A Case of Empirical Study in the History of Statistics: Yule's Analysis of Pauperism

Shuyuan Nie

College of Mathematical and Science, Luoyang Normal University, Luoyang, China

Abstract—**The relationship of Yule's analysis about pauperism with empirical study has been less recognized. The article explores the historical trace of research pauperism by Yule, which could be summarized as follows: (1) value the social problem from the innovative view of statistics; (2) arrange the available data and clearly state the source of material; (3) preprocess raw data to form complete series; (4) apply statistical theory to empirically analyze pauperism and deduce the accurate conclusions. The study results show that Yule's analysis of pauperism is no other than a case of empirical study in the history of statistics, which furthermore exactly embodied the interaction and integration between theory and practice.**

Keywords-empirical study; pauperism; George Udny Yule; correlation; history of statistics

I. INTRODUCTION

Nowadays, empirical study has been a kind of scientific research method. It conducts to obtain objective material through much observing, testing and investigating the object and conclude the essential property and development regularity from the individual to the general. The main process of empirical study includes ascertaining the problem, gathering the corresponding data, choosing the scientific tool, empirically analyzing the original material, estimating and evaluating the conclusions etc. At present there are lots of literatures to explore the focus problems by the method of empirical study in the area of economics, sociology, finance, medicine, statistics, biology, and so on, but it has scarce literature to discuss the historical empirical case. From the view of gathering, preprocessing and statistically analyzing the data, this paper systematically elaborates that analysis of pauperism by English statistics George Udny Yule (1871-1951) is no other than a concrete case of empirical study.

II. THE HISTORICAL PROCESS OF YULE'S ANALYSIS OF PAUPERISM

A. The Reason of Yule's Study Pauperism

Yule began to pay close attention to Poor Law Statistics because of a well-known debate between Mr. Charles James Booth (1840-1916) and Mr. Charles Stewart Loch (1849-1923). Booth was the keen philanthropist and sociologist to study *Labour and life of the people in London* and the old age pension question. While Loch was acknowledged as the secretary of the Charitable Organization Society (usually abbreviated to C.O.S.), which was an institution that aimed at pursuing a disinterested and efficient allocation pattern of philanthropic relief and avoiding waste to the maximum. At that time, there existed essentially two kinds of relief pattern, i.e. in-door relief and out-door relief. According to the Poor Law Amendment Act of 1834, "in-door relief" pointed that the pauperisms have to earn their relief through working in the poor house. On the contrary, the out-door recipients would be distributed to money without their earning it in the poor house.

Facing the relationship between relief method and poverty, the arguments of Booth and Loch were entirely different. Booth believed that there was no general relation between the proportion of out-relief and the total percentage of pauperism. That is to say, the way of the recipient received out-door or in-door relief would not make a difference of him being a pauper. After authoritatively investigated the living appearance of the pauper and explored the management and distribution for relief supplies, Loch claimed that if increasing the amount of out-door-relief, more people would mainly use "easy money" and the quantity of pauperism would apparently rise. So an anti-out-relief should be advocated to decrease the total pauperism in time. But Yule deemed:

If one variable be correlated with the rate of change of another, the two variables themselves will, also be correlated.

On the basis of this idea, Yule concluded that either Booth or Loch was right. Yule agreed with Loch, and he commenced to strictly analyze the original data from the view of statistics and proved his conclusion by the aid of the notion of correlation and regression.

B. Yule's Elaboration about the Source of Data

It is known to all that the authenticity and reliability of data is the premise and foundation of empirical analysis. When settling the problem with the help of empirical technology nowadays, we usually put the firsthand data after the paper in the form of appendix. If the information is too much to list laboriously, it often marks out directly the source to be consulted by the interested reader. Such thinking was also extremely valued by Yule. Yule had already unambiguously elaborated the exact source of entire data in his papers and listed out the detailed data tables as much as possible.

