2014

Strong power and subexponential laws for an ordered list of trajectories of a Markov chain

Vladimir V. Bochkarev
vbockarev@mail.ru

Eduard Yu. Lerner

Follow this and additional works at: http://repository.uwyo.edu/ela

Recommended Citation
DOI: https://doi.org/10.13001/1081-3810.1917

This Article is brought to you for free and open access by Wyoming Scholars Repository. It has been accepted for inclusion in Electronic Journal of Linear Algebra by an authorized editor of Wyoming Scholars Repository. For more information, please contact scholcom@uwyo.edu.
STRONG POWER AND SUBEXPONENTIAL LAWS FOR AN
ORDERED LIST OF TRAJECTORIES OF A MARKOV CHAIN∗

VLADIMIR V. BOCHKAREV† AND EDUARD YU. LERNER‡

Abstract. Consider a homogeneous Markov chain with discrete time and with a finite set of states $E_0, \ldots, E_n$ such that the state $E_0$ is absorbing and states $E_1, \ldots, E_n$ are nonrecurrent. The frequencies of trajectories in this chain are studied in this paper, i.e., “words” composed of symbols $E_1, \ldots, E_n$ ending with the “space” $E_0$. Order the words according to their probabilities; denote by $p(t)$ the probability of the $t$th word in this list. As was proved recently, in the case of an infinite list of words, in the dependence of the topology of the graph of the Markov chain, there exists either the limit $\ln p(t)/\ln t$ as $t \to \infty$ or that of $\ln p(t)/t^{1/D}$, where $D \in \mathbb{N}$ (weak power and subexponential laws). As appeared, in the latter case the decreasing order of the function $p(t)$ is always subexponential (the strong subexponential law). In the first case, this paper describes necessary and sufficient conditions of the power order (the strong power law). These conditions are fulfilled, in particular, if the graph of the Markov chain that corresponds to states $E_1, \ldots, E_n$ is strongly connected.

Key words. Substochastic matrices, Markov chains, Directed graphs, Strong power laws.

AMS subject classifications. 15B48, 60J10.

1. Introduction. The nature of the power law and the spheres of its applicability has been an interest of mathematicians in recent decades [6, 8, 18]. For real networks, one has proposed several models describing the occurrence of the power law; the most known one is the preferential attachment model [1]. In linguistics, mechanisms of the occurrence of Zipf and Heaps laws were thoroughly studied in the time of B. Mandelbrot [15, 16]. Papers containing empirical studies and mathematical models also appear regularly (see, for example, [14] and references therein; for the mathematical motivation of this paper see [7]). However, there are no commonly accepted explanations of the fact that in reality with some values of parameters, the power law does not adequately describe processes under consideration [6]. Here we try to answer this question, considering probabilities of the occurrence of various trajectories in a homogeneous Markov chain.

∗Received by the editors on December 31, 2013. Accepted for publication on July 10, 2014. Handling Editor: Bryan L. Shader. This work was supported by the Russian Foundation for Basic Research, Grant N 12-06-00404-a.
†Institute of Physics, Kazan Federal University, Kremlievskaya str.18, Kazan, 420018, Russia (vbockarev@mail.ru).
‡Institute of Computer Mathematics and Information Technologies, Kazan Federal University, Kremlievskaya str.18, Kazan 420018, Russia (eduard.lerner@gmail.com).
Our model has occurred when studying a huge data set of the Google Books repository [17]. Usually one describes frequencies of word occurrences with the help of power law asymptotics [2]. But note that the power law is irrelevant in hieroglyphic scripts [14].

As the initial model explaining the power law of the decrease in frequencies of the occurrences of English words, we consider the model of the word generation process consisting in the sequential independent random addition of various symbols (letters and the space), each of which has a fixed probability (the monkey model). This model has a long history, but only recently, the power character of the asymptotics of the sorted list of word frequencies has been strictly justified [3, 7].

In this paper, we study one natural generalization of this model, namely, the model with the Markov connection of neighboring symbols. Such model was studied by B. Mandelbrot [16]; however, he has mainly considered a particular case of the occurrence of the power asymptotics. As appeared, in the dependence of the matrix of transition probabilities, the ordered list of frequencies of all possible trajectories of a Markov chain can have essentially different asymptotics.

Thus, let us consider a homogeneous Markov chain with discrete time and with a finite set of states $E_0, \ldots, E_n$ such that

\begin{equation}
\text{the state } E_0 \text{ is absorbing,}
\end{equation}

\begin{equation}
\text{states } E_1, \ldots, E_n \text{ are nonrecurrent.}
\end{equation}

The goal of this work is to study frequencies of trajectories in this chain, i.e., “words” composed of symbols $E_1, \ldots, E_n$ ending with the “space” $E_0$.

Let us order words (trajectories) according to their probabilities; denote by $p(t)$ the probability of the $t^{th}$ word in this list. In this paper, we prove that in a typical case the asymptotics of the function $p(t)$ has a power character, and define its exponent from the matrix of transition probabilities of the chain minus the absorbing state. If this matrix is reducible, then with some specific values of transition probabilities, the power asymptotics become logarithmic. But if this matrix is rather sparse, then probabilities quickly decrease; namely, the rate of the asymptotics has a subexponential order. One can easily calculate the order of the potentiated power function. The calculation of the constant in the exponent at the power function is more difficult, but we have succeeded in obtaining an explicit formula for it.

Having completed the main part of this paper (see [4]), we became aware of the paper [9], where under the same conditions one proves the existence of limits $\ln p(t)/\ln t$ as $t \to \infty$ (which coincides with our case of the power order of the asymptotics, as well as with the case of the power asymptotics, where correction data change is slower) and the limit $\ln p(t)/t^{1/D}$, where $D \in \mathbb{N}$ (which coincides with our case of the subex-
ponential order of the asymptotics). Therefore, one can consider results obtained in this paper as a strengthening of results of [9], where one has proved weak power and weak subexponential asymptotics. Note that earlier in [4] for the subexponential case, we considered only necessary and sufficient conditions of the exponential decreasing order ($D = 1$). We have proved the subexponential order in a general case only after getting acquainted with results of [9]. We calculate constants in the subexponent by explicit formulas, while the corresponding constants in [9] are calculated by recurrent formulas (which can be easily reduced to explicit ones). In the conclusion of this paper, we compare results obtained by us and those of [9] in detail.

2. The exact statement of main result.

2.1. Definitions and denotations. Let $P_0$ be a (stochastic) transition probability matrix of the Markov chain with the state set (1.1), and let $P$ be its (substochastic) submatrix corresponding to states $E_1, \ldots, E_n$. Denote by $G_0$ the directed pseudograph with the set of vertices $\{0, \ldots, n\}$, whose arcs $(i, j)$ are defined by inequalities $p_{ij} > 0$. Conditions (1.1) are equivalent to the fact that the graph $G_0$ is (weakly) connected, and $\{0\}$ is the only collection of vertices that has no arcs leading to its complement. Let $G$ be the subgraph of the graph $G_0$ with the set of vertices $\{1, \ldots, n\}$ including all arcs of the initial graph $G_0$ between these vertices (the subgraph generated by vertices $\{1, \ldots, n\}$). Let $H$ be a subgraph of the graph $G_0$ generated by some set of vertices. Then we denote by $P_H$ the corresponding submatrix of the matrix $P_0$: $P_H = (p_{ij})_{i,j \in V(H)}$. Thus, for example, $P_G \equiv P$. In addition, we set $P_H(\beta) = (p_{ij}^{\beta})_{i,j \in V(H)}$.

Recall that a strongly connected component (SCC) is a maximal complete subgraph such that any pair of its vertices is mutually connected. Denote by $G'$ the digraph obtained from the graph $G_0$ by identifying vertices and arcs that belong to the same SCC of the initial graph $G_0$ (in [13] this graph is called the condensation). In this paper, the graph $G'$ is connected and 0 is the only vertex having no outgoing arcs. Recall that [13] the graph $G'$ is acyclic.

