On weak convergence of randomly selected partial sums

By M. CSÖRGŐ[1)] in Ottawa (Canada) and S. CSÖRGŐ in Szeged

In honour of Professor Béla Szőkefalvi-Nagy on his sixtieth birthday

1. Introduction. Let $\xi_1, \xi_2, \ldots$ be a sequence of random variables (r.v.'s) defined on a probability space $(\Omega, \mathscr{B}, P)$ and suppose that the partial sums $S_n = \xi_1 + \xi_2 + \cdots + \xi_n$ obey the central limit theorem, say with the positive norming factors a_n, so that the distribution of $a_n^{-1} S_n$ is asymptotically the unit normal. Now let ν_n be a sequence of positive integer valued r.v.'s defined on the same probability space. Beginning with the early work of ANSCOMBE (1952), [1], several authors have dealt with the convergence problem of $a_{\nu_n}^{-1} S_{\nu_n}$ (see e.g. [15], [13], [5] and [6]) and, in general, with the following problem: if we are given that a sequence of r.v.'s already satisfies an asymptotic law, then under what conditions should the same sequence, but indexed by ν_n, satisfy the same law (see [17], [12], [8] and [9]). On the other hand, these results have established the background for studying the problem of weak convergence of randomly selected partial sum processes on function spaces, and this work has begun with BILLINGSLEY ([12], 1962). This paper is going to deal with this latter approach, trying to provide a general procedure.

2. Weak convergence of randomly selected partial sum type processes on the space D

Let $D = D[0, 1]$ be the space of functions with discontinuities only of the first kind. Under Prohorov's metric [14] or under Skorokhod's metric [18] with Billingsley's modification of it [3] D is a complete and separable metric space. Let $\mathscr{D}$ be the σ-algebra generated by the open sets of D. If for each $n \geqq 0$, X_n is a measurable mapping from $(\Omega, \mathscr{B})$ to $(D, \mathscr{D})$, that is, in Billingsley's terminology (which will be followed throughout, [3]), X_n is a random function of D, and $\mathscr{P}_n$ denotes the induced image law of X_n on $(D, \mathscr{D})$, then we say X_n converges in distribution to X_0 with the

[1)] Work supported by a National Research Council of Canada Grant at Carleton University, Ottawa.

induced image law $\mathscr{P}_0$, written $X_n \xrightarrow{\mathscr{D}} X_0$, if for all real valued continuous bounded functions g on D $\lim_{n\to\infty} \int_D g\,d\mathscr{P}_n = \int_D g\,d\mathscr{P}_0$ holds. The sequence X_n is called tight if for every positive ε there exists a compact set K in $\mathscr{D}$ such that $\mathscr{P}_n(K) > 1-\varepsilon$, $n=1, 2, \ldots$. The following two theorems (Theorems 15. 2 and 15. 1, [3]), of which the first one characterises the notion of tightness in D and the second one the convergence in distribution (weak convergence), will be used in the sequel.

Theorem A. *The sequence X_n is tight if and only if these two conditions hold:*
(i) *For each positive η there exists a d such that*

$$P\{\sup_t |X_n(t)| > d\} < \eta, \quad n \geqq 1.$$

(ii) *For each positive ε and η, there exist a δ, $0<\delta<1$, and an integer n_0 such that $P\{w'(X_n, \delta)>\varepsilon\}<\eta$, $n\geqq n_0$, where*

$$w'(X_n, \delta) = \inf_{(1,\delta,\{t_i\})} \max_{1\leqq i\leqq r} w_{X_n}([t_{i-1}, t_i)) \tag{1}$$

with $w_{X_n}([t_{i-1}, t_i)) = \sup_{u,v\in[t_{i-1},t_i)} |X_n(u)-X_n(v)|$, *and the infimum extends over all finite sets (t_i) of points satisfying*

$$0 = t_0 < t_1 < \ldots < t_r = 1, \quad t_i - t_{i-1} > \delta, \qquad i = 1, 2, \ldots, r.$$

