THE RELATIONS BETWEEN STABILITY AND HOMOGENEITY*

By

L. v. Bortkiewicz

The idea of investigating the stability of statistical frequencies from the standpoint of the theory of probability goes back to the French mathematician Bienaymé. From various examples taken from social and moral statistics, he was the first to establish the fact that, almost without exception, the stability in question was essentially less than the "classical norm," that is, less than the expectation which is associated with the classical scheme of independent trials with a constant underlying probability. In order to explain this discrepancy between theory and observation, Bienaymé used a modification of the traditional procedure which was characterized by the assumption that between neighboring trials in a time ordered sequence a sort of dependence existed. Though interesting in itself and among other things adopted by Cournot as his own, we shall replace this method in what follows by another, originating from Lexis, which has the advantage of a wider usefulness, in that it can be applied not only

*Translated by A. R. Crathorne. Read before the American Statistical Association at Cleveland, Ohio, December 30, 1930.

to undulatory but to evolutory sequences.[1]

Let us assume that for a series of z successive time intervals, say years, we have found that some event (accident, death, marriage, crime) has happened x_1, x_2, ... times, and that the corresponding number of "trials," that is the numbers of persons observed, are s_1, s_2, ... so that the quotients $y_1 = \frac{x_1}{s_1}$, $y_2 = \frac{x_2}{s_2}$, ... represent a time ordered sequence of relative frequencies. Instead of assuming, as the traditional theory demands, that each term y_k of this series corresponded to a common fundamental probability p, weighted with accidental errors, Lexis assumed that each value y_k was associated with a distinct probability p_k.

As a result of this, the expected amplitude of the fluctuations of the values y_k increased, and the greater the variations in the p_k's the greater the amplitude. Under the simplifying hypothesis s_k = const. ($= s$), the corresponding standard deviation σ is defined by

$$\sigma^2 = \frac{1}{z}\sum_{k=1}^{z}(y_k - y)^2, \quad y = \frac{1}{z}\sum_{k=1}^{z} y_k$$

For the case of a constant p we may write

$$E(\sigma^2) = \frac{z-1}{z} \cdot \frac{p(p-1)}{s} \tag{1}$$

where E denotes "expectation." In the Lexis procedure with a variable p_k, using the notation

$$\frac{z-1}{z} \cdot \frac{p(1-p)}{s} = u^2, \quad \frac{1}{z}\sum_{k=1}^{z} p_k = p, \quad p_k - p = \varepsilon_k, \; \frac{1}{z}\sum_{k=1}^{z} \varepsilon_k^2 = \omega^2,$$

[1] Bienayme, in the journal "L'Institute," Vol. 7 (1831), pages 187-189, and in "Journal de la Societe de Statistique de Paris," 17e (1876), pages 199-204. A. Cournot, Exposition de la theorie des chances et des probabilities, Paris, 1843, Nos. 79 and 117.
W. Lexis, "Uber die Theorie der Stabilitat statistischer Reihen," in the Jahrbuch fur Nationalokonomie und Statistik, Vol. 32 (1879), pages 60 . ., reprinted in Abhandlungen zur Theorie der Bevölkesungs und Moralstatistik, Jena, 1903, pages 170-212.

the corresponding relation

$$E(\sigma^2) = u^2 + \frac{zs - z + 1}{zs} \omega^2 \tag{2}$$

can be derived.[1]

In the following numerical examples the numbers of observations s_k are never less than some ten thousands, while $z = 10$. Hence, as far as these and similar examples are concerned, the numerical results are not appreciably altered if, instead of (2), we use

$$E(\sigma^2) = u^2 + \omega^2 \tag{3}$$

However, a certain inaccuracy arises, if, in the application of formula (3) to the raw data, one has disregarded the fundamental assumption that s_k is constant and in the expression for u^2 has replaced s by the arithmetic mean of the z values s_k. If, however, the latter differ little from one another, such a procedure gives rise to no great discrepancy. Lexis called the quantities u and ω in formula (3) the two "fluctuation components," which combine (according to the law of composition of forces) to give the expected total fluctuation. The quantity u gives expression to the effect of the "accidental causes" in the sense of the theory of probability, and this effect grows less and less with increasing s until it vanishes for $s = \infty$. For this reason Lexis called u the normal component. He also used the term "unessential fluctuation componont." On the other hand, ω depends on the variations of the fundamental probability, that is on the underlying general conditions, and in this sense was designated by Lexis as the physical component. We may also

[1] One does not find formula (2) in Lexis's work. He was satisfied at this point with a rather inexact method yielding an approximate result. However, this did not affect the essential part of his discussion.

call it the essential component.