C. The Pretreatment to the Original data by Yule

It is common knowledge in the field of statistics that the researcher must carefully examine and scientifically preprocess the material before statistical analysis, whether it is the actual data obtained by direct observation and measurement or it is the secondhand data. Yule particularly

emphasized the pretreatment of the corresponding data about pauperism. For example, because the two variables of the rate of pauperism and out-relief ratio were represented by the interval range, if the variable value in some unions was exactly an integer, such as a union had three out-paupers to one in-pauper, Yule first divided it into two halves and respectively counted 0.5 in the rows 2-3 and 3-4. In literature [1]-[6], Yule repeatedly emphasized that he had applied the above pretreatment method to the procession of establishing the correlation table of pauperism and out-relief. Moreover, Yule flexibly adopted all kinds of methods to deal with different circumstances. When working out the percentage ratio of the pauperism, out-relief ratio and proportion of old in literature [6], Yule kept the nearest unit percentage for three variables. But if a ratio was just right a dividing line between two groups, e.g. 55 percent, the handling method used by Yule was to examine one decimal place of the figure, then the ratio would go into the corresponding group. However if the ratio was exactly 55 percentage without approximation, Yule expanded the original data to another digit or more and recalculated the ratio to digits as many as possible, thereby obtained a new figure slightly different from 55 percentage. It was through such preprocessing that Yule had constructed gradually sundry data tables about pauperism.

Yule also highly emphasized the test and elimination of outlier before studying the corresponding data. The primary instances could be enumerated as follows: First, when the correlation tables of the rate of total pauperism and the given proportion of out-relief were tabulated, the Welsh unions were excluded because there were several examples with over 100 out-paupers to one in-pauper, which were apparently different with the general cases. Yule regarded the Welsh unions as an exceptional ensample. Second, when calculating the corresponding means of successive rows in the tables of pauperism and out-relief in 1871 and 1891, Yule omitted the few rows of the numbers being not sufficient. Third, because the available data wasn't accurate, the case of the metropolitan unions was considered separately, i.e. all the correlation tables were separated into two kinds. There existed some analogous circumstances in Yule's papers. These basic works, including the above transformation of the data not being directly used to analyze and the deletion of outlier, were just the pretreatment of Yule's empirical study.

D. The Empirical Analysis of Pauperism

1) The Controversy with Booth and Preliminary Discussion about the Correlation of Pauperism with Out-relief: When he figured out that the correlation coefficient of pauperism and out-relief was individually up to 0.262 and 0.388 in 1871 and 1891, but the corresponding probable error was just 0.025 and 0.022, Yule began to investigate the reason why Booth had a wrong conclusion. Yule found, for the unions of England and Wales in part II of the Aged Poor, coincidently compared with the individual union, Booth had the highest and the lowest rate of pauperism with the mean of the whole union, which was significant only in the condition of extremely high correlation. To dissect this case through concrete data, Yule divided each of Booth's original lists in

half and respectively averaged two halves. The results showed that for the variable—the percentage cost of out-relief to out and in together, the difference of the average of the top and the lower half was very obvious in ten groups, but Booth wasn't aware of it. For example, in group II the difference was 8.8, while Booth didn't consider appearing "any definite relation''; The maximum difference value was up to 13.4 in group X, but Booth declared that the relativity was "not very clearly marked". Yule formed a table with all data in 10 groups, where every figure was compared with Booth's conclusion.

As a retort to Yule and also for the sake of explaining the meaning of general relation, Booth drew two curves of the pauperism and out-relief separately for 50 agricultural unions, which didn't indicate the characteristic of rising and falling together. For this case, Yule not only firstly stressed from the view of statistical theory that small groups were rather untrustworthy and dangerous, but also excavated the real situation by means of detailed figures. Yule found, the fifty unions chose by Booth were so special that the first half nearly all came from the southern and eastern counties, but the other half almost entirely came from the northern counties. However, the income of laborers in two districts had obvious distinction. The former was 10s. to 13s. one week and the latter was up to 16s. to 17s. under the same condition. It was this significant difference in wealth that hided the form of management and distribution. Through applying these actual data, Yule denied Booth's special instance and unearthed his vague idea.

