

THE GENERAL LIMIT THEOREM IN THE THEORY OF PROBABILITY¹

WOLFGANG STERNBERG

Let x_1, x_2, \dots, x_n be independent chance variables with any distribution functions V_1, V_2, \dots, V_n .

We put² for all $i=1, 2, \dots, n$

$$M[x_i] = \int \xi dV_i(\xi) = a_i \quad (\text{mean value}),$$

$$\text{St. D. } (x_i) = \int (\xi - a_i)^2 dV_i(\xi) = b_i \quad (\text{standard deviation}),$$

$$\int |\xi - a_i|^3 dV_i(\xi) = c_i \quad (\text{absolute moment of the 3rd order}).$$

(The term "standard deviation" is used by some authors for $b_i^{1/2}$ instead of b_i .)

Denoting mean value and standard deviation of the sum $x_1 + \dots + x_i$ ($i=2, \dots, n$) by A_i and B_i respectively, it is known that

$$A_i = a_1 + \dots + a_i, \quad B_i = b_1 + \dots + b_i.$$

For the distribution function W_i of the sum $x_1 + \dots + x_i$ the recursion formula

$$(1) \quad W_i(x) = \int W_{i-1}(x - \xi) dV_i(\xi), \quad i > 1,$$

holds true.

Let a_i and B_n be as follows:

$$(2) \quad a_i = 0,$$

$$(3) \quad B_n = 1.$$

These conditions are only preliminarily stated in order to abbreviate the explanation. From (2) we have that $A_i=0$, in particular, $A_n=0$. (The case $b_i=0$ may be excluded, since then x_i assumes the value 0 only with the probability 1 and thus does not appear in the sum $x_1 + \dots + x_n$.)

¹ Presented to the Society, October 28, 1939.

² The following integrals are Stieltjes integrals. The lower and upper limits always are $-\infty$ and $+\infty$ respectively.

We have to prove that the function $W_n(x)$ lies uniformly in x near the Gaussian function having also the mean value 0 and the standard deviation 1, that is,

$$\phi(x/2^{1/2}) = \frac{1}{\pi^{1/2}} \int_{-\infty}^{x/2^{1/2}} e^{-t^2} dt,$$

if very general suppositions with respect to the V are made. We suppose that the quotients c_i/b_i are "very small" for all i and state the general limit theorem in the following form:

To every positive ϵ there is a positive δ , so that uniformly in x

$$| W_n(x) - \phi(x/2^{1/2}) | < \epsilon$$

holds true if the conditions

$$(4) \quad c_i/b_i < \delta, \quad i = 1, 2, \dots, n,$$

are satisfied.

Our proof is suggested by that of A. Khintchine (see below). We remark first of all that

$$(5) \quad b_i < \delta, \quad i = 1, 2, \dots, n,$$

follows from (4) according to the well known inequality $b_i^3 < c_i^2$. (Of course we can suppose $\delta < 1$.) Since $B_n = 1$ but the b_i are small, it is obvious that n must be a large number.

Now we note that the function of two variables

$$\phi(x/(2y)^{1/2}) = \frac{1}{\pi^{1/2}} \int_{-\infty}^{x/(2y)^{1/2}} e^{-t^2} dt = U(x, y),$$

that is, the Gaussian function with the variable standard deviation y represents a solution of the differential equation of the conduction of heat

$$(6) \quad \frac{\partial \phi}{\partial y} = \frac{1}{2} \frac{\partial^2 \phi}{\partial x^2}.$$

Here and in what follows it is always supposed that $y \geq \eta$, η being a positive small but constant magnitude. The equation (1) suggests that we compare the two expressions $U(x, B_i)$ and

$$\int U(x - \xi, B_{i-1}) dV_i(\xi).$$

(Later the variable y will be put equal to B_{i-1} .) According to Taylor's

formula we have

$$U(x, y + b_i) = U(x, y) + b_i \frac{\partial U(x, y)}{\partial y} + R_1,$$

$$R_1 = \frac{b_i^2}{2} \frac{\partial^2 U(x, y + \theta b_i)}{\partial y^2}, \quad 0 < \theta < 1.$$

Let $M(\eta)$ be an upper limit of the absolute values of the derivatives of the 2d and 3d order of $U(x, y)$ in the half-plane $y \geq \eta$; then

$$|R_1| < b_i^2 M(\eta)/2.$$

On the other hand,

$$U(x - \xi, y) = U(x, y) - \xi \frac{\partial U(x, y)}{\partial x} + \frac{\xi^2}{2} \frac{\partial^2 U(x, y)}{\partial x^2}$$

$$- \frac{\xi^3}{6} \frac{\partial^3 U(x - \theta_1 \xi, y)}{\partial x^3}, \quad 0 < \theta_1 < 1.$$