The first of the two components u and ω can be easily calculated directly with sufficient approximation. The usual method is to substitute for the unknown p in the expression for u^2 the value y, the arithmetic mean of the frequencies y_k, obtaining

$$u^2 = \frac{z-1}{z} \cdot \frac{y(1-y)}{s} \tag{4}$$

As for the second component ω, it is calculated by the indirect method of substituting σ^2 for $E(\sigma^2)$ in (3) and then ω is found from $\omega^2 = \sigma^2 - u^2$. This method, however, assumes that $\sigma > u$, or what is the same thing, that the dispersion coefficient, $Q = \frac{\sigma}{u}$, is greater than 1. In his older papers, Lexis distinguished between subnormal, normal and supernormal dispersion, according to whether Q was distinctly less than 1, approximately equal to 1, or distinctly greater than 1, and found that in social and moral statistics the subnormal dispersion never occurred and the normal rarely. Supernormal dispersion was the rule. So Lexis based his scheme of a varying underlying probability on the case of supernormal dispersion. In fact, from formula (3), we have

$$E(Q^2) = 1 + \left(\frac{\omega}{u}\right)^2 . \tag{5}$$

which says that the variations in the underlying probability lead us to expect values of Q greater than unity.[1]

Notwithstanding the fact that Q was usually greater than unity, Lexis did not consider this a proof that his scheme ade-

[1]Under the influence of accidental causes, Q may be less than unity not only for constant, but also for varying underlying probabilities, and this circumstance must be considered in the determination of ω. It would carry us too far afield to go further into this matter.

quately described the actual facts. In addition to this he was more concerned with the fact that in experience Q showed a tendency to decrease with decreasing number of "trials," that is with decreasing s. Indeed, in a series of examples, Lexis had shown that a value of Q which was decidedly greater than unity when calculated for an entire country, decreased to nearly 1 when the data for the single administration districts of the same country were used. Lexis considered such behavior of Q as entirely in harmony with his scheme.

If we write formula (5) in the form

$$(6) \qquad E(Q^2) = 1 + s\,\frac{z\,\omega^2}{(z-1)p(1-p)},$$

we see that the excess of Q^2 over and above 1 is in expectation directly proportional to s. This was the explanation of the decrease of Q with decreasing s, for as Lexis said, we have no ground to expect that s being large or small had any bearing on the value of ω.

It is this last point about which the criticism of Lexis's dispersion theory centers. Notwithstanding the endeavors of Lexis to fit his theory to statistical reality, we can show that the facts were against him as far as his assumption that ω is fundamentally independent of s is concerned. If this assumption were true, then formula (6) tells us distinctly how Q decreases with diminishing s. We learn from experience that as a rule this decrease in Q is less than that given by the formula; from which it follows that the essential component, ω, has a tendency to increase with decreasing s.

If we desire to investigate just what happens in reality, a certain complication arises, because we are never able to compare groups which differ among one another as to s, but not as to p (or y). In order to eliminate to some extent the variations of p we consider the ratio of ω to p. Let $\frac{\omega}{p} = \beta$, and call β the *relative* essential component to distinguish it from the *absolute* essential component ω. Formula (6) then becomes the following:

$$E(Q^2) = 1 + sp \frac{z\beta^2}{(z-1)(1-p)} \tag{7}$$

The product sp can be considered as the expected number of "successes." For a constant s_k (= s) we have

$$E(x_k) = sp_k, \qquad E\left(\frac{1}{z}\sum_{k=1}^{z} x_k\right) = sp$$

and, letting $s = \frac{1}{z}\sum s_k$, the last relation is true with sufficient approximation for a variable s_k provided the variation is not too pronounced. Let $sp = m$. Often, as in the examples which follow, p is so small that we can consider ($1-p$) as equal to 1. Formula (7) then becomes

