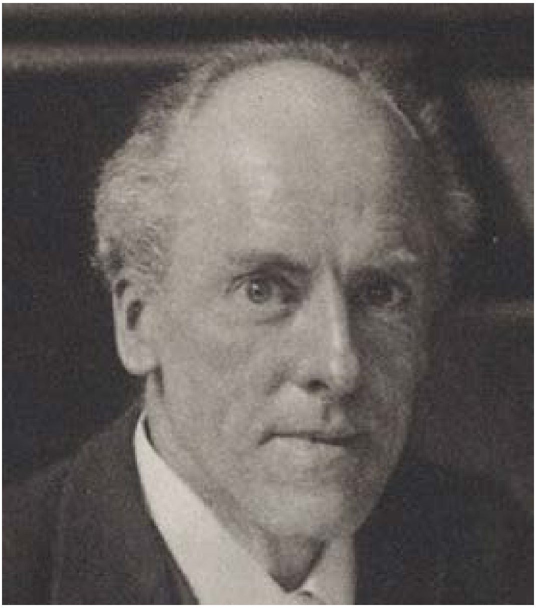
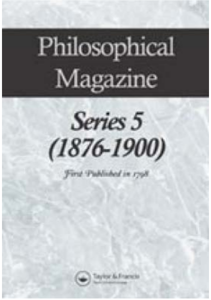
Karl Pearson

On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling

The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science
Series 5, vol. 50, n°302, 1900





X. *On the Criterion that a given System of Deviations from the Probable in the Case of a Correlated System of Variables is such that it can be reasonably supposed to have arisen from Random Sampling.* By KARL PEARSON, *F.R.S., University College, London**.

THE object of this paper is to investigate a criterion of the probability on any theory of an observed system of errors, and to apply it to the determination of goodness of fit in the case of frequency curves.

(1) *Preliminary Proposition.* Let $x_1, x_2 \ldots x_n$ be a system of deviations from the means of n variables with standard deviations $\sigma_1, \sigma_2 \ldots \sigma_n$ and with correlations $r_{12}, r_{13}, r_{23} \ldots r_{n-1,n}$.

Then the frequency surface is given by

$$Z = Z_0 e^{-\frac{1}{2}\left\{S_1\left(\frac{R_{pp}}{R}\frac{x_p^2}{\sigma_p^2}\right)+2S_2\left(\frac{R_{pq}}{R}\frac{x_p}{\sigma_p}\frac{x_q}{\sigma_q}\right)\right\}}, \quad . \quad . \quad . \quad . \quad \text{(i.)}$$

where R is the determinant

$$\begin{vmatrix} 1 & r_{12} & r_{13} \ldots r_{1n} \\ r_{21} & 1 & r_{23} \ldots r_{2n} \\ r_{31} & r_{32} & 1 \ldots r_{3n} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ r_{n1} & r_{n2} & r_{n3} \ldots 1 \end{vmatrix}$$

and R_{pp}, R_{pq} the minors obtained by striking out the pth row and pth column, and the pth row and qth column. S_1 is the sum for every value of p, and S_2 for every pair of values of p and q.

Now let

$$\chi^2 = S_1\left(\frac{R_{pp}}{R}\frac{x_p^2}{\sigma_p^2}\right)+2S_2\left(\frac{R_{pq}}{R}\frac{x_p x_q}{\sigma_p \sigma_q}\right). \quad . \quad . \quad . \quad \text{(ii.)}$$

Then: $\chi^2 =$ constant, is the equation to a generalized "ellipsoid," all over the surface of which the frequency of the system of errors or deviations $x_1, x_2 \ldots x_n$ is constant. The values which χ must be given to cover the whole of space are from 0 to ∞. Now suppose the "ellipsoid" referred to its principal axes, and then by squeezing reduced to a sphere, $X_1, X_2, \ldots X$ being now the coordinates; then the chances of a system of errors with as great or greater frequency than

* Communicated by the Author.

that denoted by χ is given by

$$P = \frac{\left[\iiiint \dots e^{-\frac{1}{2}\chi^2} dX_1\, dX_2 \dots dX_n\right]_\chi^\infty}{\left[\iiiint \dots e^{-\frac{1}{2}\chi^2} dX_1\, dX_2 \dots dX_n\right]_0^\infty},$$

the numerator being an n-fold integral from the ellipsoid χ to the ellipsoid ∞, and the denominator an n-fold integral from the ellipsoid 0 to the ellipsoid ∞. A common constant factor divides out. Now suppose a transformation of coordinates to generalized polar coordinates, in which χ may be treated as the ray, then the numerator and denominator will have common integral factors really representing the generalized "solid angles" and having identical limits. Thus we shall reduce our result to

$$P = \frac{\int_\chi^\infty e^{-\frac{1}{2}\chi^2} \chi^{n-1} d\chi}{\int_0^\infty e^{-\frac{1}{2}\chi^2} \chi^{n-1} d\chi}. \quad \dots \quad \text{(iii.)}$$

This is the measure of the probability of a complex system of n errors occurring with a frequency as great or greater than that of the observed system.

(2) So soon as we know the observed deviations and the probable errors (or σ's) and correlations of errors in any case we can find χ from (ii.), and then an evaluation of (iii.) gives us what appears to be a fairly reasonable criterion of the probability of such an error occurring on a random selection being made.

