A GENERALIZATION OF THE LAW OF LARGE NUMBERS

By Hilda Geiringer

It is well known that the law of large numbers can be established for dependent as well as for independent chance variables by using Tchebycheff's inequality [1] and assuming that the variance of the sum of the variables tends towards infinity less rapidly than n^2.

In recent years v. Mises has introduced the notion of *statistical functions* [2] and has shown that, under certain assumptions the law of large numbers is still valid if, instead of the arithmetic mean of the n observations $x_1, \cdots, x_n$ a statistical function of these observations is considered. For example in the very special case, where the n collectives which have been observed are *identical* k-valued *arithmetic* distributions with probabilities $p_1, \cdots, p_k$ corresponding to the attributes $c_1, \cdots, c_k$ and with observed relative frequencies $n_1/n, \cdots, n_k/n$ one obtains the result: It is to be expected for every $\epsilon > 0$ with a probability P_n converging towards one as $n \rightarrow \infty$, that $|f(n_1/n, \cdots, n_k/n) - f(p_1, \cdots, p_k)| < \epsilon$ under very general conditions concerning the function f.

In the present paper we shall generalize these new results so that they will apply also to collectives which are not independent.

1. Lemma concerning alternatives. Let us consider the *n-dimensional collective* consisting of a *sequence of n trials* and let us assume that the n trials are alternatives, i.e. for each trial there are only two possible results which we denote by "success," "failure," by "occurrence," "non-occurrence" or by "1," "0." The total result of the n trials is expressed by n numbers each equal to 0 or 1. Let $v(x_1, x_2, \cdots, x_n)$ be the probability of obtaining the result x_1 at the first trial, x_2 at the second one, $\cdots, x_n$ at the last one ($x_\nu = 0, 1$; $\nu = 1, \cdots, n$). In the same way we introduce $v_{12}(x, y) = \sum_{x_3, \cdots, x_n} v(x, y, x_3, \cdots, x_n)$ and generally $v_{\mu\nu}(x, y)$ as the probability that the μth result equals x, the νth equals y, ($\mu \neq \nu$), and finally let $v_\mu(x) = \sum_y v_{\mu\nu}(x, y)$ be the probability that the μth result equals x. In particular let us write

$$v_\mu(1) = p_\mu, \quad v_{\mu\nu}(1, 1) = p_{\mu\nu}, \quad (\mu, \nu = 1, \cdots, n; \mu \neq \nu)$$

p_μ being the probability of success in the μth trial and $p_{\mu\nu}$ the probability of simultaneous success both in the μth and νth trials.

The *variance* s_n^2 of the sum $(x_1 + \cdots + x_n)$ is easily found:

$$s_n^2 = \text{Var}\,(x_1 + \cdots + x_n) = \sum_{x_1,\cdots,x_n} (x_1 + \cdots + x_n - p_1 - \cdots - p_n)^2 v(x_1, \cdots, x_n)$$
$$= \sum_{x_1,\cdots,x_n} (x_1 - p_1)^2 v(x_1, \cdots, x_n) + \cdots$$
$$+ 2 \sum_{x_1,\cdots,x_n} (x_1 - p_1)(x_2 - p_2) v(x_1, \cdots, x_n) + \cdots$$
$$= \sum_{x_1} (x_1 - p_1)^2 v_1(x_1) + \cdots + 2 \sum_{x_1,x_2} (x_1 - p_1)(x_2 - p_2) v_{12}(x_1, x_2) + \cdots$$
$$= p_1(1 - p_1) + \cdots + p_n(1 - p_n) + 2(p_{12} - p_1 p_2) + \cdots + 2(p_{n-1,n} - p_{n-1} p_n).$$

Thus:

$$(1) \qquad s_n^2 = \text{Var}\,(x_1 + \cdots + x_n) = \sum_{\nu=1}^{n} p_\nu(1 - p_\nu) + 2 \sum_{\mu,\nu=1}^{n} (p_{\mu\nu} - p_\mu p_\nu).$$

The first sum on the right is $\leqq n/4$; the second one consists of $N = \frac{1}{2}n(n - 1)$ terms, therefore we cannot be sure that it tends toward zero after division by n^2.