We denote by $a = (a_0, \ldots, a_n)$ the initial distribution of probabilities on the state set. Without loss of generality, we assume that

\[
(2.1) \quad \text{every state is accessible.}
\]

Condition (2.1) means that for each state $E_i$ there is a time $t$ such that there is a positive probability of being at state $E_i$ at time $t$. In what follows we sometimes deal with initial distributions, for which Condition (2.1) is not assumed to be fulfilled; we specify all such cases separately.
Let us associate an arbitrary path \( c = (i_1, \ldots, i_m) \) in the graph \( G_0 \) with the weight
\[
\tilde{\Pr}(c) = p_{i_1 i_2} \cdots p_{i_{m-1} i_m}.
\]
Instead of a path in the graph, it is often more convenient to consider an ordered set of states of the chain \( w = (E_{i_1}, \ldots, E_{i_m}) \). We call this set a word, if \( a_{i_1} > 0 \), \( E_{i_m} = E_0 \), and \( E_{i_{m-1}} \neq E_0 \). In other words, we understand a word as a sequence of states reached by the system from the start of the walk till the absorption by the state \( E_0 \). We determine the word probability \( \Pr(w) \), taking into account the initial distribution:
\[
(2.2) \quad \Pr(w) = a_{i_1} p_{i_1 i_2} \cdots p_{i_{m-1} i_m}.
\]
One can easily prove that the set of all words with the measure \( \Pr \) forms a discrete probability space (i.e., the sum of probabilities of all words equals one).

We understand the length \( L \) of a word \( w \) as the number of states in it, excluding the last absorbing state \( E_0 \). A simple cycle is a closed path without repeated vertices, except the vertices which are used to start and end the cycle. We denote by \( C \) the set of all simple cycles in the graph \( G \). Let \( W' \) be the set of all words with unrepeated states. For any \( w' \in W' \) we denote by \( C(w') \) the set of all simple cycles that intersects the path in the graph \( G \) corresponding to the word \( w' \).

Let us sort all words in the nonincreasing order of their probabilities. Evidently, both the value \( p(t) = \Pr(w_t) \) (the probability of the \( t \)th word in this ordered list) and the “inverse” function of \( p(t) \), \( Q(q) \), \( q \in (0, 1) \), (that equals the number of words whose probability is less than \( q \)) are defined. We are interested in the asymptotics of the function \( p(t) \) for \( t \to \infty \) (or, equivalently, that of the function \( Q(q) \) for \( q \to 0 \).

We use the standard \( O \)-symbolics and we denote by \( \Theta \) the asymptotic order and we denote by \( \Omega \) the lower estimate of the order ([12, Section 9.2]):
\[
\begin{align*}
 f(x) &= \Omega(g(x)) \iff |f(x)| > C|g(x)| \text{ for some } C > 0, \\
 f(x) &= \Theta(g(x)) \iff f(x) = \Omega(g(x)) \text{ and } f(x) = O(g(x)).
\end{align*}
\]

2.2. The statement of the main theorem.

THEOREM 2.1. Three cases are possible:

A. If the graph \( G \) is acyclic, then the function \( p(t) \) is finitary (i.e., the number of all possible words is finite).

B. If the graph \( G \) contains a vertex which is common for two different simple cycles (if there is at least one SCC on \( G \) which is not a cycle), then \( p(t) = \Omega(t^{-1/\beta}) \), where \( \beta \) is a real number, with which the maximal modulo eigenvalue of the matrix \( P_G(\beta) \) equals one. Note that such \( \beta \) exists, is unique, and belongs to the interval \( (0, 1) \). Moreover, \( p(t) = o(t^{-1/\beta'}) \) for any \( \beta' > \beta \).

In addition, the exact power order (i.e., the equality \( p(t) = \Theta(t^{-1/\beta}) \)) is attained if and only if any simple path in the graph \( G' \) contains at most one
vertex (a SCC $H$ of the graph $G$) such that the matrix $P_H(\beta)$ has the unit eigenvalue.

C. If the graph $G$ contains cycles, and each vertex of the graph $G$ belongs to no more than one simple cycle (every SCC on $G$ is a cycle), then $p(t) = \Theta(\exp(-\sqrt{\nu}t))$; here $\nu$ is determined by the formula

$$
1/\nu = \sum_{w' \in W'} 1/D! \prod_{c \in C(w')} -1/\ln \tilde{F}(c),
$$

where $D = \max_{w' \in W'} |C(w')|$. 

**Remark 2.2.** The item A of the Theorem is trivial (we give it here only for the sake of completeness). It follows from the fact that in an acyclic graph, the length of any word does not exceed $n$.

**Remark 2.3.** The parameter $\nu$ of the exponential asymptotics (as distinct from the order of power case) depends not only on the matrix of transition probabilities, but also on the set of states $v$ such that $a_v > 0$ (for more details see Remark 4.2).

**Remark 2.4.** As was proved earlier [3, 7], if states are chosen independently and the probability of each one is $p_i$, then for $n > 1$ the function $p(t)$ has a power asymptotic; its exponent determined from the equation $\sum_{i=1}^{n} P_i^\beta = 1$ equals $1/\beta$. This is a particular case of Theorem 2.1.B, where the matrix $P$ consists of nonzero elements and has equal rows. Raising all elements of the matrix $P$ to the power $\beta$, we obtain a stochastic matrix; it is well known that the maximal eigenvalue of a stochastic matrix equals one.

2.3. **Examples.** The graph shown in diagram b) in Fig. 2.1 has only one SCC with vertices $\{1, 2\}$ (we do not take into account the trivial cycle from the absorbing state to itself). This component contains cycles $(1, 2, 1)$ and $(1, 1)$, therefore, the function $p(t)$ has a power asymptotic. For example, if all probabilities of transitions from states $E_1$ and $E_2$ equal $1/2$, then one can easily calculate that $\beta = \log_2(1+\sqrt{5})/2$. The graph shown in Fig. 2.2 has two SCCs $H_1$ and $H_2$ (we do not take into account the trivial cycle from the absorbing state to itself), and both of them belong to one and the same path in the graph $G'$. Moreover, the graph $H_1$ contains a vertex which is common for two different simple cycles. We use Theorem 2.1.B. If probabilities of all transitions from states $E_1$, $E_2$, $E_3$, $E_4$ equal $1/3$, then one can easily calculate that $\beta = \log_2 2$. With this value of $\beta$ matrices $P_{H_1}(\beta)$ and $P_{H_2}(\beta)$ have the unit eigenvalue (all their elements equal $1/2$). Therefore, the power asymptotics do not take place, i.e., $p(t) = \Omega(t^{-\log_2 3})$ and $p(t) = o(t^{-\delta})$ for any $\delta < \log_2 3$, but $p(t) \neq \Theta(t^{-\log_2 3})$.

The graph shown in diagram c) in Fig. 2.1 contains two simple cycles-loops, and in the graph $G$ there is a path going through all vertices. Therefore, the decreasing order of $p(t)$ equals $\exp(-\sqrt{\nu}t)$. If probabilities of all transitions from the state $E_1$ are equal
Fig. 2.1. Examples of graphs $G_0$ of a Markov chain with three states $E_0, E_1, E_2$ (the vertex that corresponds to the absorbing state $E_0$ is pictured at the bottom). In case a) the function $p(t)$ is finitary. In case b) the asymptotics of the function $p(t)$ has a power order. In case c) the asymptotics of the function $p(t)$ has the order $\exp(-\sqrt{\nu t})$. In cases d) and e) the function $p(t)$ has an exponential decreasing order. Note that the classification depends only on the graph $G$ (the upper part of the figure), provided that states $E_1$ and $E_2$ are nonrecurrent.