Theorem B. *For X_n to converge in distribution to X_0 it is necessary and sufficient that the finite dimensional distributions of it should converge to those of X_0 and that X_n should be tight.*

The random functions of D we are going to be concerned with are of the form:

$$X_n(t) = X_n(t, \omega) = a_n^{-1} X(nt, \omega), \quad 0 \leqq t \leqq 1, \quad n \geqq 1, \tag{2}$$

where $X(u, \omega)$, for each fixed ω in Ω, is a right continuous function of u on $[0, \infty)$ having left-hand limits at each point, and for u fixed it is measurable with respect to $(\Omega, \mathscr{B})$; the sequence a_n of positive numbers increases monotonically to ∞ with n, and it is also slowly oscillating in the sense of Karamata. This latter notion means that a_n is of the form $n^\alpha L(n)$ with α positive and $L([cn])/L(n) \to 1$ as $n \to \infty$ for every positive c.

The most immediate examples of the form of (2) are the partial sum processes $a_n^{-1} S_{[nt]}$, $S_0=0$, and several other processes of D can be brought into this form when technicalities of certain proofs so require. As to the latter we mention a fortchoming paper by one of us [10], concerning the weak convergence of the random sample size empirical process.

In this exposition the possibility of deducing the weak convergence of X_{ν_n} from that of X_n of (2) is examined. Towards this end the following lemma is essential.

Lemma. *If X_n of (2) is tight and if the sequence of positive integer valued r.v.'s ν_n is such that $\nu_n/n \xrightarrow{P} \nu$, where ν is an arbitrary positive r.v then X_{ν_n} is also tight.*

Proof. The special form of X_n of (2) implies that conditions (i) and (ii) of Theorem A are satisfied if and only if

(i)* For each positive η there exists a d such that

$$P\{\sup_{0\leqq v\leqq n} |X(v)| > da_n\} < \eta, \quad n \geqq 1.$$

(ii)* For each positive ε and η, there exists a δ, $0<\delta<1$, and an integer n_0 such that

$$P\{\inf_{(n,\delta,\{v_i\})} M_r(n) > \varepsilon a_n\} < \eta, \quad n \geqq n_0,$$

where $M_r(n) = \max\limits_{1\leqq i\leqq r} \sup\limits_{v,u\in[v_{i-1},v_i)} |X(u)-X(v)|$, and $\inf\limits_{(n,\delta,\{v_i\})}$ stands for the infimum over the finite sets of points (v_i) satisfying

$$0 = v_0 < v_1 < \ldots < v_r = n, \quad v_i - v_{i-1} > n\delta, \quad i = 1, 2, \ldots, r.$$

In order to prove our lemma, we simply have to verify conditions (i)* and (ii)* for X_{ν_n}, that is to say (i)* and (ii)* with n replaced by ν_n in them.

First we verify condition (ii)*. Let ε and η be fixed positive numbers and let $\theta=\eta/3$. Choose $0<a<b$ so that $P\{a \leqq \nu < b\} > 1-\theta$. Without loss of generality assume that $\varepsilon<a$, and choose $n_1=n_1(\varepsilon, \theta)$ so that $P\left\{\left|\frac{\nu_n}{n}-\nu\right|>\varepsilon\right\}<\theta$ for $n\geqq n_1$. For arbitrary δ and $n\geqq n_1$

$$P\{\inf_{(\nu_n,\delta,\{v_i\})} M_r(\nu_n) > \varepsilon a_{\nu_n}\} \leqq P\{\inf_{(\nu_n,\delta,\{v_i\})} M_r(\nu_n) > \varepsilon a_{\nu_n}, \tag{3}$$

$$n(a-\varepsilon) \leqq \nu_n \leqq n(b+\varepsilon)\} + 2\theta.$$

Now for each fixed $\omega\in\Omega$ and ν_n in the indicated range above

$$\inf_{(n(b+\varepsilon),\delta,\{v_i\})} M_r\big(n(b+\varepsilon)\big) \geqq \inf_{(\nu_n,\delta,\{v_i\})} M_r(\nu_n),$$

and $a_{[n(a-\varepsilon)]}\leqq a_{\nu_n}$, so the last probability of relation (3) is less than or equal to

$$P\{\inf_{(n(b+\varepsilon),\delta,\{v_i\})} M_r\big(n(b+\varepsilon)\big) > \varepsilon a_{[n(a-\varepsilon)]}\}. \tag{4}$$

