worked out, but fell on the other side. He should hesitate in reference to the conclusion that the side of the mean on which the median lay indicated the tendency of the development, because that would mean that the policy adopted in Yorkshire, Lancashire, and Cheshire tended to increase the pauperism of those districts. Taking Mr. Charles Booth's figures, from which he understood that Mr. Yule had worked, and comparing 1881 with 1891, he found that the criterion rose very considerably. But it was quite possible that one hundred unions was too small a number to yield trustworthy results.²

Mr. C. S. LOCH said that he was not familiar with the methods adopted by Mr. Yule. He was greatly interested in the subject; but, so far as he had dealt with it statistically, he had worked only on the rules of simple arithmetic. The effect of increase or decrease of population upon statistics of pauperism had often puzzled him, and he felt sure that many deductions had been made respecting the latter which were fallacious. Results sometimes seemed satisfactory which were due primarily to the diminution or increase of population, and not to the causes assigned to account for them. The same remark applied also to local statistics of ratings and other matters. It was to be hoped that expert statisticians might enable them to deal more satisfactorily with a point of this kind. He was much interested in what the last speaker had said about one hundred unions in the north of England. The results were just what he would have expected. The deviation would be less, and ought to be less. Where there was least pauperism there would actually be least deviation. He believed that Mr. Yule's figures expressed total numerical results irrespective of causes;

² A further examination of the figures shows that this discrepancy really does arise from the smallness of the number of unions, and a calculation of the median by a method better adapted to this small number, gives a result in accordance with those obtained by Mr. Yule.—A.W.F.

but it was interesting to attempt, if possible, to isolate causes. With the aid of local knowledge it might be possible to work carefully over the whole area of England, and group unions on a logical and scientific method from this point of view; and either Mr. Yule's method or Mr. Galton's, which seemed to him extraordinarily clear, might be applied so as to find some numerical expression of the cause. No one who had tried to work out these figures by arithmetical means only, could help feeling that instruments of much greater precision were wanted. As a very important contribution towards this result he welcomed Mr. Yule's paper very heartily. He hoped that it might be followed by others that would in an equal degree force them to strenuous thinking and a reconsideration of statistical methods.

Mr. W. F. SHEPPARD said that he had two criticisms to make on Mr. Yule's method. In the first place, he was not satisfied as to the *à priori* grounds on which the kind of curve used by Mr. Yule was taken to be the appropriate curve for the particular data. It seemed to him that the predominance of the normal symmetrical curve, both in theory and in actual statistics, made it desirable to test the data first for such a curve, and then to consider whether the discrepancies might not be due to paucity of observations. The total number of unions was not very great, and this would account for a good deal of irregularity. Secondly, in the absence of these *à priori* grounds for supposing the proper curve to be a curve of the particular kind adopted, the median seemed to be a better criterion of the degree of pauperism than the mode. The median could be obtained, within small limits of error, from the actual observations, whereas in determining the maximum ordinate from the observations an appreciable error might be caused by inaccuracy in the amount of pauperism in a few unions. Thus, for the year 1860, an alteration of the relative position of some ten unions, by perhaps a very slight alteration in their actual pauperism, would make a great change in the position of the maximum ordinate.

Mr. J. A. BAINES differed from Mr. Bailey with regard to this paper, for he thought that it made more widely known to a class that did not habitually deal with higher mathematics, the value of such methods when adapted to actual statistics. As the last speaker but one (Mr. Loch) had truly observed, the results might not hold good for a special area, but the avowed object of the paper was to reduce figures for local areas to one general total, and to show the progress or retrocession of pauperism for the whole of England. No doubt it was easy to test general results by a group of unions, but the methods set forth in this paper seemed almost the only means of combining the whole body of the statistics into one concrete form. He had often experienced the impossibility of getting accurate results from interpolated figures, and that was an additional argument, if indeed one was needed, for the contention so often and strongly put forward by the Society for a quinquennial census, which he hoped they would never cease striving

for. This paper dealt almost entirely with methods, and he had hoped that Mr. Loch, than whom there was no one more conversant with statistics of pauperism, was going to tell them whether, in his view, the results of this mathematical reduction were in harmony with the tendency in England generally. The way in which the widely spread curve of 1850 had shrunk to the narrow and concentrated curve of 1891, seemed to him to prove most clearly that there had been remarkable inclination towards greater uniformity of administration, that is, the great improvement shown in the ratios dealt with in the paper was due to the much more uniform system on which the unions all over the country were administering relief; this feature being far more striking in the curves than the variation in the actual numbers.