Fig. 2.2. An example of the graph $G_0$ of a Markov chain with five states $E_0, E_1, E_2, E_3, E_4$. The function $p(t)$ is bounded by two functions having a power asymptotic, however, their degrees are different (arbitrarily close). The asymptotics of the function $p(t)$ itself does not necessarily have a power order, if matrices of transition probabilities of graphs $H_1$ and $H_2$ coincide or so do the corresponding exponents $\beta$.

to $1/3$, while those from the state $E_2$ are equal to $1/2$, then we can easily calculate that $\nu = 2 \ln 2 \ln 3$. The graph shown in diagram d) in Fig. 2.1 contains two analogous cycles, but in the graph $G$ there is no path described in the previous example; this means that the decrease of the function $p(t)$ has an exponential asymptotic. The graph shown in diagram e) has one simple cycle, and the asymptotic is also exponential. Assume that $a_1 > 0$ and $a_2 > 0$; then there are 4 words with nonrepeating states, namely, $(E_1, E_0), (E_2, E_0), (E_1, E_2, E_0), (E_2, E_1, E_0)$. Now assume that for Markov
chains with graphs shown in diagrams d) and e) all probabilities of transitions from states $E_1, E_2$ equal $1/2$; then one can easily calculate that in both cases d) and e) $\nu = \ln \sqrt{2}$.


3.1. A spectral substochastic lemma. Prior to proving Theorem 2.1.B, let us prove the unique existence of the exponent $\beta$ in this case. Consider an arbitrary substochastic matrix $P = (p_{ij})_{i,j=1}^n$ with the following properties (in conditions given below, indices $i, j$ belong to $\{1, \ldots, n\}$):

$$0 \leq p_{ij} \leq 1 \text{ for all } i,j;$$

$$\sum_{j=1}^n p_{ij} \leq 1 \text{ for all } i \text{ (the substochasticity);}$$

the matrix $P$ is not nilpotent;

for any principal submatrix of the matrix $P$ there exists a row such that the sum of its elements in this submatrix is strictly less than 1.

Note that with $P \equiv P_G$ the latter property is equivalent to the nonrecurrence of all states (except the absorbing one) [10]; the matrix $P$ is nilpotent if and only if the graph $G$ is acyclic.

Recall that for matrices with nonnegative elements (nonnegative matrices) the next theorem [11, Theorem 3, Chapter XIII] is valid. Namely, “A non-negative matrix $A = (a_{ij})_{i,j=1}^n$ always has a non-negative characteristic value $r$ such that moduli of all characteristic values of $A$ do not exceed $r$. To this maximal characteristic value $r$ there corresponds a non-negative characteristic vector $Ay = ry \ (y \geq 0, y \neq 0)$.” Note that both the matrix $A$ and that $A^T$ (the symbol $T$ is the transposition sign) may have no positive eigenvector (a vector all whose components are strictly positive). Later we discuss existence conditions for such a vector.

Recall that the symbol $P(\beta)$ denotes the matrix $(p_{ij}^\beta)_{i,j=1}^n$ (here $0^\beta = 0$ for any $\beta$), while $G$ stands for a directed graph with $n$ vertices, whose arcs correspond to nonzero elements of the matrix $P$.

**Lemma 3.1.** For any matrix $P$ in form (3.1) there exists unique $\beta \in \mathbb{R}$ such that the maximal characteristic value of the matrix $P(\beta)$ equals 1, while $0 \leq \beta < 1$. The inequality $\beta > 0$ is equivalent to the existence in the graph $G$ of two different simple cycles that go through one and the same vertex.

**Proof.** Denote by $s_i$ the sum $\sum_{j=1}^n p_{ij}$. Let $s = \min_i s_i$ and $S = \max_i s_i$. It is known that [11, Remark on p. 68] the maximal characteristic value $r$ of any nonnegative matrix satisfies the inequality $s \leq r \leq S$. Denote by $r(\psi)$ (here $\psi \geq 0$) the maximal eigenvalue of the matrix $P(\psi)$, let $s(\psi) = \min_i \sum_{j=1}^n p_{ij}^\psi$ and $S(\psi) = \max_i \sum_{j=1}^n p_{ij}^\psi$. 

Let us prove the uniqueness of the choice of $\beta$ from the lemma condition and the validity of the inequality $0 \leq \beta < 1$. Recall that the matrix $P$ is called indecomposable if the oriented graph $G$ is strongly connected. In a general case, the decomposition of a graph into SCCs corresponds to the normal form of the matrix obtained from the initial one by renumbering its rows (and, correspondingly, columns). In the normal form (see [11, p. 75]), the diagonal is occupied by square blocks corresponding to collections of vertices that belong to one and the same SCC; the matrix elements located above these blocks equal zero. Therefore, sequentially decomposing the determinant by the group of rows that correspond to SCCs, we obtain that the characteristic polynomial of the matrix $P(\psi)$ equals the product of characteristic polynomials of each of diagonal blocks. As a consequence of this fact, the eigenvalues of the matrix is the union of the eigenvalues in the individual blocks, so $r(\psi)$ coincides with the maximal eigenvalue of blocks. However, according to Formula (3.1), for square submatrices that correspond to each of these blocks, the value $s$ is strictly less than one. In addition, not all blocks are zero, otherwise the matrix $P$ is nilpotent. For the block $H$

\begin{equation}
    P_H(0) = \text{the adjacency matrix of the graph } H,
\end{equation}

so $s(0) \geq 1$ for at least one of blocks. It is known that [11, p. 63] indecomposable nonnegative matrices with unequal values of $s$ and $S$ satisfy the strict inequality $s < r < S$. Consequently, $r(1) < 1$ and $r(0) \geq 1$.

Evidently, $p_{ij}^\psi$ decreases as $\psi$ increases, if $p_{ij} > 0$. It is known that [11, Theorem 6, Chapter XIII] if some elements of a nonnegative indecomposable matrix decrease, then its maximal characteristic value strictly decreases. Therefore, $r(\psi)$ is a decreasing function. We have proved the uniqueness of the choice of $\beta$ and the validity of the inequality $0 \leq \beta < 1$.

Let us prove the last assertion of the lemma. In the normal form of the matrix $P$, we consider the block containing the vertex that belongs to two different cycles. For this block we introduce analogs of values $s(\psi)$ and $S(\psi)$; we denote them by $s'(\psi)$ and $S'(\psi)$, correspondingly. The considered block, by definition, is an indecomposable matrix. Consequently (see (3.2)), $s'(0) \geq 1$ and $S'(0) \geq 2$. Hence, for the matrix $P(0)$ we get $r(0) > 1$, which implies that in this case the desired value of $\beta$ (by condition of the lemma) is strictly positive.

It remains to prove that if no vertex in the graph $G$ belongs to two cycles, then the desired value of $\beta$ equals zero. Really, the considered diagonal blocks either are trivial (i.e., consisting of one element) or correspond to nontrivial SCCs of the graph $G$. A nontrivial component, by definition, contains a cycle going through all its vertices. In our case this cycle cannot be self-intersecting, because in this case there would exist a vertex belonging to two cycles. Therefore, the SCC consists of a single simple
cycle. But this means that for the corresponding block, \( S'(0) = s'(0) = 1 \). Since the eigenvalues of \( P(0) \) are the union of the eigenvalues of diagonal blocks, we obtain \( r(0) = 1 \).

Evidently, Lemma 3.1, taking into account the nonrecurrence of states of the Markov chain, implies the existence of the exponent \( \beta \) in the interval \((0, 1)\), provided that assumptions of Theorem 2.1.B are fulfilled.

### 3.2. A positive eigenvector of the matrix \( P(\beta)^T \) and the initial distribution of the Markov chain.

Let us mention the following fact.

**Corollary 3.2.** Assume that under conditions of Lemma 3.1, \( \beta > 0 \) and the normal form of the matrix \( P \) contains several blocks representing SCCs \( H \) of the graph \( G \) such that characteristic numbers of matrices \( P_H(\beta) \) equal one. Then each of these graphs \( H \) contains a vertex that belongs to two (or more) different simple cycles.

Let us now consider the case when the matrix \( P(\beta)^T \) has a positive eigenvector corresponding to the unit eigenvalue. Redefining the standard necessary and sufficient conditions for the existence of a positive eigenvector (see [11, Theorem 7, Chapter XIII]), we obtain the following assertion.

**Proposition 3.3.** Let assumptions of Lemma 3.1 be fulfilled and \( \beta > 0 \). The matrix \( P(\beta)^T \) has a positive eigenvector corresponding to the unit eigenvalue if and only if in the graph \( G' \) vertices without incoming arcs, and only they, correspond to SCCs \( H \), for which matrices \( P_H(\beta) \) have the unit characteristic value.