2) Profound Analysis about the Relativity of Two Variables: To testify further the correlation of pauperism and out-relief, Yule analyzed carefully all kinds of data about pauper in literature [2]. He had firstly calculated that the mean pauperism rose from 23 to 30 percent when the ratio out-relief increased from 0-1 to 6-7 in the year's count, and it rose from 13 to 20 percent in the day's count. The strong relativity was stated clearly. Then Yule reckoned and compared some concrete figures including the correlation coefficient and probable error for various contradistinctive conditions, such as all ages versus males over 65, year's count versus day's count, all unions in England versus excluding the Metropolis unions only. Yule further studied the empirical result from the following aspects:

First, although varied in magnitude, all coefficients were positive. That is to say, the rate of total pauperism was notably positive correlated with the ratio of out-relief, just only the relativity had a difference between the strong and the weak.

Second, according to a series of data, it was sure that the Metropolis unions were very particular. So choosing the case of excluding the Metropolis unions only as a subject, Yule contrasted the result in the day's count and in the year's count for males over 65. Because the coefficient 0.31 in the former apparently exceeded the coefficient 0.24 in the latter, Yule laid stress on the stronger relativity in the day's count. But Yule pointed out that the difference between 0.31 and 0.24 was not large enough to agree with the conclusion of Booth—the anti-out-relief policy should be supported in the day's count and the reverse it shouldn't be supported in the year's count. Moreover, Yule also compared the case of all ages with the

case of males over 65 in the day's count, two coefficients 0.31 and 0.43 both displayed the stronger correlation between rate of pauperism and out-relief, which provided evidence for Yule criticizing some ambiguous ideas of Booth.

3) Revelation on the Whole Numerical History of English Pauperism by the Method of Frequency-curves: On the basis of the theory of correlation and multiple regression and the figures of the percentage of the population in receipt of relief in England and Wales on January 1 of each census year, including 1850, 1860, 1870, 1881 and 1891, Yule not only reckoned almost entire vital statistic for the above five years, such as modes and means, standard deviation, three moment-coefficients μ_2 - μ_4 , criterion β_1 , β_2 and k etc, but also fitted the corresponding equations to the skew frequency-curves by the aid of criterion. At the same time, Yule calculated range, skewness of the distribution and the constants p, q, n and c of the theoretical point-binomial distributions. All detailed data could be consulted from table II-VI in P. 348-349 of the appendix of Yule's research literature [3]. Furthermore, Yule drew respectively five sets of graphics for each year mentioned above. Every graphic contained and compared three kinds of curves, which were the original frequency-polygons, the fitted skew-curves and the point-binomial. It was just through the calculation of these particular numbers and the comparison of these tables and diagrams that Yule had found and summarized the essential changes of pauperism and the corresponding numerical measures of the changes for the whole English country.

First, because all the values of criterion k remained positive from 1850 to 1891, every skew-curve belonged to the forth kind equation. But the values of k varied in magnitude and the form of equations was different. Observing the alteration of criterion k, it was clear that the value 0.6 in 1850 quickly rose to 0.9 in 1860, and there was no apparent change from 1860 to 1870. Then it dropped to 0.7 in 1881 and further dropped to 0.4 in 1891. The overall trend of k was decreasing and tending to zero. So the distribution seemed gradually turning to generalized binomial curve, which was the simple binomial type and the value of k equaled zero. The statistics which had a similar change feature with criterion k were the range and constants n, c in point-binomial. Only just the alteration of constant c was in the same direction along k, but constant n and range were in the opposite direction. Both the decrease of c and the increase of n and range since 1850 further proved that the skew-distribution was slowly passing over to the binomial type and ultimately towards the normal form.

Second, there was an obvious law on the whole for almost all the constants listed above from 1850 to 1891. That was, it had a considerable change from 1850 to 1860, and a clear stagnation in the following decade. After that it had again an apparent alteration between 1870 and 1881 and had a slightly further change in the same direction in the last decade. Meanwhile all the figures were divided into two classes. In one class including mode, mean, moment-coefficients, standard deviation etc, the difference of 1850-1860 and 1870-1881 had the same symbol, but in another class including

criterion and skewness etc, the symbol of the difference was opposite.

