The construction of linear order statistics with the help of pseudo-random numbers

By BÉLA GYIRES (Debrecen)

Introduction

The authors paper [3] was devoted to one of the important questions of modern mathematical statistics, to the limit theorems of linear order statistics. A part of the results obtained proves helpful in finding a method for the construction of order statistics with given limit distribution. In these constructions pseudo-random numbers can also play a role. The present paper is devoted to this question.

The paper falls into three parts. The first part is devoted to finding pseudo-random numbers suitable for our purpose.

In the second part we define linear order statistics and we formulate two theorems due to the author, upon which the results of the third part will be built.

The five theorems of the third part deal with the construction, based on pseudo-random numbers, of order statistics with given limit distribution.

In the hole paper, a fundamental role is played by the convergence in the weak sense of random variables. For different definitions of this notion see [2], 37—38, 58. As in [2], weak convergence will be denoted by $\Rightarrow$.

1. F. Riesz has given the following generalization of the well-known ergodic theorem of G. Birkhoff ([4], 224):

Let a measurable set Ω be given, of finite or infinite measure, the corresponding measure and integral being defined according to Lebesgue, or more generally, by means of a distribution of positive masses. That being the case, let us designate by T a point-transformation which is single-valued (but not necessarily one-to-one) from Ω onto itself; and let us suppose that T conserves measure in the sense that, E being a measurable set, TE its transform, and E' the set of points P whose images appear in TE, the sets E' and TE have the same measure. Then, if $f_1(P)$ is an integrable function and $f_k(P)=f_1(T^{k-1}P)$, the arithmetic mean of the functions $f_1, \dots, f_n$ converges almost everywhere, as $n\to\infty$, to an integrable function $\varphi(P)$ which is invariant (almost everywhere) under T.

Let us add finally that in the case where Ω is of finite measure, it follows by a term by term integration (which is permitted in this case because of the uniform integrability of the terms)

$$\int_\Omega \varphi(P)\,dP=\int_\Omega f_1(P)\,dP.$$

Moreover $\varphi(P)$ is almost everywhere a constant if and only if T is ergodic.

If Ω and T are known, then this theorem enables us to generate pseudo-random numbers, playing an important role in Monte Carlo methods.

Choosing $\Omega=[0, 1)$, J. N. FRANKLIN has shown [1] that for a natural number $N>1$ and a number $\Theta\in[0, 1)$ the transformation

$$Tx = Nx+\Theta-[Nx+\Theta] \tag{1}$$

satisfies the conditions of Riesz' ergodic theorem and it is ergodic. Thus he obtained the following result:

Let $f(x)$ be an arbitrary integrable function in the sense of Lebesgue. Let $x_0\in[0, 1)$, and let a sequence $x_0, x_1, \ldots$ be formed according to

$$x_n = Nx_{n-1}+\Theta-[Nx_{n-1}+\Theta], \tag{2}$$

where N is a fixed integer >1 and Θ is a fixed number from $[0, 1)$. Then for almost all x_0

$$\frac{1}{n}\sum_{k=0}^{n-1} f(x_k) \to \int_0^1 f(x)\,dx, \quad n\to\infty. \tag{3}$$

When $N=1$ in (1), in the classical paper ([5], 313) H. WEYL showed that for every value of the sequence (2) is equidistributed if and only if Θ is irrational. The sequence $x_0, x_1, \ldots$ is said to be equidistributed if, for every fixed $[a, b)\subset[0, 1)$

$$\frac{1}{n}\sum_{\substack{x_k\in[a, b)\\(k=0, 1, \ldots, n-1)}} 1 \to b-a, \quad \text{if} \quad n\to\infty.$$

H. Weyl has also shown ([5], 314) that if for $N=1$ the numbers (2) are equidistributed, then for any function $f(x)$ defined on $[0, 1)$ and integrable in the sense of Lebesgue (3) holds. (As a matter of fact, Weyl has established this result for Riemann-integrable functions, his proof, however, carries over without change to the Lebesgue-integrable case.)

Making suitable choices for $f_1(P)$ and for $f(x)$, one proves without difficulty that the numbers $T^{k-1}P$ $(k=1, 2, \ldots)$ occurring in Riesz' theorem and the numbers (2) of Franklin's theorem are equidistributed with probability 1 on the set Ω and on $[0, 1)$ respectively.

From the point of view of mathematical statistics these random numbers are only pseudo-random numbers. As a matter of fact, mathematical statistics requires from random numbers not only the validity of the above criterion of uniformity, but also statistical independence. This latter condition, however, is not satisfied for our random numbers, since the choice of the first number determines the subsequent ones.