$$E(Q^2) = 1 + \frac{z}{z-1} m\beta^2 \tag{8}$$

The question as to whether there is a connection between s and ω is now changed to an investigation of the relationship between m and β. In undertaking such an investigation empirically, we compare as to the behavior of m and β a statistical aggregate considered as a total with its component parts considered as partial aggregates. Let the number of the partial aggregates be n, and let the corresponding values of m and β as well as u, ω and σ be indicated by the subscript i, which can also serve as the ordinal number of the partial aggregate. For the total aggregate, let $i = 0$. The symbols $s_{i,k}$, $x_{i,k}$, $y_{i,k}$, $p_{i,k}$, are the s, x, y, p of the i th partial aggregate and the k th time interval. We also use the notation

$$s_i = \frac{1}{z}\sum_{k=1}^{z} s_{i,k}\,, \quad x_i = \frac{1}{z}\sum_{k=1}^{z} x_{i,k}\,,$$

$$y_i = \frac{1}{z}\sum_{k=1}^{z} y_{i,k}\,, \quad p_i = \frac{1}{z}\sum_{k=1}^{z} p_{i,k}\,.$$

from which we have

$$s_o = \sum_{i=1}^{n} s_i\,, \qquad x_o = \sum_{i=1}^{n} x_i\,,$$

$$y_o = \frac{1}{s_o}\sum_{i=1}^{n} s_i y_i\,, \quad p_o = \frac{1}{s_o}\sum_{i=1}^{n} s_i p_i\,.$$

We have also the following relations:

$$m_i = s_i p_i\,, \quad \sigma_i^2 = \frac{1}{z}\sum_{k=1}^{z}(y_{i,k} - y_i)^2,$$

$$u_i^2 = \frac{z-1}{z} \cdot \frac{p_i(1-p_i}{s_i}\,, \quad \omega_i^2 = \frac{1}{z}\sum_{k=1}^{z} e_{i,k}^2$$

$$\text{where} \quad e_{i,k} = p_{i,k} - p_i\,, \quad \beta_i = \frac{\omega_i}{p_i}\,,$$

$$E(\sigma_i^2) = u_i^2 + \omega_i^2\,, \quad Q_i = \frac{\sigma_i}{u_i}\,,$$

and using the notation $\frac{e_{i,k}}{p_{i,k}} = \varepsilon_{i,k}$ we have further

$$\beta_i^2 = \frac{1}{z}\sum_{k=1}^{z} \varepsilon_{i,k}^2$$

Finally, corresponding to formula (8), we have

$$E(Q_i^2) = 1 + \frac{z}{z-1} m_i \beta_i^2 \tag{9}$$

We shall now apply these formulas to statistics on the frequency of suicides in Germany for the decade 1902-1911. The numbers of "trials," $s_{i,k}$, are here the populations of the regions in question; the "successes," $x_{i,k}$, are the numbers of suicides for each year. The relative frequencies, $y_{i,k}$, are found by dividing the numbers of suicides by the corresponding populations. Like various other kinds of social phenomena, the suicides in pre-war German statistics were grouped according to states, the provinces of Prussia, right Rhenish Bavaria and left Rhenish Bavaria being included as states. In this way we have forty territories of very unequal size. For the decade 1902-1911, the mean population of the territories ranged from a maximum of 6,587,000 (Rhine Province) to a minimum of 45,000 (Schaumburg-Lippe). The maximum average number of suicides per annum was 1453 (Saxony) and the minimum 7 (Schaumburg-Lippe). Corresponding to the purpose of the investigation, these suicide figures x_i , which can be considered as approximations to m_i , were arranged in descending order, with $x_1 = 1453$ and $x_{40} = 7$.