For the special purpose we have in view, let us evaluate the numerator of P by integrating by parts; we find

$$\int_\chi^\infty e^{-\frac{1}{2}\chi^2}\chi^{n-1} d\chi = \Big[\chi^{n-2} + (n-2)\chi^{n-4} + (n-2)(n-4)\chi^{n-}$$

$$+ \dots + (n-2)(n-4)(n-6)\dots(n-2r-2)\chi^{n-2r}\Big] e^{-\frac{1}{2}\chi^2}$$

$$+ (n-2)(n-4)(n-6)\dots(n-2r)\int_\chi^\infty e^{-\frac{1}{2}\chi^2}\chi^{n-2r-1}$$

$$= (n-2)(n-4)(n-6)\dots(n-2r)\Big[\int_\chi^\infty e^{-\frac{1}{2}\chi^2}\chi^{n-2r-1} d\chi$$

$$+ e^{-\frac{1}{2}\chi^2}\left\{\frac{\chi^{n-2r}}{n-2r} + \frac{\chi^{n-2r+2}}{(n-2r)(n-2r+2)} + \frac{\chi^{n-2r+4}}{(n-2r)+(n-2r+2)(n-2r+}\right.$$

$$\left.\dots + \frac{\chi^{n-2}}{(n-2r)(n-2r+2)\dots(n-2)}\right\}\Big].$$

Further,

$$\int_0^\infty e^{-\frac{1}{2}\chi^2}\chi^{n-1}d\chi=(n-2)(n-4)(n-6)\ldots(n-2r)\int_0^\infty e^{-\frac{1}{2}\chi^2}\chi^{n-2r-1}d\chi.$$

Now n will either be even or odd, or if n be indefinitely great we may take it practically either.

Case (i.) n odd. Take $r=\dfrac{n-1}{2}$. Hence

$$P=\frac{\displaystyle\int_\chi^\infty e^{-\frac{1}{2}\chi^2}d\chi+e^{-\frac{1}{2}\chi^2}\left\{\frac{\chi}{1}+\frac{\chi^3}{1.3}+\frac{\chi^5}{1.3.5}+\ldots+\frac{\chi^{n-2}}{1.3.5\ldots\overline{n-2}}\right\}}{\displaystyle\int_0^\infty e^{-\frac{1}{2}\chi^2}d\chi}. \quad \ldots \quad \text{(iv.)}$$

But

$$\int_0^\infty e^{-\frac{1}{2}\chi^2}d\chi=\sqrt{\frac{\pi}{2}}.$$

Thus

$$P=\sqrt{\frac{2}{\pi}}\int_\chi^\infty e^{-\frac{1}{2}\chi^2}d\chi$$
$$+\sqrt{\frac{2}{\pi}}\,e^{-\frac{1}{2}\chi^2}\left(\frac{\chi}{1}+\frac{\chi^3}{1.3}+\frac{\chi^5}{1.3.5}+\ldots+\frac{\chi^{n-2}}{1.3.5\ldots\overline{n-2}}\right). \quad \text{(v.)}$$

As soon as χ is known this can be at once evaluated.

Case (ii.) n even. Take $r=\frac{1}{2}n-1$. Hence

$$P=\frac{\displaystyle\int_\chi^\infty e^{-\frac{1}{2}\chi^2}\chi d\chi+e^{-\frac{1}{2}\chi^2}\left\{\frac{\chi^2}{2}+\frac{\chi^4}{2.4}+\frac{\chi^6}{2.4.6}+\ldots+\frac{\chi^{n-2}}{2.4.6\ldots\overline{n-2}}\right\}}{\displaystyle\int_0^\infty e^{-\frac{1}{2}\chi^2}\chi d\chi}$$

$$=e^{-\frac{1}{2}\chi^2}\left(1+\frac{\chi^2}{2}+\frac{\chi^4}{2.4}+\frac{\chi^6}{2.4.6}+\ldots+\frac{\chi^{n-2}}{2.4.6\ldots\overline{n-2}}\right). \quad \text{(vi.)}$$

The series (v.) and (vi.) both admit of fairly easy calculation, and give sensibly the same results if n be even moderately large. If we put $P=\frac{1}{2}$ in (v.) and (vi.) we have equations to determine $\chi=\chi_0$, the value giving the "probability ellipsoid." This ellipsoid has already been considered by Bertrand for $n=2$ (probability ellipse) and Czuber for $n=3$. The table which concludes this paper gives the values of P for a series of values of χ^2 in a slightly different case. We can, however, adopt it for general purposes, when we only want a rough approximation to the probability or improbability of a given system of deviations. Suppose we have n correlated variables and we desire to ascertain whether an outlying observed set is really anomalous. Then we calculate χ^2 from (ii.); next we take $n'=n+1$ to enter our table, *i. e.* if we have 7 correlated quantities we should look in the column marked 8. The row χ^2 and the column $n+1$ will give the value of P, the probability of a system of deviations as great or greater than the outlier in question. For many practical purposes, the rough interpolation which this table affords will enable us to ascertain the general order of probability or improbability of the observed result, and this is usually what we want.

If n be very large, we have for the series in (v.) the value $e^{\frac{1}{2}\chi^2}\int_0^\chi e^{-\frac{1}{2}\chi^2}d\chi$ *, and accordingly

$$\mathrm{P}=\sqrt{\frac{2}{\pi}}\int_0^\infty e^{-\frac{1}{2}\chi^2}d\chi=1.$$

Again, the series in (vi.) for n very large becomes $e^{\frac{1}{2}\chi^2}$, and thus again $\mathrm{P}=1$. These results show that if we have only an indefinite number of groups, each of indefinitely small range, it is practically certain that a system of errors as large or larger than that defined by any value of χ will appear.

Thus, if we take a very great number of groups our test becomes illusory. We must confine our attention in calculating P to a finite number of groups, and this is undoubtedly what happens in actual statistics. n will rarely exceed 30, often not be greater than 12.

(3) Now let us apply the above results to the problem of the fit of an observed to a theoretical frequency distribution. Let there be an $(n+1)$-fold grouping, and let the observed frequencies of the groups be

$$m'_1,\ m'_2,\ m'_3\ \ldots\ m'_n,\ m'_{n+1},$$

and the theoretical frequencies supposed known *a priori* be

$$m_1,\ m_2,\ m_3\ \ldots\ m_n,\ m_{n+1}\ ;$$

then $\mathrm{S}(m)=\mathrm{S}(m')=\mathrm{N}=$ total frequency.

