London has little bearing upon that of the United Kingdom, since it so largely depends upon intra-migration.—G.B.L.]

II.—*Application of the Method of Percentiles to Mr. Yule's Data on the Distribution of Pauperism.* By FRANCIS GALTON, F.R.S.[1]

WHEN any large group of statistical cases is sorted into a hundred classes equal in number, and progressively increasing in value, the dividing values between the classes are called Percentiles;[2] or, if into ten classes, they are called Deciles; or if into four classes, they are called Quartiles. The fiftieth percentile, the fifth decile, and the second quartile are consequently the same as the median. All other deciles, &c., are calculated on the same principle as the median, and in the way that Mr. Yule has already explained. Still, it will be convenient to workers to possess a form by which a complete set of deciles may be compactly worked out, so Table A is printed as an illustration. Its headings sufficiently explain the process, with perhaps the following exceptions: In the first column the entries at either end of the series have been thrown into the separate lumps of below 1·75 and above 10·25, while instead of the remaining entries appearing as in Table I (p. 347), under the form of 2·0, 2·5, &c.. they are written as 1·75 to 2·25; 2·25 to 2·75, and so on, those being the values to which the former figures really refer. The small number of cases that fell exactly on the dividing lines had been sorted, as I am informed by Mr. Yule, as equally as might be into the adjacent compartments. The multiplier 0·5 is introduced into column F, because the intervals between the successive lines refer to differences of 0·5 per cent. Usually I calculate the 0·5th and 9·5th deciles in addition to those given in the table. If the ends of the series were abrupt, the 0·1st and the 9·9th might also appear. The deciles being merely a condensed method of description, those additional values should be selected that are most suitable to the intercomparison of the particular series that are to be dealt with. The 2·5th decile is identical with the 1st *quartile* and the 7·5th with the 3rd; they are not used here. The quartile (irrespectively of its + or − sign) in a normal (symmetrical) series, is identical with its "*probable error.*" Table B contains the deciles 1 to 9, for each of the five years. The

[1] For Mr. Galton's remarks in the discussion which followed Mr. Yule's paper, see p. 350.

[2] The principles of the method of percentiles, and the ogival curve by which I commonly expressed them, are fully explained in my *Natural Inheritance*, p. 46 (Macmillan, 1889); the particular form under which percentiles or deciles are here used, was employed in a paper by me on "Assigning Marks for Bodily Efficiency," see *Report British Association*, 1889, p. 475.

Distribution of Pauperism.

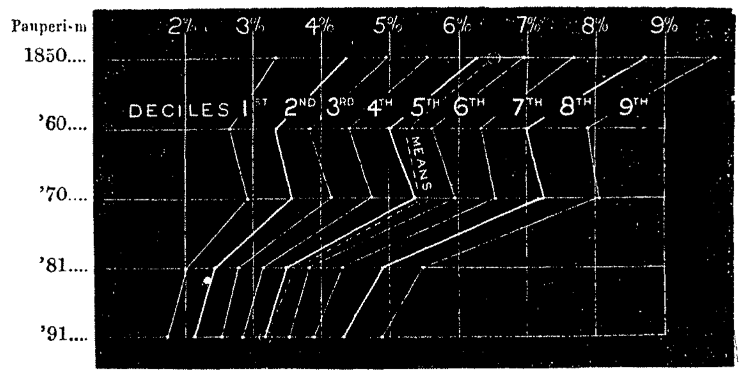

diagram is drawn from the contents of Table B, to afford a graphic view of the changes during the period under discussion. Table C contains three selections for each year from Table B, namely the median, and the values of the deciles 2 and 8 after the median has been subtracted from them; it contains two simple deductions from these, and it also utilises the mean (its value having been taken from Mr. Yule's Paper). The result is that four entries for each year give all the necessary data for such quantitative comparisons as statisticians want, between that and the remaining years. Two other entries which have been deduced from the above are added for convenience, making six entries in all for each year. The diagram and the small Table C appear to me simpler and more distinct, and to be as useful and accurate in the present instance for the ordinary purposes of the statistician, as are the five plates and the results of the elaborate calculations that accompany Mr. Yule's paper. The former were compiled with ease; they show the changes of the median and mean from year to year; they measure the *spread* of each series by means of the difference between the 2nd and 8th deciles (which forms an excellent measure of the spread), and they give a fairly good indication of the symmetry or skewness of the curve (by the *ratio* which is entered in the bottom line) between the deviations of deciles 2 and 8. In a symmetrical curve those deviations being equal, the ratio is 1; in a skew-curve it is less than 1 when the mode is to the left of the median, greater than 1 when it is to the right. Besides this, the data contain materials for applying the empirical method of Professor Yule by which, in numerous cases, the mode may be determined, namely that mode $= \{\text{mean} - 3 \times (\text{mean} - \text{median})\}$ or, using u to express the entries in the second line of Table C, mode $=$ mean $- 3u$, but, as the mean $=$ median $+ u$, this may be changed into mode $=$ median $- 2u$. Anyhow the line u gives a second measure of asymmetry. Whenever the series to be dealt with are more symmetrical than these, there is yet another result obtainable from such data as those in Table C, namely, the power of approximately reconstructing the entire curve. It is effected by the help of the normal curve, and succeeds in not a few cases with curious exactness. The deciles of deviations appropriate to the normal curve, which have to be used for this purpose, are given in the bottom line of Table B. I have, for curiosity's sake, attempted a reconstruction on this principle of the absent deciles in C, for the year 1891, which is, on the whole, the most symmetrical of the set, and find the errors of fit to be 0·19, 0·07, 0·03, 0·01, 0·0, 0·01, 0·03, 0·07, 0·09.

