Journal of Mathematical Sciences, Vol. 88, No. 1, 1998

ON THE SUMS OF COORDINATE RANDOM VARIABLES FOR A CERTAIN CLASS OF SUBSTITUTIONAL DYNAMICAL SYSTEMS

A. N. Livshits

UDC 519.2

In this note the question about the limit distributions for sequences of the sums of the coordinate random variables, generated by some adic and substitutional symbolic dynamical systems is studied. The weak convergence is proved for some classes of subsequences of a sequence of appropriate normed sums, by the limits of which all the limit distributions are exhausted. Bibliography: 5 titles.

In [1], primitive substitutions were considered such that the spectra of their matrices obey some relations. In [1], the asymptotic behavior of the sums of certain functions of coordinates for fixed points of substitutions from this class was described. Our purpose is to prove the existence of limit distributions for some sequences of properly normed sums of coordinate random variables corresponding to these functions [2, 3] on the phase space of dynamical systems (substitutional or adical [4]) generated by the substitutions considered in [1].

The result of [1] is as follows. Let $Z_* = \{u_i\}_{i=1}^k$ be an alphabet, and $\mathcal{A} = \{A_i\}_{i=1}^k$ be a collection of words in it. Let $\omega_{\mathcal{A}}$ be a map of a set $(Z_*)^+ = \bigcup_{l>0} Z_*^l$ (considered as a set of nonempty words) into itself: $\omega_{\mathcal{A}}(u_{i_1}\dots u_{i_l}) = A_{i_1}\dots A_{i_l}$. The map $\omega_{\mathcal{A}}|_{Z_*}$ is called a substitution. The corresponding matrix $(g_{ij}) : g_{ij} = |A_i|_j$, $1 \le i,j \le k$, is denoted by $G_{\mathcal{A}}$ [4]. Suppose that $G_{\mathcal{A}}$ is primitive and (see [1]) its eigenvalues $\lambda_1, \dots, \lambda_s$ (some of them may be multiple) obey the relation $\lambda > \lambda_1 > \max_{2\le i\le s}|\lambda_i|$. Now if $f : Z_* \to R$ is a function such that $\sum f(i)\mu_i = 0$, where $(\mu_i)_{i=1}^k$ is an eigenvector of ${}^tG_{\mathcal{A}}$ corresponding to the Perron–Frobenius eigenvalue λ, $\alpha+1$ is the multiplicity of the root λ_1 of the minimal polynomial of $G_{\mathcal{A}}$, and w, l', i, $1 \le i \le k$, are such that $\omega_{\mathcal{A}}^{l'}(u_i) = u_i w$, then there exists a real-valued Hölder nowhere differentiable function $F_{f,i}(x)$ such that if $\omega_{\mathcal{A}}^{l'\cdot\infty}(u_i) = a_0 a_1 \dots$, then

$$\sum_{i=0}^{N-1} f(a_j) = N^{\frac{\lg\lambda_1}{\lg\lambda}} (\log N)^\alpha F_{f,i}(N) + \Psi_{f,i}(N),$$

where $|\Psi_{f,i}(N)| = o\left(N^{\frac{\lg\lambda_1}{\lg\lambda}} (\log N)^\alpha\right)$, and for any x we have $F_{f,i}(\lambda^{l'}x) = F_{f,i}(x)$.

Now let an arbitrary positive integer M be fixed. Consider a discrete random variable ξ_M with finite set of values and with probabilities

$$P_M(\beta) = \mathrm{Fr}\left\{ l \ge 0 \;\middle|\; \sum_{j=l}^{j=l+M-1} f(a_j) \times \left(M^{\frac{\log\lambda_1}{\log\lambda}} (\log M)^\alpha\right)^{-1} = \beta \right\}$$

(Fr means frequency). Since the matrix $G_{\mathcal{A}}$ is primitive, the frequencies used in the definition always exist and do not depend on the choice of (i,l). Moreover, there exist frequencies of appearance for every fixed word of length M [5]. Denote $M^{\frac{\log\lambda_1}{\log\lambda}}(\log M)^\alpha$ by H_M, and for an arbitrary word $B = b_1 \dots b_j$ let $f(B) = \sum_{l=1}^{j} f(b_l)$. Because of the single-ergodicity of substitutional dynamical systems, it is clear that this random variable is just the sum of M coordinate random variables connected with a function proportional to the function f of the null coordinate of the point of the phase space of a substitutional or adical dynamical system. This construction is mentioned in [2], and it is considered at greater length in [3] for the Rudin–Shapiro substitution. The question on the set of all limit points of the set $\{P_M\}$ (in the weak topology) is of natural interest. An answer is given by the theorem below. Let

$$\Pi = \left\{ l \;\middle|\; \exists\, a \in Z_*,\ w \in (Z_*)^+,\ \omega_{\mathcal{A}}^l(a) = aw \quad \text{or} \quad \exists\, a' \in Z_*,\ w' \in (Z_*)^+ : \omega_{\mathcal{A}}^l(a') = w'a' \right\}$$

