

On the Statistics of Interdependent Events

English reading proof · Version 0.2

F. Eggenberger and G. Pólya

16 October 2026

Zürich

[279] In most applications of probability theory, whether in physical, biological or social statistics, only “independent” events are considered. The n individual cases in a statistical series are called “independent” of one another when the chances in each case remain unaffected by the outcome of the other $n - 1$. We may, for example, combine all births in Germany during a calendar month into a series and record at each birth whether the newborn child is a boy or a girl. In this respect the individual births are independent of one another; at least, the available numerical material gives no reason to suppose that the chances of a boy being born in the second half of the month are in any way affected by the proportion of boys in the first half — as most readers will believe even without a statistical investigation. But we may also combine into a statistical series all the people who board a railway train in Germany during a calendar month and record for each traveller whether he died during the journey as a result of a railway accident. The individual events in this series are not independent of one another. For the lives of the occupants of the same train are to a high degree “bound together”: the death of one person in a railway accident must be regarded as an extraordinary deterioration in the chances of all his fellow travellers.

The theoretical treatment of events that are not independent would surely be very important for every field in which probability theory is applied, but it is in fact very difficult. An extraordinarily large number of different structures of mutual dependence are conceivable; one must select a structure that, first, represents the structure of the observed series approximately correctly and, second, lends itself sufficiently to calculation for the investigation to be carried through to numerical results.

The present paper investigates a form of probabilistic interdependence [*Wahrscheinlichkeitsverkettung*] * that broadly represents, for example, the structure of epidemic mortality and of industrial and transport accidents correctly — at any rate much more correctly than the assumption of independent probabilities. Several numerical spot checks, one of which is reported below, strongly support this view. The form of probabilistic interdependence investigated — which might briefly be described as “increase in probability through success” [*Chancenvermehrung durch Erfolg*] † — is certainly also not unlike the dependence that distributes the individual members of a plant species among the different locations within a uniform natural vegetation cover. The final sections of our paper are devoted to a general investigation, prompted by this observation, of the statistical distribution of point populations [*Punktgesamtheiten*] in space. Such distributions have a certain interest for physics (radioactive decay, Brownian motion). From the mathematical point of view, one is inevitably led to the problem considered here when one

*Translator’s note. *Verkettung* means linking into a chain. The title’s “interdependent” includes the paper’s dependence on the cumulative outcome of earlier trials. In Section 1 the authors distinguish this from Markoff’s “simple chain”, where the next trial depends on the immediately preceding one.

†Translator’s note. “Success” denotes the outcome counted as 1, rather than a favourable event: it can represent a fatal accident. In the urn, a draw increases the representation of its own colour and therefore its chance of recurring. This is the mechanism now described as reinforcement. The authors retain the explanatory phrase “increase in probability through success”.

wishes to bring the two simplest and oldest problems of probability theory — repeated draws from an urn with and without replacement of the balls — “under a common denominator”.

The theoretical considerations are due to the second-named author, and the practical execution of the applications to the first-named author.

1. The structure of probabilistic interdependence

[280] We consider an ordered sequence of quantities $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \dots$ subject to chance.¹ If the chances on which the outcome of $\mathbf{x}_n$ depends are fixed in advance, unaffected by the outcome of all the others, then $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \dots$ are called *independent* of one another. If the chances of $\mathbf{x}_{n+1}$ are determined by the outcome of $\mathbf{x}_n$ ($n \geq 1$), we have a *simple chain* of quantities, in the terminology of Markoff, to whom we owe a very extensive theory of this simplest case of probabilistic dependence.² The next simplest case would be that in which the chances of $\mathbf{x}_{n+1}$ are determined by the total result of the preceding chance outcomes, that is, by the value taken by the sum $\mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_n$. Unfortunately, a general theory of this kind of dependence is not yet available; a special case will be treated here and first illustrated by the urn scheme.

At the beginning of the game an urn contains R red and S black balls, $R + S = N$ balls in all. A ball is drawn from the urn, and *in place of the ball drawn, $1 + \Delta$ balls of the same colour are put into the urn.* Another ball is then drawn and the same operation repeated. After the n th draw the urn therefore contains $N + n\Delta$ balls. If r red and s black balls ($r + s = n$) have been drawn in the first n draws, the urn contains $R + r\Delta$ red and $S + s\Delta$ black balls, and the probabilities of drawing a red or a black ball respectively on the $(n + 1)$ th draw are

$$\frac{R + r\Delta}{N + n\Delta} = \frac{\varrho + r\delta}{1 + n\delta}, \quad \frac{S + s\Delta}{N + n\Delta} = \frac{\sigma + s\delta}{1 + n\delta}. \quad (1)$$

Here, for brevity, we have put

$$\frac{R}{N} = \varrho, \quad \frac{S}{N} = \sigma, \quad \frac{\Delta}{N} = \delta. \quad (2)$$

For $\Delta = 0$ we have the familiar problem of balls drawn with replacement, and for $\Delta = -1$ that of balls drawn without replacement.

We can express this more abstractly as follows: the random variables $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \dots$ can take only two values, 1 and 0. Let the probability that $\mathbf{x}_1 = 1$ (a red draw) be ϱ , and the probability that $\mathbf{x}_1 = 0$ (a black draw) be σ , with $\varrho + \sigma = 1$. Let

$$\begin{aligned} & \frac{\varrho + (\mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_n)\delta}{1 + n\delta}, \\ & \frac{\sigma + (n - \mathbf{x}_1 - \mathbf{x}_2 - \dots - \mathbf{x}_n)\delta}{1 + n\delta}. \end{aligned} \quad (3)$$

be the probabilities that $\mathbf{x}_{n+1} = 1$ and $\mathbf{x}_{n+1} = 0$, respectively. If we put

$$\mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_n = r, \quad s = n - r, \quad (4)$$

¹In Italian, instead of “a quantity dependent on chance”, one can say, somewhat more aptly, “variabile casuale” [random variable]. Compare G. Castelnuovo, *Calcolo delle probabilità* (Milan–Rome–Naples, 1919), §30. In what follows, quantities dependent on chance are distinguished by bold type.