Further, if $e=m'-m$ give the error, we have

$$e_1+e_2+e_3+\ \ldots\ +e_{n+1}=0.$$

Hence only n of the $n+1$ errors are variables; the $n+1$th is

* Write the series as F, then we easily find $d\mathrm{F}/d\chi=1+\chi\mathrm{F}$, whence by integration the above result follows. Geometrically, $\mathrm{P}=1$ means that if n be indefinitely large, the nth moment of the tail of the normal curve is equal to the nth moment of the whole curve, however much or however little we cut off as "tail."

determined when the first n are known, and in using formula (ii.) we treat only of n variables. Now the standard deviation for the random variation of e_p is

$$\sigma_p = \sqrt{\mathrm{N}\left(1-\frac{m_p}{\mathrm{N}}\right)\frac{m_p}{\mathrm{N}}}, \quad . \quad . \quad . \quad \text{(vii.)}$$

and if r_{pq} be the correlation of random error e_p and e_q,

$$\sigma_p\sigma_q r_{pq} = -\frac{m_p m_q}{\mathrm{N}} \quad . \quad . \quad . \quad . \quad . \quad \text{(viii.)}$$

Now let us write $\frac{m_q}{\mathrm{N}} = \sin^2\beta_q$, where β_q is an auxiliary angle easily found. Then we have

$$\sigma_q = \sqrt{\mathrm{N}}\sin\beta_q\cos\beta_q, \quad . \quad . \quad . \quad . \quad \text{(ix.)}$$

$$r_{pq} = -\tan\beta_q\tan\beta_p. \quad . \quad . \quad . \quad . \quad \text{(x.)}$$

We have from the value of R in § 1

$$\mathrm{R} = \begin{vmatrix} 1 & -\tan\beta_2\tan\beta_1 & -\tan\beta_3\tan\beta_1 & \ldots & -\tan\beta_n\tan\beta_1 \\ -\tan\beta_1\tan\beta_2 & 1 & -\tan\beta_3\tan\beta_2 & \ldots & -\tan\beta_n\tan\beta_2 \\ -\tan\beta_1\tan\beta_3 & -\tan\beta_2\tan\beta_3 & 1 & \ldots & -\tan\beta_n\tan\beta_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -\tan\beta_1\tan\beta_n & -\tan\beta_2\tan\beta_n & -\tan\beta_3\tan\beta_n & & 1 \end{vmatrix}$$

$$= (-1)^n \tan^2\beta_1\tan^2\beta_2\tan^2\beta_3 \ldots \tan^2\beta_n \times$$

$$\begin{vmatrix} -\cot^2\beta_1 & 1 & 1 & \ldots\ldots & 1 \\ 1 & -\cot^2\beta_2 & 1 & \ldots\ldots & 1 \\ 1 & 1 & -\cot^2\beta_3 & \ldots\ldots & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & 1 & & -\cot^2\beta_n \end{vmatrix}$$

$$= \tan^2\beta_1\tan^2\beta_2\tan^2\beta_3 \ldots \tan^2\beta_n \times \mathrm{J}, \text{ say.}$$

Similarly,

$$\mathrm{R}_{11} = (-1)^{n-1}\tan^2\beta_2\tan^2\beta_3 \ldots\ldots \tan^2\beta_n \times \mathrm{J}_{11},$$

$$\mathrm{R}_{12} = (-1)^{n-1}\tan\beta_1\tan\beta_2\tan^2\beta_3 \ldots \tan^2\beta_n \times \mathrm{J}_{12}.$$

Hence the problem reduces to the evaluation of the determinant J and its minors.

If we write

$$\eta_q = \cot^2 \beta_q = \frac{N}{m_q} - 1 \quad . \quad . \quad . \quad . \quad . \qquad \text{(xi.)}$$

$$J = \begin{vmatrix} -\eta_1 & 1 & 1 & \dots\dots & 1 \\ 1 & -\eta_2 & 1 & \dots\dots & 1 \\ 1 & 1 & -\eta_3 & \dots\dots & 1 \\ \cdot & \cdot & \cdot & \cdot\cdot\cdot\cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot\cdot\cdot\cdot & \cdot \\ 1 & 1 & 1 & \dots\dots & -\eta_n \end{vmatrix}$$

Clearly,

$$J_{12} = - \begin{vmatrix} 1 & 1 & 1 & \dots\dots & 1 \\ 1 & -\eta_3 & 1 & \dots\dots & 1 \\ 1 & 1 & -\eta_4 & \dots\dots & 1 \\ \cdot & \cdot & \cdot & \cdot\cdot\cdot\cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot\cdot\cdot\cdot & \cdot \\ 1 & 1 & 1 & \dots\dots & -\eta_n \end{vmatrix}$$

$$= (-1)^{n-1}(\eta_3+1)(\eta_4+1)\dots(\eta_n+1).$$

Generally, if $\lambda = (\eta_1+1)(\eta_2+1)(\eta_3+1)\dots(\eta_n+1)$,

$$J_{pq} = (-1)^{n-1}\frac{\lambda}{(\eta_p+1)(\eta_q+1)} \cdot \quad . \quad . \qquad \text{(xii.)}$$

But $\qquad J_{11} - \eta_2 J_{12} + J_{13} + J_{14} + \dots + J_{1n} = 0.$

Hence $\quad J_{11} = (1+\eta_2)J_{12} - J_{12} - J_{13} - J_{14} \dots - J_{1n}$

$$= \frac{(-1)^{n-1}\lambda}{1+\eta_1}\left(1 - \frac{1}{1+\eta_2} - \frac{1}{1+\eta_3} - \frac{1}{1+\eta_4} \dots - \frac{1}{1+\eta_n}\right).$$