2. Let the matrices with real elements

$$A_\nu = \begin{pmatrix} a_{11}^{(\nu)} \ldots a_{1\nu}^{(\nu)} \\ a_{21}^{(\nu)} \ldots a_{2\nu}^{(\nu)} \\ . \;\; \ldots \;\; . \end{pmatrix} \qquad (\nu = 1, 2, \ldots) \tag{4}$$

be given. Let us define the random variable $\eta_j^{(\nu)}$ $(j=1, 2, \ldots)$ on the matrix A_ν as follows:

If $\alpha_1, \ldots, \alpha_s$ $(1 \leqq s \leqq \nu)$ are pairwise different natural numbers and $k_1, \ldots, k_s$ are arbitrary different numbers from the numbers $1, \ldots, \nu$, then

$$P(\eta_{\alpha_1}^{(\nu)} = a_{\alpha_1 k_1}^{(\nu)}, \ldots, \eta_{\alpha_s}^{(\nu)} = a_{\alpha_s k_s}^{(\nu)}) = \frac{1}{\nu(\nu-1)\ldots(\nu-s+1)}.$$

From this definition we infer that $\eta_j^{(\nu)}$ is a uniformly distributed discrete random variable, namely

$$P(\eta_j^{\nu} = a_{jk}^{(\nu)}) = \frac{1}{\nu} \qquad (k = 1, \ldots, \nu).$$

Definition 1. By the linear order statistics generated by the random variables $\eta_1^{(m+n)}, \ldots, \eta_m^{(m+n)}$ *we mean the random variable*

$$\xi_{m,n} = \eta_1^{(m+n)} + \ldots + \eta_m^{(m+n)}$$

with a non-negative integer n.

Definition 2. By the linear order statistics generated by the matrices (4) *we mean the ensemble of the random variables*

$$\xi_{m,n} \qquad (m = 1, 2, \ldots;\ n = 0, 1, 2, \ldots).$$

Definition 3. The linear order statistics generated by the matrices (4) *are asymptotic, if for any natural number m there exists a random variable* ξ_m *such that*

$$\xi_{m,n} \Rightarrow \xi_m, \quad n \to \infty,$$

Definition 4. The linear order statistics generated by the matrices (4) *are doubly asymptotic, if there exists a random variable* ξ *such that*

$$\xi_{m,n} \Rightarrow \xi, \quad \text{if} \quad n \to \infty, \quad m \to \infty.$$

We are going also to speak about asymptotically ξ_m-distributed $(m=1, 2, \ldots)$, and about doubly asymptotically ξ-distributed linear order statistics respectively.

Clearly, the asymptotic order statistics generated by the matrices (4) are doubly asymptotically distributed if and only if $\xi_m \Rightarrow \xi$, $m \to \infty$.

Let $\mathscr{E}_1$ be the set of the uniformly distributed discrete random variables and if $\eta^{(\nu)} \in \mathscr{E}_1$, the index ν denotes that the probabilities belongs to $\eta^{(\nu)}$ are $\frac{1}{\nu}$. Let $\mathscr{E}_2$ be the set of those random variables η, which can be represented in the form

$$\eta^{(\nu)} \Rightarrow \eta, \quad \nu \to \infty, \quad \eta^{(\nu)} \in \mathscr{E}_1.$$

The author has proved the following two theorems ([3], theorems 1.5., 1.6.):

Theorem A. *If* $\eta_j \in \mathscr{E}_2$ $(j = 1, 2, \ldots)$, *i.e. if*

$$\eta_j^{(\nu)} \Rightarrow \eta_j, \quad \nu \to \infty, \quad \eta_j^{(\nu)} \in \mathscr{E}_1,$$

then the linear order statistics generated by the matrices (4) *determined by the values of the random variables* $\eta_j^{(\nu)}$ $(\nu=1, 2, \ldots;\ j=1, 2, \ldots)$, *are asymptotically equal to the sums of the random variables* $\eta_1, \ldots, \eta_m$ $(m=1, 2, \ldots)$ *independent from each other.*

Theorem B. *If* $\eta_j \in \mathcal{E}_2$ $(j = 1, 2, \ldots)$, *i.e. if*

$$\eta_j^{(\nu)} \Rightarrow \eta_j, \quad \nu \to \infty, \quad \eta_j^{(\nu)} \in \mathcal{E}_1,$$

then the linear order statistics generated by the matrices (4) *determined by the values of the random variables* $\eta_j^{(\nu)}$ $(\nu=1, 2, \ldots;\ j=1, 2, \ldots)$ *are doubly asymptotically of distribution* η *if and only if*

$$\eta_1 + \ldots + \eta_m \Rightarrow \eta, \quad m \to \infty.$$

3. In this section we shall construct linear order statistics having given limit distributions with the help of pseudo-random numbers.

