

We can now by aid of (v) express η_{rs} . We have:

$$\eta_{rs}^2 = r^2 + \frac{e_{12}^2}{\phi_2} + \left(e_{12} - \frac{\phi_2}{\phi_1} e_{13} \right)^2 / (\phi_1 - \phi_2^2/\phi_2) \\ + \frac{\left\{ e_{14} - \frac{\phi_4}{\phi_2} e_{12} - \frac{\phi_2}{\phi_2^2} \left(e_{12} - \frac{\phi_2}{\phi_1} e_{13} \right) / (\phi_1 - \phi_2^2/\phi_2) \right\}^2}{\phi_2 - \phi_4^2/\phi_2 - \frac{\phi_2^2}{\phi_1^2} / (\phi_1 - \phi_2^2/\phi_2)} + \text{etc.}$$

The conditions therefore for linear regression, or $\eta_{rs} = r$, are:

$$e_{12} = e_{13} = e_{14} = \text{etc.} = 0 \dots\dots$$

That is: $q_{12} = r\sqrt{\beta_1}$, $q_{13} = r\beta_2$, $q_{14} = r\frac{\beta_2}{\sqrt{\beta_1}}$, etc.

For parabolic regression: $e_{12} = \frac{\phi_2}{\phi_1} e_{13}$, $e_{14} = \frac{\phi_4}{\phi_2} e_{12}$, etc.,

or $q_{12} - \frac{\phi_2}{\phi_1} q_{13} = r \left(\beta_2 - \frac{\phi_2}{\phi_1} \sqrt{\beta_1} \right)$,
 $q_{14} - \frac{\phi_4}{\phi_2} q_{12} = r \left(\frac{\beta_2}{\sqrt{\beta_1}} - \frac{\phi_4}{\phi_1} \sqrt{\beta_1} \right)$, etc.

And lastly for cubical regression:

$$e_{14} - \frac{\phi_4}{\phi_2} e_{12} - \frac{\phi_2}{\phi_1^2} \left(e_{12} - \frac{\phi_2}{\phi_1} e_{13} \right) / \left(\phi_1 - \frac{\phi_2^2}{\phi_1} \right) = 0, \text{ etc.}$$

Such conditions, especially with regard to their probable errors, become less and less manageable as we proceed.

The general principle involved in the present paper has been discussed by Tchebycheff*, and more adequately by J. P. Gram†, but the former had in view the fitting or graduating of curves. He calculated quantities which correspond to our μ_s 's on the assumption that $\mu_s = 1$, i.e. that the weight of the \bar{y}_s 's are all the same or that the marginal total is a rectangle. He was thinking of fitting a curve to a curve and not fitting a curve to a swarm of points. In his case each μ_s and accordingly each β and each ϕ is expressible in terms of the total number m of subranges which he takes of equal length. There are I think simpler methods of calculating the equation to a higher order parabola in such cases‡. As far as I am aware these orthogonal regression functions have not hitherto been dealt with and they throw a good deal of light on the original equations I provided in 1905 for skew regression. I had not recognised at that time that my expressions of each order were true orthogonal functions. It will be seen that my solution does not involve equality of subranges and is not limited to any special frequency distribution.

II. Note on the "Fundamental Problem of Practical Statistics."

(*Biometrika*, Vol. XIII, p. 1.)

Some misunderstanding has arisen with regard to my paper under the above title in the last issue of this *Journal*. I believe it to be due to the critics not having read Bayes' original theorem as given by Price in the *Phil. Trans.*, Vol. LIII. Bayes takes a ball and places it at random on a table, say of breadth unity, and its distance from one side being x , its chance of falling between x and $x + \delta x$ is δx . x is thus not a chance, but a variate. He now calls a "success," the chance that any other ball placed at random on the table will be nearer to the same side than the first

* *Mémoires de l'Académie de Saint-Petersbourg*. Mémoires in 1854 and 1859. A résumé by B. Badau: *Bulletin Astronomique*, T. VIII, Paris, 1891, pp. 350, 376 et seq. See also Liouville's *Journal*, 2^e Série, T. III (1858), p. 259 et seq.