²Compare A. A. Markoff, *Wahrscheinlichkeitsrechnung* (Leipzig–Berlin, 1912), pp. 272–311.

then formulae (1) and (3) are identical. If $\delta > 0$, there is an increase in probability through success and a decrease in probability through failure; when $\delta < 0$ the reverse holds. Thus, when $\delta > 0$, both success and failure are “contagious”. The value $\delta = 0$ gives the classical, simplest case of independent events. When $\delta \geq 0$ the sequence of quantities $\mathbf{x}_1, \mathbf{x}_2, \dots$ can be continued indefinitely; when $\delta < 0$, only as long as $1 + n\delta > 0$.*

2. Calculation of probabilities and expectations

What is the probability that $\mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_n = r$? (The possible values of r are $0, 1, 2, \dots, n$.) In other words, what is the probability, under the stated conditions, of drawing r red and $s = n - r$ black balls from the urn in n draws? By the multiplication rule for probabilities, the probability that

$$\mathbf{x}_1 = 1, \mathbf{x}_2 = 1, \dots, \mathbf{x}_r = 1, \quad \mathbf{x}_{r+1} = 0, \dots, \mathbf{x}_{r+s} = 0$$

is

$$\begin{aligned} & \frac{\varrho}{1} \frac{\varrho + \delta}{1 + \delta} \frac{\varrho + 2\delta}{1 + 2\delta} \dots \frac{\varrho + (r-1)\delta}{1 + (r-1)\delta} \\ & \times \frac{\sigma}{1 + r\delta} \frac{\sigma + \delta}{1 + (r+1)\delta} \dots \frac{\sigma + (s-1)\delta}{1 + (n-1)\delta}. \end{aligned} \quad (5)$$

The probability that r specified members of $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ take the value 1 and the remaining s take the value 0 is again given by (5). For if ^[281] one forms the corresponding product according to (3), one again obtains n fractions whose denominators are the same as those in (5), while the numerators merely occur in a different order. The calculation therefore amounts simply to a permutation of the factors in the numerator (not in the denominator!). From n elements one can select r in $\binom{n}{r}$ ways; hence the required probability is

$$p_{r,s} = \frac{n!}{r!s!} \frac{\varrho(\varrho + \delta) \dots (\varrho + (r-1)\delta)}{1(1 + \delta)(1 + 2\delta) \dots (1 + (n-1)\delta)} \times \frac{\sigma(\sigma + \delta) \dots (\sigma + (s-1)\delta)}{1 + r\delta \dots 1 + (n-1)\delta}. \quad (6)$$

*Translator's note. The displayed denominator condition is the condition stated in the source. A probabilistic use also requires both colour probabilities to remain non-negative and at most 1. In the concrete drawing-without-replacement case, drawing continues while the urn contains balls; an exhausted colour has probability zero.

We now calculate the *mathematical expectation* of $\mathbf{r}(\mathbf{r}-1)(\mathbf{r}-2)\cdots(\mathbf{r}-k+1)$. Denote the quantity to be calculated by $\langle \mathbf{r}(\mathbf{r}-1)\cdots(\mathbf{r}-k+1) \rangle$.¹ According to (6),

$$\begin{aligned} & \langle \mathbf{r}(\mathbf{r}-1)\cdots(\mathbf{r}-k+1) \rangle \\ &= \sum_{r=0}^n r(r-1)\cdots(r-k+1)p_{r,n-r} \\ &= \frac{\sum_{r+s=n} \left[\begin{aligned} & r(r-1)\cdots(r-k+1) \\ & \times \frac{\frac{\varrho}{\delta}(\frac{\varrho}{\delta}+1)\cdots(\frac{\varrho}{\delta}+r-1)}{r!} \\ & \times \frac{\frac{\sigma}{\delta}(\frac{\sigma}{\delta}+1)\cdots(\frac{\sigma}{\delta}+s-1)}{s!} \end{aligned} \right]}{\frac{\frac{1}{\delta}(\frac{1}{\delta}+1)\cdots(\frac{1}{\delta}+n-1)}{n!}}. \end{aligned}$$

Thus, in the sum in the numerator, $r = 0, 1, 2, \dots, n$ and $s = n - r$ are to be taken.* The numerator is the coefficient of z^n in the product of the two series

$$\begin{aligned} & \left[\sum_{r=0}^{\infty} r(r-1)\cdots(r-k+1) \frac{\frac{\varrho}{\delta}(\frac{\varrho}{\delta}+1)\cdots(\frac{\varrho}{\delta}+r-1)}{r!} z^r \right] \\ & \times \left[\sum_{s=0}^{\infty} \frac{\frac{\sigma}{\delta}(\frac{\sigma}{\delta}+1)\cdots(\frac{\sigma}{\delta}+s-1)}{s!} z^s \right] \\ &= \left(z^k \frac{d^k}{dz^k} (1-z)^{-\varrho/\delta} \right) (1-z)^{-\sigma/\delta} \\ &= z^k \frac{\varrho}{\delta} \left(\frac{\varrho}{\delta} + 1 \right) \cdots \left(\frac{\varrho}{\delta} + k - 1 \right) \\ & \quad \times (1-z)^{-\varrho/\delta-k} (1-z)^{-\sigma/\delta} \\ &= \frac{\varrho}{\delta} \left(\frac{\varrho}{\delta} + 1 \right) \cdots \left(\frac{\varrho}{\delta} + k - 1 \right) z^k (1-z)^{-1/\delta-k} \\ &= \frac{\varrho}{\delta} \left(\frac{\varrho}{\delta} + 1 \right) \cdots \left(\frac{\varrho}{\delta} + k - 1 \right) \\ & \quad \times \sum_{n=k}^{\infty} \frac{(\frac{1}{\delta}+k)(\frac{1}{\delta}+k+1)\cdots(\frac{1}{\delta}+k+n-k-1)}{(n-k)!} z^n. \end{aligned}$$

Hence

$$\begin{aligned} & \langle \mathbf{r}(\mathbf{r}-1)\cdots(\mathbf{r}-k+1) \rangle \\ &= \frac{\frac{\varrho}{\delta}(\frac{\varrho}{\delta}+1)\cdots(\frac{\varrho}{\delta}+k-1)}{\frac{1}{\delta}(\frac{1}{\delta}+1)\cdots(\frac{1}{\delta}+n-1) \frac{1}{n!}}, \end{aligned}$$

¹The expectation (mathematical hope, average value) of a random variable $\mathbf{x}$ will always be denoted by $\langle \mathbf{x} \rangle$, with angle brackets $\langle \rangle$, whose use is reserved for this notation.