Also $a_n=n^\alpha L(n)$ is slowly oscillating and it can be easily computed that $a_{[n(a-\varepsilon)]}/a_{[n(b+\varepsilon)]} \to ((a-\varepsilon)/(b+\varepsilon))^\alpha$ as $n\to\infty$. Thus, if we now choose a positive number ϱ so that $\varepsilon_0 = \varepsilon\big(((a-\varepsilon)/(b+\varepsilon))^\alpha-\varrho\big)$ is also positive, then there exists an

$n_2(\geqq n_1)$ so that for $n \geqq n_2$ the probability under (4) is bounded above by

$$P\{\inf_{(n(b+\varepsilon),\delta,\{v_i\})} M_r(n(b+\varepsilon)) > \varepsilon_0 a_{[n(b+\varepsilon)]}\}.$$

Since the sequence X_n is tight, therefore, for ε_0 and θ we can now choose δ and $n_0(\geqq n_2)$ such that the last probability is less than θ which, in turn, implies that the left hand side probability of (3) is less than η.

Turning now to the proof (i)* we let $\varepsilon, \eta, \theta, a, b, \varrho, n_1$ and n_2 be as in the proof of (ii)* above, and putting $d_0 = d(((a-\varepsilon)/(b+\varepsilon))^\alpha - \varrho)$ we get immediately:

$$(5) \qquad P\{\sup_{0 \leqq v \leqq \nu_n} |X(v)| > da_{\nu_n}\} \leqq P\{\sup_{0 \leqq v \leqq [n(b+\varepsilon)]} |X(v)| > d_0 a_{[n(b+\varepsilon)]}\} + 2\theta,$$

provided n is not less than n_2. Again, since the sequence X_n is tight, for θ we can choose $d=d^*$ so large that d_0 becomes large enough to ensure that the right hand side probability of the inequality of (5) is less than θ for every n. Consequently, for the given η there exist a d^* and n_2 so that

$$(6) \qquad P\{\sup_{0 \leqq v \leqq \nu_n} |X(v)| > d^* a_{\nu_n}\} < \eta, \quad n \geqq n_2.$$

Thus, the only question now whether such a d should also exist which would make relation (6) hold for all n. Since the space D is complete and separable, each single probability measure on $(D, \mathscr{D})$ is tight and so are, therefore, the ones induced by $X_1, X_2, \ldots, X_{n_2-1}$. Now the characterization theorem of the compact subsets of D (Theorem 14. 3, [3]) implies the existence of d_i so that

$$P\{\sup_{0 \leqq v \leqq \nu_i} |X(v)| > d_i a_{\nu_i}\} < \eta, \qquad i = 1, \ldots, n_2 - 1,$$

and relation (6) holds for every $n \geqq 1$ with $d = \max(d^*; d_1, \ldots, d_{n_2-1})$ instead of d^* in it. This completes the proof of Lemma.

Having proved this lemma, our programme now only requires us to be able to deduce the convergence of the finite dimensional distributions of X_{ν_n} from those of X_n. On the bases of recent literature, concerning the limiting distributions of sequences of r.v.'s with random indices, this can be done several ways. We are going to demonstrate two possibilities here which, we believe, are the most important ones available from the point of view of applications. They are based on a recent paper of GUIAŞU [12]. As to other ways of possible approach we refer to a forthcoming work of FISCHLER [11].

For a random function X of D let $T_X = \{t \in [0, 1]: P\{X(t) \neq X(t-)\} = 0\}$.