Translated from *Zapiski Nauchnykh Seminarov POMI*, Vol. 216, 1994, pp. 104–109. Original article submitted December 20, 1993.

and R be the greatest common divisor of all numbers from Π. Given an arbitrary positive number γ, a sequence of natural numbers $M_l \to \infty$ is called a sequence of type γ if there exists a sequence of natural numbers p_l such that $M_l \lambda^{-rp_l} \to \gamma$ as $l \to \infty$. It is evident that if $M'_l \to \infty$, then there exist γ and a sequence of natural numbers l_k such that $\{M'_{l_k}\}$ is a sequence of type γ, hence our theorem describes the limit points of the set $\{P_M\}$ completely, as is seen from its formulation.

Theorem. *If $\{M_k\}$ is a sequence of type γ, then the sequence of distributions P_{M_k} weakly converges. The limit distribution is completely specified by γ.*

Proof. Let $r' \in \Pi$. Suppose that for some a' we have $\omega_A^{r'}(a') = w'a'$ (the case $\omega_A^{r'}(a) = aw$ can be considered in a similar way). Consider a sequence $\dots b_{-s} \dots b_{-1} = \omega_A^{r'\infty}(a')$. As in the theorem from [1], one can prove the existence of a function $F^*_{f,i}$ such that

$$\sum_{j=-N}^{-1} f(b_j) = H_N F^*_{f,i'}(N) + \Psi^*_{f,i'}(N), \quad \text{where} \quad |\Psi^*_{f,i'}(N)| = o(H_N), \quad F^*_{f,i}(\lambda^{r'} x) = F^*_{f,i}(x).$$

It is clear that there exists a natural number R such that if for some $M > 1$ we have $M = \gamma_0 \lambda^{lr'}$, $1 \le \gamma_0 < \lambda^{r'}$, then for any $a \in Z_*$ the inequality $M < \left|\omega_A^{(l+R)r'} a\right|$ holds. Let such a number R be fixed. Denote $(l+R)r'$ by $q = q(M)$. Consider the following two systems $\{\overline{P}^{r'}_{M,b}\}$, $\{\overline{P}^{r'}_{M,bb'}\}$, $b, b' \in Z_*$ of discrete measures on the real line:

$$\overline{P}^{r'}_{M,b}(\beta) = \frac{\#\{(w, w', w'') : \omega_A^q(b) = ww'w'',\ |w'| = M,\ f(w') = \beta\}}{\left|\omega_A^q(b)\right| - M + 1},$$

$$\overline{P}^{r'}_{M,bb'}(\beta) = \frac{\#\{\,(w, v, v', w') : v, v' \text{ nonempty}; |v| + |v'| = M,\ f(vv') = \beta, wv = \omega_A^q(b); v'w' = \omega_A^q(b')\,\}}{M-1}.$$

It is known [5] that the frequency of appearance of the symbol b in $\omega_A^{r'\infty}(a')$ is equal to μ_b. Let $\mathrm{Fr}(bb')$ stand for the frequency of the pair bb' in the sequence $\omega_A^{r'\infty}(a')$. (This frequency is the same for every trajectory or half-trajectory of the dynamical system.) Let

$$\begin{aligned} D_M &= \sum_{u_i \in Z_*} \mu_i\Big(\left|\omega_A^q(u_i)\right| - M + 1\Big) + (M-1) \sum_{b,b' \in Z_*} \mathrm{Fr}(bb') \\ &= \sum \mu_i \left|\omega_A^q(u_i)\right| = \lambda^q. \end{aligned}$$

For the probability measure P_M corresponding to a random variable ξ_M, we have (we suppose that B is a subset of the real line):

$$P_M(H_M^{-1} B) = D_M^{-1}\Bigg(\sum_{u_i \in Z_*} \mu_i\Big(\left|\omega_A^q(u_i)\right| - M + 1\Big)\, \overline{P}^{r'}_{M,u_i}(B) + +(M-1) \sum_{b,b' \in Z_*} \mathrm{Fr}(bb')\, \overline{P}^{r'}_{M,bb'}(B)\Bigg). \quad (1)$$