*Translator's note. The journal prints " $s = r - n$ ". The reading text corrects this to $s = n - r$, as required by $r + s = n$, formula (4), and the summation condition immediately above.

$$\begin{aligned}
& \langle \mathbf{r}(\mathbf{r} - 1) \cdots (\mathbf{r} - k + 1) \rangle \\
&= \frac{\varrho(\varrho + \delta)(\varrho + 2\delta) \cdots (\varrho + (k - 1)\delta)}{1(1 + \delta)(1 + 2\delta) \cdots (1 + (k - 1)\delta)} \\
&\quad \times n(n - 1) \cdots (n - k + 1). \quad (7)
\end{aligned}$$

Formula (7) can also be obtained by other, more combinatorial considerations. In particular, for $k = 1, 2$, (7) gives the expected value of $\mathbf{r}$ and the variance about this expectation [*Quadrat der mittleren Abweichung*]:[†]

$$\langle \mathbf{r} \rangle = \varrho n. \quad (8)$$

$$\begin{aligned}
\langle (\mathbf{r} - \langle \mathbf{r} \rangle)^2 \rangle &= \langle \mathbf{r}^2 \rangle - \langle \mathbf{r} \rangle^2 \\
&= \langle \mathbf{r}(\mathbf{r} - 1) \rangle + \langle \mathbf{r} \rangle - \langle \mathbf{r} \rangle^2 \\
&= \frac{\varrho(\varrho + \delta)}{1(1 + \delta)} n(n - 1) + n\varrho - n^2\varrho^2 \\
&= n\varrho(1 - \varrho) \frac{1 + n\delta}{1 + \delta}. \quad (9)
\end{aligned}$$

For $\delta = 0$, or for $\delta = -1/N$, $\varrho = R/N$, $\sigma = S/N$ – compare (2) – these yield the familiar classical results, which may serve as a check.

3. The limiting case of rare events

We turn to circumstances in which the number of observed cases n is very large, but ϱ is so small – that is, the characteristic with probability ϱ is so rare – that the expected number $\langle \mathbf{r} \rangle = \varrho n$ is only small. For example, let n be the number of people who begin a railway journey within a month, and ϱ the probability that the journey undertaken ends in a fatal accident. Put

$$n\varrho = h, \quad n\delta = d > 0, \quad (10)$$

[282] and suppose h and d to be moderate and n very large; or, in an appropriate mathematical abstraction, hold h and d fixed and let n tend to infinity. We first calculate the limiting value of the probability $p_{0,n}$. According to (6),

$$\begin{aligned}
p_{0,n} &= \frac{\sigma(\sigma + \delta) \cdots (\sigma + (n - 1)\delta)}{1(1 + \delta) \cdots (1 + (n - 1)\delta)} \\
&= \left(1 - \frac{\varrho}{1}\right) \left(1 - \frac{\varrho}{1 + \delta}\right) \left(1 - \frac{\varrho}{1 + 2\delta}\right) \cdots \left(1 - \frac{\varrho}{1 + (n - 1)\delta}\right).
\end{aligned}$$

Substituting, according to (10),

$$\varrho = \frac{h}{n}, \quad \sigma = 1 - \frac{h}{n}, \quad \delta = \frac{d}{n}, \quad (11)$$

[†]Translator's note. Literally, “the square of the mean deviation”. Formula (9) defines the expected squared deviation from the expectation, now called the variance. Its square root is the standard deviation. The phrase here therefore concerns the root-mean-square deviation, rather than the mean absolute deviation.

we obtain

$$\begin{aligned}
 p_{0,n} &= \left(1 - \frac{h}{n}\right) \left(1 - \frac{h}{n+d}\right) \cdots \left(1 - \frac{h}{n+(n-1)d}\right), \\
 \lg p_{0,n} &= - \left(\frac{h}{n} + \frac{h}{n+d} + \cdots + \frac{h}{n+(n-1)d} \right) \\
 &\quad - \frac{1}{2} \left[\left(\frac{h}{n}\right)^2 + \cdots + \left(\frac{h}{n+(n-1)d}\right)^2 \right] - \cdots, \\
 \lg p_{0,n} &= -\frac{h}{d} \left(\frac{d}{n} + \frac{d}{n} \frac{1}{1+d/n} + \cdots + \frac{d}{n} \frac{1}{1+(n-1)d/n} \right) - R_n. \quad (12)
 \end{aligned}$$

The remainder R_n becomes infinitesimal as $n \rightarrow \infty$; it is of order $1/n$, as is easily verified. The expression in brackets on the right of (12) is to be regarded as a sum of n rectangles with base d/n . If they are placed side by side on the abscissa axis so that their bases fill the interval between abscissae 0 and d , their upper left corners lie on the curve $y = 1/(1+x)$. Hence*

$$\begin{aligned}
 \lim_{n \rightarrow \infty} \lg p_{0,n} &= -\frac{h}{d} \lim_{n \rightarrow \infty} \sum_{\nu=1}^n \frac{d}{n} \frac{1}{1+(\nu-1)d/n} \\
 &= -\frac{h}{d} \int_0^d \frac{dx}{1+x} \\
 &= -\frac{h}{d} \lg(1+d).
 \end{aligned}$$

$$\lim_{n \rightarrow \infty} p_{0,n} = (1+d)^{-h/d} = P_0. \quad (13)$$

From (6) and (11) it follows that

$$\begin{aligned}
 p_{r,n-r} &= p_{0,n} \frac{n(n-1) \cdots (n-r+1)}{1 \cdot 2 \cdots r} \\
 &\quad \times \frac{\varrho}{\sigma + (n-1)\delta} \frac{\varrho + \delta}{\sigma + (n-2)\delta} \cdots \frac{\varrho + (r-1)\delta}{\sigma + (n-r)\delta} \\
 &\quad \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{r-1}{n}\right) \\
 &\quad \times h(h+d) \cdots (h+(r-1)d) \\
 &= p_{0,n} \frac{r! (1+d - \frac{h+d}{n})(1+d - \frac{h+2d}{n}) \cdots (1+d - \frac{h+rd}{n})}{r!}.
 \end{aligned}$$

For fixed r , this formula, together with (13), gives

$$\begin{aligned}
 \lim_{n \rightarrow \infty} p_{r,n-r} &= (1+d)^{-h/d} \frac{1}{r!} \frac{h}{1+d} \frac{h+d}{1+d} \cdots \frac{h+(r-1)d}{1+d} \\
 &= \frac{1}{r!} h(h+d)(h+2d) \cdots (h+(r-1)d) (1+d)^{-h/d-r} = P_r. \quad (14)
 \end{aligned}$$

*Translator's note. The source's $\lg$ denotes the natural logarithm in this derivation, as the logarithmic series and the integral $\int dx/(1+x)$ show. Its notation is retained.