Whence, comparing J with J_{11}, it is clear that:

$$J = (-1)^n\lambda\left(1 - \frac{1}{1+\eta_1} - \frac{1}{1+\eta_2} - \frac{1}{1+\eta_3} - \frac{1}{1+\eta_4} \dots - \frac{1}{1+\eta_n}\right).$$

Now

$$S\left(\frac{1}{1+\eta}\right) = S\left(\frac{m}{N}\right), \text{ by (xi.)}, = \frac{N - m_{n+1}}{N} = 1 - \frac{m_{n+1}}{N}.$$

Thus:

$$J = (-1)^n \lambda \frac{m_{n+1}}{N}.$$

Similarly:

$$J_{pp} = (-1)^{n-1} \frac{\lambda}{1+\eta_p}\left(\frac{m_p}{N} + \frac{m_{n+1}}{N}\right).$$

Thus:

$$\frac{R_{pp}}{R} = -\frac{J_{pp}}{J} \cot^2 \beta_p = \cot^2 \beta_p \frac{m_p}{N}\left(1 + \frac{m_p}{m_{n+1}}\right);$$

or from (vii.)

$$\frac{R_{pp}}{R} \frac{1}{\sigma_p{}^2} = \frac{1}{m_p} + \frac{1}{m_{n+1}} \quad \ldots\ldots\ldots \quad \text{(xiii.)}$$

Again:

$$\frac{R_{pq}}{R} = -\cot \beta_p \cot \beta_q \frac{J_{pq}}{J} = \cot \beta_p \cot \beta_q \frac{m_p m_q}{N m_{n+1}}.$$

and:

$$\frac{R_{pq}}{R} \frac{1}{\sigma_p \sigma_q} = \frac{1}{m_{n+1}} \quad \ldots\ldots \quad \text{(xiv.)}$$

Thus by (ii.):

$$\chi^2 = S_1 \left\{\left(\frac{1}{m_p} + \frac{1}{m_{n+1}}\right) e^2_p\right\} + 2S_2 \left\{\frac{1}{m_{n+1}} e_p e_q\right\}$$

$$= S_1 \left(\frac{e^2_p}{m_p}\right) + \frac{1}{m_{n+1}} \left\{ S_1 (e_p) \right\}^2.$$

But

$$S_1(e_p) = -e_{n+1},$$

hence:

$$\chi^2 = S\left(\frac{e^2}{m}\right), \quad \ldots\ldots \quad \text{(xv.)}$$

where the summation is now to extend to all $(n+1)$ errors, and not merely to the first n.

(4). This result is of very great simplicity, and very easily applicable. The quantity

$$\chi = \sqrt{S\left(\frac{e^2}{m}\right)}$$

is a measure of the goodness of fit, and the stages of our investigation are pretty clear. They are:—

(i.) Find χ from Equation (xv.):

(ii.) If the number of errors, $n'=n+1$, be odd, find the improbability of the system observed from

$$P = e^{-\frac{1}{2}\chi^2}\left(1+\frac{\chi^2}{2}+\frac{\chi^4}{2.4}+\frac{\chi^6}{2.4.6}+\ldots+\frac{\chi^{n'-3}}{2.4.6\ldots n'-3}\right).$$

If the number of errors, $n'=n+1$, be even, find the probability of the system observed from

$$P = \sqrt{\frac{2}{\pi}}\int_\chi^\infty e^{-\frac{1}{2}\chi^2}d\chi$$

$$+\sqrt{\frac{2}{\pi}}e^{-\frac{1}{2}\chi^2}\left(\frac{\chi}{1}+\frac{\chi^3}{1.3}+\frac{\chi^5}{1.3.5}+\ldots+\frac{\chi^{n-3}}{1.3.5\ldots n'-3}\right).$$

(iii.) If n be less than 13, then the Table at the end of this paper will often enable us to determine the general probability or improbability of the observed system without using these values for P at all.

(5). Hitherto we have been considering cases in which the theoretical probability is known *à priori*. But in a great many cases this is not the fact; the theoretical distribution has to be judged from the sample itself. The question we wish to determine is whether the sample may be reasonably considered to represent a random system of deviations from the theoretical frequency distribution of the general population, but this distribution has to be inferred from the sample itself. Let us look at this somewhat more closely. If we have a fairly numerous series, and assume it to be really a random sample, then the theoretical number m for the whole population falling into any group and the theoretical number m_s as deduced from the data for the sample will only differ by terms of the order of the probable errors of the constants of the sample, and these probable errors will be small, as the sample is supposed to be fairly large. We may accordingly take:

$$m = m_s + \mu,$$

where the ratio of μ to m_s will, as a rule, be small. It is only at the "tails" that μ/m_s may become more appreciable, but here the errors or deviations will be few or small *.

* A theoretical probability curve without limited range will never at the extreme tails exactly fit observation. The difficulty is obvious where the observations go by units and the theory by fractions. We ought to take our final theoretical groups to cover as much of the tail area as amounts to at least a unit of frequency in such cases.

Now let χ_s be the value found for the sample, and χ the value required marking the system of deviations of the observed quantities from a group-system of the same number *accurately* representing the general population.

Then:

$$\chi^2 = S\left\{\frac{(m'-m)^2}{m}\right\} = S\left\{\frac{(m'-m_s-\mu)^2}{m_s+\mu}\right\}$$

$$= S\left\{\frac{(m'-m_s)^2}{m_s}\right\} - S\left\{\frac{\mu(m'^2-m_s^2)}{m_s}\right\} + S\left\{\left(\frac{\mu}{m_s}\right)^2\frac{m'^2}{m_s}\right\},$$

if we neglect terms of the order $(\mu/m_s)^3$.