Therefore, in view of (2) and (6)

$$\int U(x - \xi, y) dV_i(\xi) = U(x, y) + b_i \frac{\partial U(x, y)}{\partial y} + R_2,$$

where

$$R_2 = - \frac{1}{6} \int \xi^3 \frac{\partial^3 U(x - \theta_1 \xi, y)}{\partial x^3} dV_i(\xi);$$

thus

$$|R_2| < c_i M(\eta)/6.$$

It follows that

$$(7) \quad U(x, y + b_i) = \int U(x - \xi, y) dV_i(\xi) + R$$

in which $R = R_1 - R_2$; accordingly

$$(8) \quad |R| \leq |R_1| + |R_2| < M(\eta) \{b_i^2/2 + c_i/6\} < [M(\eta)/2] \{b_i^2 + c_i\}.$$

Now note that all $b_i < \eta$ because of (5), since of course $\delta < \eta$ may be supposed. On the contrary there are numbers $B_i > \eta$, as for instance $B_n = 1 > \eta$. There must be consequently a number B_r satisfying $\eta < B_r < 2\eta$. Now we put $y = B_{i-1}$ in (7); and we suppose $i \geq r+1$ in order that $y \geq B_r > \eta$. Besides we write with regard to what follows

$x + 2\alpha$ instead of x , α representing a positive parameter. Then it follows from (7) that

$$(9) \quad U(x + 2\alpha, B_i) = \int U(x + 2\alpha - \xi, B_{i-1}) dV_i(\xi) + R$$

in which R satisfies (8). Putting the difference

$$(10) \quad W_i(x) - U(x + 2\alpha, B_i) = W_i(x) - \phi\left(\frac{x + 2\alpha}{(2B_i)^{1/2}}\right) = D_i(x),$$

$$i = 1, 2, \dots, n,$$

and subtracting (9) from (1), we get the fundamental formula, valid for all x ,

$$(11) \quad D_i(x) = \int D_{i-1}(x - \xi) dV_i(\xi) - R$$

together with (8), where $i = r + 1, r + 2, \dots, n$.

In order to get further estimates we need the elementary inequalities of Tschebyscheff, which are valid for any distribution function F with vanishing mean value and standard deviation not greater than B :

$$F(\alpha) \geq 1 - B/\alpha^2, \quad F(-\alpha) \leq B/\alpha^2, \quad \alpha > 0.$$

From these inequalities it easily follows for two distribution functions F_1, F_2 with vanishing mean values, the standard deviations of which are both not greater than B , and for all x , that

$$(12) \quad F_1(x) - F_2(x + 2\alpha) \leq B/\alpha^2, \quad \alpha > 0.$$

For, if $x \leq -\alpha$, one has $F_1(x) \leq F_1(-\alpha) \leq B/\alpha^2$; thus (12) follows because $F_2 \geq 0$. If $x > -\alpha$, one concludes $F_2(x + 2\alpha) \geq F_2(\alpha) \geq 1 - B/\alpha^2$; therefore

$$F_1(x) - F_2(x + 2\alpha) \leq 1 - F_2(x + 2\alpha) \leq 1 - (1 - B/\alpha^2),$$

that is, (12). On the foundation of (11) and (12) the proof may be carried out. Putting $i = r$ one has, according to the definition of $D_i(x)$ and because both $W_r(x)$ and $U(x, B_r)$ have the mean value 0 and the standard deviation $B_r < 2\eta$, on account of (12),

$$(13) \quad W_r(x) - U(x + 2\alpha, B_r) = D_r(x) < 2\eta/\alpha^2 \quad \text{for all } x, \quad \alpha > 0;$$

whence, denoting by g_i the upper limit of $D_i(x)$, we conclude from (13) $g_r < 2\eta/\alpha^2$ and from (11), for any $i \geq r + 1$, $D_i(x) \leq g_{i-1} + |R|$; thus

$$g_i \leq g_{i-1} + [M(\eta)/2] \{b_i^2 + c_i\}$$

and through induction from i to $i+1$

$$(14) \quad g_n \leq g_r + \frac{M(\eta)}{2} \sum_{i=r+1}^n \{b_i^2 + c_i\} < \frac{2\eta}{\alpha^2} + \frac{M(\eta)}{2} \sum_{i=1}^n \{b_i^2 + c_i\}.$$

On the other hand one has

$$b_i < \delta, \quad c_i/b_i < \delta, \quad i = 1, 2, \dots,$$

because of (5) and (4). Therefore

$$\sum_{i=1}^n b_i^2 < \delta \sum b_i = \delta B_n = \delta, \quad \sum_{i=1}^n c_i < \delta \sum b_i = \delta.$$

It follows that $g_n < 2\eta/\alpha^2 + \delta M(\eta)$; thus

$$D_n(x) = W_n(x) - \phi\left(\frac{x + 2\alpha}{2^{1/2}}\right) < 2\eta/\alpha^2 + \delta M(\eta).$$

Therefore

$$(15) \quad W_n(x) < \phi(x/2^{1/2}) + 2\alpha/(2\pi)^{1/2} + 2\eta/\alpha^2 + \delta M(\eta).$$