As a check on formula (14), we calculate the sum of the limiting probabilities $P_0, P_1, P_2, \dots, P_r, \dots$. We have

$$\begin{aligned} \sum_{r=0}^{\infty} P_r &= (1+d)^{-h/d} \sum_{r=0}^{\infty} \frac{\frac{h}{d}(\frac{h}{d}+1) \cdots (\frac{h}{d}+r-1)}{1 \cdot 2 \cdots r} \left(\frac{d}{1+d}\right)^r \\ &= (1+d)^{-h/d} \left(1 - \frac{d}{1+d}\right)^{-h/d} = 1. \end{aligned}$$

For a further check on (14), we calculate the expectation

$$\begin{aligned} &\langle \mathbf{r}(\mathbf{r}-1) \cdots (\mathbf{r}-k+1) \rangle \\ &= \sum_{r=0}^{\infty} r(r-1) \cdots (r-k+1) P_r \\ &= (1+d)^{-h/d} \\ &\quad \times \left[z^k \frac{d^k}{dz^k} \sum_{r=0}^{\infty} \frac{\frac{h}{d}(\frac{h}{d}+1) \cdots (\frac{h}{d}+r-1)}{1 \cdot 2 \cdots r} z^r \right]_{z=d/(1+d)} \\ &= (1+d)^{-h/d} \left(\frac{d}{1+d}\right)^k \\ &\quad \times \frac{h}{d} \left(\frac{h}{d}+1\right) \cdots \left(\frac{h}{d}+k-1\right) \left(1 - \frac{d}{1+d}\right)^{-h/d-k}. \end{aligned}$$

$$\langle \mathbf{r}(\mathbf{r}-1) \cdots (\mathbf{r}-k+1) \rangle = h(h+d)(h+2d) \cdots (h+(k-1)d). \quad (15)$$

We could also have obtained the same result directly from (7) by passing to the limit. In particular, either from (15) or by passing to the limit in (8) and (9), we obtain the following values for the expectation $\langle \mathbf{r} \rangle$ and the variance ^[283] about it:

$$\langle \mathbf{r} \rangle = h. \quad (16)$$

$$\langle (\mathbf{r}-h)^2 \rangle = h(1+d). \quad (17)$$

So far we have considered the case $d > 0$, that is, the case of an increase in probability through success. For $d = 0$ we have

$$\lim_{d \rightarrow 0} P_0 = \lim_{d \rightarrow 0} (1+d)^{-h/d} = e^{-h} = P_0^*,$$

and, generally, compare (14),

$$\lim_{d \rightarrow 0} P_r = \frac{h^r e^{-h}}{r!} = P_r^*. \quad (18)$$

Thus, for $d = 0$, the formulae derived include the familiar case of *independent* rare events, discovered by Poisson and recognised in its significance for statistics by L. von Bortkiewicz.¹

¹L. von Bortkiewicz, *Das Gesetz der kleinen Zahlen* (Leipzig, 1898). In our notation we follow R. von Mises, this journal, vol. 1, pp. 121–124.

4. Application

The probabilities $P_0, P_1, P_2, \dots, P_r, \dots$ given by formula (14) arise, in brief, from the following probabilistic structure: the body of observations is very large, the characteristic under investigation is very rare, and the individual observed cases are linked in such a way that the more strongly the characteristic is represented in one part of the observations, the more strongly it is also to be expected in the remainder. The observed cases might, for instance, be the deaths in a particular country within a month, and the particular characteristic under investigation death from a rare epidemic disease such as smallpox or cholera. The registration of an especially large number of smallpox cases in the first half of the month *prejudges* what is to be expected in the second half of the month *in the same direction*. Railway accidents or industrial accidents, for example those resulting from an explosion, spread, statistically speaking, in much the same way as a contagious disease, since they generally claim several lives. We may expect the statistical figures for phenomena of this kind to agree qualitatively with the course of the probabilities $P_0, P_1, P_2, \dots, P_r, \dots$

Now for a quantitative test! Take, for example, the deaths from smallpox in Switzerland during the years 1877–1900, that is, during 288 months; compare Table 1, original p. 285. There were 1,584 such deaths in all, giving a monthly average of 5.5. If we put

$$h = 5.5, \quad (19)$$

we can calculate the probabilities $P_0^*, P_1^*, \dots, P_r^*, \dots$ from formula (18) ($P_0^*, P_1^*, \dots, P_r^*, \dots$ correspond to *independent* rare events). The quantity $288P_r^*$ is the expected (theoretical) number, among the 288 observed months, of months in which exactly r deaths should occur; compare column III of Table 1. The actual number of these months is in column II of Table 1. Column VI of Table 1 contains the numbers $288(P_0^* + P_1^* + \dots + P_r^*) = 288 \sum_{\nu=0}^r P_\nu^*$, set beside the corresponding observed numbers in column V. Thus, at the intersection of column V and row r (where $r = 0, 1, 2, \dots$) stands the number of months, among the 288 observed, in which *no more than* r deaths of the kind in question occurred. The disagreement between the theory that assumes the independence of the individual cases and leads to the probabilities $P_0^*, P_1^*, \dots, P_r^*, \dots$ and the available facts is apparent either from a comparison of columns II and III, or from one of columns V and VI, or, finally, from the following consideration. If the events are independent, the variance ought to be h (compare formula (17) for $d = 0$; in our case $h = 5.5$). Multiplying $(r - 5.5)^2$ by the number of months in which exactly r deaths occurred (compare column II, row r), summing from $r = 0$ to $r = 60$, and dividing the resulting sum by 287, that is, the total number of months minus 1, gives 83.5888, a value considerably greater than 5.5.*