Hence:

$$\chi^2-\chi_s^2 = -S\left\{\frac{\mu}{m_s}\frac{m'^2-m_s^2}{m_s}\right\} + S\left\{\left(\frac{\mu}{m_s}\right)^2\frac{m'^2}{m_s}\right\}.$$

Now χ_s must, I take it, be less than χ, for otherwise the general population distribution or curve would give a better fit than the distribution or curve actually fitted to the sample. But we are supposed to fit a distribution or curve to the sample so as to get the "best" values of the constants. Hence the right-hand side of the above equation must be positive. If the first term be negative then it must be less than the second, or the difference of χ and χ_s is of the order, not of the first but of the *second* power of quantities depending on the probable errors of the sample. On the other hand, if the first term be positive, it means that there is negative correlation between $\frac{\mu}{m}$ and $\frac{m'^2-m_s^2}{m_s}$, or that when the observed frequency exceeds the theoretical distribution given by the sample $(m'>m_s)$, then the general population would fall below the theoretical distribution given by the sample $(m<m_s)$, and *vice versa*. In other words the general population and the observed population would always tend to fall on opposite sides of the sample theoretical distribution. Now this seems impossible; we should rather expect, when the observations exceeded the sample theoretical distribution, that the general population would have also excess, and *vice versa*. Accordingly, we should either expect the first term to be negative, or to be very small (or zero) if positive. In either case I think we may conclude that χ only differs from χ_s by terms of the order of the squares of the probable errors of the constants of the sample distribution. Now our argu-

ment as to goodness of fit will be based on the general order of magnitude of the probability P, and not on slight differences in its value. Hence, if we reject the series as a random variation from the frequency distribution determined from the sample, we must also reject it as a random variation from a theoretical frequency distribution differing by quantities of the order of the probable errors of the constants from the sample theoretical distribution. On the other hand, if we accept it as a random deviation from the sample theoretical distribution, we may accept it as a random variation from a system differing by quantities of the order of the probable errors of the constants from this distribution.

Thus I think we can conclude, when we are dealing with a sufficiently long series to give small probable errors to the constants of the series, that :—

(i.) If χ_s^2 be so small as to warrant us in speaking of the distribution as a random variation on the frequency distribution determined from itself, then we may also speak of it as a random sample from a general population whose theoretical distribution differs only by quantities of the order of the probable errors of the constants, from the distribution deduced from the observed sample.

(ii.) If χ_s^2 be so large as to make it impossible for us to regard the observed distribution as a sample from a general population following the law of distribution deduced from the sample itself, it will be impossible to consider it as a sample from any general population following a distribution differing only by quantities of the order of the probable errors of the sample distribution constants from that sample distribution.

In other words, if a curve is a good fit to a sample, to the same fineness of grouping it may be used to describe other samples from the same general population. If it is a bad fit, then this curve cannot serve to the same fineness of grouping to describe other samples from the same population.

We thus seem in a position to determine whether a given form of frequency curve will effectively describe the samples drawn from a given population to a certain degree of fineness of grouping.

If it serves to this degree, it will serve for all rougher groupings, *but it does not follow* that it will suffice for still finer groupings. Nor again does it appear to follow that if the number in the sample be largely increased the same curve will still be a good fit. Roughly the χ^2's of two samples appear to vary for the same grouping as their total contents. Hence if a curve be a good fit for a large sample it will be good for a small one, but the converse is not true, and a larger sample may show that our theoretical frequency gives only an approximate law for samples of a certain size. In practice we must attempt to obtain a good fitting frequency for such groupings as are customary or utile. To ascertain the ultimate law of distribution of a population for any groupings, however small, seems a counsel of perfection.

(6) *Frequency known or supposed known a priori.*

Illustration I.

The following data are due to Professor W. F. R. Weldon, F.R.S., and give the observed frequency of dice with 5 or 6 points when a cast of twelve dice was made 26,306 times :—

No. of Dice in Cast with 5 or 6 Points.	Observed Frequency, m'.	Theoretical Frequency, m.	Deviation, e.
0	185	203	− 18
1	1149	1217	− 68
2	3265	3345	− 80
3	5475	5576	−101
4	6114	6273	−159
5	5194	5018	+176
6	3067	2927	+140
7	1331	1254	+ 77
8	403	392	+ 11
9	105	87	+ 18
10	14	13	+ 1
11	4	1	+ 3
12	0	0	0
	26306	26306	

The results show a bias from the theoretical results, 5 and 6 points occurring more frequently than they should do. Are the deviations such as to forbid us to suppose the results due to random selection? Is there in apparently true dice a real bias towards those faces with the maximum number of points appearing uppermost?

We have :—

Group.	e^2.	e^2/m.	Group.	e^2.	e^2/m.
0	324	1·59606	7	5929	4·72807
1	4624	3·79951	8	121	0·30903
2	6400	1·91330	9	324	3·72414
3	10201	1·82945	10	1	0·07346
4	25281	4·03013	11	9	9·00000
5	30976	6·17298	12	0	·00000
6	19600	6·69628			
			Total...	...	43·87241

Hence $\chi^2 = 43\cdot87241$ and $\chi = 6\cdot623{,}625$.

As there are 13 groups we have to find P from the formula:

$$P = e^{-\frac{1}{2}\chi^2}\left(1 + \frac{\chi^2}{2} + \frac{\chi^4}{2\,.\,4} + \frac{\chi^6}{2\,.\,4\,.\,6} + \frac{\chi^8}{2\,.\,4\,.\,6\,.\,8} + \frac{\chi^{10}}{2\,.\,4\,.\,6\,.\,8\,.\,10}\right),$$

which leads us to

$$P = \cdot000016,$$

or the odds are 62,499 to 1 against such a system of deviations on a random selection. With such odds it would be reasonable to conclude that dice exhibit bias towards the higher points.

Illustration II.

If we take the total number of fives and sixes thrown in the 26,306 casts of 12 dice, we find them to be 106,602 instead of the theoretical 105,224. Thus $\frac{106{,}602}{12 \times 26{,}306} = \cdot3377$ nearly, instead of $\frac{1}{3}$.