For the whole of Germany, we have $x_0 = 13173$, y_0 $214 \cdot 10^{-6}$ (that is an average number of 214 suicides per annum for each million population). The ten values $y_{0,k}$ vary between $204 \cdot 10^{-6}$ and $223 \cdot 10^{-6}$ These fluctuations are markedly greater than one expects from the classical norm. The calculation of the dispersion-quotient gives $Q_0 = 3.14$, and, as the Lexis theory demands, is greater than any one of the 40 values of Q_i.[1] These values give 2.03 as a maximum and 0.75 as a

[1] A study of suicides and of homicides in the United States yields much the same general results as those shown here for suicides in Germany. (Note by the translator.)

minimum. Fixing attention on the eight smallest values of x_i, we find an average value of 1.02 for Q_i, and of the eight values, three are larger and five less than 1. So in this example the dispersion becomes very nearly 1 by narrowing the observation field.

But we have still to find out whether Q_i decreases with x_i according to the measure of decrease that one would expect under the hypothesis that β_i is fundamentally independent of x_i. To decide this question, we let β_i = const. $= \beta$, including $\beta_0 = \beta$, and substitute also x_i for m_i in formula (9). We have then on the one hand in expected values

$$Q_0^2 = 1 + \frac{z}{z-1} x_0 \beta^2$$

and on the other hand

$$\frac{1}{n} \sum_{i=1}^{n} Q_i^2 = 1 + \frac{z}{z-1} \cdot \frac{x_0}{n} \beta^2$$

from which follows

$$\frac{1}{n} \sum_{i=1}^{n} Q_i^2 = 1 + \frac{1}{n} (Q_0^2 - 1)$$

However, in our example, we find

$$\frac{1}{n} \sum_{i=1}^{n} Q_i^2 = 1.56, \qquad 1 + \frac{1}{n} (Q_0^2 - 1) = 1.22$$

and the difference 0.34 cannot be ascribed to chance for it is three times the probable error (the determination of which we cannot now take up). We must, then, assume that the average of the values β_i, for $i = 1$ to 40, is greater than β_0. Why this is so we shall see in the following discussion.

We consider now the mutual relationship between the deviations $\varepsilon_{i,k}$ and $\varepsilon_{j,k}$ which refer to two arbitrary territories N_i and N_j, and we build up according to the formula for a correlation coefficient the expression

$$\gamma_{i,j} = \frac{1}{z}\sum_{k=1}^{z} \frac{\varepsilon_{i,k}\,\varepsilon_{i,k}}{\beta_i \beta_j}$$

The number of combinations of the subscripts i and j is $\frac{n(n-1)}{2}$, so there are that many values $\gamma_{i,j}$. Finally we construct a weighted arithmetic mean of these values according to the formula,

$$\gamma = \frac{\sum_{i=1}^{n}\sum_{j=i+1}^{n} m_i m_j \beta_i \beta_j \gamma_{i,j}}{\sum_{i=1}^{n}\sum_{j=i+1}^{n} m_i m_j \beta_i \beta_j}$$

The expression γ serves to characterize the mutual relationship of time ordered series of fundamental probabilities $p_{i,k}$, hence also of relative frequencies $y_{i,k}$, which may be considered as approximations to $p_{i,k}$. If we give the name "syndromy" to such an array of simultaneously distinct fundamental probabilities (or relative frequencies), we may call γ a "coefficient of syndromy." For $\gamma = 1$, we shall speak of "isodromy," for $1 > \gamma > 0$, of "homodromy," for $\gamma = 0$, of "paradromy," and for $\gamma < 0$, of "antidromy." We may include the last three cases, namely $\gamma < 1$, under the name "anisodromy."

With the help of γ we can exhibit the relation between β_0 on the one hand and the n values $\beta_1, \beta_2, \ldots \beta_n$ on the other hand as follows:

$$(10) \quad m_0^2 \beta_0^2 = \sum_{i=1}^{n} m_i \beta_i^2 + \gamma \left\{ \left(\sum_{i=1}^{n} m_i \beta_i\right)^2 - \sum_{i=1}^{n} m_i^2 \beta_i^2 \right\}$$

Since $m_0 = \sum_{i=1}^{n} m_i$, we find for $\gamma = 1$, from (10)

$$\beta_0 = \frac{\sum_{i=1}^{n} m_i \beta_i}{\sum_{i=1}^{n} m_i} \tag{11}$$

and for $\gamma < 1$

$$\beta_0 < \frac{\sum_{i=1}^{n} m_i \beta_i}{\sum_{i=1}^{n} m_i} \tag{12}$$

Hence, only in the case of isodromy is the assumption justified that the relative essential fluctuation component for the total aggregate is as large as that for the partial aggregates. In every other case, namely for anisodromy, the relative essential component for the total aggregate falls below the level for the partial aggregates more and more as γ becomes less and less.