**Corollary 3.4.** Assume that under conditions of Theorem 2.1.B (there is at least one SCC on \( G \) which is not a cycle) the matrix \( P(\beta)^T \) has a positive eigenvector corresponding to the unit eigenvalue. Then we can choose a vector \( a = (a_1, \ldots, a_n) \) such that \( a_k = 0 \) for all vertices with less than two incoming arcs, and the probability of reaching any vertex is greater than zero.

**Proof of Corollary 3.4.** Consider graphs \( H \) mentioned in Proposition 3.3. According to Corollary 3.2, in each of them there exists a vertex which belongs to two cycles. Assume that \( a_v > 0 \) for all such vertices \( v \), and \( a_v = 0 \) otherwise. Then the probability to reach any vertex of graphs \( H \) is greater than zero, because all these vertices are located in one and the same SCC. Proposition 3.3 implies that all the rest SCCs are also reachable with nonzero probabilities. But then we can get, with nonzero probabilities, to all vertices of the graph \( G \).
4. The power law in the case of the existence of a positive eigenvector.

4.1. The choice of the initial distribution. We need some more auxiliary assertions about power inequalities for the function $p(t)$. Note that Lemma 4.1 is valid even without assumptions on the existence and positiveness of the eigenvector of the matrix $P(\beta)^T$. We use it for proving both the main result of this section (in the framework of the mentioned assumption), and its corollaries (in a more general case).

**Lemma 4.1.** A. Let $\delta > 0$. With some initial distribution $a$ (not necessarily satisfying Condition (2.1)) we obtain $p_a(t) = \Omega(t^{-\delta})$ (hereinafter the subscript indicates the initial distribution under consideration). Then with any initial distribution $a'$ satisfying Condition (2.1), we have $p_{a'}(t) = \Omega(t^{-\delta})$.

B. Let $\delta > 0$. Assume that with some initial distribution $a, a = (a_1, \ldots, a_n)$, satisfying Condition (2.1) it holds $p_a(t) = O(t^{-\delta})$. Then with any initial distribution $a'$ we have $p_{a'}(t) = O(t^{-\delta})$.

As a corollary, we obtain that if $p_a(t) = \Theta(t^{-\delta})$ with some initial distribution $a$ satisfying (2.1), then it is also valid for all initial distributions satisfying (2.1).

**Remark 4.2.** If the order of the asymptotics is not power, then the assertion analogous to Lemma 4.1, generally speaking, is not true. Namely, the order of the asymptotics of the function $p(t)$, possibly, depends on the initial distribution. Thus, when calculating the Markov chain that corresponds to the (last) diagram e) in Fig. 2.1, we obtain the exponential order of the asymptotics of the function $p(t)$ with the exponent $\nu = \ln \sqrt{2}$. Here we assume that $a_1 > 0, a_2 > 0$. But if $a = (1, 0)$ in this chain, then, as one can easily prove, the asymptotic is exponential with $\nu = \ln 2$.

We denote a Markov chain with an initial distribution $a$ as $\text{MCh}_a$, we denote probabilities of words $w$ in this Markov chain by $\text{Pr}_a(w)$. By definition, all words in $\text{MCh}_a$ begin in the set $E(a) = \{E_i : a_i > 0\}$, and we denote the corresponding set of vertices by $I(a) = \{i : a_i > 0\}$. The idea of the proof consists in associating words in $\text{MCh}_a$ with those in $\text{MCh}_a'$, and then in estimating the function $p$.

**Proof of Lemma 4.1.** A. Evidently, we can select particular path $(i', i_1, \ldots, j)$ for each $j, j \in I(a)$ such as $i' \in I(a')$; we denote this path by $\pi(j)$. We associate each word $w$ in $\text{MCh}_a$, beginning with $E_j$, with a word $w'$ in $\text{MCh}_a'$ by adding the prefix $(E_{i'}, E_{i_1}, \ldots, E_j)$. Evidently, $\text{Pr}_{a'}(w') = \text{Pr}_a(w)c(j)$, where $c(j) = \text{Pr}(\pi(j))a_{i'} / a_j$. It is possible that several words in $\text{MCh}_a$ correspond to one and the same word in $\text{MCh}_a'$. However, in the associated list, this word may appear no more than $n$ times because there exists only $n$ prefixes (consequently there exists no more than $n$ variants of prefixes that begin with $E_{i'}$).
Consider the sorted list of first $t$ words $(w_1, w_2, \ldots, w_t)$ in $\text{MCh}_a$ and associate them with words $(w'_1, \ldots, w'_t)$ in $\text{MCh}_a'$ (some of them, possibly, coincide). It is obvious that for any list of different words $(w'_1, \ldots, w'_t)$ in $\text{MCh}_a'$ occurs $p_{a'}(t) \geq \min_{1 \leq i \leq t} \Pr_{a'}(w'_i)$. We get $p_{a'}(t) \geq \min_{1 \leq i \leq nt} \Pr_{a'}(w'_i) \geq p_a(nt) \min_{j \in I(a')} c(j) > \text{const } t^{-\delta}$.

Proof of Lemma 4.1.B is quite similar (it uses the inequality $p_{a'}(t) \leq c p_a([t/n])$).

4.2. The key Lemma.

In Lemma 4.1, the “inversed” function $Q$ can be considered instead of the function $p$, where $Q(q)$ equals the number of words whose probability is less than $q$. The assertion of Lemma 4.1 is equivalent to an analogous one for $Q(q)$ with $1/\delta$ in place of $\delta$. Really, the graph of the function $p(t)$ demonstrates that the inequality $p(t) < c t^{-\delta}$ (with all $t \geq 1$) is equivalent to $Q(q) < (q/c)^{-1/\delta} = \text{const } q^{-1/\delta}$ (or, respectively, $Q(q) > \text{const } q^{-1/\delta}$) with all (sufficiently small) values of $q$.

**Lemma 4.3.** Assume that a graph $G$ has a vertex that belongs to two different simple cycles, $\beta$ is chosen in accordance with Lemma 3.1, and the matrix $P(\beta)^T$ has a positive eigenvector $e$ corresponding to the unit eigenvalue. Then $p(t) = \Theta(t^{-1/\beta})$.

**Proof** (cf. the proof in [3]). As noted above, the assertion about the power asymptotics of the function $p(t)$ is equivalent to an analogous assertion for the function $Q$. Let us prove it now.

We understand an incomplete word as the initial part of a path $(i_1, \ldots, i_m)$ such that $a_{i_m} > 0$; we define the “probability” of an incomplete word by the same formula (2.2). For positive $x$, we introduce functions $Q_k(x)$, $k = 1, \ldots, n$, which equal the number of incomplete words ending with the symbol $E_k$ whose “probabilities” are not less than $x$. Evidently, $Q_k(x) = 0$ with $x > 1$. We also need functions $\tilde{Q}_k(x)$: $\tilde{Q}_k(x) = Q_k(x) + 1$, $k = 1, \ldots, n$.

Let us prove that $Q_k(x) = \Theta(x^{-\beta})$ as $x \to 0$. Evidently, such power estimate from above (from below) for the function $Q_k(x)$ is equivalent to an analogous estimate for $Q_k(x)$.

Put
\[
\chi_0(x) = \begin{cases} 1 & \text{for } x \leq 1, \\ 0 & \text{for } x > 1. \end{cases}
\]

The definition implies the following important recurrent correlation:
\begin{equation}
Q_k(x) = \sum_{m \cdot p_{m,k} > 0} Q_m(x/p_{m,k}) + \chi_k(x),
\end{equation}

where \( \chi_k(x) = \begin{cases} 
\chi_0(x/a_k), & a_k > 0, \\
0, & \text{otherwise}.
\end{cases} \)

In particular, the following inequality is valid:
\begin{equation}
Q_k(x) \geq \sum_{m \cdot p_{m,k} > 0} Q_m(x/p_{m,k}), \quad k = 1, \ldots, n.
\end{equation}

Let us now use Lemma 4.1, which gives some freedom of the choice of the initial distribution. Choosing \( a_k \) as is described in Corollary 3.4, for all vertices \( k \) with one incoming arc \((m, k)\) we get \( Q_k(x) = Q_m(x/p_{m,k}) \).