Theorem 1. *Let X and the sequence X_n be random functions of the space D, X_n having the form as in* (2). *Assume:*

(a) $\nu_n/n \xrightarrow{P} \nu$, *where the sequence ν_n and the r.v. ν are as in Lemma;*

(b) $X_n \xrightarrow{\mathscr{D}} X$;

(c) *For an arbitrary positive integer k, all arbitrary real numbers $c_1, c_2, \ldots, (c_k \neq 0)$ and arbitrary time points $t_1, t_2, \ldots, t_k \in T_X$, the random variables $Y_n = \sum_{i=1}^{k} c_i X_n(t_i)$ and $Y = \sum_{i=1}^{k} c_i X(t_i)$ satisfy (at every continuity point x of $P\{Y \leqq x\}$)*

$$\lim_{n\to\infty} P\{Y_n \leqq x \mid A\} = P\{Y \leqq x\},$$

for every $A \varepsilon \mathcal{K}_\nu$, $P\{A\} > 0$, where $\mathcal{K}_\nu$ is the σ-algebra generated by ν;

(d) *For every positive ε and η and every $A \in \mathcal{K}_\nu$, $P\{A\} > 0$, there exist a positive real number $c = c(\varepsilon, \eta)$ and a natural number $n_0 = n_0(\varepsilon, \eta, A)$ such that for every $n \geqq n_0$*

$$P\{\max_{n(1-c) \leqq m \leqq n(1+c)} |X(nt) - X(mt)| > \varepsilon a_n | A\} < \eta,$$

at every fixed $t \in T_X$. Then $X_{\nu_n} \xrightarrow{\mathcal{D}} X$.

Proof. In the light of our Lemma and Theorem B we only have to deal with the convergence of the finite dimensional distributions of X_{ν_n}. If we now observe

$$P\left\{\max_{n(1-c) \leqq m \leqq n(1+c)} \left|\sum_{i=0}^{k} c_i X(nt_i) - \sum_{i=1}^{k} c_i X(mt_i)\right| > \varepsilon a_n \middle| A\right\} \leqq$$

$$\leqq \sum_{i=1}^{k} P\{\max_{n(1-c) \leqq m \leqq n(1+c)} |X(nt_i) - X(mt_i)| > \varepsilon\, a_n/k\, |c_i|\, |A\},$$

then the conditions of Theorem 3 of GUIAŞU [12] are satisfied for the sequence $Y_n = \sum_{i=1}^{k} c_i X_n(t_i) = a_n^{-1} \sum_{i=1}^{k} c_i X(nt_i)$ if one also notes that Theorem 3 of [12] holds for a sequence of random variables Y_n of the form $Y_n = a_n^{-1} Z_n$ with its condition (C6) modified to the extent that in it one writes Z_i and Z_n instead of Y_i and Y_n respectively, and εa_n instead of ε. As a consequence of the Cramér—Wold device ([3], p. 49), our theorem is now proved.

Remarks. Lemma and Theorem 1 remain valid if the sequence ν_n and ν are such that $\nu_n/f(n) \xrightarrow{P} \nu$, where $f(n)$ are constants going to infinity. Incidentally, the Lemma itself would still hold if $a_n = n^\alpha L(n)$ is monotone decreasing or, if α is negative, independently again of a_n being increasing or decreasing. Condition (d) of Theorem 1 with $A = \Omega$ is the classical Anscombe condition [1].

Applications of Theorem 1

1) Let $\nu=\theta$, a positive constant, and $X_n(t)=S_{[nt]}/\sigma\sqrt{n}$ $(\sigma>0)$. Billingsley ([3], Theorem 17. 1) proves that $X_n \xrightarrow{\mathcal{D}} W$ implies $X_{\nu_n} \xrightarrow{\mathcal{D}} W$, where W is the Brownian motion on D. This result is a special case of Theorem 1, for when ν is a constant then $\mathcal{K}_\nu=\{\emptyset, \Omega\}$ and condition (c) with $A=\Omega$ is implied by (b). Also, condition (b) implies tightness of X_n, which, in turn, implies condition (d) with $A=\Omega$. This completes the proof of Billingsley's Theorem 17. 1 and the above procedure also shows that the assumption $X_n \xrightarrow{\mathcal{D}} W$ in his theorem can be replaced by $X_n \xrightarrow{\mathcal{D}} X$, where X is not necessarily the Brownian motion.

