Reply to Prof. Pearsons criticisms

BY

J. C. KAPTEYN.

In Nos 1 and 2 of *Biometrika,* Vol. IV, Prof. Pearson referring to my paper „Skew frequency curves in biology and statistics" [1]):

1st. maintains that in my theory I have followed Edgeworth without acknowledging his priority;

2nd. refutes my or Edgeworth's theory.

As to the first point: I must plead guilty *in part* and I offer Prof. Edgeworth my apologies. I confess to have overlooked his papers. I may perhaps adduce as an attenuating circumstance that these papers have been also overlooked in the bibliographies of both Prof. Ludwig and of Davenport, the only bibliographies on the subject with which I am acquainted.

On my request Prof. Edgeworth kindly sent me a reprint of his papers in the Journal of the statistical society. In Vol. 61 Part 4 the author, in accordance with what is said on pages 10—12 of my paper, remarks in substance as follows:

When a certain character (x) is distributed according to

1) P. Noordhoff, Groningen, 1903.

a normal frequency curve, then other characters proportional to any function φ (x) of that character will be generally distributed according to an asymmetrical frequency curve.

This remark is undoubtedly correct and, rightly translated into a mathematical formula, would lead to the following equation

$$(1) \qquad y = \frac{h}{\sqrt{\pi}} F'(x)\, e^{-h^2 [F(x) - M]^2}$$

F (x) being the solution for z of the equation $x = \varphi(z)$.

Now this equation is no other than the fundamental equation at which I arrived in my paper (p. 16).

If, notwithstanding this, I still feel justified in claiming my part in the ownership of this formula, it is on the ground that Prof. Edgeworth's remark, correct though it be, is still not equivalent to a general theory.

It does not prove that (1) must be the general equation of frequency curves. Prof. Edgeworth expressly says that it is not (l.c. p. 8). Nor does the theory, and this is the all important point, connect these curves in any way with the causes, which give rise to them.

The general theory involves the solution of this problem (and its reverse):

„On certain quantities x, which at starting are equal, „there come to operate certain causes of deviation, the „effect of which depends in a given way on the value „of x. What will be the frequency-curve produced?"

It is this problem which I treated in my paper and of which the general solution is given p. 15—16. It leads to the identical equation (1), when the effect of the causes is proportional to

$$\frac{1}{F'(x)}$$

The difference in the significance of the result, however, is evident.

Prof. Pearson overlooks the difference. He completely ignores the general problem which constitutes the real subject of my paper and says (p. 199). „He (i. e. Kapteyn) „*assumes* that some quantity obeys the normal distribution" whereas there is no question of such an assumption either in the enunciation of the problem or in its solution.

I am sorry to state that this is not the only inexact representation of the contents of my „Skew Curves". This is particularly disappointing in a paper which shows good evidence of the fact that the author has largely profited by the exposition of the theory which he refutes.

After perusing this refutation I strongly felt that it would be right to abstain from any reply, safe that on the question of priority.

Any trained mathematician would, without difficulty, judge for himself.

After a while, however, I came to consider that naturalists and most of the other persons mainly interested in the matter, can hardly be expected, as a rule, to be sufficiently well trained in mathematics to see for themselves were the truth lies. Thus real advantage might be gained by not letting the matter rest.

It is this consideration that made me resolve, and this brings me to my second point, to devote at least a few lines to a direct reply to the criticisms brought forward against my theory.

For the purpose in view, however, no *detailed* reply is at all necessary. It will be sufficient to show:

I. That Prof. Pearson actually adopts my theory (which he refutes) as the only rigorous and general one;

II. That Pearson's formulae, even now that he has tried to derive them from *our* equation (1) may, at the very best, be accepted as *empirical* representations.

These statements must seem startling. Still nothing is easier than to show their correctness; in fact Prof. Pearson has gone to great pains in destroying his own theory.

[In what follows the pages quoted refer to Prof. Pearsons paper in parts I and II of the present volume of this Journal].