Mr. N. A. HUMPHREYS said that no one felt more strongly than himself the necessity of a more frequent census, but at the same time he thought several speakers had suggested difficulties with regard to the arithmetical method of determining the proportion of pauperism to population which were scarcely valid. It was quite true that during a current decennium it was impossible to make satisfactory local estimates of population, but when a decennium was completed, there was no great difficulty in estimating the population of defined areas for any given year with quite sufficient accuracy to serve as a basis for calculating ratios of pauperism, and for testing the effects of variations in poor law administration.

Mr. R. H. HOOKER, referring to the statement on p. 340, that the distributions appeared to be slowly passing over to the simple binomial type, in which the curve was limited at one end and tailed off to infinity at the other, pointed out that inspection of the diagrams of pauperism led one to expect this result, which the author had arrived at by calculation. For as the general tendency of the mode was to approach the ordinate of no pauperism, the rates less than the mode were all crowded into a smaller space, and the tendency of the curve to cut the axis instead of to touch it, *i.e.*, to be limited towards the left, became greater, while there was still room for them to extend indefinitely to the right. This passage from one type to another led to the consideration of the method of determining the class of curve to be used. This was ascertained by examination of the value of κ . If κ was "approximately zero" the curve would be of the binomial type; if it was positive, the curve would be of the limited range type. The question thus was, What amount would be considered "approximately zero?" and what indication would one have that a wrong kind of curve had been taken? Would there be any abnormal result? As the curve actually seemed to be passing from one type to another, the point was of practical interest.

On p. 330 it was mentioned that the ratio of skewness was zero for a symmetrical curve where the mode and the mean coincided, and was never greater than unity. A geometrical interpretation of a skewness unity would be useful, and he would like to know if he was right in deducing that the skewness would be unity in the

improbable case, for instance, of pauperism being the exception, *i.e.*, if the maximum number of unions had a pauperism of less than $\frac{1}{4}$ per cent., so that the mode coincided with the lowest value of the variable.

With regard to the author's endeavour to ascertain the form of the curve for 1886, he should not himself expect any reliable result from an interpolation based upon the usual estimates of the population in intercensal periods. These were notoriously uncertain; in fact, had the author obtained a reasonable value for κ , he (Mr. Hooker) would have been inclined to take this result as indicating that the estimate of population for 1886 was a correct one. He was, however, rather disposed to think that the distribution of population, as between urban and rural districts, might be materially different in 1886 to what it was both in 1881 and 1891, in which two years it was probably somewhat similar. He was inclined to think that such a different distribution would result from the fact that 1881 was near a year of maximum trade prosperity, and 1891 still more so, while 1886 was a year of minimum trade prosperity. The population estimates for 1886 being thus probably peculiarly inaccurate, it would follow that the estimates of pauperism would be so also. In 1870 the country was beginning to revive from a depression, and in this year pauperism was practically a maximum, so that a very different type of curve to that for 1881 would naturally be expected. These facts, Mr. Hooker thought, would account in part, possibly only to a slight extent, for the great difference in the type of curve between 1870 and 1881, and the great similarity between 1881 and 1891. 1886 having been somewhat similar, as regards trade, to 1870, it would therefore have been interesting if it had been possible to work out the values of σ and other particulars for 1886, to ascertain whether the curve showed any tendency (however slight) to revert to the form of 1870.

Sir RAWSON W. RAWSON, K.C.M.G., C.B., said that the Society and statistical science generally were greatly indebted to Mr. Yule for showing how valuable and important, under careful management, mathematical processes were in testing and correcting the masses of statistical information which it is the province of the Society to collect and apply. He thought it was an advantage to the Society that the Council had admitted this paper for discussion, in order that other persons might be encouraged to use similar efforts to apply mathematical processes scientifically to statistical investigations. He also hoped it would encourage Mr. Francis Galton to give them an explanation of his method of working out problems of the same character.

The CHAIRMAN (The Right Hon. the Earl of Verulam) then moved a very cordial vote of thanks to Mr. Yule for the great care, labour, and accuracy he had displayed in bringing forward this very difficult subject. He would not attempt to deal with the details of the paper, but it was obvious what value such a paper must have, by showing the effect of a given policy upon the amount

of pauperism prevailing in the country, so that the lessons taught by such statistics could be acted upon by those who had the management of affairs; and without such methods of ascertaining the actual facts, obviously their policy could not be directed on the best possible lines. With regard to the statistics given by Mr. Flux in connection with the one hundred unions in Lancashire, Yorkshire, and Cheshire, it struck him that the reason for the variation he spoke of might possibly be that certain industries had ceased and others had sprung up in the respective districts. They often heard of towns being depleted of population in consequence of the cessation of a normal industry, the working out of mines, or the transference of industries from one place to another. One important point which Mr. Yule had very properly urged, was the necessity for more frequent censuses. He had no doubt that, to be of any real value, a quinquennial census certainly was desirable, and he was inclined to say that an annual one would be most valuable.