Professor Weldon has suggested to me that we ought to take $26{,}306(\cdot3377 + \cdot6623)^{12}$ instead of the binomial $26{,}306(\frac{1}{3} + \frac{2}{3})^{12}$ to represent the theoretical distribution, the difference between $\cdot3377$ and $\frac{1}{3}$ representing the bias of the dice. If this be done we find:

Group.	m'.	m.	e.	e^2/m.
0	185	187	− 2	·021,3904
1	1149	1146	+ 3	·007,8534
2	3265	3215	+ 50	·777,6050
3	5475	5465	+ 10	·018,2983
4	6114	6269	−155	3·991,8645
5	5194	5115	+ 79	1·220,1342
6	3067	3043	+ 24	·189,2869
7	1331	1330	+ 1	·000,7519
8	403	424	− 21	1·040,0948
9	105	96	+ 9	·841,8094
10	14	15	− 1	·666,6667
11	4	1	+ 3	9
12	0	0	0	0

Hence: $\chi^2 = 17\cdot775{,}7555$.

This gives us by the first formula in (ii.) of art. 4:

$$P = \cdot1227;$$

or the odds are now only 8 to 1 against a system of deviations as improbable as or more improbable than this one. It may be said accordingly that the dice experiments of Professor Weldon are consistent with the chance of five or six points being thrown by a single die being ·3377, but they are excessively improbable, if the chance of all the faces is alike and equal to 1/6th.

Illustration III.

In the case of runs of colour in the throws of the roulette-ball at Monte Carlo, I have shown * that the odds are at least 1000 millions to one against such a fortnight of runs as occurred in July 1892 being a random result of a true roulette. I now give χ^2 for the data printed in the paper referred to, *i. e.*:

4274 Sets at Roulette.

Runs	1	2	3	4	5	6	7	8	9	10	11	12	Over 12
Actual ...	2462	945	333	220	135	81	43	30	12	7	5	1	0
Theory ...	2137	1068	534	267	134	67	33	17	8	4	2	1	0

From this we find $\chi^2 = 172{\cdot}43$, and the improbability of a series as bad as or worse than this is about $14{\cdot}5/10^{30}$! From this it will be more than ever evident how little chance had to do with the results of the Monte Carlo roulette in July 1892.

(7) *Frequency of General Population not known a priori.*

Illustration IV.†

In my memoir on skew-variation (Phil. Trans. vol. clxxxvi. p. 401) I have fitted the statistics for the frequency of petals in 222 buttercups with the skew-curve

$$y = {\cdot}211225x^{-{\cdot}322}(7{\cdot}3253 - x)^{3{\cdot}142}.$$

The possible range is from 5 to 11 petals, and the frequencies are :—

No. of Petals...	5	6	7	8	9	10	11
Observation ...	133	55	23	7	2	2	0
Theory	136·9	48·5	22·6	9·6	3·4	0·8	0·2

These lead to $\chi^2 = 4{\cdot}885{,}528$; whence we find for the probability of a system of deviations as much or more removed

* 'The Chances of Death,' vol. i. : The Scientific Aspect of Monte Carlo Roulette, p. 54.

† Illustrations IV. and V. were taken quite at random from my available data. Other fits with skew-curves may give much worse results, others much better, for anything I can as yet say to the contrary.

from the most probable

$$P = \cdot 5586.$$

In 56 cases out of a hundred such trials we should on a random selection get more improbable results than we have done. Thus we may consider the fit remarkably good.

Illustration V.

The following table gives the frequencies observed in a system recorded by Thiele in his *Forelaesinger over almindelig Iagttagelseslaere*, 1889, together with the results obtained by fitting a curve of my Type I. The rough values of the moments only were, however, used, and as well ordinates used measure areas :—

Groups.	Observed m'.	Curve m_1.	e.	e^2.	e^2/m.
1	0	·18	− ·18	·0324	·18
2	3	·68	− 2·32	5·3824	7·9153
3	7	13·48	+ 6·48	41·9904	3·1150
4	35	45·19	+10·19	103·8361	2·2977
5	101	79·36	−21·64	468·2896	5·9008
6	89	96·10	+ 7·10	50·4100	·5245
7	94	90·90	− 3·10	9·6100	·1058
8	70	71·41	+ 1·41	1·9881	·0278
9	46	48·25	+ 2·25	5·0625	·1049
10	30	28·53	− 1·47	2·1609	·0757
11	15	14·94	− ·06	·0036	·0002
12	4	6·96	+ 2·96	8·7616	1·2523
13	5	2·88	− 2·12	4·4944	1·5605
14	1	1·06	+ ·06	·0036	·0035
15	0	·34	+ ·34	·1156	·3400
16	0	·10	+ ·10	·0092	·0960
17	0	·00	+ 0	·0	·0
Total ...	500	500·36*	+ ·36	...	23·5000

Thus gives $\frac{1}{2}\chi^2 = 11\cdot 75 = \eta$, say.

Then

$$P = e^{-\eta}\left(1 + \frac{\eta}{1} + \frac{\eta^2}{\underline{|2}} + \frac{\eta^3}{\underline{|3}} + \frac{\eta^4}{\underline{|4}} + \frac{\eta^5}{\underline{|5}} + \frac{\eta^6}{\underline{|6}} + \frac{\eta^7}{\underline{|7}}\right).$$

Substituting and working out we find

$$P = \cdot 101 = \cdot 1, \text{ say.}$$

Or, in one out of every ten trials we should expect to differ from the frequencies given by the curve by a set of deviations as improbable or more improbable. Considering that we should get a better fit of our observed and calculated frequencies by (i.) reducing the moments, and (ii.) actually

* Due to taking ordinates for areas and fewer figures than were really required in the calculations.

calculating the areas of the curve instead of using its ordinates, I think we may consider it not very improbable that the observed frequencies are compatible with a random sampling from a population described by the skew-curve of Type I.