Third, continually considering the degree of the statistic's change, Yule found that from 1850 to 1891, the mode dropped from 5.82 percent of pauperism to 2.99, the union mean decreased from 6.51 percent to 3.29, standard deviation ranged from 2.51 percent to 1.25. All of three constants had been nearly halved. These specific figures, combining with the changes in the average pauperism, would offer more abundant and accurate information about pauperism question.

Finally, Yule had calculated the misfit percentage of the skew-curves to the original frequency-polygons, which were respectively 11.4, 10, 10, 9.3 and 4.5 percent in five years. Comparing with the misfit scope i.e. 7-9 percent on a larger number of observations, Yule affirmed the effect of fitting. The process was no other than the test and evaluation in the current empirical study. Certainly, Yule had also analyzed some extraordinary changes of certain statistics. Limited to the length of the paper, it didn't retell here. Moreover, Yule had attempted to explain simply the causes of the changes in literature [3], which were the theme of literature [6] and will be discussed in the following part.

4) Further Investigation into the Causes of Changes in Pauperism with the Multiple Regression Equation: When studying the theory of regression and correlation in literature [3], Yule had already established three regression equations of pauperism, out-relief and earnings, which were listed as follows:

Pauperism of males over 65 = 13.92 percent + 0.728 out-relief ratio

Out-relief ratio = 1.27 percent + 0.155 pauperism of males over 65

Pauperism = 11.64 percent - 0.5 earnings in shillings

But Yule deemed that these equations had an overt disadvantage. It was possible to be given a double interpretation. One idea was that the changes of pauperism were the direct consequence of the changes in proportion of out-relief, while the other was that the changes of two variables were simultaneous association of economic and social changes, where the changes of pauperism were not directly influenced by the change of administration. Then a most important problem was put forward: which interpretation could reflect the fact better?

To answer this question, Yule sorted out the concrete frequency tables for different groups and periods for every possible pair of four variables, which were pauperism, out-relief ratio, proportion of old and population. Yule further calculated the fundamental constants for above variables, including mean, standard deviation, and all forty-eight correlation coefficients etc. Based on these detailed and key statistics figures, Yule first qualitatively analyzed the main relationship of pauperism with other factors. For example, the feature of concomitant variety of pauperism and out-relief ratio was fairly obvious in all cases, while a close relativity between pauperism and proportion of old or population was not clear. Moreover, after briefly compared the difference of

the change in pauperism and out-relief ratio for all cases, Yule concluded that out-relief was not the sole factor influencing pauperism. Therefore, it once again emphasized here that the multiple regression equation was necessary and vital. Certainly, similar to the modern statistical technique, Yule also paid much attention to quantitative analysis, because qualitative analysis was simple and intuitive, but somewhat vague and uncertain. So Yule spent more time and energy to establish the multiple regression equation of pauperism with social and economic changes. It took the case of rural group in 1871-1881 as an example, where the equation was

Change in pauperism = -27.07 + 0.299 change in out-relief ratio+ 0.271 change in proportion of old + 0.064 change in population.

Starting from these regression equations, Yule began to explore the causes of the change in pauperism from three aspects: changes not accounted for; changes in proportion of old and population; changes in out-relief ratio.

First, the equation indicated that there was a decrease of 27 percent unaccounted changes in pauperism by any variable used. Yule horizontally and vertically compared the percentage of unaccounted changes in different status, which varied in magnitude and direction. But through an averaging method of "balance sheets", Yule calculated that mean of unaccounted changes was a little steady during two decades, which was respectively -10.9 and -11.7 percent. So Yule judged it was due to undoubtedly some stable economic changes tending to decrease pauperism. By the contrast in numerical measure and the corresponding curves, Yule confirmed, it was the change in in-door pauperism that afforded a measure of the unaccounted change. Not only the change in in-relief and the unaccounted change had the same sign, but also the magnitude was very close in 1871-1881.