Equality (1) is a consequence of the relations

$$\dots b_{-s} \dots b_{-1} = \omega_A^{r'\infty}(a') = \dots \omega_A^q(c_{-s}) \dots \omega_A^q(c_{-1}),$$

where the corresponding frequencies for the sequence $\dots c_{-s} \dots c_{-1}$ are the same as in formula (1), the right-hand side of which corresponds to enumeration of possible variants of arrangement of the symbols $a_l \dots a_{l+M-1}$ from the definition of ξ_M in the words $\omega_A^q(c_{-i})$. Let φ be a finite continuous function on the real line with modulus of continuity $\Omega(\delta)$. Let $\overline{\Psi}^*_{f,i'}(N)$ stand for $\sup\limits_{0<x<N} \left|\Psi^*_{f,i'}(x)\right|$. We also suppose that $\mathrm{Fr}(bb') > 0$.

Lemma. *There exist continuous functions $g_b^{r'}(x)$, $g_{bb'}^{r'}(x)$ defined on the interval $[1, \lambda^{r'})$ and such that the following inequalities hold:*

$$\left| \int_{-\infty}^{\infty} \varphi(H_M^{-1} x)\, d\overline{P}_{M,b}^{r'} - g_b^{r'}(\gamma_0) \right| \le \Omega\left(\frac{c_1 \overline{\Psi}_{f,i'}^*(c_2 M)}{H_M} \right) + \Omega\left(\frac{c_3}{\log M} \right) + c_4 \frac{H_M}{M} + c_5\, \Omega\left(\frac{c_6}{M} \right),$$

$$\left| \int_{-\infty}^{\infty} \varphi(H_M^{-1} x)\, d\overline{P}_{M,bb'}^{r'} - g_{bb'}^{r'}(\gamma_0) \right| \le \Omega\left(\frac{c_1^* \overline{\Psi}_{f,i'}^*(c_2^* M)}{H_M} \right) + \Omega\left(\frac{c_3^*}{\log M} \right) + c_4^* \frac{H_M}{M} + c_5^*\, \Omega\left(\frac{c_6^*}{M} \right).$$

where $c_1, \dots, c_6$; $c_1^*, \dots, c_6^*$ are some constants.

Proof. We shall prove the existence of $g_{bb'}$ and the second estimate. The first case can be considered in a similar way. Select the minimal number $j > 1$, for which $c_{-j-1} c_{-j} = bb'$. Denote the sum $\sum_{p=-j}^{-1} |\omega_A^q(c_p)|$ by $T_q = T_{q(M)}$. Then

$$I_M = \int_{-\infty}^{\infty} \varphi(H_M^{-1} x)\, d\overline{P}_{M,bb'}^{r'} = \frac{1}{M-1} \sum_{t=-T_q-M+1}^{-T_q-1} \varphi\Big(H_m^{-1} \sum_{t'=t}^{t'=t+M-1} f(b_{t'}) \Big).$$

Since j does not depend on q, there exists a constant $E > 0$ such that $T_q + M < EM$. But

$$\begin{aligned} \sum_{t'=t}^{t'=t+M-1} f(b_{t'}) &= \sum_{t'=t}^{t'=-1} f(b_{t'}) - \sum_{t'=t+M}^{t'=-1} f(b_{t'}) \\ &= H_{-t} F_{f,i'}^*(-t) - H_{-t-M} F_{f,i'}^*(-t-M) + \Psi_{f,i'}^*(-t) - \Psi_{f,i'}^*(-t-M). \end{aligned}$$

Hence,

$$\left| I_M - \frac{1}{M-1} \sum_{t=-T_q-M+1}^{t=-T_q-1} \varphi\left(\frac{H_{-t}}{H_M} F_{f,i'}^*(-t) - \frac{H_{-t-M}}{H_M} F_{f,i'}^*(-t-M) \right) \right| \le \Omega\left(\frac{c_1^* \overline{\Psi}_{f,i'}^*(c_2^* M)}{H_M} \right).$$

Then

$$\left| \frac{H_{-t}}{H_M} - \left(\frac{-t}{M} \right)^{\frac{\lg \lambda_1}{\lg \lambda}} \right| + \left| \frac{H_{-t-M}}{H_M} - \left(-\frac{t}{M} - 1 \right)^{\frac{\lg \lambda_1}{\lg \lambda}} \right| < c_3^* (\log M)^{-1}.$$