Mr. YULE, in reply, said he would deal first with a point raised by Mr. Galton, and again by Mr. Sheppard, as to the position of the mode. The latter speaker (Mr. Sheppard) appeared to think the mode corresponded to the greatest ordinate actually occurring; of course this was not so, it corresponded to the maximum ordinate of the fitted curve. Being thus determined (in a somewhat similar way to the mean) from the totality of the observations, the mode was entirely unaffected by the shifting of a few units from one side to the other near the actual maximum, and Mr. Sheppard's criticisms on this point were entirely beside the mark. Such a shifting of a few units would however be quite sufficient to entirely upset the judgment of one's eye, and for that reason he (Mr. Yule) did not consider that any great weight could be attached to simple estimates of the mode such as Mr. Galton had made.

With the use of percentiles, advocated by Mr. Galton, he preferred not to deal at length on the present occasion; the chief fault of the method in his eyes was that it did not lend itself to theoretical treatment. The value of a given form of diagram was always more or less a matter for personal opinion, but it seemed to him that five frequency distributions (without the fitted curves at all) really gave much more than five decile systems.³

With Mr. Bailey's remark, that the cases of pauperism and mortality were not parallel, he was in complete agreement. It was however quite as important to have a concise statement of facts in the first case, where changes were rapid, as in the second case, where changes were slow. The "law" of pauperism was a law of distribution round an average.

Mr. Flux's criticism, that an investigation of causes was most required, seemed to him perfectly just. But for such an investigation it was necessary to deal with a distribution in three dimensions, *i.e.*, a frequency-surface, not a frequency-curve. He (Mr.

³ Some future remarks on the use of deciles and normal curves will be found on p. 396, in a note on Mr. Galton's memorandum, p. 392.

Yule) had used this method for one investigation⁴ in pauperism, and hoped to be able to do further work in the same direction. The method of the present paper simply served to illustrate, without reference to causation, the breadth and character of the changes that had taken place; it did not profess to do more.

Mr. Flux's further remarks on the compactness of the distribution in a group of unions from one district were of great interest, and in complete agreement with what he (Mr. Yule) had noticed. A group, for instance, from a single registration division nearly always gave a much smaller standard deviation than that given by the whole country. The geographical limitations of pauperism in England were very marked.

Passing on to Mr. Sheppard's more theoretical criticisms, he differed from that speaker on one point of fact: the normal curve was by no means predominant in statistics, it was (on the contrary) decidedly exceptional. As regards the skew-curves employed, Mr. Sheppard did not appear to have followed the method. The normal curve was given its chance; *i.e.*, the criterion was calculated, and if it had been near zero the normal curve would have been used. But the criterion was distant from zero by forty to eighty times its probable error! Under any circumstances how could a normal curve give a distribution sharply limited at one end, as these are at zero pauperism?

He (Mr. Yule) also failed to understand why the median was any better "criterion of the degree of pauperism" than the average; the one was no more the type than the other.

Mr. Humphreys had suggested that it was quite sufficiently accurate in most cases to interpolate for population; but surely the process was very uncertain for small areas like unions, largely affected by migration? For the country as a whole it might possibly serve.

With regard to Mr. Hooker's queries, by the statement that the criterion must be small, was meant that it should not be large compared with its own probable error.⁵ There would be no very noticeable differences between one type of curve and another, when the criterion was small; no abnormal results would be obtained by using the wrong form under these circumstances.

The skewness would be unity, for example, in the type of binomial curve, with the maximum at one end of the range.⁶

The questions as to the condition of 1886, suggested by Mr. Hooker, were interesting, but it seemed impossible to arrive at any conclusions owing to the wretched population difficulty. Redistribution actually discovered for 1886, showed no tendency to revert to the form of 1870, but the reverse. The standard deviation found for 1886 was smaller than that for 1891! But of course no weight could be attached to these results.

⁴ "On the Correlation of Total Pauperism with proportion of Out-relief." "Economic Journal." December, 1895.

⁵ The phrase in Sec. 6, to which Mr. Hooker referred, has been altered since the proofs were issued.

⁶ Curve II, fig. v, plate 9, "Skew Variation, &c." Karl Pearson. "Phil. Trans.," 186 A. 1895.