In the suicide example under consideration we have homodromy, which is reasonable, since the fluctuations in suicide frequency in the single states are influenced in part by factors which are not local but general for all Germany. Somewhat tedious calculations give $\gamma = 0.38$. At the same time we find $\beta_0 = 0.0246$ approximately, while the average for β_i, $i = 1$ to 40 is 0.0392.

If now we group the 40 states into five groups so that states numbered 1 to 8 form the first group, states numbered 9 to 16 the second, and so on, we find as average values of β_i, 0.0354, 0.0358, 0.0485, 0.0528 and 0.0767. The quantities β_i then show a tendency to increase as x_i (or m_i) decreases.

If, as in this example, the total aggregate is a "natural unit," we should expect to have homodromy in the vast majority of cases. On the other hand, we should expect paradromy if the total aggregate is an "artificial unit," that is, one made up by throwing together entirely unrelated groups. As an illustration of paradromy we take the array of marriage frequencies for the six cities, Barcelona, Birmingham, Boston, Leipzig, Melbourne and Rome, for the decade 1899-1908. By marriage frequence we mean the ratio of the number married (twice the number of marriages) to population.

For the six cities taken as a whole, with a total population of about three million, the marriage frequence $y_{o,k}$ varies between 18.00 and 19.02 per cent with an average of 18.38 per cent. The dispersion coefficient Q_o is 3.17. For the six cities taken singly in the above order, each with a population of about half a million, the values of Q_i are 2.69, 4.32, 4.17, 2.88, 3.76 and 2.72, with an average 3.42, somewhat higher than Q_o This result is a direct contradiction of the statement of Lexis that a narrowing field of observation reduces the value of Q. Lexis, without giving the matter much thought, worked with the hypothesis that isodromy, or at least a decided homodromy, always existed. In our example, however, we have paradromy, if not antidromy, for we find γ to be -0.054. Corresponding to this, we have β_o less than each of the values β_1 to β_6, for β_o approximates 0.0167 while β_i, $i = 1$ to 6, lies between 0.0334 and 0.0563. The quadratic mean of these quantities is 0.0450.

It is of prime interest to investigate for paradromy the theoretical relation of β_o to the quadratic mean of the values β_1, β_2 . . . β_n and of Q_o to the quadratic mean of Q_1, Q_2 . . . Q_n, for the case m_i = const. = m. In this case, $m_o = nm$, and if O is substituted for γ in (10) we have

$$\beta_o^2 = \frac{1}{n^2}\sum_{i=1}^{n}\beta_i^2 \quad \text{, whence} \quad \beta_o = \frac{1}{\sqrt{n}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\beta_i^2}$$

At the same time, we find on the one hand, from (9), the expected value

$$Q_o^2 = 1 + \frac{z}{z-1} m_o \beta_o^2 ,$$

or

$$Q_o^2 = 1 + \frac{z}{z-1} \frac{m}{n} \sum_{i=1}^{n} \beta_i^2$$

and on the other hand

$$\frac{1}{n} \sum_{i=1}^{n} Q_i^2 = 1 + \frac{z}{z-1} \frac{m}{n} \sum_{i=1}^{n} \beta_i^2$$

whence

$$Q_o = \sqrt{\frac{1}{n} \sum_{i=1}^{n} Q_i^2}$$

In the marriage frequence example, where the quantities m_i, though not equal, differ very little from one another, we have the values already found

$$\beta_o = 0.0167 \text{ and } \qquad Q = 3.17$$

to compare with the values

$$\frac{1}{\sqrt{n}} \sqrt{\frac{1}{n} \sum_{i=1}^{n} \beta_i^2} = 0.0184$$

and

$$\sqrt{\frac{1}{n} \sum_{i=1}^{n} Q_i^2} = 3.49$$

The differences $0.0167 - 0.0184 = -0.0017$ and $3.17 - 3.49 = -0.32$ are explained partly by the fact that the assumption $m_i = \text{const.}$ is not exactly in accord with the facts, and partly because paradromy is really not present as assumed, but only a weak antidromy. This last should, however, be considered as due to chance. The artificial character of a total aggregate shows itself in paradromy.