Theorem 1. *Let* T_j *be a measure-preserving ergodic transformation of the measurable set* $\Omega \subset R_1$ *with finite measure into itself or into a part of itself, and let*

$$\int_\Omega dx = 1.$$

Let, moreover, $f_j(x)$ *be an function being Lebesgue integrable on the set* Ω. *Then the linear order statistics generated by the matrix* (4) *formed with the elements*

$$a_{jk}^{(\nu)} = f_j(T_j^{k-1} x_j), \quad x_j \in \Omega \quad (j, k = 1, 2, \ldots)$$

are asymptotically

$$f_1(\zeta_1) + \ldots + f_m(\zeta_m) \qquad (m = 1, 2, \ldots)$$

distributed with probability one, where $\zeta_1, \ldots, \zeta_m$ *are independent, on the set* Ω *equidistributed random variables.*

Proof. Let $x \in \Omega$ and T be a measure-preserving ergodic transformation of Ω into itself or into a subset. Let $f(x)$ be Lebesgue-integrable on the set Ω. Let the random variable η_n be defined by

$$P(\eta_n = f(T^{k-1}x)) = \frac{1}{n} \qquad (k = 1, \ldots, n).$$

In order to prove our theorem, by Theorem A. it is sufficient to show that

$$P(\eta_n \Rightarrow f(\zeta),\ n \to \infty) = 1, \tag{5}$$

where ζ is a random variable equidistributed on the set Ω.

Together with $f(x)$ the functions $\cos(tf(x))$ and $\sin(tf(x))$, $t \in R_1$ are also Lebesgue-integrable on Ω; thus by the ergodic theorem of F. Riesz the relations

$$\frac{1}{n} \sum_{k=1}^{n} \{\cos[tf(T^{k-1}x)] + i \sin[tf(T^{k-1}x)]\} = \frac{1}{n} \sum_{k=1}^{n} \exp[itf(T^{k-1}x)] \to$$
$$\to \int_\Omega \cos[tf(x)]\,dx + i \int_\Omega \sin[tf(x)]\,dx = \int_\Omega e^{itf(x)}\,dx, \quad n \to \infty, \quad t \in R_1$$

is satisfied with probability one on the set Ω, and this is just the statement expressed by formula (5).

At the same way by the theorem of J. N Franklin and H. Weyl respectively we can proof the following two theorems:

Theorem 2. *If $f_j(x)$ is a Lebesgue-integrable function defined on the interval* $[0, 1)$, $x_j \in [0, 1)$, $\Theta_j \in [0, 1)$, $N_j > 1$ *is a natural number,*

$$x_{jk} = N_j x_{jk-1} + \Theta_j - [N_j x_{jk-1} + \Theta_j] \qquad (k = 1, 2, \dots;\ x_{j0} = x_j),$$

then the linear order statistics generated by the matrices (4) *formed with the help of the elements*

$$a_{jk}^{(\nu)} = f_j(x_{jk-1})$$

are asymptotically with probability one

$$f_1(\zeta_1) + \dots + f_m(\zeta_m) \qquad (m = 1, 2, \dots)$$

distributed, where $\zeta_1, \dots, \zeta_m$ are independent random variables equidistributed in the interval $[0, 1)$.

Theorem 3. *If $f_j(x)$ is a Lebesgue-integrable function defined on the interval* $[0, 1)$, $x_j \in [0, 1)$, $\Theta_j \in [0, 1)$, *irrational,*

$$x_{jk} = x_{jk-1} + \Theta_j - [x_{jk-1} + \Theta_j] \qquad (k = 1, 2, \dots;\ x_{j0} = x_j),$$

then the linear order statistics generated by the matrices (4) *having elements*

$$a_{jk}^{(\nu)} = f_j(x_{jk-1})$$

are asymptotically

$$f_1(\zeta_1) + \dots + f_m(\zeta_m) \qquad (m = 1, 2, \dots)$$

distributed, where $\zeta_1, \dots, \zeta_m$ are independent random variables, uniformly distributed in the interval $[0, 1)$.

On the basis of our Theorem B already quoted, it is also possible to construct with the help of the three theorems just formulated linear order statistics which are (with probability one) doubly asymptotically of given distribution.

Let $\mathscr{C}$ denote the set of those continuous distribution functions $F(x)$, which are strictly monotonely increasing in some interval (a, b), and satisfy $F(a)=0$, $F(b)=1$. $a=-\infty$, $b=\infty$ is also possible. Let the inverse of $y=F(x)$, existing in the interval (a, b) be $F^{-1}(y)$.