[284] To test the theory of an increase in probability through success against this case, we use formula (17) to determine d from the equation

$$h(1 + d) = 83.5888$$

(the calculation of the number 83.5888 has just been explained). We obtain

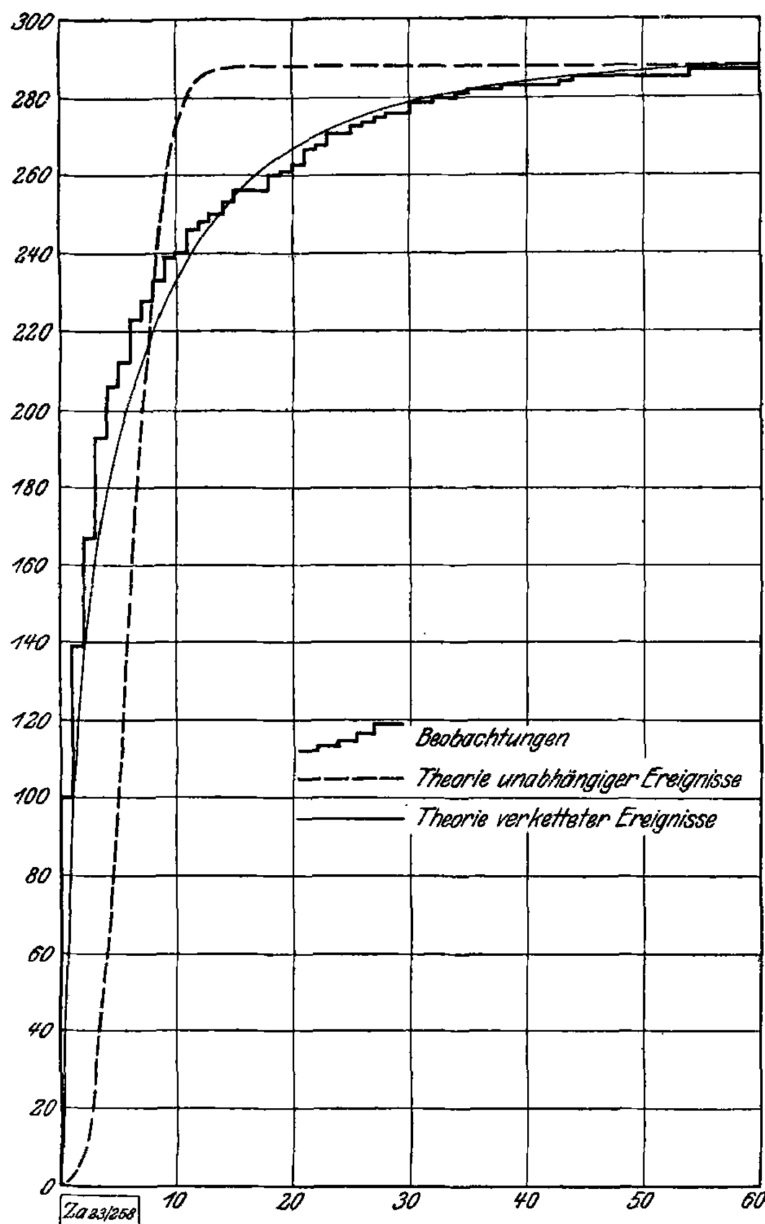
$$d = 14.20. \quad (20)$$

With the numerical values (19) and (20), we calculate the probabilities $P_0, P_1, P_2, \dots$ according to formula (14). Comparing columns II and IV, and still more columns V and VII, of Table 1 shows, at first

*Translator's note. The 63 rows of the printed Table 1 give 288 months and 1,584 deaths, as stated, but evaluating the variance by the procedure described here gives $26690/287 = 92.9965$ rather than 83.5888. This would give $d = 15.91$ after rounding, using (17), instead of 14.20. The authors' stated variance, parameter, theoretical table columns and graph are retained together; the source of the arithmetic discrepancy is unresolved.

sight, an entirely satisfactory quantitative agreement between the theoretical expected numbers in IV and VII and the corresponding observed numbers in II and V.

Of course, the result that better agreement was obtained with the P_r than with the P_r^* could also be attributed to the circumstance that the P_r^* contain only one adjustable parameter, namely h , whereas the P_r contain two, namely h and d . We incline, however, to the view that the better agreement arises not merely from this external circumstance but rather from the much better agreement of the probabilistic picture with reality when an increase in probability through success is assumed than when independence of the cases is assumed. Tests of the theory against other statistical material gave, in part, even more favourable results. The first-named author will publish the details elsewhere.



Comparison of theory and observation in the Swiss smallpox statistics. (The ordinate at abscissa x gives the number of months in which the number of deaths is less than x or equal to x .)

Graph key, translated from the original: stepped line — observations; dashed line — theory of independent events; continuous curve — theory of interdependent events.

5. Uniform and stable distribution of a point population in space

Whether the space in question is one-, two- or three-dimensional makes only an insignificant difference to the following discussion. For definiteness, let us consider the two-dimensional case: the points are distributed in a plane.

Let a plane region F of area f be given. By saying that a point P is *scattered* over the region F , we mean the following: if F' is any subregion of F , with area f' , the probability that the point P lies in the subregion F' is f'/f . It is now clear what is meant by several points being scattered over F independently of one another. If $r + \nu$ points are scattered independently over F , the probability that r of the $r + \nu$ points fall in the subregion F' and ν outside it is

$$\binom{r + \nu}{\nu} \left(\frac{f'}{f} \right)^r \left(1 - \frac{f'}{f} \right)^\nu. \quad (21)$$

[285] **Table 1. Deaths from smallpox in Switzerland in the years 1877–1900.**

Columns II–IV give numbers of months; V–VII give cumulative numbers. VIII is VI minus V; IX is VII minus V. The original column numbers are retained. A dash reproduces a printed dash. Decimal points replace decimal commas.

	I	II	III	IV	V	VI	VII	VIII	IX
0		100	1.2	100.4	100	1.2	100.4	−98.8	0.4
1		39	6.5	36.3	139	7.7	136.7	−131.3	−2.3
2		28	17.8	23.5	167	25.5	160.2	−141.5	−6.8
3		26	32.6	17.5	193	58.1	177.7	−134.9	−15.3
4		13	44.9	13.8	206	103.0	191.5	−103.0	−14.5
5		6	49.4	11.3	212	152.4	202.8	−59.6	−9.2
6		11	45.2	9.5	223	197.6	212.3	−25.4	−10.7
7		5	35.5	8.1	228	233.1	220.4	5.1	−7.6
8		5	24.5	7.0	233	257.6	227.4	24.6	−5.6
9		6	15.0	6.1	239	272.6	233.5	33.6	−5.5
10		1	8.2	5.3	240	280.8	238.8	40.8	−1.2
11		6	4.1	4.7	246	284.9	243.5	38.9	−2.5
12		2	1.9	4.2	248	286.8	247.7	38.8	−0.3
13		2	0.8	3.7	250	287.6	251.4	37.6	1.4
14		3	0.3	3.3	253	287.9	254.7	34.9	1.7
15		3	0.1	3.0	256	288.0	257.7	32.0	1.7
16		—	—	2.7	256	288.0	260.4	32.0	4.4
17		—	—	2.4	256	288.0	262.8	32.0	6.8
18		4	—	2.2	260	288.0	265.0	28.0	5.0
19		1	—	2.0	261	288.0	267.0	27.0	6.0
20		2	—	1.8	263	288.0	268.8	25.0	5.8
21		4	—	1.6	267	288.0	270.4	21.0	3.4
22		1	—	1.5	268	288.0	271.9	20.0	3.9
23		3	—	1.3	271	288.0	273.2	17.0	2.2
24		—	—	1.2	271	288.0	274.4	17.0	3.4
25		2	—	1.1	273	288.0	275.5	15.0	2.5
26		1	—	1.0	274	288.0	276.5	14.0	2.5
27		1	—	0.9	275	288.0	277.4	13.0	2.4
28		1	—	0.8	276	288.0	278.2	12.0	2.2
29		—	—	0.8	276	288.0	279.0	12.0	3.0
30		3	—	0.7	279	288.0	279.7	9.0	0.7
31		—	—	0.6	279	288.0	280.3	9.0	1.3