Of the two quantities Q and β, only the latter can be considered as a proper measure of the stability of a statistical frequency—more exactly, of the corresponding fundamental probability. And, since on account of formulas (11) and (12), the total aggregate can never show a higher value of β than the average for the partial aggregates (because the upper limit for γ is 1), we obtain a glimpse of the question of the connection between stability and homogeneity.

The idea of homogeneity as we here understand it has reference to the result of the decomposition of a statistical aggregate according to some attribute or complex of attributes. The aggregate may consist of s elements, say s human beings and the decomposition may yield N sub-aggregates containing s', s'' . . . elements. Let some event A be observed x times in the total aggregate and x', x'', . . . times in the sub-aggregates. If we find the relative frequencies

$$y = \frac{x}{s}, \qquad y' = \frac{x'}{s'}, \qquad y'' = \frac{x''}{s''}, \ldots\ldots$$

then, on account of the two identities, $s' + s'' + \ldots = s$, and $x' + x'' + \ldots = x$, we have the relation

$$y = \frac{s'y' + s''y'' + \cdots\cdots}{s' + s'' + \cdots\cdots}$$

The "general frequency" then appears as the weighted arithmetic mean of the "special frequencies," y', y'', . . .

The theory of probabilities, with more or less assurance, furnishes us a criterion for deciding whether or not the deviations of the quantities y', y'', . . . from y are due to chance.

If they are not due to chance we say that the total aggregate "reacts" to the decomposition in question and that the attribute or complex of attributes which governs the decomposition is "relevant." If they are due to chance, we say that the total aggregate does not react to the decomposition and that the attribute is "indifferent."

According to the standpoint of the theory of probability, the relative frequencies y, y', y'' . . . as also the quotients $\frac{S'}{S}$, $\frac{S''}{S}$, . . . can be considered as approximations of distinct probabilities. If we designate the two series of probabilities thus inferred by p, p', p'', . . . and g', g'', . . . respectively, we find

$$p = g'p' + g''p'' + \cdots \tag{13}$$

and the character of the attribute in question as relevant or indifferent finds expression in the fact that the "special probabilities" p', p'', . . . either differ from one another or are all equal to p, the "general probability."

For every ample enough complex of attributes we can imagine the decomposition going on and on by applying one attribute of the complex after another. Finally a point is reached where the sub-aggregates no longer react to further decomposition, or, expressed otherwise, the supply of relevant attributes is exhausted, and the probabilities p', p'', . . . which are associated with these sub-aggregates are called "elementary probabilities." In this case we say that the sub-aggregates themselves are "completely homogeneous" with reference to the event A.

The total aggregate—still in reference to A—is the more diversified the more the elementary probabilities p', p'', . . . differ among themselves, that is, the more they differ from p. It is reasonable to take as a measure of this diversity the expression

δ , defined by

$$\delta^2 = g'(p'-p)^2 + g''(p''-p)^2 + \cdots \tag{14}$$

Diversity and homogeneity are antithetical notions; the more undiversified the aggregate, the more it is homogeneous, and vice versa.