The value $T_q / \lambda^{lr'}$ has a limit Q as $l \to \infty$, and the deviation from it is of order H_{T_q}/T_q or, what is the same, H_M/M, because, c being fixed, the sequence $|\omega_A^q(c)|$ is a recursion, whereas the corresponding polynomial is a characteristic polynomial for G_A. Taking into account the properties of $F_{f,i'}^*$ we get $F_{f,i'}^*(-t) = F_{f,i'}^*\left(-\frac{t}{\lambda^{lr'}} \right)$. Thus, it remains to estimate the sum

$$\frac{1}{M-1} \sum_{t=-T_q-M+1}^{-T_q-1} \varphi\left(\left(\frac{-t}{M} \right)^{\frac{\lg \lambda_1}{\lg \lambda}} F_{f,i'}^*\left(\frac{-t}{\lambda^{lr'}} \right) - \left(-\frac{t}{M} - 1 \right)^{\frac{\lg \lambda_1}{\lg \lambda}} F_{f,i'}^*\left(-\frac{t}{\lambda^{lr'}} - \frac{M}{\lambda^{lr'}} \right) \right).$$

This sum (with an error roughly estimated by the last summand in the lemma) can be replaced by the integral

$$I(\gamma_0) = \int_{-Q-\gamma_0}^{-Q} \varphi\left(\left(\frac{-y}{\gamma_0} \right)^{\frac{\lg \lambda_1}{\lg \lambda}} F_{f,i'}^*(-y) - \left(-\frac{y}{\gamma_0} - 1 \right)^{\frac{\lg \lambda_1}{\lg \lambda}} F_{f,i'}^*(-y-\gamma_0) \right) dy. \tag{2}$$

The value of this integral is taken as the value of the function $g_{bb'}^{r'}(\gamma_0)$.

Considering the first case of the lemma, one should slightly change limits of integration. Let $j' > 1$ stand for the minimal number for which $c_{-j'} = b$, T_q' stand for the sum $\sum_{p=-j+1}^{-1} |\omega_A^q(c_p)|$, Q' stand for $\lim_{l\to\infty} \frac{T_q'}{\lambda^{lr'}}$, and Δ stand for $\lim_{l\to\infty} \frac{|\omega_A^q(b)|}{\lambda^{lr'}}$. We get for $g_b^{r'}(\gamma_0)$ the limits of integration from $-Q' - \Delta$ to $-Q' - \gamma_0$. The lemma is proved.

It is easily seen from (2) that the modulus of continuity of the functions $g_{bb'}^{r'}(\gamma_0)$ $(g_b^{r'}(\gamma_0))$ can be in a natural way majorized by a function which can be expressed in terms of moduli of continuity of φ and $F_{f,i'}^*$.

Now using formula (1) and the fact that the coefficients on the right-hand side of (1) for large M are also close to the values of some functions of γ_0, one can easily prove the following assertion. There exist a positive sequence δ_M ($\delta_M \to 0$ as $M \to \infty$) and a function $\delta(\varepsilon)$ ($\delta(\varepsilon) \to 0$ as $M \to \infty$) such that if $r' \in \Pi$ and $M_2/M_1 = (1+\varepsilon)\lambda^{r'l'}$, $l' > 0$ is an integer, $M_1 > M$, then

$$\left| \int \varphi \, dP_{M_1} - \int \varphi \, dP_{M_2} \right| < \delta_M + \delta(\varepsilon).$$

Since r is the greatest common divisor of a finite collection of elements of Π, the theorem is proved.

Translated by V. Sudakov.

REFERENCES

1. J.-M. Dumont and A. Thomas, "Systèmes de numeration et fonctions fractales relatifs aux substitutions," *Theor. Comput. Sci.*, **65**, No. 2, 153–169 (1989).
2. A. N. Livshits, "On a stationary sequence of random variables," Preprint POMI P-5-92 (1992).
3. A. N. Livshits, "On the limit distributions and the asymptotics of extremal values for some sequences of random variables," *Zap. Nauchn. Semin. POMI*, **216**, 86–103.
4. A. N. Livshits, "A sufficient condition for the weak intermixing of substitutions and stationary adical transformations," *Mat. Zametki*, **44**, No. 6, 785–793 (1988).
5. M. Queffelec, "Substitution dynamical systems—spectral analysis," *Lect. Notes Math.*, **1294** (1987).