The positive number ϵ being given, one determines first α through $2\alpha/(2\pi)^{1/2} = \epsilon/3$, then η through $2\eta/\alpha^2 = \epsilon/3$, with which $M(\eta)$ is determined, and finally δ through $\delta M(\eta) \leq \epsilon/3$. It follows that

$$(16) \quad W_n(x) < \phi(x/2^{1/2}) + \epsilon,$$

uniformly in x . Using

$$\phi\left(\frac{x - 2\alpha}{(2B_i)^{1/2}}\right) - W_i(x) = \Delta_i(x)$$

one finds in the same way

$$\Delta_n(x) < 2\eta/\alpha^2 + \delta M(\eta)$$

and

$$(17) \quad W_n(x) > \phi(x/2^{1/2}) - \epsilon.$$

The proof is thus complete.

One should compare this with the proof given by A. Khintchine (*Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 2, no. 4, Berlin, 1933), which is based on works of I. Petrowsky and A. Kolmogoroff. We have chosen other assumptions in order to make the exposition as simple as possible, while Khintchine strives for wide generality so that nothing is supposed by him as to moments of the third order.

A so-called upper- and lower-function, the introduction of which is denoted by Khintchine as the fundamental idea of the Petrowsky-Kolmogoroffian proof, is not used by us.

We now generalize the investigation and omit the special conditions (2) and (3).

We use the transformation

$$(18) \quad \frac{x_i - a_i}{B_n^{1/2}} = u_i, \quad i = 1, \dots, n.$$

If mean value, standard deviation, and absolute moment of the third order of u_i are denoted by \bar{a}_i , \bar{b}_i , and \bar{c}_i respectively, it becomes

$$(2') \quad \bar{a}_i = 0, \quad \bar{b}_i = b_i/B_n, \quad \bar{c}_i = c_i/(B_n)^{3/2}$$

and accordingly

$$\frac{\bar{c}_i}{\bar{b}_i} = \frac{c_i}{b_i(B_n)^{1/2}},$$

and

$$(3') \quad \bar{B}_n = \text{St. D. } (u_1 + \dots + u_n) = \bar{b}_1 + \dots + \bar{b}_n = 1.$$

The assumptions (4) must be replaced by

$$(4') \quad \frac{c_i}{b_i B_n^{1/2}} < \delta, \quad i = 1, \dots, n.$$

Denoting by \bar{W}_n the distribution function of the sum $u_1 + \dots + u_n$, we thus have according to (16), (17) uniformly in u

$$(19) \quad \bar{W}_n(u) = \phi(u/2^{1/2}) + \sigma, \quad |\sigma| < \epsilon.$$

But it easily follows in view of

$$u_1 + \dots + u_n = (x_1 + \dots + x_n - A_n)/B_n^{1/2}$$

that the distribution function $W_n(x)$ of $x_1 + \dots + x_n$ is identical with $\bar{W}_n(u)$ if the relation

$$(20) \quad u = (x - A_n)/B_n^{1/2}$$

between u and x is valid.

According to (19) and $W_n(x) = \bar{W}_n(u)$ we have uniformly in x

$$(21) \quad W_n(x) = \phi\left(\frac{x - A_n}{(2B_n)^{1/2}}\right) + \sigma, \quad |\sigma| < \epsilon.$$

We can thus say: *To every positive ϵ there is a positive δ , so that (21) holds true, if the assumptions (4') are satisfied.*

We call special attention to the case

$$0 < g < b_i; \quad c_i < H,$$

g and H being independent of i . One has $gn < B_n$ and as for the rest $b_i < c_i^{2/3} < H^{2/3}$ and $B_n < nH^{2/3}$, so that B_n is exactly of the order n . It follows that

$$\frac{c_i}{b_i B_n^2} < \frac{H}{g g^{1/2} n^{1/2}}.$$

Therefore (4') can be satisfied, if n is sufficiently large.

In the same way one can discuss sets of chance variables $x_{11}, x_{12}, \dots, x_{1n_1}, \dots, x_{k1}, x_{k2}, \dots, x_{kn_k}, \dots, n_k$ increasing indefinitely with k . It is supposed that the quotients

$$\frac{c_{k1}}{b_{k1} B_{n_k}^{1/2}}, \dots, \frac{c_{kn_k}}{b_{kn_k} B_{n_k}^{1/2}}$$

approach 0 as a limit, if k increases indefinitely. That means that if any $\delta > 0$ is given one can choose k sufficiently large so that *all* these quotients are less than δ . Then the distribution function W_{n_k} of $x_{k1} + \dots + x_{kn_k}$ approaches $\phi((x - A_{n_k}) / (2B_{n_k})^{1/2})$ as a limit.

The conditions for the quotients are in particular satisfied if $B_{n_k} > B > 0$ for every k (B being a constant number) and the quotients $c_{k1}/b_{k1}, \dots, c_{kn_k}/b_{kn_k}$ approach 0, as k increases indefinitely.

NEW YORK, N. Y.