Illustration VI.

In the current text-books of the theory of errors it is customary to give various series of actual errors of observation, to compare them with theory by means of a table of distribution based on the normal curve, or graphically by means of a plotted frequency diagram, and on the basis of these comparisons to assert that an experimental foundation has been established for the normal law of errors. Now this procedure is of peculiar interest. The works referred to generally give elaborate analytical proofs that the normal law of errors is the law of nature—proofs in which there is often a difficulty (owing to the complexity of the analysis and the nature of the approximations made) in seeing exactly what assumptions have been really made. The authors usually feel uneasy about this process of deducing a law of nature from Taylor's Theorem and a few more or less ill-defined assumptions; and having deduced the normal curve of errors, they give as a rule some meagre data of how it fits actual observation. But the comparison of observation and theory in general amounts to a remark —based on no quantitative criterion—of how well theory and practice really do fit! Perhaps the greatest defaulter in this respect is the late Sir George Biddell Airy in his text-book on the 'Theory of Errors of Observation.' In an Appendix he gives what he terms a "Practical Verification of the Theoretical Law for the Frequency of Errors."

Now that Appendix really tells us *absolutely nothing* as to the goodness of fit of his 636 observations of the N.P.D. of Polaris to a normal curve. For, if we first take on faith what he says, namely, that positive and negative errors may be clubbed together, we still find that he has *thrice smoothed* his observation frequency distribution before he allows us to examine it. It is accordingly impossible to say whether it really does or does not represent a random set of deviations from a normal frequency curve. All we can deal with is the table he gives of observed and theoretical errors and his diagram of the two curves. These, of course, are not his proper data at all: it is impossible to estimate how far his three smoothings counterbalance or not his multiplication of errors by eight. But as I understand Sir George Airy, he would have considered such a system of errors as he gives on his p. 117 or in his diagram on p. 118 to be sufficiently represented by a normal curve. Now I have investigated his 37 groups of errors, observational and theoretical. In order to avoid so many different groups, I have tabulated his groups in $\cdot 10''$ units, and so reduced them to 21. From these 21 groups I have found χ^2 by the method of this paper. By this reduction of groups I have given Sir George Airy's curve even a better chance than it has, as it stands. Yet what do we find? Why,

$$\chi^2 = 36 \cdot 2872.$$

Or, using the approximate equation,

$$P = \cdot 01423.$$

That is to say, only in one occasion out of 71 repetitions of such a set of observations on Polaris could we have expected to find a system of errors deviating as widely as this set (or more widely than this set) from the normal distribution. Yet Sir George Airy takes a set of observations, the odds against which being a random variation from the normal distribution are 70 to 1, to prove to us that the normal distribution applies to errors of observation. Nay, further, he cites this very improbable result as an experimental confirmation of the whole theory! "It is evident," he writes, "that the formula represents with all practicable accuracy the observed Frequency of Errors, upon which all the applications of the Theory of Probabilities are founded: and the validity of every investigation in this Treatise is thereby established."

Such a passage demonstrates how healthy is the spirit of scepticism in all inquiries concerning the accordance of theory and nature.

Illustration VII.

It is desirable to illustrate such results a second time. Professor Merriman in his treatise on Least Squares * starts in the right manner, not with theory, but with actual experience, and then from his data deduces three axioms. From these axioms he obtains by analysis the normal curve as the theoretical result. But if these axioms be true, his data can only differ from the normal law of frequency by a system of deviations such as would reasonably arise if a random selection were made from material actually obeying the normal law. Now Professor Merriman puts in the place of honour 1000 shots fired at a line on a target in practice for the U.S. Government, the deviations being grouped according to the belts struck, the belts were drawn on the target of equal breadth and parallel to the line. The following table gives the distribution of hits and the theoretical frequency-

* 'A Textbook on the Method of Least Squares,' 1891, p. 14.

distribution calculated from tables of the area of the normal curve*.

Belt.	Observed Frequency.	Normal Distribution.	e.	$\frac{e^2}{m}$.
1	1	1	0	0
2	4	6	− 2	·667
3	10	27	−17	10·704
4	89	67	+22	7·224
5	190	162	+28	4·839
6	212	242	−30	3·719
7	204	240	−36	5·400
8	193	157	+36	8·255
9	79	70	+ 9	1·157
10	16	26	−10	3·846
11	2	2	0	0
	1000	1000	$\chi^2 = 45{\cdot}811$	

Hence we deduce: $P = {\cdot}000{,}00155$.

In other words, if shots are distributed on a target according to the normal law, then such a distribution as that cited by Mr. Merriman could only be expected to occur, on an average, some 15 or 16 times in 10,000,000 trials. Now surely it is very unfortunate to cite such an illustration as the foundation of those axioms from which the normal curve must flow! For if the normal curve flows from the axioms, then the data ought to be a probable system of deviations from the normal curve. But this they certainly are not. Now it appears to me that, if the earlier writers on probability had not proceeded so entirely from the mathematical standpoint, but had endeavoured first to classify experience in deviations from the average, and then to obtain some measure of the actual goodness of fit provided by the normal curve, that curve would never have obtained its present position in the theory of errors. Even today there are those who regard it as a sort of fetish; and while admitting it to be at fault as a means of generally describing the distribution of variation of a quantity x from its mean, assert that there must be some unknown quantity z of which x is an unknown function, and that z really obeys the normal law! This might be reasonable if there were but few exceptions to this universal law of error; but the difficulty is to find even the few variables which obey it, and these few are not those usually cited as illustrations by the writers on the subject!

* I owe the work of this illustration to the kindness of Mr. W. R. Macdonell, M.A., LL.D.

Illustration VIII.