Putting $p_{\mu\nu} - p_\mu p_\nu = \alpha_{\mu\nu}^{(n)}$ we see immediately:

(a) *A necessary and sufficient condition for* $\lim_{n\to\infty} s_n/n = 0$ *is*

$$(2) \qquad \lim_{n\to\infty} 1/n^2 \sum_{\mu,\nu=1}^{n} \alpha_{\mu\nu}^{(n)} = 0.$$

Denoting by σ_μ^2 the variance of $v_\mu(x)$ and by $r_{\mu\nu}$ the correlation coefficient of $v_{\mu\nu}(x, y)$ we have

$$\alpha_{\mu\nu}^{(n)} = p_{\mu\nu} - p_\mu p_\nu = r_{\mu\nu}\sigma_\mu\sigma_\nu .$$

We see that $\alpha_{\mu\nu}^{(n)}$ takes values between $-1/4$ and $+1/4$ and our conditions (2) postulates that the sum of these positive and negative terms tends towards infinity less rapidly than n^2. As to the meaning of the signs of these terms we see that a term $\alpha_{\mu\nu}^{(n)}$ will be $\gtreqless 0$, according as $p_{\mu\nu}/p_\nu \gtreqless p_\mu$. This means: the fact that the νth event has presented itself makes the occurrence of the μth event either more probable; or it is without influence on it; or it makes it less probable. And we see that s_n/n tends toward zero, only if there is a certain "equalization" or "stabilization" of positive and negative mutual influence. If in particular for a pair of values μ, ν, $r_{\mu\nu} = +1$, that is $v_{\mu\nu}(0, 1) = v_{\mu\nu}(1, 0) = 0$, the events must either both occur or both fail and $p_\mu = p_\nu$. If $r_{\mu\nu} = -1$ we have $v_{\mu\nu}(0, 0) = v_{\mu\nu}(1, 1) = 0$ the simultaneous occurrence is impossible and likewise the simultaneous failure, and $p_\mu + p_\nu = 1$. If we have $p_{\mu\nu} = 0$ (case of mutually exclusive events) then $p_\mu + p_\nu \leqq 1$.

Since $s_n^2 \geqq 0$ and $\sum_{\nu=1}^{n} p_\nu(1 - p_\nu) = \sum_{\nu=1}^{n} \sigma_\nu^2 \leqq n/4$ we conclude from (1) that $\sum_{\mu,\nu=1}^{n} \alpha_{\mu\nu}^{(n)} \geqq -n/8$ and we obtain the following simple *sufficient* condition for the validity of (2):

(*b*) *Let us denote by* m_n *the number of all combinations* $\mu, \nu (\mu \leqq n; \nu \leqq n; \mu \neq \nu)$, *such that, however large n may be,* $\alpha_{\mu\nu}^{(n)} > \epsilon$, *where* ϵ *is a given positive number; then* $\frac{1}{n^2} \sum_{\mu,\nu=1}^{n} \alpha_{\mu\nu}^{(n)}$ *converges toward zero if* $\lim_{n \to \infty} m_n/n^2 = 0$.

We have in fact

$$-\frac{n}{8} \leqq \sum_{\mu,\nu=1}^{n} \alpha_{\mu\nu}^{(n)} \leqq m_n + (N - m_n)\epsilon$$

and dividing by n^2 we find that $\frac{1}{n^2} \sum_{\mu,\nu=1}^{n} \alpha_{\mu\nu}^{(n)}$ is enclosed between $\frac{-1}{8n}$ and $m_n/n^2 + \epsilon \frac{N - m_n}{n^2}$ which both tend toward zero. Roughly speaking this condition implies that for "almost all" combinations of indices μ, ν, the $\alpha_{\mu\nu}^{(n)}$ converge toward "negative or vanishing correlation."