Another form of reconstruction is by the extremely rude and scarcely defensible method, but still a sometimes serviceable one, of looking upon skew-curves as made up of the halves of two different normal curves pieced together at the mode. One must be guided by the run of the deciles to know whether this process would be tolerably successful in any particular case. On trying it, again for curiosity's sake, with the present series for all the five years, there was of course no error for the 2nd, 5th, and 8th deciles, while as regards the other six ($6 \times 5 = 30$ cases) I found the five largest misfits to be 0·33, 0·22, 0·22, 0·20, 0·14, and the five smallest to be 0·00, 0·00, 0·01, 0·02, 0·2. The average misfit of these six deciles in all five years was 0·09, but if the misfits were divided among all the 9 deciles, that is among 45 cases, the average would be 0·06.

We have now to utilise the diagram and Table C for arriving at conclusions. The first obvious fact is that the year 1870 is quite out of the run of the other four years. What that run is can be seen by laying a narrow strip of paper on the diagram so as to entirely cover all that lies between 1860 and 1881, then to connect with lines drawn on this paper the corresponding ends of the deciles above and below; also to substitute in brackets ( ), as below, the interpolated values for 1870 in the place of the observed ones. The lines and figures for the complete system will now run consistently; they show a steady decrease in the median percentage of pauperism (as determined by the returns from unions), the figures being 6·26, 5·00, (4·26), 3·52, 3·20. There is also a general decline in its variability throughout the kingdom, the measures of the relative *spread* of the series being 4·31, 3·67, (3·08), 2·49, 2·20. Moreover the skewness diminishes regularly as shown by the line of u in Table C, which gives the figures 25, 20, (18), 16, 9. It is similarly shown by the increasing approach to 1·0 of the values in the bottom line (with however one exception which also affects the interpolated value), viz., 78, 83, (80), 77, 88.

When the strip of paper is removed and the year 1870 is compared with the rest, the temporary increase of pauperism from what may be called its normal rate of change, as expressed by the interpolated values, is very striking, namely, from 4·26 to 5·39, or by 1·13 per cent. In spread, the increase is from 3·08 to 3·68, or in the ratio of 1 to 1·2. The skewness diminishes at the same time in the ratio of $\{5{\cdot}39 - 2 \times (0{\cdot}18)\}$ to $\{5{\cdot}39 - 2 \times (0{\cdot}06)\} =$ 503 to 527, or as 0·95 to 1. It follows that the distribution of pauperism as the years go by tends to assume a symmetrical and normal form. In other words, the causes of its variability tend to become of the same general character as those which occasion errors in ordinary observations.

I do not propose to analyse the causes of the march of these figures, having no special knowledge of the subject of pauperism, which I regret, as the year 1870 would apparently reward a competent attempt at analysis. These remarks are merely offered as an example of the method of deciles, with the hope of proving that it gives the quantitative results that are chiefly desired by statisticians, and a very graphic map of changes, by remarkably simple means.

TABLE A.