† *Thesis*: "Om Bækkendviklinger bestemte ved Hjaelp af de mindste Kvadraters Methode." Kjøbenhavn, 1879.

‡ *Biometrika*, Vol. II, pp. 12-16.

placed ball and a "failure" that it should be greater. The chance therefore of p successes and q failures now happening is

$$x^p (1-x)^q \times dx \times \frac{p+q}{x}.$$

It is solely the fact that all possible values of the variate x are made *a priori* equally likely that makes the chance of a success x , equal to the variate itself. Those who criticise Bayes after reading his actual paper, say that he ought not to have made the chances of a ball being placed anywhere on the table equally likely. He makes in fact his distribution of the variate x a straight line—a somewhat unusual form of frequency distribution*. My answer to that objection to Bayes' work was that you can make the distribution of that variate—i.e. position on the table—any continuous curve you please as Bayes' Theorem with Bayes' results will flow from it equally well. Against this position my critics raise the cry that the chance is no longer x of a success and $1-x$ of a failure. Of course not, because that depends on horizontality of frequency distribution and it was merely fortuitous that for that case Bayes' variate x corresponded to a chance. In other cases the chance is a function of the variate x and not x itself. But if the critics say: Then this is not what we mean by Bayes' Theorem, I would reply: Quite so, but it is what Bayes meant by his own Theorem, and it probably fits much better the type of cases to which we are accustomed to apply it than what you mean by Bayes' Theorem.

Let me illustrate this point. An event has happened p times and failed q times; what is the chance that in $r+s$ further trials it will occur r times and fail s ? The critics say: this is our Bayes' Theorem, not your Bayes' Theorem. I reply that it is both, but that your way of solving it is not Bayes'. Perhaps Bayes saw further than some of the critics who have not troubled to read his original paper. What Bayes said was this, the event will happen when there is an excess (or it may be defect) of a certain variate, but I do not know what is the limiting value of this variate *a priori*. Look at Bayes' billiard table from a more modern standpoint. Men will sicken from a disease when their resistance falls below x an *a priori* unknown value of the variate. Bayes took the chance of this limiting value lying between x and $x+dx$ to be dx , if the total range be taken as unity. He ought to have taken it $\phi(x)dx$, where $\phi(x)$ is arbitrary. He took the chance of occurrence of the disease to be x , when he ought to have taken it $\int_0^x \phi(x) dx = P_x$; and of failure $1-x$ instead of $1-P_x$.

But had he taken these better values he would have reached finally precisely the same result as he did by his equal distribution of ignorance. Bayes made every value of his variate x equally likely. He ought to have given them a perfectly arbitrary frequency distribution. *A priori* all degrees of immunity are not equally likely to be the limiting value in the case of a disease. The generalised Bayes as thus envisaged has a very wide application to vital statistics; in fact it seems to me to entirely replace the other sort of Bayes' Theorem suggested by his critics.

Nay I would go further, and say that it is Bayes' Theorem in Bayes' sense that we need in most questions of prediction of the future from the past. If, for example, two men play a set of games and A has won p and B q games and we consider the chance in the following $r+s$ games of A's winning r and B's winning s , then I believe that A's winning may be accurately considered as depending on the excess of a certain variate x which is a function of A's skill relative to B's. *A priori* we do not know what the value of x is for which A will win. All we can say is that when relative to B he exhibits a certain excess of skill he will win, but that we must not assume with Bayes that all limit-values of x are equally likely. We must take any frequency curve for the possible distribution of x .

I believe that in most cases such a variate may be hypothecated and if it can the objection to Bayes that he made all positions of his balls on the table "equally likely" can be removed, and if removed one fundamental objection to his theorem as he stated it, i.e. in terms of excess or defect of a variate, disappears.

K. P.

* It is in fact the "rectangle point" $\beta_1=0$, $\beta_2=1.8$, just as limited a distribution as the Gaussian $\beta_1=0$, $\beta_2=3$.