On the other hand the sum of all positive and negative terms in $\sum_{\mu,\nu=1}^{n} \alpha_{\mu\nu}^{(n)}$ cannot become less than $-n/8$. Therefore, if "almost all" positive terms are supposed to tend towards zero it follows that also almost all negative terms tend toward zero. Thus we obtain the *sufficient* condition (*c*) which is neither more nor less general than (*b*):

(*c*) *The sum* $\frac{1}{n^2} \sum_{\mu,\nu=1}^{n} \alpha_{\mu\nu}^{(n)}$ *tends towards zero as* $n \to \infty$, *if "almost all" the individual terms* $\alpha_{\mu\nu}^{(n)} = p_{\mu\nu} - p_\mu p_\nu$ *tend toward zero.* Or more exactly, the sum in question tends toward zero if $| \alpha_{\mu\nu}^{(n)} | \leqq \epsilon$ for every ϵ and sufficiently large n with the exception of μ_n terms where $\lim_{n \to \infty} \mu_n/n^2 = 0$. That is "convergence towards independence" for almost all combinations μ, ν of indices. Let us, for example, assume that all the p_ν are $\neq 0$ and all the $p_{\mu\nu} = 0$, then all the $\alpha_{\mu\nu}^{(n)}$ are certainly < 0 and (*b*) is fulfilled; but it is easily seen (3) that in this case $p_1 + p_2 + \cdots p_n \leqq 1$. Therefore all the products $p_\mu p_\nu$ (with the possible exception of a finite number) tend toward zero, and (*c*) holds as well.

2. Statistical functions. Suppose n observations have given the results $x_1, x_2, \cdots, x_n$. Let us assume for the sake of simplicity that they are all bounded between two real numbers A and B. To each real x corresponds the number $n\, S_n(x)$ of observations with a result $\leqq x$. $S_n(x)$ is a monotone non-decreasing step function with n steps, each of height $1/n$; however several steps may coincide at the same point. We have

$$(1) \qquad S_n(x) = 0 \quad \text{if} \quad x < A \quad \text{and} \quad S_n(x) = 1 \quad \text{if} \quad x \geqq B.$$

$S_n(x)$ is called by v. Mises the *partition* (Aufteilung) of the n observations. $S_n(x)$ coincides with the well known cumulative frequency distribution if the attributes $c_\kappa (\kappa = 1, \cdots k)$ and the corresponding relative frequencies $n_1/n, \cdots n_k/n$ are given.

A *statistical function* is a function of the $x_1, x_2, \cdots, x_n$ *which depends only on* $S_n(x)$, the partition of the n results. It will be denoted by $f\{S_n(x)\}$. If the c_κ and the n_κ/n are given then statistical function means simply "function of the relative frequencies" and it becomes a function of k variables. In $f\{S_n(x)\}$ the partition $S_n(x)$ takes the place of the independent variable. Such a statistical function has the following properties: (*a*) It is a symmetric function of the $x_1, x_2, \cdots, x_n$. That is, it is independent of the succession of the n results. (*b*) It is "homogeneous" in the following sense: If instead of n observations we have nl observations and if at the same time each x_ν is replaced by lx_ν then the statistical function is not changed.[1] Examples of statistical functions are the *moments*

$$\frac{1}{n}\sum_{\nu=1}^{n} x_\nu^r = \int x^r\, dS_n(x) = M_r^0$$

or, if $M_1^0 = \alpha$, the moments about the mean α:

$$\frac{1}{n}\sum_{\nu=1}^{n} (x_\nu - \alpha)^r = \int (x - \alpha)^r\, dS_n(x) = M_r, \text{ etc.}$$

The independent variable in $f\{S_n(x)\}$ is a partition; but in addition we shall define $f\{P(x)\}$ where $P(x)$ is a certain bounded distribution which is not necessarily a partition. A distribution $P(x)$ is called bounded if

(1′) $$P(x) = 0 \quad \text{if} \quad x < A \quad \text{and} \quad P(x) = 1 \quad \text{if} \quad x \geqq B.$$

If this is true for a sequence $P_1(x), P_2(x), \cdots$ with the same A and B then the sequence is called *uniformly bounded.* Let us now consider a bounded partition $P(x)$ which in every point of continuity of $P(x)$ is the limit as $n \to \infty$ of a sequence of bounded partitions $S_n(x)$. *As* $S_n(x)$ *converges toward* $P(x)$, if $f\{S_n(x)\}$ converges towards a limit L which does not depend on the limiting process $S_n(x) \to P(x)$ then that limit shall be denoted by $f\{P(x)\}$; it will be called the *value of the statistical function at the "point"* $P(x)$ and $f\{S_n(x)\}$ will be called *continuous* at $P(x)$. The definition of continuity can be given also in the following way: Corresponding to every $\epsilon > 0$ exists an $\eta > 0$ such that