On page 210 (and several other places) the equation:

$$(2) \qquad \frac{1}{y}\frac{dy}{dx} = \frac{-x}{\sigma_0^2 \, f\left(\frac{x}{\sigma_0}\right)}$$

is stated by Prof. Pearson to represent the generalised „probability curve for an infinite number of cause groups". [1] On page 211 again he asserts that all discussion of asymmetrical frequency must turn on this equation. Only, in accordance with p. 178, it is here written, with a slightly different notation:

$$(3) \qquad \frac{1}{y}\frac{dy}{dx} = \frac{-x}{\sigma_0^2 \, F(x)}$$

According to Prof. Pearson this equation leads at once to his (Prof. Pearson's) generalised probability curves by expanding $f\left(\frac{x}{\sigma_0}\right)$ in a series of ascending porwers of $\left(\frac{x}{\sigma}\right)$ (p. 210, 211) „A very few terms of the expansion, „however, suffice for describing practical frequency distri„bution" (p. 211). According to p. 204 and 212 Prof. Pearson's curves stop at *three* terms, in fact he puts (p. 204 and 212)

$$(4) \qquad F(x) = a_0 + a_1 x + a_2 x^2.$$

Now this equation (2) or (3), which thus is stated to be

1) The express condition of very numerous causes of deviation has been adhered to throughout in my „skew curves". Considerations based on a supposed very restricted number of causes can be easily shown to be illusory in nearly every case of asymmetric frequency.

the true general equation of the frequency curves is *not* Prof. Pearson's equation, but simply the equation of the curves of Edgeworth-Kapteyn [1]), The identity is only hidden by the fact that it is the *differential* equation, whereas I derived at once the equation in its finite form.

Everybody may convince himself of the fact by simply differentiating equation (1). In order to accomodate to Prof. Pearsons notation in equation (3) he has only to substitute:

$$\begin{array}{lll} f(x) & \textit{for} & F(x) - M \; {}^{2)} \\ \dfrac{1}{2\sigma_0^2} & " & h^2 \end{array}$$

so that this equation becomes:

$$(5) \qquad y = \frac{1}{\sigma_0 \sqrt{2\pi}} f^1(x)\, e^{-\frac{1}{2\sigma_0^2}[f(x)]^2}$$

and further to introduce Prof. Pearsons abbreviation (p. 178)

$$(6) \qquad F(x) = \frac{x f^1(x)}{f(x)\,[f^1(x)]^2 - \sigma_0^2 f^{11}(x)}$$

This proves point I.

As to point II.

One would naturally imagine that, if it be true, as shown just now, that Prof. Pearson derives his own curves from those given by myself, both curves must be identical; the only possible difference being that my formulae must be rigorous, whereas Prof. Pearson's, in which only a few terms of Maclaurin's series are used, must be only more or less approximate.

---

1) If Prof. Edgeworth has no objection I will gladly adopt this denomination applied to them by Prof. Pearson.

2) This does *not* mean, as Prof. Pearson erroneously supposes, that we choose the mode as the origin (p. 178).

As a matter of fact this identity, or approximate identity, will not exist at all but in a few very exceptional cases.

As a consequence thereof Pearsons formulae will lose their rational character.

The reason is that to substitute the expression

$$(7) \qquad a_0 + a_1 x + a_2 x^2$$

for $F(x)$ and for a long range of values of x, is permissible (even as an approximation) only in quite exceptional cases.

If it be permissible to substitute the expression (7) for $F(x)$, why not for

$$\frac{x}{F(x)}$$

which would make the equation (3) still simpler, or, simplest of all, why not take (7) for the ordinates of the frequency-curves themselves.

The only *possible* answer is, that experience shows that Pearsons assumption leads to equations which can be made to represent tolerably a great number of observed frequency-curves, whereas the other assumptions do not.

But this is equivalent to admitting that Pearsons curves are purely *empirical*; which is just what I maintain.

It settles Point II. [1]

In Conclusion. As Prof. Pearson now derives his own theory from mine, it need not be said that every objection raised by him against my general theory bears directly on his own.

Of the objections contained in his paper against the *special* case (causes proportional to some power of $x + \varkappa$)

---

1) The same reasoning still holds of course in the case that more terms of a Maclaurin expansion are included.

more fully developed by me, some are as astonishing as the points here treated. I will say nothing about them, however, though I do not admit the validity of a single one of them. This only may be pointed out, that Prof. Pearsons statement (p. 178) that this form „has „been suggested by Kapteyn as a general form of the „Skew frequency-curves" is erroneous. [1])

Quite recently some cases have been submitted to me which are evidently *not* contained in my special form. To meet such cases I have developed the general theory somewhat more fully in a paper now ready for press.

GRONINGEN, Januari 1906.

1) See for instance pp. 18 and 29 of my paper.