The reader may ask : Is it not possible to find material which obeys within probable limits the normal law ? I reply, yes ; but this law is not a universal law of nature. We must hunt for cases. Out of three series of personal equations, I could only find one which approximated to the normal law. I took 500 lengths and bisected them with my pencil at sight. Without entering at length into experiments, destined for publication on another occasion, I merely give the observed and normal distribution of my own errors in 20 groups.

Group.	Observation.	Theory.	Group.	Observation.	Theory.
1	1	2·3	11	53	57·0
2	3	3·4	12	50·5	47·1
3	11	6·9	13	28·5	34·0
4	14·5	13·1	14	27	22·7
5	21·5	22·2	15	13·5	13·5
6	30	33·6	16	7·5	7·0
7	47	47·5	17	0	3·5
8	51·5	57·8	18	1	1·6
9	72	63·2	19	0	·6
10	65·5	62·7	20	2	·3

Calculating χ^2 in the manner already sufficiently indicated in this paper, we find

$$\chi^2 = 22{\cdot}0422.$$

We must now use the more complex integral formula for P, and we find

$$P = {\cdot}2817.$$

Or, in every three to four random selections, we should expect one with a system of deviations from the normal curve greater than that actually observed.

I think, then, we may conclude that my errors of judgment in bisecting straight lines may be fairly represented by a normal distribution. It is noteworthy, however, that I found other observers' errors in judgment of the same series of lines were distinctly skew.

(8) We can only conclude from the investigations here considered that the normal curve possesses no special fitness for describing errors or deviations such as arise either in observing practice or in nature. We want a more general theoretical frequency, and the fitness of any such to describ a given series can be investigated by aid of the criterion discussed in this paper. For the general appreciation of the probability of the occurrence of a system of deviations defined by χ^2 (or any greater value), the accompanying table has been calculated, which will serve to give that probability closely enough for many practical judgments, without the calculations required by using the formulæ of art. 4.

Table of Values of P for Values of χ^2 and n'; χ^2 from 1 to 70, n' from 3 to 20 *.

χ^2 \ n'	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.
1	·606,531	·801,253	·909,796	·962,566	·985,612	·994,829	·998,249	·999,438	·999,828	·999,950	·999,986	·999,997	·999,999	1·	1·	1·	1·	1·
2	·367,879	·572,407	·735,759	·849,146	·919,699	·959,839	·981,012	·991,466	·996,340	·998,494	·999,406	·999,772	·999,917	·999.969	·999,990	·999,996	·999,999	·999,999
3	·223,130	·391,633	·557,825	·699,994	·808,847	·885,010	·934,357	·964,303	·981,424	·990,734	·995,544	·997,942	·999,074	·999,605	·999,830	·999,938	·999,972	·999,997
4	·135,335	·261,470	·406,006	·549,422	·676,676	·779,783	·857,123	·911,418	·947,347	·969,923	·983,436	·991,197	·995,466	·997,743	·998,903	·999,489	·999,763	·999,899
5	·082,085	·171,799	·287,298	·415,882	·543,813	·659,965	·757,576	·834,310	·891,178	·931,168	·957,979	·975,195	·985,813	·992,128	·995,754	·997,772	·998,860	·999,433
6	·049,787	·111,611	·199,148	·306,220	·423,190	·539,750	·647.232	·739,919	·815,263	·873,366	·916,082	·946,154	·966,491	·979,749	·988,095	·993,187	·996,197	·997,930
7	·030,197	·071,888	·135,888	·220,631	·320,847	·428,870	·536,632	·637,110	·725,544	·799,074	·857,613	·902,142	·934,711	·957,640	·973,260	·983,539	·990,125	·994.203
8	·018,316	·046,012	·091,578	·156,236	·238,103	·332,594	·433,470	·534,146	·628,837	·713,304	·785,131	·843,601	·889,327	·923,783	·948,867	·966,547	·978,637	·986,671
9	·011,109	·029,291	·061,099	·109,064	·173,578	·252,656	·342,296	·437,274	·532,104	·621,892	·702,931	·772,944	·831,051	·877,518	·913,414	·940,262	·959,743	·973,479
10	·006,738	·018,567	·040,428	·075,236	·124,652	·188,574	·265,026	·350,486	·440,493	·530,388	·615,960	·693,935	·762,183	·819,740	·866,628	·903,611	·931,906	·952,946
15	·000,553	·001,817	·004,701	·010,363	020,256	·036,000	·059,145	·090,810	·132,061	·182,371	·241,436	·307,227	·378,154	·450,691	·524,638	·594,156	·661,967	·721,272
20	·000,045	·000,170	·000,499	·001,250	·002,769	·005,570	·010,336	·017,913	·029,253	·045,341	·067,086	·095,212	·130,141	·171,934	·220,220	·274,231	·332,819	·394,580
25	·000,004	·000,016	·000,050	·000,139	000,341	·000,759	·001,554	·002,971	·005,345	·009,117	·014,822	·023,084	·034,566	·049.943	·069,824	·094,710	·124,915	·160,542
30	·000,000	·000,001	·000,005	·000,015	·000,039	·000,095	·000,211	·000,439	·000,857	·001,585	·002,792	·004,710	·007,632	·011,921	·018,002	·026,345	·037,446	·051,798
40	·000,000	·000,000	·000,000	·000,000	·000,000	·000,001	·000,003	·000,008	·000,017	·000,036	·000,072	·000,138	·000,255	·000,453	·000,778	·001,294	·002,087	·003,272
50	·000,000	·000,000	000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000,5	·000,001	·000,003	·000,006	·000,012	·000,023	·000,042	·000,075	·000,131
60	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,001	·000,002	·000,004
70	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000	·000,000

* I have to thank Miss Alice Lee, D.Sc., for help in the calculation of part of this table. The certainty, *i. e.* the 1· in columns 16 to 20, denotes, of course, something greater than ·999,9995, *i. e.* unity to six figures.