In Relation (4.1), we can put a sign \( \leq \), replacing \( \chi_k \) by one. Consequently, for vertices with at least two incoming arcs, \( Q_k(x) \leq \sum_{m \cdot p_{m,k} > 0} Q_m(x/p_{m,k}) + (l - 1), \) where \( l \) is the number of terms in the sum. Therefore,
\begin{equation}
\tilde{Q}_k(x) \leq \sum_{m \cdot p_{m,k} > 0} \tilde{Q}_m(x/p_{m,k}), \quad k = 1, \ldots, n.
\end{equation}

Let the eigenvector \( e \) mentioned in the condition of the lemma have components \((e_1, \ldots, e_n)\). One can easily verify that functions \( f_k(x) = e_k x^{-\beta}, \quad k = 1, \ldots, n, \) satisfy the following set of functional equations:
\begin{equation}
\tilde{f}_k(x) = \sum_{m \cdot p_{m,k} > 0} \tilde{f}_m(x/p_{m,k}), \quad k = 1, \ldots, n.
\end{equation}

Now let \( M \) be the minimum of positive elements of the matrix \( P \), and let \( M' \) be the maximum of its non-unit elements. Fix \( y \) such that \( Q_k(y) > 0 \) for all \( k \). Evidently that on the segment \([My, y]\) the function \( Q_k(y) \) is monotone and positive (more exactly, on this segment it takes on a finite number of natural values). This means that one can find positive constants \( c_1 \) and \( c_2 \) independent of \( k \) such that inequalities \( Q_k(x) \geq c_1 f_k(x) \) and \( Q_k(x) \leq c_2 f_k(x), \quad k = 1, \ldots, n, \) are valid with \( My \leq x \leq y \). But then Formulas (4.2, 4.3, 4.4) imply that the same inequalities (with the same constants \( c_1 \) and \( c_2 \)) are valid with \( x \in [M'My, y] \) and, consequently, with all \( x \leq y \). The estimate \( Q_k(x) = \Theta(x^{-\beta}) \) for \( x \leq y \) is proved.

Since \( Q(x) = \sum_{m \cdot p_{m,k} > 0} Q_m(x/p_{m0}) \), we obtain that \( Q(x) = \Theta(x^{-\beta}) \) for sufficiently small \( x \). \( \square \)
4.3. The proof of assertions about $\Omega$ and the $o$-asymptotics.

Hereinafter, “Let assumptions of Theorem 2.1.B be fulfilled” will be understood as “the graph $G$ contains a vertex which is common for two different simple cycles” (or “there is at least one SCC on $G$ which is not a cycle”).

**Corollary 4.4.** Let assumptions of Theorem 2.1.B be fulfilled. Then $p(t) = \Omega\left(t^{1/\beta}\right)$, where $\beta$ is a real number such that the maximal modulo eigenvalue of the matrix $P_G(\beta)$ equals one.

**Proof.** The idea of the proof consists in the application of Lemma 4.1.A. But first we need to find at least one initial distribution for which our power estimate from below is valid.

Consider $\beta$ defined in the condition of Corollary 4.4 (recall that in view of Lemma 3.1, it exists and is positive and unique). In the normal form, the matrix $P_G(\beta)$ has blocks that represent SCCs $H$ such that the maximal modulo eigenvalue of the matrix $P_H(\beta)$ equals one. Assume that conditions of Proposition 3.3 are violated. Then in some path in the graph $G'$, one such block does not correspond to the first vertex in the path. Without loss of generality, we can assume that no arc enters the initial vertex of the path under consideration. We delete this vertex from the graph $G'$ and the corresponding SCC from the graph $G$. Consider the “truncated” Markov chain with the obtained graph. Evidently, as above, it satisfies Conditions (1.1) and assumptions of Theorem 2.1.B; moreover, for the matrix of transition probabilities, the value of $\beta$ remains the same.

Repeating this operation several times, we can make the matrix of the obtained graph $\tilde{G}$ satisfy conditions of Proposition 3.3. Fixing the initial distribution $a$ for the Markov chain with the graph $\tilde{G}$, we fix the corresponding distribution $a$ for the Markov chain with the graph $G$; however, in this case, we never reach deleted vertices. By applying Lemma 4.3 (which is proved above) and using Lemma 4.1.A, we obtain the assertion of Corollary 4.4.

**Corollary 4.5.** Let assumptions of Theorem 2.1.B be fulfilled. Then $p(t) = o\left(t^{1/\beta'}\right)$ for any $\beta' > \beta$.

**Proof.** The idea of the proof consists in the application of Lemma 4.1.B. But first we perform the operation opposite to that in the proof of the previous lemma. Namely, we add to the graph $G$ additional SCCs so as to make the obtained Markov chain satisfy the condition of Lemma 4.3 with some exponent $\beta''$ lesser than $\beta'$.

Let $k$ be the number of vertices in the graph $G'$ which have no incoming arcs, let $v$ be one such vertex, and let $H_v$ be the SCC of the graph $G$ corresponding to it. Let us add to $G$ some subgraphs $\tilde{H}_v$ which have the form shown in the upper part
of diagram b) in Fig. 2.1, then an arc from the added subgraph will lead to one of vertices in $H_v$. As a result, we will obtain a graph with $n + 2k$ vertices.

Consider a Markov chain with $n + 2k$ non-absorbing states, whose matrix of transition probabilities $\tilde{P}$ is obtained from the matrix $P$ by adding $k$ pairs of rows that correspond to subgraphs $\tilde{H}_v$. Each pair corresponds to a diagonal $2 \times 2$ block in the form $P_2 = \begin{pmatrix} r & s \\ t & 0 \end{pmatrix}$, where $0 < r, s, t < 1$, $r + s = 1$, numbers $r, s, t$ are the same for all blocks. Let us choose numbers $r, s, t$ so as to make the maximal eigenvalue of the matrix $P_2(\beta'')$ equal one (for some $\beta'': \beta < \beta'' < \beta'$). To this end, it suffices to choose $x$ such that $r\beta''x + s\beta'' = 1$ (since $r\beta'' + s\beta'' > 1$, the desired value of $x$ is less than one), and then set $t = x^{1/\beta''}$.

Let us now consider the Markov chain with the transition probability matrix (between non-absorbing states) $\tilde{P}$. Evidently, the matrix $\tilde{P}(\beta'')$ satisfies conditions of Proposition 3.3, whence by Lemma 4.3 and Lemma 4.1.B we get $p_{a}(t) = O(t^{-\beta''})$ for any initial distribution $a$ of this Markov chain. In particular, this is also valid for $I(a) \in V(G)$, and in this case, we never reach vertices of added graphs $\tilde{H}(v)$. Thus, for the initial Markov chain we have $p(t) = O(t^{-\beta''}) = o(t^{-\beta'})$.

5. Completion of the proof of Theorem 2.1.B.

5.1. Sequential and parallel connections of graphs of Markov chains.
It remains to establish necessary and sufficient conditions for the power asymptotics. Sufficient but not necessary conditions are given by assumptions of Lemma 4.3. In order to complete the proof of Theorem 2.1.B with the help of Lemma 4.3, we need two more auxiliary assertions.

Let us first consider the case of a “parallel” connection of graphs $G_1$ and $G_2$ of Markov chains (we denote the Markov chains themselves by $\text{MCh}_{G_1}$ and $\text{MCh}_{G_2}$); we identify the absorbing states of these graphs.

![Diagram](image_url)

**Fig. 5.1.** The construction of $\text{MCh}_{G}$ by the “parallel” connection of graphs of $\text{MCh}_{G_1}$ and $\text{MCh}_{G_2}$. Arcs that earlier led from $G_1$ and $G_2$ to their “own” absorbing states, now lead to the common absorbing state $E_0$.