Second, Yule not only directly explained the relationship of the changes in pauperism, proportion of old age, population and prosperity per head, but also illustrated the essential difference of positive association and negative correlation. Then Yule deduced the quantity relationship between the percentage of the change in pauperism and proportion of old age. Based on this primary formula, Yule described quantitatively that the partial regression coefficient of the change in pauperism on proportion of old age turned a small figure to a relatively large one from 1871-1881 to 1881-1891. With regard to the partial regression coefficient of pauperism on population, the alternation of sign can more embody the difference of each group. It was almost negative in the urban group and conversely positive in the rural group. However, the changes in pauperism on the whole exhibited little correlation with changes in proportion of old age or population.

The last one, which was the most important point, discussed the relationship of the change in pauperism with the change in out-relief ratio. According to table C of P. 259 in literature [6], the partial regression coefficients of pauperism on out-relief ratio were positive at all events without any exception. This result implied the change in rate of total pauperism invariably exhibited the strong relativity with the change in out-relief ratio. This conclusion was highly consistent with the previous results in Yule's study. To judge whether the changes of two variables were due to a direct influence of some extraneous causes, Yule drew four curves on the same graphics to engender their respective fluctuation, which were the rate of total pauperism, in-door pauperism, marriage-rate and vagrancy. Through carefully observing the ripples and the trend of rising and falling of four curves, Yule clearly brought forward:

The greatest changes in pauperism and out-relief ratio must have been effected by some factors other than the economic, administration seems the only alternative.

Yule successfully overcame the trouble presented in the beginning of this section. Then Yule began to explore the influencing degree of out-relief ratio on pauperism, because the partial regression coefficients differed very considerably in magnitude. Yule not only drew straight lines of the partial regression of pauperism on out-relief for different groups to illustrate the impacting difference in two decades, but also discussed the significance of the change in each variable and the corresponding contribution to the total average change in pauperism. Further Yule calculated the change in mean pauperism credited to the change in out-relief ratio by two methods. At last, Yule calculated that the percentage of mean decreased in total pauperism and mean credited to change in out-relief ratio were respectively 33.5 and 21.3 in the decade 1871-1881 for the whole England. That is to say, about five-eighths of the decrease of pauperism was due to the changed administration. But in the second decade, the averaging process was not satisfactory, which didn't imply a weaker efficiency of anti-out-relief because it had barely changed policy during 1881-1891.

III. CONCLUSIONS

In summary, all results about pauperism obtained by Yule derived from the actual figures and mathematical model. Yule not only calculated all kinds of correlation coefficients to indicate the strong relativity of pauperism with out-relief ratio, but also further proved that about five-eighths of the decrease of pauperism credited to alterations of anti-out-relief during 1871-1881. He even reckoned different statistics constants and drew curves to numerically measure the pauperism history of the whole England. The procedure of studying pauperism by Yule fit well with the thinking of empirical study. In fact, being the first-person to analyze social problem using strict statistical method, Yule always emphasized the close combination of theory and practice. That is to say, whether for the whole research thinking or the particular research process, the analysis of pauperism by Yule was just an interesting case of empirical study.

REFERENCES

[1] Yule, G. U.: On the correlation of total pauperism with proportion of out-relief I [J]. The Economic Journal, 1895, 5: 603-611.

[2] Yule, G. U.: On the correlation of total pauperism with proportion of out-relief II [J]. The Economic Journal, 1896a, 6: 613-623.

[3] Yule, G. U.: 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 [J]. Journal of the Royal Statistical Society, 1896b, 59: 318-357.

[4] Yule, G. U.: On the Significance of Bravais' Formulae for Regression, &c., in the Case of Skew Correlation [J]. Proceedings of the Royal Society of London, 1896 – 1897, 60: 477-489.

[5] Yule, G. U.: On the theory of correlation [J]. Journal of the Royal Statistical Society, 1897a, 60: 812-854.

[6] Yule, G. U.: An investigation into the causes of changes in pauperism in England, chiefly during the last two intercensal decades (Part I) [J]. Journal of the Royal Statistical Society, 1899, 62: 249-295.

[7] Yule, G. U.: Note on the teaching of the theory of statistics at university college [J]. Journal of the Royal Statistical Society, 1897b, 60: 456-458.