(2) $$|f\{S_n(x)\} - f\{P(x)\}| < \epsilon$$

for all values of n and for every bounded $S_n(x)$ *such that* at every point of continuity of $P(x)$

(3) $$|S_n(x) - P(x)| \leqq \eta.$$

In this case $f\{S_n(x)\}$ is called continuous at the point $P(x)$. Thus a statistical function is defined for bounded partitions and for certain bounded distributions which are not themselves partitions. If the continuity defined by (2) and (3) exists for a sequence $P_1(x), P_2(x), \cdots$ of bounded distributions with the same η

---

[1] This condition of homogeneity is fulfilled e.g. for $\sqrt{x_1 x_2 \cdots x_n}$ but not for $x_1 x_2 \cdots x_n$.

corresponding to a given ϵ, we call the statistical function *uniformly* continuous at the points $P_1(x)$, $P_2(x)$, $\cdots$.

3. The general law of large numbers. The generalization of the law of large numbers which we have in mind can be demonstrated in a way analogous to the demonstration given by v. Mises in the case of independent collectives if we introduce the results of paragraph 1 in order to estimate the variance. We shall consider here only one dimensional, bounded collectives in order to make clearer what is the essential of the generalization.

A sequence of dependent collectives $P_1(x)$, $P_2(x)$, $\cdots$, $P_n(x)$ can be given in the following manner. Let $P(x_1, x_2, \cdots, x_n)$ be the probability that the result of the first observation is $\leqq x_1$, of the second $\leqq x_2$, $\cdots$, of the nth $\leqq x_n$. This distribution will be said to be *bounded* in (A, B) if $P = 1$ when all the x_ν are $\geqq B$ and $P = 0$ if at least one of these arguments is less than A. From this n-dimensional distribution we deduce n one dimensional distributions

$$(1) \qquad P_1(x) = P(x, B, \cdots, B), \qquad P_2(x) = P(B, x, B, \cdots, B), \cdots, P_n(x) = P(B, \cdots, B, x)$$

where $P_\nu(x)$ is the probability that the νth observation be $\leqq x$. The $P_\nu(x)$ are uniformly bounded in (A, B) which is a consequence of $P(x_1, x_2, \cdots, x_n)$ having been assumed to be bounded in this interval. In an analogous way we deduce from $P(x_1, x_2, \cdots, x_n)$ the $\frac{1}{2}n(n-1)$ uniformly bounded two dimensional distributions

$$(2) \quad P_{12}(x, y) = P(x, y, B, \cdots B), \qquad P_{13}(x, y) = P(x, B, y, B, \cdots B), \cdots .$$

Here $P_{\mu\nu}(x, y)$ is the probability that the μth result is $\leqq x$, the νth result $\leqq y$, and we have $P_{\mu\nu}(x, y) = P_{\nu\mu}(y, x)$. Of course we have also

$$(1') \qquad P_1(x) = P_{12}(x, B) = P_{13}(x, B) = \cdots = P_{1n}(x, B)$$

$$P_2(x) = P_{12}(B, x) = P_{23}(x, B) = \cdots = P_{2n}(x, B) \text{ etc.}$$

If we put in (2) $x = y$ we obtain $P_{\mu\nu}(x, x) = P_{\nu\mu}(x, x)$ and we introduce

$$(3) \qquad P_{\mu\nu}(x, x) = P_{\nu\mu}(x) = P_{\mu\nu}(x)$$

the probability that both the μth and the νth observation is $\leqq x$. Then $P_{\mu\nu}(x)$ equals zero if $x < A$ and equals one if $x \geqq B$, and this is valid with the same A and B for all the distributions $P_{\nu\mu}(x)$.

Now if $p_1, p_2, \cdots, p_n$ are the probabilities of success for n general alternatives Tchebycheff's Lemma asserts that the probability W that the average $(x_1 + x_2 + \cdots + x_n)/n$ of n observations differs by more than η from its expectation $(p_1 + p_2 + \cdots + p_n)/n$ is subject to the following inequality

$$(4) \qquad W \leqq \frac{1}{\eta^2} \operatorname{Var}\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) = \frac{s_n^2}{\eta^2 n^2}.$$

Here s_n^2 is given by (1) of paragraph 1.