2) If the summands $\xi_1, \xi_2, \ldots$ of $S_{[nt]}$ are independent and identically distributed with zero mean and variance σ^2 and $\nu_n/f(n) \xrightarrow{P} \nu$, where ν is a positive r.v., then (again with $X_n(t)=S_{[nt]}/\sigma\sqrt{n}$) $X_{\nu_n} \xrightarrow{\mathcal{D}} W$, (Theorem 17. 2, [3]). This theorem of Billingsley is a generalization of Donsker's theorem, and it is also implied by Theorem 1 as follows: Donsker's theorem says that $X_n \xrightarrow{\mathcal{D}} W$, which implies that X_n is tight and this, in turn, ensures the tightness of X_{ν_n} via Lemma. Condition (c) of Theorem 1 is a mixing condition in the sense of Rényi [16], and for Y_n it can be verified exactly the same way as for one sum of independent, identically distributed r. v.'s. As to condition (d) we refer to Lemma 3 of Blum, Hanson and Rosenblatt [5], which implies that the conditional probability there can be considered only with $A=\Omega$, which then becomes the classical Anscombe condition for sums of independent, identically distributed r.v.'s, and, with this, Theorem 17. 2 of [3] now follows.

For a function $s(t)$ of D let $h(s(t)) = s(t)-ts(1)$. Then, with X_n as in 2) of Applications above and $\sigma^2=1$ we have $h(X_{\nu_n}) \xrightarrow{\mathcal{D}} W^0$, W^0 the Brownian bridge on D. In [7] we indicated a direct proof of this and used it to prove the random sample size Kolmogorov—Smirnov theorems. As already mentioned earlier, the weak convergence of the random sample size empirical process itself will be proved in [10].

The way we have proved Billingsley's Theorem 17. 2, [3], suggests the following version of Theorem 1.

Theorem 2. *Let X and the sequence X_n be random functions of the space D, X_n having the form as in* (2). *Assume conditions* (a), (b), (c) *of Theorem* 1 *and its condition* (d) *with $A=\Omega$. Assume also:*

(e) *For every positive ε and c and every $A\in\mathcal{K}_\nu$, $P\{A\}>0$, we have*

$$\limsup_{n\to\infty} P\{A_n^t|A\} = \limsup_{n\to\infty} P\{A_n^t\}$$

at every fixed $t\in T$, where A_n^t is the event

$$\{\max_{n(1-c)\leq m\leq n(1+c)} |X(nt)-X(mt)| > \varepsilon a_n\}.$$

Then $X_{\nu_n} \xrightarrow{\mathcal{D}} X$.

Proof. It is sufficient to note that condition (d) with $A=\Omega$ of Theorem 1 and condition (e) together imply condition (d) of Theorem 1. Thus Theorem 2 follows from Theorem 1.

We note that condition (e) of Theorem 2 holds any time each set in the tail σ-field of the sequence X_n has probability 0–1 (Theorem 2, [4]).

Applications of Theorem 2. Let $X_n(t)=S_{[nt]}/\sqrt{n}$, where the summands $\xi_1, \xi_2, \ldots$ of $S_{[nt]}$ have mean zero and variance one, but are not necessarily independent and identically distributed r.v.'s. SREEHARI (Theorems 2. 2 and 3. 1, [19]) proves that if conditions (1) $X_n \xrightarrow{\mathcal{D}} W$, where W is the Brownian motion on D, (2) condition (a) of Theorem 1 holds, (3) the random variables Y_n of condition (c) of Theorem 1 in terms of $X_n(t)=S_{[nt]}/\sqrt{n}$ satisfy

$$P\{Y_n \leqq x \mid A\} - P\{Y_n \leqq x\} \to 0, \quad \text{as} \quad n \to \infty,$$

for every $A \in \mathcal{K}_v$, $P\{A\}>0$, and (4) for every positive ε and c and every $A \in \mathcal{K}_v$, $P\{A\}>0$, we have (cf. Remark of [19], p. 437)