Theorem 4. *If the random variable ξ has an expectation and its distribution function $F(x)$ belongs to $\mathscr{C}$, the $F^{-1}(x)$ is Lebesgue-integrable on* $[0, 1]$. *If, moreover $\Theta \in [0, 1)$ is irrational and*

$$x_k = x_{k-1} + \Theta - [x_{k-1} + \Theta], \quad x_0 \in [0, 1), \tag{6}$$

then the random variables η_n defined by

$$P\big(\eta_n = F^{-1}(x_k)\big) = \frac{1}{n} \qquad (k = 0, 1, \dots, n-1)$$

weakly converge for $n \to \infty$ to the random variable ξ, i.e. $\xi \in \mathscr{E}_2$.

PROOF. In view of

$$\int_0^1 F^{-1}(y)\,dy = \int_{-\infty}^{\infty} x\,dF(x),$$

$F^{-1}(y)$ is Lebesgue-integrable on [0, 1]. Since moreover, $\zeta = F(\xi)$ is uniformly distributed in the interval [0, 1], we have $\eta_n \Rightarrow F^{-1}(\zeta) = \xi$, $n \to \infty$, by what has been said in the proof of Theorem 1.

Theorem 4 implies the following corollaries, which we have obtained earlier by another method ([2], Corollary 2.2., 2.3):

Corollary 1. If the random variable ξ with an absolutely continuous distribution function has an expectation and its density function is positive on an interval (a, b) (with possibly, $a = -\infty$ and or $b = \infty$) and zero outside this interval, then $\xi \in \mathscr{E}_2$.

Corollary 2. The random variables with normal, Chi-square, Student $(n > 1)$, exponential distribution are elements of the set $\mathscr{E}_2$.

By basing our considerations on Theorem 2 rather than on Theorem 3, we are able to obtain a theorem analogous to Theorem 4.

Theorem 5. *If the random variable ξ has an expectation and its distribution function $F(x)$ belongs to $\mathscr{C}$, $\Theta \in [0, 1)$, $N > 1$ a natural number,*

$$x_k = Nx_{k-1} + \Theta - [Nx_{k-1} + \Theta], \quad x_0 \in [0, 1),$$

and if the random variable η_n is defined by

$$P[\eta_n = F^{-1}(x_k)] = \frac{1}{n} \qquad (k = 0, 1, \ldots, n-1),$$

then

$$P(\eta_n \Rightarrow \xi,\ n \to \infty) = 1,$$

i.e.

$$\xi \in \mathscr{E}_2.$$

With the help of Theorems 3. and 4. we are able to construct, making use of the pseudo-random numbers (6), also linear order statistics which are asymptotically Chi-square distributed. To construct such statistics is advantageous also in view of the well-known fact that the sum of independent Chi-square distributed random variables is again Chi-square distributed, and the degrees of freedom are added up.

Let $F^{-1}(x)$ be the inverse of the function

$$F(y) = \frac{1}{\sqrt{2\pi}} \int_0^y \frac{e^{-(x/2)}}{\sqrt{x}}\,dx, \quad y > 0.$$

$F(y)$ is the distribution function of the Chi-square distributed random variable with degree of freedom one, in the case $y > 0$.

Let ξ and η be continuously distributed random variables, and

$$\xi_1, \ldots, \xi_m, \quad \eta_1, \ldots, \eta_n$$

samples with respect to these random variables arranged in a monotonically increasing order, with $\xi_1, \dots, \xi_m$ having the rank $r_1, \dots, r_m$ respectively. Since by Theorems 3. and 4. the random variable

$$F^{-1}(x_{r_1}) + \dots + F^{-1}(x_{r_m})$$

formed with the help of the pseudo-random numbers (6) is for $n \to \infty$ Chi-square distributed with m degree of freedom, with the help of Chi-square tables and on the basis of a great number of observations we are able to reach a decision concerning the adoption or the rejection of the hypothesis

$$H_0 : P(\xi < x) = P(\eta < x).$$

References

[1] J. N. FRANKLIN, On the equidistribution of pseudo-random numbers. *Quart. Appl. Math.* **16** (1958), 183—188.
[2] B. V. GNEDENKO—A. N. KOLMOGOROV, Limit distribution of sums of independent random variables. (Russian.) *Moscow,* 1949. English transl., Addison—Wesley, *Cambridge, Mass.,* 1954.
[3] B. GYIRES, On limit distribution theorems of linear order statistics. *Publ. Math. (Debrecen)* **21** (1974), 95—112.
[4] F. RIESZ, Sur la théorie ergodique. *Comm. Math. Helv.* **17** (1945), 221—239.
[5] H. WEYL, Über die Gleichverteilung von Zahlen mod. Eins. *Math. Ann.* **77** (1916), 313—352.

(Received June 2, 1973.)