Let us introduce the average $\bar{P}_n(x)$ of the $P_\nu(x)$:

$$(5) \qquad \bar{P}_n(x) = [P_1(x) + P_2(x) + \cdots + P_n(x)]/n$$

and let Q_n be the probability that at any point of continuity of $\bar{P}_n(x)$ the inequality

$$(6) \qquad | S_n(x) - \bar{P}_n(x) | > \eta$$

holds. Our aim will be to show that for every η under certain restrictions regarding the given collectives, Q_n tends toward zero as n tends toward infinity.

For a fixed point x' the probabilities $P_\nu(x) = p_\nu$ and $P_{\mu\nu}(x) = p_{\mu\nu}$ are constants and we put $\bar{P}_n(x) = \bar{p}_n = (p_1 + p_2 + \cdots p_n)/n$. The probability that in x'

$$(7) \qquad | S_n(x') - \bar{P}_n(x') | > \eta/2$$

is then, according to (4) smaller than $(s_n^2)_{x'}/(\frac{1}{2}\eta)^2 n^2$. Here we denote by $(s_n^2)_{x'}$ the value of s_n^2 in x' (as given by (1) in paragraph 1).

Now we divide the interval (A, B) in N parts in such a way that in every one of the N intervals e.g. in (x', x'') the variation

$$(8) \qquad \delta = \bar{P}_n(x'') - \bar{P}_n(x') \leqq \eta/2.$$

If there is at x' (or at x'') a step of $\bar{P}_n(x)$ we take the limit which $\bar{P}_n(x)$ approaches as $x \longrightarrow x'$ (or x'') from the interior of the interval. In order to obtain such a division we need only divide the total variation 1 of $\bar{P}_n(x)$ in $2/\eta$ equal parts and project these points of division on $\bar{P}_n(x)$, disposing however in a suitable way of horizontal parts of $\bar{P}_n(x)$. The abscissae of these points form the endpoints of the N intervals. If there is a step of $\bar{P}_n(x)$ at an endpoint of one of these intervals the variation in both the adjacent intervals can only be diminished. It is further possible that the two ends of an interval coincide $x' = x''$, this will be so if $\bar{P}_n(x)$ has for x' a step $> \eta/2$. In any case we have a division in $N \leqq 2/\eta$ intervals such that all the points of continuity of $\bar{P}_n(x)$ are enclosed in them and in each of these intervals (8) is valid.

Let us now assume that in the left end point x' of the rth interval (x', x'') the inequality

$$(9) \qquad | S_n(x') - \bar{P}_n(x') | \leqq \eta/2$$

is valid. Then we have for *every* x between x' and x''

$$(10) \qquad | S_n(x) - \bar{P}_n(x) | \leqq \eta/2 + \delta \leqq \eta.$$

Because, since $S_n(x)$ and $\bar{P}_n(x)$ are both monotone, the difference $S_n(x') - \bar{P}_n(x')$ cannot increase by more than $\delta \leqq \eta/2$ as x varies from x' to x''. Therefore if (6) is valid for any point x in this interval then (7) must be valid for the left end point x' of this interval and the probability q_r of this latter inequality is less than or equal to $4(s_n^2)_{x'}/\eta^2 n$.

But there are N intervals with the left endpoints $x_1', x_2', \cdots x_N'$ and **the** probability that (6) may be valid in any point belonging to any one of these intervals is $\leqq q_1 + q_2 + \cdots + q_N$. Denoting by s_n^2 the greatest of the N variances $(s_n^2)_{x_1'}$, $(s_n^2)_{x_2'}$, $\cdots$, $(s_n^2)_{x_N'}$ we have for Q_n (which is the probability that (6) may be valid at any point of continuity of $P(x)$) the inequality

$$(11) \qquad Q_n \leqq q_1 + q_2 + \cdots + q_N \leqq \frac{4N}{\eta^2 n^2} s_n^2 \leqq \frac{8}{\eta^3} \frac{s_n^2}{n^2}.$$

Therefore Q_n tends toward zero for every η if s_n/n tends toward zero.

