A New Graphic Method of Using the Normal Probability Curve.

BY

BUCKNER SPEED.*

Engineer in Charge, Physical Laboratory, Western Electric Co.

If the logarithms of the ordinates of the normal probability curve are platted to the natural values of x, the ends of the familiar curve droop down to infinity and the curve becomes a parabola, as is seen by writing the equation of the normal probability curve

$$y = \frac{n}{s\sqrt{2\pi}} e^{-(x^2/2s^2)}$$

and taking logarithms of both sides

$$\log y = \log \frac{n}{s\sqrt{2\pi}} - \frac{x^2}{2s^2}$$

or

$$\log y = k - \frac{x^2}{2s^2}$$

and putting $u = \log. y$

$$u = k - \frac{x^2}{2s^2}$$

a parabola with its vertex at $\begin{matrix} u = k \\ x = 0 \end{matrix}$.

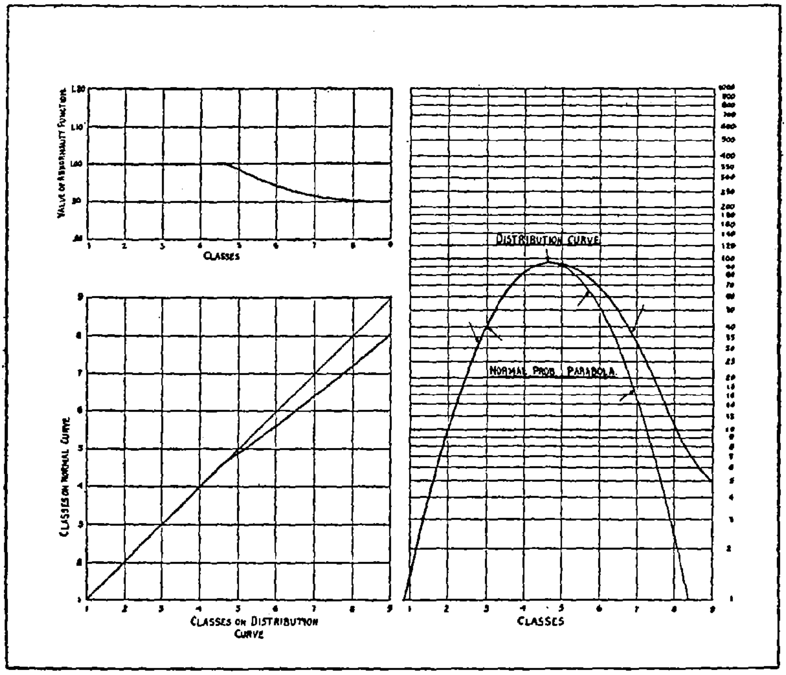
The advantage of platting distribution curves on paper ruled with equal spacing of the vertical lines and logarithmic spacing of the horizontal lines, is that if the distribution curve is normal it is a parabola, and if after platting a moderate number of individuals we draw in the parabola fitting the tops of the ordinates, and then continue platting more and more individuals, the same parabola continues to fit the tops of the ordinates; it merely moves up, the parameter of the parabola remaining constant. In fact the parameter of the parabola is a measure of the spread of the distribution.

* Presented by Albert W. Whitney.

Since five points determine a parabola, it is possible to use fewer classes; in fact, seven or nine classes are usually enough. One set of parabolic templates then will fit all normal distributions.

If we wish to study non-normal distribution curves, the utility of the equal spaced x, logarithmic y paper is seen.

Let me take the distribution of English words into classes of one, two, three, four, etc., letters per word. As we plat the curve we may stop at any time and try a parabolic template. We see that the characteristic shape develops early. We note that as we go on



taking words, we do not have to change templates, the same one is merely moved up.

We find, however, that while the left limb of the curve and its top is easy to fit, the right limb is too high.

Let us fit the parabola as best we can, and draw it in, and count the individuals included between the distribution curve and the parabola, and replat them as a residual curve. Again we have a parabola, but of different parameter, and with its axis to the right of the axis of the first parabola.

We have resolved the distribution curve into two normal probability curves. Many irregular distribution curves may be resolved into two or more normal curves, and in certain cases such resolution may have real meaning, but not always, just as we may resolve certain periodic curves by the Fourier Analysis into a series of sine curves which do not have any real meaning at all—I am thinking of pulse blood pressure curves.

Suppose we have a number of objects to be classified according to some measurable characteristic, and let us suppose the distribution curve is not normal, but that when we measure some other characteristic in the same lot of individuals, we do get a normal curve, or let us suppose that (since we do not suspect any progressive or cumulative change in mode or plurality of modes) we suspect there may be some characteristic which will distribute the objects normally. Then we may find what function of that characteristic the characteristic is which we used, as follows.

Plat and obtain a non-normal distribution curve for the classification according to the characteristic chosen. Fit the parabola that will most nearly cover the most normal half of the experimental curve, and draw in this parabola. The method is indicated in the figure.

The abnormal distribution curve suggests the following. Suppose we make the statement that if a sufficient number of objects have been produced by causes which, while variable, independently, during the production of the objects have not varied cumulatively; these objects may be sorted into class groups whose class magnitudes (not the number of objects in the class) are arranged in arithmetical progression, and the number of objects in the class groups will form a distribution curve capable of being resolved into as many normal probability curves as there were independent causes, whatever be the characteristic which was chosen for measurement.

The question lies in the use of the word "arithmetical progression." It may be that this is only a special case, and that the more general case is that of a series in which each difference is an equal part of the class magnitude at that point, that is to say, a logarithmic axis of x, where the total range of class magnitudes is but a small fraction of the mean class magnitude. This becomes the arithmetical series in the limit.

In the more intractable case, where the causes are changing cumulatively, the logarithmic platting of either y, or y and x has not seemed to be of so much help.

Mr. W. G. Housekeeper, of the Western Electric Physical Laboratory, has collaborated with me in this work, which was suggested by his exhaustive study and analysis of the measurements of the electrical and photometric characteristics of the miniature electric lamps manufactured by the Western Electric Company for use in telephone switchboards.