**Lemma 5.1.** Assume that Markov chains with graphs $G_1$ and $G_2$ with some
initial distributions (satisfying Condition (2.1)) for \( p_1(t) \) and \( p_2(t) \) (probabilities of the \( t^{th} \) word in the corresponding sorted list) satisfy correlations \( p_1(t) = O(t^{-\delta_1}) \) and \( p_2(t) = O(t^{-\delta_2}) \), where \( \delta_1, \delta_2 > 0 \). Assume that for the Markov chain with the function \( p(t) \), any word represents either a word from the first Markov chain or that of the second one; its graph \( G \) represents a non-connected union of graphs \( G_1 \) and \( G_2 \), while the corresponding transition probabilities remain the same (see Fig. 5.1). Then with any initial distribution the following correlation is valid:

\[
(5.1) \quad p(t) = O(t^{-\delta}), \text{ where } \delta = \min\{\delta_1, \delta_2\}.
\]

Proof. By Lemma 4.1.B it suffices to prove Inequality (5.1) with some concrete initial distribution \( a \) satisfying Condition (2.1). Let us choose it as \((a' + a'')/2\), where \( a' \) and \( a'' \) are initial probability distributions in the first and second Markov chains, correspondingly. Then probabilities of all words in the aggregated Markov chain are 2 times less than probabilities of the same words in calculations of \( p_1(t) \) and \( p_2(t) \). The list of the first \( t \) words of our Markov chain, sorted in the non-increasing order of their probabilities, consists of the initial part of the analogous list of the first MCh alternated with the initial part of the second MCh; consequently, this list contains a word of either first or second MCh with the index \( \lceil t/2 \rceil \). We have

\[
(5.2) \quad p(t) \leq \max\{p_1(\lceil t/2 \rceil), p_2(\lceil t/2 \rceil)\}
\]

(we could have again divide the right-hand side by 2, but even the weakened variant of the inequality suits us).

By assumption there exist positive constants \( c_1 \) and \( c_2 \) such that

\[
(5.3) \quad p_1(t) < c_1 t^{-\delta_1}, \quad p_2(t) < c_2 t^{-\delta_2} \text{ for all } t.
\]

Let us choose a constant \( c \) such that \( c t^{-\delta} > \max\{2^{\delta_1} c_1 t^{-\delta_1}, 2^{\delta_2} c_2 t^{-\delta_2}\} \) for all positive integers \( t \). Using (5.2) and (5.3), we obtain \( p(t) < c t^{-\delta} \).

Remark 5.2. Evidently, Lemma 5.1 can be extended by induction to the case of the “parallel” connection of MCh\(_{G_1} \), MCh\(_{G_2} \), ..., MCh\(_{G_m} \).

Let us now consider the case when graphs of Markov chains are connected “sequentially”. Consider the graph \( G \) obtained from the union of graphs \( G_1 \) and \( G_2 \) of Markov chains by redirecting at least some arcs that earlier led from \( G_1 \) to the absorbing state, and now do to the graph \( G_2 \). Denote the set of these arcs by \( E_{12} \). Assume that one can reach any vertex of the graph \( G_2 \) along the path that goes through the proper arc from \( E_{12} \), and transition probabilities in MCh\(_G \) are equal to the corresponding probabilities in MCh\(_{G_1} \) and MCh\(_{G_2} \) (see Fig. 5.2).
Fig. 5.2. The construction of MChG by a “sequential” connection of graphs of MChG$_1$ and MChG$_2$. Arcs that earlier led from G$_1$ to their “own” absorbing states form two groups; arcs of the first group lead to the common absorbing state E$_0$, those of the second one do to the graph G$_2$. All arcs that earlier led from G$_2$ to their “own” absorbing states now lead to the common absorbing state E$_0$.

**Lemma 5.3.** Assume that Markov chains with graphs G$_1$ and G$_2$ with some initial distributions (satisfying Condition (2.1)) for $p_1(t)$ and $p_2(t)$ (probabilities of the $t$th word in the corresponding sorted list) fulfill correlations $p_1(t) = O(t^{-\delta_1})$ and $p_2(t) = O(t^{-\delta_2})$, where $\delta_1, \delta_2 > 0$. Let the Markov chain with the function $p(t)$ correspond to the graph $G$ representing the union of graphs G$_1$ and G$_2$ with additional arcs going from the graph G$_1$ to that G$_2$ so that any vertex of the graph G$_2$ is attainable through the path consisting of these arcs. Then Formula (5.1) is valid with $\delta_1 \neq \delta_2$. Correlation (5.1) is false if the initial distribution satisfies Condition (2.1), while $\delta_1 = \delta_2$ and $p_1(t) = \Omega(t^{-\delta_1})$, $p_2(t) = \Omega(t^{-\delta_2})$.

**Proof.** As the initial distribution in MChG, we consider a distribution $a$ concentrated at vertices of the graph G$_1$ and satisfying Condition (2.1) for it. Evidently, for MChG, Condition (2.1) is also valid; further considerations are related to the corresponding function $p(t)$.

Note that the assertion of Lemma 4.1 remains valid, even if the probability $a_0$ that the initial state is absorbing differs from zero. Moreover, in this case, in order to make the sum of probabilities of all words equal one, it is convenient to add to the sorted list of all possible words one more word, the empty one, whose probability equals $a_0$ (this, naturally, does not affect the asymptotic properties of considered functions).

Assume that the constant $c_1$ in Inequality (5.3) is defined for the initial distribution $a'$ in MChG$_1$, coinciding with the distribution $a$. We assume that the initial distribution $a''$ in MChG$_2$ is concentrated at end vertices $E_{12}$ and at the absorbing state. Moreover, values $a''_i$ equal probabilities of reaching the corresponding states in MChG with the initial distribution $a$. Taking into account the remark in the previous paragraph, we assume that the constant $c_2$ in Inequality (5.3) is defined just for the initial distribution $a''$. In addition, if earlier $p_i(t) = \Omega(t^{-\delta_i})$, $i = 1, 2$, then we denote by $c'_1, c'_2 > 0$ constants such that $p_1(t) > c'_1 t^{-\delta_1}$ and $p_2(t) > c'_2 t^{-\delta_2}$.
As was noted earlier (before the proof of Lemma 4.3), the assertion about the power estimates of the function \( p(t) \) is equivalent to an analogous assertion for the function \( Q(q) \). Let us first consider the case \( \delta_1 \neq \delta_2 \).

First of all, note that any word \( w \) in the initial Markov chain is representable in the form \((w_1, w_2)\), where \( w_i, i = 1, 2 \), are words of the Markov chain with the graph \( G_i \). Here, as one can easily see, \( \text{Pr}_{G_i}(w) = \text{Pr}_{G_1}(w_1) \text{Pr}_{G_2}(w_2) \) (the subscript at the symbol \( \text{Pr} \) indicates the graph of the Markov chain, where we consider the word).

Evidently, \( \text{Pr}_{G_i}(w_i) = p_i(t_i) \), where \( t_i \) is the number of the word \( w_i \) in the corresponding list. Assuming that \( \delta_1 > \delta_2 \), we get (below \( t_1, t_2 \) run over all possible positive integer values):

\[
Q(q) = \left| \{(t_1, t_2) : p_1(t_1)p_2(t_2) \geq q\} \right| \leq \left| \{(t_1, t_2) : c_1 t_1^{-\delta_1} c_2 t_2^{-\delta_2} \geq q\} \right| = \left| \{(t_1, t_2) : t_1^{\delta_1} t_2^{\delta_2} \leq (c_1 c_2)/q\} \right| \leq \sum_{t_1 = 1}^{\infty} \left( (c_1 c_2)/q \right)^{-\delta_1/\delta_2} t_1^{\delta_1/\delta_2} = \text{const} \, q^{-1/\delta_2}.
\]

In the case \( \delta_1 = \delta_2 = \delta \), analogous considerations lead to the inequality

\[
Q(q) \geq \left| \{(t_1, t_2) : t_1 t_2 \leq (c_1 c_2)/q\} \right|.
\]

According to the Dirichlet formula for the divisor function \([19, \text{Chapter XII}]\), the number of points with positive integer coordinates, whose product does not exceed \( N \), equals \( N \ln N + (2\gamma - 1)N + O(\sqrt{N}) \), where \( \gamma \) is the Euler constant. Therefore, the inequality \( Q(q) \leq \text{const} \, q^{-1/\delta} \) can be fulfilled with small \( q \) with no positive constant, which was to be proved. \( \square \)

### 5.2. Completion of the proof of Theorem 2.1.B.

We prove that \( p(t) = \Theta(t^{-1/\beta}) \) under conditions of Theorem 2.1.B by induction with respect to the length of the maximal path in the graph \( G' \). If the graph \( G' \) consists of unconnected vertices, then the assertion of the Theorem follows from Remark 5.2 and Lemma 4.3. Otherwise, we represent the graph \( G \) as a “sequential” connection of the graph \( G_1 \) consisting of SCCs corresponding to initial vertices of the graph \( G' \) (vertices without incoming arcs), and the graph \( G_2 \) consisting of the part of the graph \( G \). Applying Lemma 5.3 (and the induction hypothesis for the graph \( G_2 \)), we obtain \( p(t) = O(t^{-1/\beta}) \). Consequently (see Corollary 4.4), \( p(t) = \Theta(t^{-1/\beta}) \).