But according to (2) in paragraph 1, s_n/n tends toward zero if for every x in (A, B)

$$(12) \qquad \lim_{n\to\infty} \frac{1}{n^2} \sum_{\mu,\nu=1}^{n} [P_{\mu\nu}(x) - P_\mu(x)P_\nu(x)] = 0.$$

Considering the definition of continuity of a statistical function we have obtained the following result:

As in (1′), (2), (3) *and* (5) *let* $P_{\mu\nu}(x, y)$ *be two dimensional distributions* ($\mu, \nu = 1, \cdots, n; \mu \neq \nu$), *uniformly bounded in* (A, B); $P_{\mu\nu}(x, B) = P_\mu(x)$; $P_{\mu\nu}(x, x) = P_{\mu\nu}(x)$ *and* $\bar{P}_\nu(x) = 1/\nu(P_1(x) + P_2(x) + \cdots + P_\nu(x))$.

If the variable partition $S_n(x)$ *is bounded in* (A, B) *and if* $f\{S_n(x)\}$ *is uniformly continuous at the "points"* $\bar{P}_1(x)$, $\bar{P}_2(x)$, $\cdots$ *then the probability that*

$$(13) \qquad |f\{S_n(x)\} - f\{\bar{P}_n(x)\}| > \epsilon$$

tends toward zero for every ϵ *as* $n \to \infty$, *provided* (12) *is uniformly valid for every* x *in* (A, B).

4. Examples. Let us illustrate by simple examples.

1) In order to define the $P_\nu(x)$ etc. mentioned in our theorem we define the n-dimensional distribution $P(x_1, x_2, \cdots x_n)$ used at the beginning of paragraph 3 by indicating the probability density

$$(1) \qquad \begin{aligned} \mu(x_1, x_2, \cdots, x_n) &= C_n[1 - x_1x_2 \cdots x_n] \quad \text{in the "unit cube"}, \\ &= 0 \qquad\qquad\qquad\qquad\quad \text{elsewhere.} \end{aligned}$$

The corresponding probability distribution is

$$(2) \qquad P(x_1, x_2, \cdots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} \mu(x_1, x_2, \cdots, x_n)\, dx_1 \cdots dx_n.$$

By putting

$$(3) \qquad C_n = \frac{2^n}{2^n - 1},$$

we see that $P(x_1, x_2, \cdots, x_n)$ equals unity if all the arguments are $\geqq 1$ and it equals zero if one of these arguments is less than 0. Therefore $P(x_1, x_2, \cdots, x_n)$ is bounded in the unit cube.

From (1) we deduce the two-dimensional densities

$$(4)\qquad \begin{aligned} v_{\mu\nu}(x, y) &= C_n\left(1 - \frac{xy}{2^{n-2}}\right) && \text{in the unit square,} \\ &= 0 && \text{elsewhere} \end{aligned}$$

and the distributions

$$(5)\qquad P_{\mu\nu}(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} v_{\mu\nu}(x, y)\, dx\, dy.$$

We see that

$$\begin{aligned} P_{\mu\nu}(x, y) &= C_n xy\left(1 - \frac{xy}{2^n}\right) && \text{in the unit square} \\ &= 0 && \text{if } x \text{ or } y \leqq 0 \\ &= 1 && \text{if } x \text{ and } y \geqq 1 \end{aligned}$$

and e.g. for $x \geqq 1$, $0 < y < 1$ we have $P_{\mu\nu}(x, y) = P_{\mu\nu}(1, y)$ etc. Thus the $P_{\mu\nu}(x, y)$ are completely given.

It follows from (3) that $-C_n/2^n = 1 - C_n$; therefore putting $C_n = C$ we have in (0, 1)

$$(6)\qquad \begin{aligned} P_{\mu\nu}(x, x) = P_{\mu\nu}(x) &= Cx^2 + (1 - C)x^4 \\ P_\nu(x) &= Cx + (1 - C)x^2 \end{aligned}$$

therefore

$$(7)\qquad P_{\mu\nu}(x) - P_\mu(x)P_\nu(x) = C(1 - C)x^2(1 - x)^2$$

is < 0 for every x in (0, 1) since $C > 1$. For $x \leqq 0$, $P_{\mu\nu}(x)$ and $P_\nu(x)$ both equal zero and for $x \geqq 1$ they both equal 1. Therefore our conditions of paragraph 1 are fulfilled. We see that C_n tends towards unity as $n \to \infty$, therefore for every x in (0, 1) $P_{\mu\nu}(x) - P_\mu(x)P_\nu(x)$ tends towards zero, we have "convergence towards independence" but by no means independence.