In order to apply this view of homogeneity, now considered for itself, to the procedure and the examples which we have brought forward in the discussion of stability, we must disregard the time fluctuations of the probabilities in question. That is, we do not use the quantities $p_{i,k}$ but fix attention on the probabilities p_i which refer to an individual time interval of n partial intervals—say a decade. By carrying out repeatedly the decomposition according to formula (13), the quantities p_i , p_o not included may be expressed in the form

$$p_i = g_i' p_i' + g_i'' p_i'' + \cdots$$

where p_i' , p_i'' . . . are elementary probabilities. Corresponding to formula (14), we have

$$\delta_i^2 = g_i'(p_i'-p_i)^2 + g_i''(p_i''-p_i)^2 + \cdots \tag{15}$$

If we designate the proportion of the i th partial aggregate to the total aggregate by c_i , that is, if we let $\frac{S_i}{S_o} = c_i$, we find

$$p_o = \sum_{i=1}^{n} c_i p_i$$

and at the same time

$$(16) \qquad \delta_0^2 = \sum_{i=1}^{n} \left\{ c_i g_i' (p_i' - p_0)^2 + c_i g_i'' (p_i'' - p_0)^2 + \cdots \right\}$$

The number of summands in (16) is nN, since there are n partial aggregates and each of these is a totality of N sub-aggregates. It may easily occur that some of the nN elementary probabilities are equal and this is expected in connection with elementary probabilities which are associated with similar sub-aggregates. But even in the most extreme case, where the elementary probabilities are equal without exception, we cannot say that the probabilities p_i are all alike. This can occur only when the values g_i', g_i'', . . . are independent of i. This highly improbable case is excluded from our discussion. We have then

$$(17) \qquad \sum_{i=1}^{n} c_i (p_i - p_0)^2 > 0$$

From (15) and (16), we have the following:

$$g_i' (p_i' - p_0)^2 + g_i'' (p_i'' - p_0)^2 + \cdots = \delta_i^2 + (p_i - p_0)^2$$

$$\delta_0^2 = \sum_{i=1}^{n} c_i \delta_i^2 + \sum_{i=1}^{n} c_i (p_i - p_0)^2$$

so that, on account of (17)

$$\delta_0^2 > \sum_{i=1}^{n} c_i \delta_i^2$$

and *a fortiori*

$$(18) \qquad \delta_0 > \sum_{i=1}^{n} c_i \delta_i$$

The total aggregate is then under all circumstances less homogeneous than the partial aggregates are on the average

This statement might possibly correspond to the every-day meaning of the word "homogeneity," which carries with it no precise quantitative idea. Indeed, when we consider that in the case of the total aggregate we have to take into account not only the lack of homogeneity within the partial aggregates, but also the diversity with which the partial aggregates may make up the whole, we are inclined to say that the total aggregate is less homogeneous than any of its parts. With that idea, however, we do not hit upon the right thing as far as our mathematical criterion of homogeneity is concerned. The inequality (18) says only, that the average of the values δ_1, δ_2, . . . δ_n is less than δ_0, not that each one is less than δ_0.

In our foregoing discussion of stability as measured by the relative essential fluctuation component, we found that for the total aggregate the stability was higher than the average for the partial aggregates, except for the case of isodromy, which in practice rarely occurs. Hence, there exists between homogeneity and stability an antagonistic relation—small homogeneity goes hand in hand with great stability. For example, the provinces into which a country may be divided will show, on the average, a greater homogeneity and at the same time a lesser stability in reference to an event A than will the country taken as a whole.

Again, the districts into which the provinces may be divided will on the average show a greater homogeneity associated with a still smaller stability. We can say that in general the homogeneity increases with the narrowing of the field of observation, while the stability decreases.

Is this to be considered as a warning against the all too popular diversification of statistical material which is being more and more accepted in research methods? Not in the least. That would be an obsolete point of view, as if the problem of statistics consisted in a search for most stable values. Rather does the opposition between homogeneity and stability give direction to business practice, especially to that branch of business which is in such close touch with statistics, namely insurance, where stability is of prime importance. It has been known for a long time that it contributes to the even tenor of the business side if the risks are as heterogeneous as possible. It is of advantage if the insured persons or things are spread relatively widely according to geographical and other points of view, instead of concentrating on a limited territory or few kinds of risks.

Accordingly, even if this thesis, that an antagonistic relation exists between homogeneity and stability, seems surprising and strange, we find on closer consideration that the theory agrees with a practice which has instinctively grasped the true situation. It is now twelve years since I had the first opportunity to explain at greater length than here the foregoing developed ideas and with the verifying data to present them to my colleagues.[1] As far as I know, only one of these has taken a definite stand in the matter. This is John Maynard Keynes.[2] He makes the charge against me, that instead of clearing up a very simple matter, I have befogged it with a profusion of mathematical formulas

[1] Homogeneitat und Stabilitat in der Statistik, in the Skandinavisk Aktuarietidskrift, 1918, pages 1-81, Upsala.