Let us prove the necessity of conditions for the power asymptotics in Theorem 2.1.B. Assume the contrary. Consider a path in the graph \( G' \) with exactly two vertices corresponding to graphs \( H_1 \) and \( H_2 \) for which \( P_{H_1}(\beta) \) and \( P_{H_2}(\beta) \) have unit characteristic values. We can choose \( H_1 \) such that any path in the graph \( G' \) beginning at \( H_1 \) contains no more than one such vertex of \( H_2 \). Really, otherwise there exists
a path $G'$ beginning at $H_2$ that contains a vertex of $H_3$, where $P_{H_3}(\beta)$ has the unit characteristic value. Then we can choose for $H_1$ the former graph $H_2$ (and do $H_3$ for $H_2$), and so on till the desired condition is fulfilled.

Consider an initial distribution $\alpha$ (not necessarily satisfying Condition (2.1)) concentrated at vertices of the graph $H_1$. Let $\tilde{G}$ be the part of the graph $G$ reachable from these vertices. According to Lemma 4.1, the necessity of conditions of the power order for $\text{MCh}_G$ automatically implies its necessity for $\text{MCh}_G$.

The graph $\tilde{G}$ is representable as a “sequential” connection of the graph $G_1 \equiv H_1$ and the graph $G_2$ consisting of the rest of the graph $\tilde{G}$. As was proved above, for the graph $G_2$ it holds $p_2(t) = \Theta(t^{-1/\beta})$. Analogous inequalities $p_1(t) = \Theta(t^{-1/\beta})$ for the graph $G_1 \equiv H_1$ are proved in Lemma 4.3. Applying the final part of Lemma 5.3, we conclude that conditions of the power order with the exponent $-1/\beta$ cannot be fulfilled for $\text{MCh}_G$ and, consequently, for $\text{MCh}_G$. 

6. Proof of Theorem 2.1.C. Evidently, one can associate any word $w$ with a word $w'$ obtained from $w$ by deleting cycles, and a collection of nonnegative numbers $(k_1, \ldots, k_{|C(w')|})$, where $k_i$ is the number of bypasses of the $i$th cycle in the path corresponding to the word $w$. Thus, for example, the word for the MCh shown in Fig. 2.1 c) $(E_1, E_1, E_1, E_2, E_2, E_0)$ which corresponds to the word $w' = (E_1, E_2, E_0)$ and the vector (2, 1); the word $(E_1, E_1, E_1, E_0)$ for the MCh in Fig. 2.1 d) corresponds to the word $w' = (E_1, E_0)$ and the vector (2) (here $|C(w')| = 1$); the word for the MCh in Fig. 2.1 e) $(E_1, E_2, E_1, E_2, E_0)$ corresponds to the word $w' = (E_1, E_2, E_0)$ and the vector (1) (here we also have $|C(w')| = 1$).

**Proposition 6.1.** Under assumptions of Theorem 2.1.C the correspondence between all words $w$ and pairs $w'$ ($w' \in W'$), together with the vector $(k_1, \ldots, k_{|C(w')|})$, is biunique.

**Proof.** According to conditions of Theorem 2.1.C, the case, when going along the path that corresponds to some word $w$ we first encounter a cycle $c$, then $c'$, and then again to $c$ is impossible; otherwise it would mean that $c$ contains a vertex belonging to two cycles. Therefore, both the word $w'$, and the order of cycles are defined uniquely. Note that one can define the letter followed by a cycle in various ways. Thus, one can insert a cycle in the word $w' = (E_1, E_2, E_3)$ for the MCh shown in diagram c) in Fig. 2.1 both after $E_1$ and after $E_2$. Moreover, since the cycle is bypassed uniquely, we obtain one and the same word $w$ (in our example, this word is $w = (E_1, E_2, E_1, E_2, E_0)$).

Let $C(w') = \{c_1, \ldots, c_m\}$, $m = |C(w')|$. Proposition 6.1 implies that one can find $\ln \Pr(w)$ as $\ln \Pr(w') + \sum_{i=1}^{m} k_i \ln \Pr(c_i)$. The number of distinct words $w$ of the mentioned type, whose probability is not less than $x$, $(x \in (0, 1))$, equals the number
non-negative integer vectors $k$ solving the inequality

\[(6.1) \ln \Pr(w) - \ln x \geq \sum_{i=1}^{m} k_i \ln \Pr(c_i). \]

Evidently, each $k_i$ can be bounded by the range from 0 to $\lfloor (\ln \Pr(w') - \ln x) / \ln \tilde{\Pr}(c) \rfloor$. However, the inequality is not necessarily fulfilled for all values that belong to the obtained integer rectangular parallelepiped, but only for their part of $1/m!$ that lies inside the simplex. To put it more precisely, the number of such words $w$ which are obtained from $w'$ differs from

\[(6.2) 1/m! \prod_{c \in C(w')} \left( \ln x / \ln \tilde{\Pr}(c) \right) \]

at most by the value of the order $O((-\ln x)^{m-1})$. Here $O((-\ln x)^{m-1})$ is the number of points on the simplex boundary which is defined either by the equality to zero of one of values $k_i$ or by the replacement of the inequality sign in (6.1) by the equality sign. The sum of (6.2) over all $w' \in W'$ is

\[Q(x) = \sum_{w' \in W':|C(w')|=D} \frac{1}{D!} \prod_{c \in C(w')} \frac{-\ln x}{-\ln \Pr(c)} + O((-\ln x)^{D-1}). \]

Let $y = (-\ln x)^{D/\nu}$. We have

\[(6.3) Q(\exp(-\sqrt[\nu]{y})) = y + O(y^{(D-1)/D}). \]

The right-hand side of the equality (6.3) is some positive integer number $t$. Expressing $y$ in terms of $t$ and applying the function $p(\cdot)$ to both sides of (6.3), we get $p(t) = \exp(-\sqrt[\nu]{t/D} + O((t^{(D-1)/D})))$. We have obtained the latter identity for “thinned” positive integers $t$, i.e., all possible values of the function $Q$. However, in view of (6.3) the difference of distinct neighboring values of the function $Q$ does not exceed $O(t^{(D-1)/D})$. This means that the obtained bound for $p(t)$ is valid for all positive integers $t$.

One can easily make sure that $\sqrt[\nu]{t/D} + O(t^{(D-1)/D}) - \sqrt[\nu]{t} = O(1)$ (here by replacing $O(t^{(D-1)/D})$ with const $t^{(D-1)/D}$ we show that the limit of the difference is finite). Consequently, $p(t) = \exp(-\sqrt[\nu]{t} + O(1)) = \Theta(\exp(-\sqrt[\nu]{t}))$. \[\Box\]

7. Conclusion. We have proved a subexponential order of the asymptotics for $p(t)$ in the case when all SCCs of the graph of an MCh are cycles. For the case when the graph of an MCh contains more nontrivial SCCs, we have established necessary and sufficient conditions for a power order of the asymptotics. These conditions are fulfilled, in particular, if the eigenvector corresponding to the unit eigenvalue...
of the matrix $P(\beta)^T$ is positive. In the latter case the proof is based on the recurrent Correlation (4.1) and the inequality for its solution. An analogous technique was used in [3] for the case of independent random variables, i.e., the case when rows of the matrix $P$ coincide.