$$P\{A_n^t \mid A\} - P\{A_n^t\} \to 0, \quad \text{as} \quad n \to \infty,$$

where A_n^t is as in condition (e) of Theorem 2 in terms of $X_n(t)=S_{[nt]}/\sqrt{n}$, hold then $X_{v_n} \xrightarrow{\mathcal{D}} W$. This result is a special case of Theorem 2, for condition (1) implies that condition (3) is of the form of condition (c) of Theorem 1. Also, condition (4) above implies the form of condition (e) of Theorem 2. Thus Theorems 2. 2 and 3. 1 of Sreehari's paper [19] follow from Theorem 2 and our proof of it also shows that the assumption $X_n \xrightarrow{\mathcal{D}} W$ in his theorems can be replaced by $X_n \xrightarrow{\mathcal{D}} X$, where X is not necessarily the Brownian motion. Examples, satisfying the conditions (3) and (4) above and, therefore, also the relevant conditions of Theorem 2, are given in [19].

References

[1] F. J. ANSCOMBE, Large-sample theory of sequential estimation, *Proc. Cambridge Philos. Soc.*, **48** (1952), 600—607.

[2] P. BILLINGSLEY, Limit theorems for randomly selected partial sums, *Ann. Math. Statist.*, **33** (1962), 85—92.

[3] P. BILLINGSLEY, *Convergence of probability measures* (Wiley, New York, 1968).

[4] D. BLACKWELL and D. A. FREEDMAN, The tail σ-field of a Markov chain and a theorem of Orey, *Ann. Math. Statist.*, **35** (1964), 1290—1295.

[5] J. BLUM, D. HANSON and J. ROSENBLATT, On the central limit theorem for the sum of a random number of independent random variables, *Z. Wahrscheinlichkeitstheorie und verw. Gebiete*, **1** (1963), 389—393.

[6] M. CSÖRGŐ, On the strong law of large numbers and the central limit theorem for martingales, *Trans. Amer. Math. Soc.*, **131** (1968), 259—275.

[7] M. Csörgő and S. Csörgő, An invariance principle for the empirical process with random sample size, *Bull. Amer. Math. Soc.*, **76** (1970), 706—710.

[8] M. Csörgő and R. Fischler, Departure from independence: the strong law standard and random-sum central limit theorems, *Acta Math. Acad. Sci. Hung.*, **21** (1970), 105—114.

[9] S. Csörgő, On limit distributions of sequences of random variables with random indices, *Acta Math. Acad. Sci. Hung.* (to appear).

[10] S. Csörgő, On weak convergence of the empirical process with random sample size, *Acta Sci. Math.* (to appear).

[11] R. Fischler, Weak convergence limit theorems with randomized indices, *Rev. Roumaine Math. Pures Appl.* (to appear).

[12] S. Guiaşu, On the asymptotic distribution of the sequences of random variables with random indices, *Ann. Math. Statist.*, **42** (1971), 2018—2028.

[13] J. Mogyoródi, A central limit theorem for the sum of a random number of random variables, *Publ. Math. Inst. Hung. Acad. Sci.*, **7** (1962), 409—424.

[14] Yu. V. Prohorov, Convergence of random processes and limit theorems in probability theory, *Theor. Probability Appl.*, **1** (1956), 157—214.

[15] A. Rényi, On the central limit theorem for the sum of a random number of random variables, *Acta Math. Acad. Sci. Hung.*, **11** (1960), 383—393.

[16] A. Rényi, On mixing sequences of sets, *Acta Math. Acad. Sci. Hung.*, **9** (1958), 215—228.

[17] W. Richter, Übertragung von Grenzaussagen für Folgen von Zufälligen Grössen auf Folgen mit Zufälligen Indizes. *Theor. Probability Appl.*, **10** (1965), 82—93.

[18] A. V. Skorokhod, Limit theorems for stochastic processes, *Theor. Probability Appl.*, **1** (1956), 261—290.

[19] M. Sreehari, An invariance principle for random partial sums, *Sankhyā* (A), **30** (1968), 433—442.

(Received September 28, 1972)