This example was based on a *symmetric density*. Let us give an example of asymmetric and *arithmetic* distributions. For the sake of simplicity let $P_1(x)$, $P_2(x)$, $\cdots$ be arithmetic distributions each with only three steps at $x = 0$, 1 and 2. As starting point we take the n-dimensional arithmetic distribution $v(x_1, x_2, \cdots x_n)$ which gives the probability that the first result equals x_1, the second x_2, $\cdots$, the nth x_n, the x_ν being equal to 0 or 1 or 2; thus $v(x_1, x_2, \cdots, x_n)$ takes 3^n values the sum of which equals unity. We deduce the two dimensional distributions $v_{\mu\nu}(x, y)$, e.g. $v_{12}(x, y) = \sum_{x_3, \cdots, x_n} v(x, y, x_3, \cdots, x_n)$, the probability that the first result equals x, the second y, and finally the $v_1(x) = \sum_y v_{12}(x, y)$, etc. According to the definitions of $P_\nu(x)$ and $P_{\mu\nu}(x)$ we have then:

$$(8) \qquad \begin{aligned} P_\nu(x) &= 0 && (x < 0) \\ &= v_\nu(0) && (0 \leqq x < 1) \\ &= v_\nu(0) + v_\nu(1) && (1 \leqq x < 2) \\ &= 1 && (2 \leqq x), \end{aligned}$$

$$(9) \qquad \begin{aligned} P_{\mu\nu}(x) &= 0 && (x < 0) \\ &= v_{\mu\nu}(0, 0) && (0 \leqq x < 1) \\ &= v_{\mu\nu}(00) + v_{\mu\nu}(10) + v_{\mu\nu}(01) + v_{\mu\nu}(11) && (1 \leqq x < 2) \\ &= 1 && (2 \leqq x). \end{aligned}$$

Now we subject $v(x_1, \cdots, x_n)$ to the following conditions: Every $v(x_1, \cdots, x_n)$ equals zero if it contains either: at least two "zeros," or: at least one "zero" and one "one," or: at least two "ones." All the other v-values are supposed to be different from zero. Then we have

$$v_{\mu\nu}(0, 0) = v_{\mu\nu}(1, 0) = v_{\mu\nu}(0, 1) = v_{\mu\nu}(1, 1) = 0$$

therefore $P_{\mu\nu}(x) = 0$ for $x < 2$ and $P_{\mu\nu}(x) = 1$ for $x \geqq 2$. On the other hand $v_\nu(0) = v(2, 2, \cdots 2, 0, 2, \cdots 2)$ and $v_\nu(1) = v(2, 2, \cdots 2, 1, 2, \cdots 2)$ therefore $P_\nu(x) \neq 0$ for $0 \leqq x < 2$ and we have thus for every finite n

$$\begin{aligned} P_{\mu\nu}(x) - P_\mu(x)P_\nu(x) &= 0 \quad \text{for } x < 0 \text{ and } x \geqq 2, \\ &< 0 \quad \text{for } 0 \leqq x < 2. \end{aligned}$$

Therefore the condition (*b*) of paragraph 1 is fulfilled and thus (12) paragraph 3 holds.

I hope to have the opportunity to discuss more general applications of this theorem later.

A generalization of the *strong* law of large numbers may be given in a similar way.

REFERENCES

[1] B. H. Camp, *The Mathematical Part of Elementary Statistics*, New York, 1934, page 256.
[2] R. v. Mises, "Die Gesetze der großen Zahl für statistische Funktione," *Monatshefte für Math. u. Physik*, 1936, p. 105–128.
[3] H. Geiringer, "Sur les variables aléatoires arbitrairement liées," *Revue de l'Union Interbalcanique*, 1938, p. 6.

Bryn Mawr College,
Bryn Mawr, Pennsylvania.