Hindawi
International Journal of Mathematics and Mathematical Sciences
Volume 2017, Article ID 9143747, 11 pages
https://doi.org/10.1155/2017/9143747

Research Article

Calculation of Precise Constants in a Probability Model of Zipf's Law Generation and Asymptotics of Sums of Multinomial Coefficients

Vladimir Bochkarev and Eduard Lerner

Kazan (Volga Region) Federal University, 18 St. Kremlevskaya, Kazan 420008, Russia

Correspondence should be addressed to Eduard Lerner; eduard.lerner@gmail.com

Received 18 January 2017; Accepted 28 March 2017; Published 7 May 2017

Academic Editor: Niansheng Tang

Let $\omega_0, \omega_1, \ldots, \omega_n$ be a full set of outcomes (symbols) and let positive p_i, $i = 0, \ldots, n$, be their probabilities ($\sum_{i=0}^{n} p_i = 1$). Let us treat ω_0 as a stop symbol; it can occur in sequences of symbols (we call them words) only once, at the very end. The probability of a word is defined as the product of probabilities of its symbols. We consider the list of all possible words sorted in the nonincreasing order of their probabilities. Let $p(r)$ be the probability of the rth word in this list. We prove that if at least one of the ratios $\log p_i / \log p_j$, $i, j \in \{1, \ldots, n\}$, is irrational, then the limit $\lim_{r\to\infty} p(r)/r^{-1/\gamma}$ exists and differs from zero; here γ is the root of the equation $\sum_{i=1}^{n} p_i^{\gamma} = 1$. The limit constant can be expressed (rather easily) in terms of the entropy of the distribution $(p_1^{\gamma}, \ldots, p_n^{\gamma})$.

1. Introduction: The Statement of the Main Theorem

1.1. Brief Literature Overview. The wide presence of power laws in real networks, biology, economics, and linguistics can be explained in the framework of various mathematical models (see, e.g., [1, 2]). According to Zipf's law [3], in a list of word forms ordered by the frequency of occurrence, the frequency of the rth word form obeys a power function of r (the value r is called the rank of the word form). One can easily explain this law with the help of the so-called monkey model.

Recall that the word forms "the"; "of"; and "and" are used most frequently in English texts. According to Zipf's law, the word "the" is used in the texts twice as much as "of" and three times as much as "and"; in other words the word form occurrence frequency obeys the power function of rank r (the position number of the word form in an ordered frequency list) whose exponent is approximately -1. It should be noted that further surveys showed that Zipf's law is roughly realised only for the most frequent words. At present, the researches try to describe the main part of the lexicon using the power law with an exponent $-\alpha$, where $\alpha > 1$. Zipf explained his law on the basis of the principle of least effort. In accordance with this principle, the authors aim to minimise the length of the text, which is required to convey their thoughts, even if this introduces ambiguities. On the other hand, readers want to minimize the effort required to understand the text [4].

Another explanation of Zipf's law was suggested by Mandelbrot who slightly modified the law by introducing translation constant [5] into the argument of the power function. The important thing for our case is that later he hypothesized the existence of more simple explanation of the Zipf law associated with a simple probability model when all symbols in the text (including white-space) appear independently of each other with certain probability. Moreover, he analysed the Markovian dependence between these symbols and wrote out the correct (in a typical case) formula on the basis of special cases to determine the parameter α by the transition probabilities matrix in the Markov model [6].

First, we will consider the model thoroughly described by Miller [7] and Li [8] for a special case of Mandelbrot's experiment in which the monkey types the keys with uniform probability. To learn some other important references on the monkey model, we recommend to read the recent article by Richard Perline and Ron Perline [9] (see also references in the next subsection).

1.2. Statement of the Main Theorem and Its Connection with Other Results. Assume that a monkey types any of 26 Latin letters or the space on a keyboard with the same probability of 1/27. We understand a word as a sequence of symbols typed by the monkey before the space. Let us sort the list of possible words with respect to probabilities of their occurrence (the empty word, whose probability equals 1/27, will go first in this list followed by 26 one-letter words whose probabilities equal $1/27^2$ and then by 26^2 possible two-letters words and so on). We can prove (see [7, 8]) that the probability $p(r)$ of a word with the rank of r satisfies the inequality

$$c_1 r^{-\alpha} < p(r) < c_2 r^{-\alpha}, \tag{1}$$

where $\alpha = \log 27/\log 26$ and $c_1, c_2 > 0$ (here and below we use the symbol log if the base of the logarithm is not significant; but for the natural logarithm we use the symbol ln).