[2] A treatise on probability, London, 1921, pages 403-405.

and new technical terms, and he believed that he could show this best by an example of my own from the field of insurance. In referring to this example, Keynes thought that the distinction made by myself in a much earlier publication between a general probability p and the special probabilities p_1, p_2, . . . was the one in question, where

$$p = \frac{z_1}{z} p_1 + \frac{z_2}{z} p_2 + \cdots$$

Keynes further expressed himself as follows:

> "If we are basing our calculations on p and do not know p_1, p_2, etc., then these calculations are more likely to be borne out by the result if the instances are selected by a method which spreads them over all the groups 1, 2, etc., than if they are selected by a method which concentrates them on group 1. In other words the actuary does not like an undue proportion of his cases to be drawn from a group which may be subject to a common relevant influence *for which he has not allowed.* If the *a priori* calculations are based on the average over a field which is not homogeneous in all its parts, greater stability of result will be obtained if the instances are drawn from all parts of the non-homogeneous total field. than if they are drawn now from one homogeneous sub-field and now from another. This is not at all paradoxical. Yet I believe, though with hesitation, that this is all that Von Bortkiewicz's elaborately supported mathematical conclusion amounts to."

Suppose, for example, that a fire insurance company insures

[1]Here z refers to a series of "equally likely events," which is broken up into groups of z_1, z_2, ... equally likely events. Hence $z = z_1 + z_2 + \cdots$

two kinds of buildings, dwellings and factories, which are classified as different grades of fire risks, for insurance premiums which are not graded. The premium is to be calculated per unit on the supposition that the risks in the two categories are divided in a definite proportion. Then, according to Keynes, a greater stability in the business is guaranteed if every year dwellings as well as factories are insured, than if in one year only dwellings and in another year only factories are insured. This is certainly true and requires no lengthy argument. But it has nothing whatever to do with my thesis of the antagonistic relation between stability and homogeneity.

To give an example which does illustrate my theory, think of three insurance companies, A, B, and C. A insures only dwelling houses, B only factories, while C insures both. The premiums in A, B. and C are different because of the different classes of risks. It is assumed in C that there is no grading of premiums. A premium per unit is charged which is calculated according to the relative number of the two risks. The premium is to be just high enough so that for a period of years, allowing for variations due to chance, the damages are just covered. In the course of this period, the danger of fire varies from year to year, showing gains in some years, losses in others. Such fluctuations of fire hazard would correspond in my scheme to the variations of the probabilities $p_{i,k}$ with respect to k, while $p_{i,k}$ is associated with A, $p_{2,k}$ with B, and $p_{a,k}$ with C. And in accord with my theory that, except in the case of isodromy, the values $p_{0,k}$, relatively speaking, show weaker variations than $p_{1,k}$ and $p'_{2,k}$ do on the average, the insurance company C would show relatively smaller fluctuations of fire damage from one year to another, resulting in a more stable business than would be shown by the average of A and B. The mixed character of the risks would be conducive to greater stability. In the case of C a certain compensation of effects would take place which the time variations of the two-sided fundamental probabilities would make manifest on the business side.[1] But Keynes says nothing of these variations. He simply missed the point of my argument and his remarks were not relevant.

It is to be hoped that the new exposition of my theory, although, or because, it is essentially shorter than the older one, will give no cause for a similar misunderstanding.

---

[1]This compensation would also appear in the more complicated case where the proportions of the risks in C are not unchangeable as is assumed in the text, but would change from year to year (the premium being adjusted accordingly). We need not go further into this matter because, in my theory, the composition of $s_{\bullet,k}$ out of the component parts $s_{i,k}$ is considered as fixed. In my examples, this composition varied, but the fluctuations were insignificant in comparison to the variations of the values $p_{i,k}$ See Skandinavisk Aktuarietidscrift, pages 69-70.

L. v. Bortkiewicz.