I	II	III	IV	V	VI	VII	VIII	IX
32	1	—	0.6	280	288.0	280.9	8.0	0.9
33	—	—	0.5	280	288.0	281.4	8.0	1.4
34	1	—	0.5	281	288.0	281.9	7.0	0.9
35	1	—	0.5	282	288.0	282.4	6.0	0.4
36	—	—	0.4	282	288.0	282.8	6.0	0.8
37	—	—	0.4	282	288.0	283.2	6.0	1.2
38	1	—	0.4	283	288.0	283.6	5.0	0.6
39	—	—	0.3	283	288.0	283.9	5.0	0.9
40	—	—	0.3	283	288.0	284.2	5.0	1.2
41	—	—	0.3	283	288.0	284.5	5.0	1.5
42	—	—	0.3	283	288.0	284.8	5.0	1.8
43	1	—	0.2	284	288.0	285.0	4.0	1.0
44	1	—	0.2	285	288.0	285.2	3.0	0.2
45	—	—	0.2	285	288.0	285.4	3.0	0.4
46	—	—	0.2	285	288.0	285.6	3.0	0.6
47	—	—	0.2	285	288.0	285.8	3.0	0.8
48	—	—	0.2	285	288.0	286.0	3.0	1.0
49	—	—	0.1	285	288.0	286.1	3.0	1.1
50	—	—	0.1	285	288.0	286.2	3.0	1.2
51	—	—	0.1	285	288.0	286.3	3.0	1.3
52	—	—	0.1	285	288.0	286.4	3.0	1.4
53	—	—	0.1	285	288.0	286.5	3.0	1.5
54	2	—	0.1	287	288.0	286.6	1.0	−0.4

[286] Table 1, continued.

I	II	III	IV	V	VI	VII	VIII	IX
55	—	—	0.1	287	288.0	286.7	1.0	−0.3
56	—	—	0.1	287	288.0	286.8	1.0	−0.2
57	—	—	0.1	287	288.0	286.9	1.0	−0.1
58	—	—	0.1	287	288.0	287.0	1.0	—
59	—	—	0.1	287	288.0	287.1	1.0	0.1
60	1	—	0.1	288	288.0	287.2	—	−0.8
61	—	—	0.1	288	288.0	287.3	—	−0.7
62	—	—	0.1	288	288.0	287.4	—	−0.6

I: number of deaths in one month, r . II: number of months in which r deaths actually occurred, M_r . III: theoretical number of months with r deaths, assuming independence, $288P_r^*$. IV: theoretical number of months with r deaths, assuming an increase in probability through success, $288P_r$. V: $\sum_{\nu=0}^r M_\nu$. VI: $288 \sum_{\nu=0}^r P_\nu^*$. VII: $288 \sum_{\nu=0}^r P_\nu$. VIII: $\sum_{\nu=0}^r (288P_\nu^* - M_\nu)$, the difference between VI and V. IX: $\sum_{\nu=0}^r (288P_\nu - M_\nu)$, the difference between VII and V.*

To characterise the distribution of a point population in a plane in terms of probability theory, one must specify the probability $W(r, F)$ that exactly r points of the population lie within a region F (with r and F arbitrary). Suppose the distribution is uniform [*gleichmäßig*], that is, $W(r, F)$ depends neither

*Translator's note. The last column is IX in the first part of the table and in the authors' explanatory key. The continued header prints VIII a second time; it is corrected to IX here. Printed dashes and rounded theoretical counts are retained, including a dash where a small probability is suppressed or a difference is zero.

on the position nor on the shape of the region, but only on its area f : $W(r, F) = W_r(f)$.[†] Thus, to characterise a uniform distribution of points in the plane, one must specify the probabilities

$$W_0(f), W_1(f), W_2(f), \dots, W_r(f), \dots \quad (22)$$

The functions (22) are subject to the conditions

$$W_r(f) \geq 0 \quad \text{for } r = 0, 1, 2, 3, \dots; \quad f > 0, \quad (\text{I})$$

$$W_0(f) + W_1(f) + W_2(f) + \dots + W_r(f) + \dots = 1. \quad (\text{II})$$

The distribution of a point population shall be called *stable* [*stabil*], in the probabilistic sense, if purely random disturbances cannot alter anything essential in it. More precisely: *the stability of the distribution of a point population consists in the fact that the overall statistical picture remains unchanged when the points of the population lying within any region F are removed from F and scattered back over F independently of one another.*[‡]

Outside F , removing and scattering back the points lying in F obviously changes nothing; we therefore need only consider what it changes in any subregion F' of F . Let the distribution be assumed uniform from the outset and specified by the probabilities (22), satisfying (I) and (II). As above, denote the areas of F and F' by f and f' .

Before the points were removed from F , the probability that exactly r points lay in F' was $W_r(f')$.

