NOTE ON KOSHAL'S METHOD OF IMPROVING THE PARAMETERS OF CURVES BY THE USE OF THE METHOD OF MAXIMUM LIKELIHOOD

By

R. J. MYERS

It has been shown by R. A. Fisher(1) that the most efficient parameters for Pearsonian curves may be found by the method of maximum likelihood. In applying this method we maximize the quantity

$$L = \sum n_k \log p_k \tag{1}$$

by varying the parameters of the curve; n_k denotes the observed frequency of the k^{th} class, and p_k is the probability of an observation falling in this class as determined from the curve and is thus a function of the parameters. Thus, in maximizing L, p_k varies as the parameters are varied, but n_k remains constant throughout since it is fixed by the given data.

Usually it is impossible to obtain a solution to the maximum likelihood equation so that some method of approximation must be used. R. S. Koshal(2) has devised a very ingenious method of approximation, which can be summarized briefly as follows. Values of L are obtained first by varying only one parameter at a time, and then by varying two parameters at the same time. When only one parameter is varied, two values of L are computed for each parameter, whereas in the case of two parameters being varied, only one value of L is computed for each combination of parameters. Thus, $2n + {}_nC_2 + 1$ or $\frac{1}{2}(n+1)(n+2)$ values of L would be needed for n parameters. With these L's the constants of n simultaneous equations involving the n corrections to the n parameters can be determined, and then the corrections themselves can readily be obtained.

In applying this method a number of interesting results were obtained. The data used was the same as used by Koshal[2] because in checking through his work there were found several serious numerical errors, especially in the computation of β. This gave a poor fit so that the method of maximum likelihood had more opportunity for improvement than if there had been no error. These data are distributed according to a Type 1 distribution, whose general equation is

$$y = y_0 (x-\alpha)^{m_1} (\beta - x)^{m_2} \tag{2}$$

The values of the parameters as obtained from the moments are

$$\alpha = .33461$$
$$\beta = 16.9885$$
$$m_1 = .69753$$
$$m_2 = 4.93202.$$

The most convenient sizes of the increments for the parameters were chosen, namely .1 for α, m_1, and m_2 and 1.0 for β.

In the case of the L's in which only one parameter is varied, Koshal selected the two L's to be computed for a particular parameter in the following manner: it should be remembered that L_{0000}, the value for the unaltered parameters, has already been computed. As an illustration let us consider the L's computed for variations of α. The criterion set up was that $L_{\overline{x+1}\,000}$ should be greater than either L_{x000} or $L_{\overline{x+2}\,000}$, where x may be —2, —1, or 0. This criterion is justified by the common sense reasoning that the maximum likelihood solution will then lie somewhere between L_{x000} and $L_{\overline{x+2}\,000}$. However, in the case of the L's in which two parameters are varied, Koshal merely selected the combination of the increments at random. Thus, for the L for α and β, Koshal computed L_{1100}. In carrying out my computations I thought it best to use the same criterion on the L's in which two parameters were varied, as was used on the L's in which only one parameter was varied. For example, I gave various values to x and y so that a number of values of L_{xy00}

were obtained. The largest of these was used in the determination of the constants as explained before. It was not necessary to give all values to x and y because a good many combinations could be discarded by inspection. For example, if L_{1100} was greater than L_{1000}, it obviously was not necessary to calculate L_{1-100}

The above process was repeated for the other L's, and the constants were then determined. From these the corrections to the parameters were obtained; these corrections gave new parameters as follows:

$$\alpha = .38399$$
$$\beta = 16.5020$$
$$m_1 = .72547$$
$$m_2 = 4.80853$$

The frequency distribution obtained from these parameters was quite a bit better than the original one as judged by both the χ^2's test and its likelihood. However, it is important to note that two of the double increment L's used in obtaining the constants were greater than the L obtained from the new parameters. This would seem to show that better results could be gotten by judicious guessing than by using this method of approximation. Another fact illustrating the roughness of approximation is that the values of the constants when computed from other of the double increment L's vary by as much as 30% from those previously used. Naturally with different values of the constants, different values for the corrections to the parameters would be obtained. Several combinations of different values of the constants were tried, and a few of the resulting frequency distributions gave higher L's than the ones obtained previously, although there were none higher than the two subsidiary L's previously mentioned. It is not unlikely that a combination of constants might be found so as to yield a higher L than either of the latter two, but there would have to be a considerable amount of manipulation in order to find this combination.

Another disadvantage of this method is the fact that a great deal of time is required to apply it. Approximately sixty hours were required to carry the calculations for the Type 1 curve.

Another interesting fact was brought out when the method of Pearson and Pairman[3] for correcting the moments for grouping was applied to the original data. The frequency distribution obtained was far better than any previously obtained as shown by the fact that the L for this distribution was highest of all; χ's for this distribution was 4.64. The time required to apply this method was considerably less than needed for Koshal's method.

Since writing this paper my attention has been directed to the recent article in the Journal (Vol. XCIII, Part II, 1934, p. 331) by W. P. Elderton and G. H. Hansmann. In this paper the writers used the same data as Koshal and fit these data by an ingenious method due to Elderton[4]. It is interesting to note that the χ's of the distribution obtained by Elderton and Hansmann is practically the same as that obtained when the method of Pearson and Pairman was used. Elderton and Hansmann also came to the conclusion that Koshal's method required more labor to bring about the same results as other methods.

BIBLIOGRAPHY

1. Fisher, R. A. "On the Mathematical Foundations of Theoretical Statistics." *Phil. Trans.*, A, vol. 222, pp. 309–368.
2. Koshal, R. S "Application of the Method of Maximum Likelihood to the Improvement of Curves Fitted by the Method of Moments." *Jour. Royal Stat. Soc.*, vol. XCVI, pp. 303–313.
3. Pairman, Eleanor and Pearson, Karl. "On Corrections for the Moment-Coefficients of Limited Range Distributions when there are Finite or Infinite Ordinates and any Slopes at the Terminals of the Range." *Biometrika*, vol. 12, pp. 231–258.
4. Elderton, W. P. "Frequency Curves and Correlation," pp. 121–122, 2nd edition.