(A.) Pauperism. Per Cent.	(B.) No. of Unions.	(C.) Sums of B from top.	(D.) Successive tenths of the total of B.	(E.) D − C (in each row).	(F.) D − C multiplied into 0·5.	(G.) D − C × (0·5) and divided by B_1 *	Interpolated Deciles: Order.	Value (G + A) †
Below 1·75	7	7	—	—	—	—	—	—
1·75 to 2·25	7	14	—	—	—	—	—	—
2·25 „ 2·75	11	25	—	—	—	—	—	—
2·75 „ 3·25	21	46	59	13	6·5	0·23	1st	3·48
3·25 „ 3·75	28	74	—	—	—	—	—	—
3·75 „ 4·25	33	107	118	11	5·5	0·12	2nd	4·37
4·25 „ 4·75	46	153	176	23	11·5	0·21	3rd	4·96
4·75 „ 5·25	55	208	235	27	13·5	0·34	4th	5·59
5·25 „ 5·75	40	248	—	—	—	—	—	—
5·75 „ 6·25	45	293	294	1	0·5	0·01	5th	6·26
6·25 „ 6·75	44	337	353	16	8·0	0·23	6th	6·98
6·75 „ 7·25	35	372	412	40	20·0	0·45	7th	7·70
7·25 „ 7·75	44	416	—	—	—	—	—	—
7·75 „ 8·25	31	447	470	23	11·5	0·43	8th	8·68
8·25 „ 8·75	27	474	—	—	—	—	—	—
8·75 „ 9·25	34	508	—	—	—	—	—	—
9·25 „ 9·75	21	529	529	0	0·0	0·00	9th	9·75
9·75 „ 10·25	11	540	—	—	—	—	—	—
Above 10·25	48	588	—	—	—	—	—	—
	588	—	—	—	—	—	—	—

* B_1 in column G means the entry in column B that lies *one line below* that on which the entry in F is standing. Thus 6·5 is divided by 28, and 5·5 by 46.

† The second decimal is approximate.

TABLE B.—*Deciles.*

	1st.	2nd.	3rd.	4th.	5th.	6th.	7th.	8th.	9th.
1850	3·48	4·37	4·96	5·59	6·26	6·98	7·70	8·68	9·75
'60	2·65	3·34	3·83	4·40	5·00	5·63	6·33	7·01	7·94
'70	2·91	3·57	4·19	4·75	5·39	5·99	6·55	7·25	8·06
'81	2·00	2·44	2·81	3·19	3·52	3·88	4·34	4·93	5·57
'91	1·72	2·17	2·54	2·88	3·20	3·54	3·92	4·37	4·96
Decile deviations in a normal curve of which that for 2·5 = − 1, and that for 7·5 = + 1.	− 1·90	− 1·25	− 0·78	− 0·38	0·00	+ 0·38	+ 0·78	+ 1·25	+ 1·90

TABLE C.

Selected Data.	1850.	1860.	1870.	1881.	1891.
Median	6·26	5·00	5·39	3·52	3·20
Deviations from Median					
of the Mean (u)	+ 0·25	+ 0·20	+ 0·03	+ 0·16	+ 0·09
of Decile 2 (v)	− 1·89	− 1·66	− 1·82	− 1·08	− 1·03
of Decile 8 (w)	+ 2·42	+ 2·01	+ 1·86	+ 1·41	+ 1·17
Spread between Deciles 2 and 8	4·31	3·67	3·68	2·49	2·20
v divided by w	0·78	0·83	0 98	0·77	0·88

Note.—The signs of the deviations are disregarded in the bottom line.

III.—*Remarks on Mr. Galton's Note.* By G. U. YULE.

BY the courtesy of Mr. Galton I am enabled to offer a few remarks on his preceding note, which was sent to me for criticism before communication to the Society. The somewhat wide range of points touched upon by Mr. Galton in this note and in his remarks at the meeting may perhaps be best considered under the separate headings of "use of normal curves," "use of skew-curves." and "use of percentiles." I will take them in the order mentioned.

I. The distributions of pauperism given in the present paper seemed to be of special interest because they were markedly skew, and showed very characteristic secular changes. These changes are measured by the constants of the fitted curves. Normal curves are not skew, and consequently hardly suitable for illustrating changes in skewness.

The attempt to use the halves of two different normal curves to represent the skew distribution seems undesirable for several reasons besides the lack of theoretical justification; (1) it really assumes a knowledge of the mode, which can only be given by a fitted skew-curve; (2) it leads to sensible frequencies of negative pauperism; (3) the skew-curve gives a much better fit than any two halves of normal curves, presents a continuous distribution round the mode, and does not lead to the absurdity of sensible negative pauperisms.

II. If it be objected, as Mr. Galton objects, that the evidence of the suitability of the theoretical curves used in the present case is somewhat doubtful, the answer must be somewhat of the following character: (1) these skew-curves have been shown to be available for statistics from the most diverse sources, giving extremely close fits where the number of observations is large; their use is consequently (empirically) justified in any case unless evidence to the contrary is strong; (2) but Mr. Galton's chief evidence against them is the position of the maximum ordinate of the observation-