For exactly r points to lie in F' after removal and scattering back, $r + \nu$ points must have been present in F originally, with $\nu \geq 0$. The probability that $r + \nu$ points were in F is $W_{r+\nu}(f)$; the probability that exactly r of these $r + \nu$ fall in F' when scattered back is given by (21). Since ν may be 0, 1, 2, 3, ..., the probability that exactly r points lie in f' after scattering back is

$$\sum_{\nu=0}^{\infty} W_{r+\nu}(f) \binom{r+\nu}{\nu} \left(\frac{f'}{f}\right)^r \left(1 - \frac{f'}{f}\right)^{\nu}.$$

[287] Thus, the stability of the distribution specified by the probabilities (22) is expressed by the equation

$$W_r(f') = \sum_{\nu=0}^{\infty} W_{r+\nu}(f) \binom{r+\nu}{\nu} \left(\frac{f'}{f}\right)^r \left(1 - \frac{f'}{f}\right)^{\nu}, \quad (\text{III})$$

$$0 < f' \leq f, \quad r = 0, 1, 2, 3, \dots$$

To find the functions (22) satisfying conditions (I), (II) and (III), put

$$W_0(f) + W_1(f)z + W_2(f)z^2 + \dots = \Phi(f, z). \quad (23)$$

[†]Translator's note. "Uniform" has the authors' explicit meaning: the distribution of the number of points in a region depends only on its area. This condition permits clustering and dependence. Independence of different regions is imposed separately in Section 6.

[‡]Translator's note. "Stable" refers here to invariance of the statistical picture under the stated removal and independent redistribution operation. The term's meaning throughout Sections 5–7 follows this definition; it should not be read as the later technical definition of a stable probability law under addition.

The series (23) has positive coefficients, compare (I), and converges for $z = 1$, compare (II), and therefore throughout $f > 0$, $|z| \leq 1$; z is regarded as a complex variable. Multiply (III) by z^r and sum over $r = 0, 1, 2, \dots$; this gives

$$\begin{aligned} \sum_{r=0}^{\infty} W_r(f') z^r \\ = \sum_{r=0}^{\infty} \sum_{\nu=0}^{\infty} z^r W_{r+\nu}(f) \binom{r+\nu}{\nu} \left(\frac{f'}{f}\right)^r \left(1 - \frac{f'}{f}\right)^{\nu}. \end{aligned} \quad (24)$$

For brevity, put

$$\frac{f'}{f} = x, \quad 0 < x \leq 1. \quad (25)$$

From (24) we obtain

$$\begin{aligned} \Phi(fx, z) \\ = \sum_{r=0}^{\infty} \sum_{\nu=0}^{\infty} \frac{(r+\nu)(r+\nu-1)\cdots(r+1)}{\nu!} W_{r+\nu}(f) (zx)^r (1-x)^{\nu} \\ = \sum_{\nu=0}^{\infty} \frac{(1-x)^{\nu}}{\nu!} \sum_{r=0}^{\infty} W_{r+\nu}(f) (r+\nu)(r+\nu-1)\cdots(r+1) (zx)^r \\ = \sum_{\nu=0}^{\infty} \frac{(1-x)^{\nu}}{\nu!} \Phi^{(\nu)}(f, zx), \end{aligned}$$

in accordance with (23), where $\Phi^{(\nu)}(x, y) = \partial^{\nu} \Phi / \partial y^{\nu}$. Thus

$$\Phi(fx, z) = \Phi(f, zx + 1 - x). \quad (26)$$

Put $\Phi(1, u + 1) = \Psi(u)$.

For the variables f, x, z , substitute either $1, f, z$, or $f, 1/f, 1 + (z - 1)f$; the first choice is appropriate when $f \leq 1$, and the second when $f \geq 1$, in view of the inequality in (25). In both cases we obtain

$$\Phi(f, z) = \Psi(fz - f). \quad (27)$$

Substituting the power-series expansion (23) for $\Phi(f, z)$ in (27) gives, by Maclaurin's formula,

$$\sum_{r=0}^{\infty} W_r(f) z^r = \sum_{r=0}^{\infty} \frac{\Psi^{(r)}(-f) f^r}{r!} z^r, \quad (28)$$

and, by comparison of coefficients,

$$W_r(f) = \frac{\Psi^{(r)}(-f) f^r}{r!} \quad (r = 0, 1, 2, \dots). \quad (29)$$

From the validity of the power-series expansion (28) in the domain $f > 0$, $|z| \leq 1$, and from conditions (I) and (II), we finally reach the following result: *if the distribution of the point population defined by the probabilities (22) is to be stable, those probabilities are given by equation (29), where $\Psi(u)$ is a function*

analytic in the half-plane $\operatorname{Re}(u) < 0$, all of whose derivatives are real and positive for real negative values of u ; furthermore, $\lim_{u \rightarrow 0} \Psi(u) = 1$ as u approaches the point 0 from the left half-plane.

As u tends to 0 through negative values, $\Psi'(u)$ always increases, since $\Psi''(u) > 0$ for $u < 0$.

[288] Let

$$\lim_{u \rightarrow 0} \Psi'(u) = D \quad (30)$$

be finite. Then $D > 0$, and by (29)

$$\begin{aligned} \sum_{r=0}^{\infty} r W_r(f) &= \sum_{r=1}^{\infty} \frac{r f^r \Psi^{(r)}(-f)}{r!} \\ &= f \sum_{m=0}^{\infty} \frac{f^m \Psi^{(1+m)}(-f)}{m!} \\ &= f \Psi'(0) = f D. \end{aligned}$$

The quantity D is the average number of points in the population per unit area — in short, the *density* [*Dichte*] of the point population.

6. Independence of subpopulations

To say that the individual parts of the point population are *independent* of one another means that the chances of the distribution in each region are unaffected by the outcome of the distribution in the other regions. Let the distribution be uniform and specified by the probabilities (22). Consider two regions F_1 and F_2 , of areas f_1 and f_2 respectively, which together make up a region F of area $f_1 + f_2 = f$. If the individual parts of the point population are independent of one another, the probability that exactly n points lie in f_2 remains $W_n(f_2)$, regardless of whether 0, 1, 2, 3, ... points lie in the neighbouring region f_1 . Consequently, the probability that m points lie in f_1 and n points in f_2 is $W_m(f_1)W_n(f_2)$, and the probability that a total of r points lie in f is

$$\begin{aligned} W_0(f_1)W_r(f_2) + W_1(f_1)W_{r-1}(f_2) + \cdots + W_r(f_1)W_0(f_2) \\ = W_r(f_1 + f_2). \end{aligned} \quad (31)$$

For the function $\Phi(f, z)$ defined in (23), this gives the relation

$$\Phi(f_1 + f_2, z) = \Phi(f_1, z)\Phi(f_2, z). \quad (32)$$

If arbitrary subpopulations are independent of one another, f_1 and f_2 can be chosen arbitrarily. As is well known, the validity of (32) for every pair f_1, f_2 with $f_1 > 0, f_2 > 0$ entails the equation

$$\Phi(f, z) = e^{\varphi(z)f}, \quad (33)$$

where $\varphi(z)$ is a certain function of z . If the distribution is also to be stable, so that $\Phi(f, z)$ is of the form (27), and (30) is also to hold, then

$$\Psi(u) = e^{Du}, \quad (34)$$

$$\Phi(f, z) = e^{D(z-1)f}.$$

From (34) and (29) we obtain

$$W_r(f) = \frac{(Df)^r e^{-Df}}{r!}. \quad (35)$$

The function (34) satisfies the conditions stated at the end of Section 5. Formula (35)* follows from the expression for P_r^* in (18) if we put

$$h = Df. \quad (36)$$

It is known that the law analogous to (35) in one or three dimensions governs, respectively, the distribution of radioactive decay events in time and that of particles undergoing Brownian motion in space.