The problem considered in the paper [9] formally is different. Let $G$ be a directed graph, and let each its arc have a positive weight, Consider all possible paths that begin at the vertex $v_1$ and end at $v_2$. Let us sort their list in the increasing order of their weights. Denote the weight of the path of number $r$ by $p_r$. Without loss of generality we assume that

$$\text{for each vertex } v \text{ of the graph } G \text{ there exists a path from } v_1 \text{ to } v_2 \text{ going through } v.$$  

Let $G$ contains nontrivial SCCs. In [9] one has proved the existence of limits

$$\lim_{r \to \infty} \frac{p_r^D}{r^D}, \text{ if all nontrivial SCCs in } G \text{ are cycles; }$$

$$\lim_{r \to \infty} \frac{p_r}{\ln r}, \text{ else.}$$

In the formula (7.2), $D$ is the maximal number of cycles which can belong to a path from $v_1$ to $v_2$.

The main idea of the proof in [9] differs from that proposed by us, namely, it consists in studying the basic case of unit weights of all arcs, where the result follows from properties of the adjacency matrix of the graph $G$. By dividing an arc into $N$ parts one can reduce the case of rational weights of arcs to the case of unit arcs, where $N$ is the least common multiple of all denominators. The case of irrational weights of arcs is obtained by the estimation of the accuracy of the rational approximation.

One can easily apply results obtained in [9] to the sorted list of trajectories of an MCh. Assume that the weight of the arc $(i,j)$ of the graph of an MCh equals $-\ln p_{ij}$, $i,j = 1,\ldots,n$. Without loss of generality, we can also reduce the case, when some vertices are origins of arcs with zero weights, to the case considered in the paper [9]. Really, no other arcs originate from such vertices and we can subtend such arcs by identifying their endpoints (see Fig. 7.1). Let us now assume that the initial distribution of the MCh is concentrated at a vertex $v_1$, and we can reach the absorbing state only from a vertex $v_2$. Then, evidently, the weight of the $r$th path in the list sorted in increasing order coincides with $-\ln p(r)$ and we can apply for the function results obtained in [9].

Moreover, results of the paper [9] are also applicable in a general case, when the initial distribution is concentrated at several states (vertices), and one can reach the absorbing state from several vertices. In this case, the total list of all words is the
Fig. 7.1. Transformation of a graph with zero weight arcs to a standard case.

union of lists of words which begin with certain initial letters (states) $E_{v_1}$ and end with letters $E_{v_2}$. The fact that multipliers $a_{v_1}$ and $p_{v_2}$ are used for determining probabilities of words does not affect the asymptotics of their logarithms. By using the asymptotics of sublists of words one can find the asymptotics of the function $\ln p(t)$ with the help of Lemma 4.2 in [9]. Therefore, in the paper [9] one has actually proved that

under conditions of Theorem 2.1.B the limit $\lim_{t \to \infty} - \ln t / \ln p(t) = \beta$ exists;
under conditions of Theorem 2.1.C the limit $\lim_{t \to \infty} (\ln p(t))^D / t = \nu$ exists.

In other words, one has proved the existence of weak power and weak subexponential asymptotics. Note that formulas for constants $\beta$ and $\nu$ obtained in [9] are less convenient than our ones given in Theorem 2.1 (in particular, in [9] the constant $\nu$ is calculated by a recurrent correlation rather than by an explicit formula). However, by easy transformations, we can reduce formulas [9] to our ones.

Let us now discuss the applicability of results obtained in Theorem 2.1 to studying the asymptotics of the sorted list of all paths in the graph that begin at the vertex $v_1$ and end at $v_2$. Without loss of generality, we can assume that there are no multiple edges, because otherwise one can divide them into several parts (see [9]). Note that we impose certain conditions on the matrix $P$ (see (3.1)), while in the paper [9] weights of arcs are arbitrary positive values. However, the multiplication of these weights by a fixed constant trivially affects the asymptotics of $p_v$. In addition, as was noted, without loss of generality, we can assume that Condition (7.1) is fulfilled. Therefore, it suffices to consider only the case when the matrix $P_G$ of inverse potentiated arc weights satisfies Condition (3.1).

Note that if in (3.1) instead of the substochasticity condition, we bound the sum of elements of rows from above with some constant (assuming that $P_{ij} < 1$), then (in accordance with a lemma analogous to Lemma 3.1) only the upper boundary of the range of $\beta$ will change. Namely, it will take on the minimal value of $\psi$ ensuring the substochasticity of the matrix $P_G(\psi)$. 
Within the comparison of our results with those obtained in [9], instead of the asymptotics of the function $p(t)$ constructed from the list of all words, we are interested in the asymptotics of an analogous function constructed from the sublist of words that begin at $v_1$ and end at $v_2$, independently of the initial distribution $a$ (neither the notion of the initial distribution, nor that of the absorbing state is defined in terms of the paper [9]). We considered such sublists in Lemma 4.3; in fact, we have proved the power asymptotics for these sublists. Evidently, the obtained results will remain valid even if we neglect $a_{v_1}$ when calculating word probabilities. Therefore, the case when the matrix $P_G(\beta)^T$ has a positive eigenvector (in particular, the case when the graph $G$ is strongly connected) is, in fact, already considered by us.

In a general case, for studying the asymptotics of $p(t)$ we used Lemmas 5.1 and 5.3 on the parallel and sequential connection of graphs of MCh. However, in fact, these lemmas can be formulated as assertions on the union and composition of lists stated in terms of [9]. Here we do not redefine auxiliary results, but give only the statement of the final Theorem.

**Theorem 7.1.** Let $G$ be an arbitrary directed graph, possibly, having loops, but having no multiple arcs. We assume that each edge has a positive weight, Condition (7.1) is fulfilled, and $p_r$ is the weight of the $r$th path from $v_1$ to $v_2$ in their list sorted in the increasing order of their weights.

A. If the graph $G$ is acyclic, then the list is finite.

B. If the graph $G$ contains nontrivial SCCs different from a cycle, then we have $\exp(-p_r) = \Omega(r^{-1/\beta})$, where $\beta$ is a real number, with which the maximal modulo eigenvalue of the matrix $P_G(\beta)$ equals one. Note that such $\beta$ exists, is unique and positive. Moreover, $\exp(-p_r) = o(r^{-1/\beta'})$ for any $\beta' > \beta$. Finally, $\exp(-p_r) = \Theta(r^{-1/\beta})$ is attained if and only if any simple path from $v_1$ to $v_2$ goes through at most one SCC $H$ such that the matrix $P_H(\beta)$ has the unit eigenvalue.

C. If the graph $G$ contains no SCCs different from a cycle, then $\exp(-p_r) = \Theta(\exp(-\sqrt[4]{\nu}))$; here $\nu$ is determined by the formula

$$1/\nu = \sum_{w \in W: |C(w)| = D} 1/D! \prod_{c \in C(w)} 1/\tilde{p}(c),$$

where $W$ is the set of all simple paths from $v_1$ to $v_2$, $C(w)$ are cycles, whose vertices are encountered in such a path $w$, $\tilde{p}(c)$ is the weight of the cycle $c$, and $D = \max_{w \in W} |C(w)|$.

One can easily see that this theorem implies results of [9], i.e., the existence of limits (7.2) and (7.3). Evidently, the converse assertion is not true.
At the end part of the paper [9], one discusses areas of further research. Let us mention one more area, namely, the generalization of the obtained results for the case of hidden Markov models. We have succeeded in studying various particular cases [5], but establishing general formulas has appeared to be a rather complicated problem.

Acknowledgment. We are grateful to Yu.A. Al’pin for remarks which were useful for improving the proof in Section 3. We are also grateful to G.R. Nugmanova for multiple discussions and remarks related to the proof of Theorem 2.1.C. We wish to thank the anonymous reviewer for their insightful and helpful suggestions.

REFERENCES