7. Distribution with an increase in probability through success

Let us give an example in which the distribution law certainly cannot be (35). To deepen the concepts of plant geography, botanists¹ have undertaken statistical surveys of the following kind. Square plots of equal area f are marked out within a uniform natural vegetation cover. Within each square all the plant species present are identified. The result of the survey states how many squares were examined in all, which species were found, and in how many squares each species occurred. Dividing the last number by the total number of squares gives an approximate value of the probability of encountering at least one individual of the species within an area f . Denote this probability by H . The experiment is carried out with different square areas f , and H must evidently ^[289] increase as f increases. The following Table 2 refers to a particular species and vegetation;¹ its second row contains the empirical values of H for seven different areas f , given in square metres in the first row. The numerators and denominators of the fractions for H are the observed numbers themselves.

Table 2.

I. f in m ²	0.0001	0.0004	0.0025	0.01	0.04	0.25	1
II. H	6/1000	19/1250	55/1040	203/1300	121/300	194/210	232/240
III. D	60.18	38.29	21.73	16.98	12.91	10.30	3.40

According to (35), the probability that a region of area f contains no point of the population is e^{-Df} ; consequently, the probability of encountering at least one individual of the population within it is

$$1 - e^{-Df} = H. \quad (37)$$

Calculating D according to (37) from the seven observations assembled in Table 2 gives the seven values in row III. These not only differ very greatly from one another (they ought to be equal), but also show

*Translator's note. Formula (35) prints e^{-D} in the journal. The reading text restores e^{-Df} : it follows by differentiating $\Psi(u) = e^{Du}$ in (29), agrees with $h = Df$ in (18), and gives $\sum_r W_r(f) = 1$. The text immediately before (37) also supplies e^{-Df} .

¹G. E. Du Rietz, *Zur methodologischen Grundlage der modernen Pflanzensoziologie*, dissertation, Uppsala, 1921.

¹*Cladonia coccifera* in Table 23, G. E. Du Rietz, work cited in note 1, p. 166.

an unmistakable trend, a systematic deviation. This shows that in the present series of observations H actually increases more slowly with f than it should according to formula (37).

After what has gone before, it is easy to explain why the distribution law (35), which has proved so successful in many physical investigations, fails in this case. The assumption of independence of the individual parts, emphasised in Section 6, does *not* hold here. It is evidently false here that the probability of finding a given number of individuals in a region F is exactly the same whether many individuals of the same species, or none, occur in the neighbouring region F' . Rather, near patches richly covered with the species, neighbouring patches in which individuals of the species are also present are more readily found than near patches entirely devoid of it: the neighbouring regions are, as it were, “infected”.

The function

$$\Psi(u) = (1 - \alpha u)^{-D/\alpha} \quad (38)$$

satisfies the conditions stated in Section 5 ($\alpha > 0$); thus, the distribution specified by the probabilities

$$\begin{aligned} W_r(f) &= \frac{f^r \Psi^{(r)}(-f)}{r!} \\ &= \frac{f^r D(D + \alpha)(D + 2\alpha) \cdots (D + (r - 1)\alpha)}{r!} \\ &\quad \times (1 + \alpha f)^{-D/\alpha - r}. \end{aligned} \quad (39)$$

is *stable* in the sense of Section 5.* If, as before in (36), we put

$$fD = h \quad \text{and} \quad f\alpha = d,$$

formula (39) becomes the expression for P_r in (14). We may suppose that the distribution law (39), since it corresponds more closely to the structure, will fit the botanical statistics cited better than the now classical distribution law (35), which is, incidentally, the limiting case of (39) for $\alpha = 0$. In place of (37) we now have the formula

$$H = 1 - (1 + \alpha f)^{-D/\alpha},$$

in which H increases *more slowly* with f than in (37).

A quantitative test, requiring very great care and effort, is in progress. The example may nevertheless have served to illustrate the main idea of this investigation even before that test has been carried out.

*Translator's note. Beside (38) and (39), the journal twice refers to Section 6. The conditions on Ψ and the definition of stability are in Section 5; both references are corrected here.

About this edition

This English reading proof translates F. Eggenberger and G. Pólya, “Über die Statistik verketteter Vorgänge”, *Zeitschrift für angewandte Mathematik und Mechanik* 3, no. 4 (1923), pp. 279–289. Original publication record.

The supplied journal page images control the text. The translation preserves the seven sections, numbered formulae (1)–(39), conditions (I)–(III), both numerical tables, the graph and six authorial footnotes. Small page labels refer to the original journal pagination. Table 1 retains the source’s printed numbers and dashes; the explanatory column key is translated in full. The graph is an extract of the original, accompanied by its English caption and key. The preceding article’s conclusion on p. 279 falls outside this edition.

Clear printed errors are corrected with explanatory translator’s footnotes: the relation between s , n and r on p. 281; the exponential in formula (35); and two section references beside formulae (38)–(39). Table 1’s continued header is normalised from its duplicated VIII to IX. Original notes retain their printed numbers; translator’s notes use symbols restarting by section. The historical terminology of uniformity and stability is explained at its first substantive use. Long calculations are reflowed over several lines, with explicit multiplication signs and grouping where a product continues on the next line; every derivation step and factor is retained. A translator’s note reports an arithmetic discrepancy between Table 1 and the variance stated in the text; the authors’ stated variance and fitted parameter are retained.

English translation and editorial checks: Codex, an AI assistant, under Bird’s editorial direction. Version 0.2, 16 October 2026. The edition is a reading proof. A bounded search identified English excerpts in David Link’s 2007 “Chains to the West”; the translation here was made from the German witness. Bibliographical references within the article are transmitted as the authors give them. No new licence is assigned to the source article or its scanned figure.