Relatively recently inequality (1) was generalized to the case of nonequiprobable letters. Let p_0 be the probability that the monkey types the space, let p_i, $i = 1, \ldots, n$, denote probabilities of choosing the ith letter from the set of n letters ($p_i > 0$, $\sum_{i=0}^{n} p_i = 1$), and let $p(r)$ be, as above, the probability of a word with a rank of r. Then, as is proved in [10, 11], the following inequality analogous to (1) takes place; namely, $\exists c_1, c_2 : 0 < c_1 < c_2$, such that

$$c_1 r^{-\alpha} < p(r) < c_2 r^{-\alpha}, \quad \text{where } \alpha = \frac{1}{\gamma} \tag{2}$$

and γ is the root of the equation $\sum_{i=1}^{n} p_i^{\gamma} = 1$ (evidently, $0 < \gamma < 1$). Note that inequality (2) is equivalent to the boundedness of the difference $-\log p(r) - \alpha \log r$.

In the case when the probability of each letter is not fixed but depends on the previous one, words represent trajectories of a Markov chain with the absorbing state ω_0 and transient states $\omega_1, \ldots, \omega_n$. Then the value $p(r)$ is the probability of the rth trajectory in the list of possible trajectories sorted in the nonincreasing order of probabilities. In this case, the asymptotic behavior of $p(r)$ does not necessarily have a power order. Namely, in this case one of the two alternatives takes place [12, 13]. The first variant is that there exists the limit

$$\lim_{r\to\infty} \frac{-\log p(r)}{r^{1/m}} = c, \quad c > 0, \tag{3}$$

where m is some positive integer constant value that depends on the structure of the transition probability matrix and the structure of states, where the initial distribution of the Markov chain is concentrated. The second variant is that independently of the initial distribution there exists the following nonzero limit (the so-called *weak* power law):

$$\lim_{r\to\infty} \frac{-\log p(r)}{\log r}. \tag{4}$$

This limit equals $1/\gamma$, where γ is now defined with the help of the substochastic matrix P of transition probabilities where the row and the column that correspond to the absorbing state ω_0 are deleted. Namely, raising all elements of the mentioned matrix to the power of γ would equate its spectral radius to 1.

These results were obtained independently in [12, 14] and later refined in [13]. Namely, as appeared, the first alternative means the subexponential order of the asymptotics; that is, in this case $\exists c_1, c_2 : 0 < c_1 < c_2$, such that

$$c_1 \exp\left(-cr^{-1/m}\right) < p(r) < c_2 \exp\left(-cr^{-1/m}\right). \tag{5}$$

The case of the second alternative is much more difficult. If the matrix P does not have the block-diagonal structure with coinciding powers such that raising elements of blocks to these powers makes the spectral radius equal 1, then one can replace the weak power law with a *strong* one. Namely, in this case the asymptotic behavior of $p(r)$ has the power order; that is, inequality (2) is valid (with "matrix" γ defined above). Therefore, inequality (2) takes place in a "typical" case of letter probabilities.

However, one more natural question still remains without an answer.

Inequality (2) means that the asymptotic form has a power *order* but does not imply the *exact* power asymptotics. In a general case, as follows from the first example given in this section, useful properties can be established neither when letters in words are Markov-dependent nor when they are independent. However, as we prove later in this paper, in a "typical" case, for words composed of independent letters, the asymptotic behavior of the function $p(r)$ is exact power. The following theorem is valid.

Theorem 1 (main). *Let at least one of the ratios* $\log p_i/\log p_j$, $i, j \in \{1, \ldots, n\}$, *be irrational and let* γ *be the root of the equation* $\sum_{i=1}^{n} p_i^{\gamma} = 1$. *Then the limit*

$$\lim_{r\to\infty} \frac{p(r)^{-\gamma}/p_0^{-\gamma}}{r} \tag{6}$$

exists and equals $H(\mathbf{p}^{\gamma})$, *where* $H(\mathbf{p}^{\gamma})$ *is the entropy of* $\mathbf{p}^{\gamma} = (p_1^{\gamma}, \ldots, p_n^{\gamma})$; *that is,* $H(\mathbf{p}^{\gamma}) = -\gamma \sum_{i=1}^{n} p_i^{\gamma} \ln p_i$.

Here and below we always write the function under consideration in the numerator and do the norming (defined analytically) function in the denominator of the fraction, whose limit is to be calculated. In intermediate calculations it may be more convenient to do the opposite, but since this results only in the trivial raising of the limit constant to the power of -1, we sacrifice the convenience of calculations for the clarity of statements of results. Evidently, the theorem asserts that under certain assumptions there exists the nonzero limit $p(r)/r^{-\alpha}$ (where $\alpha = 1/\gamma$) as $r \to \infty$. It is equal to $p_0 H(\mathbf{p}^{\gamma})^{-1/\gamma}$.

Let us describe the structure of the remaining part of the paper. In Section 2 we state the main theorem in terms of multinomial coefficients (of the Pascal pyramid). The proof of the theorem is reduced to the estimation of the limit behavior of the sum of these coefficients over some simplex. In Section 3 we prove an analog of this theorem with an integral in place of the sum. In this section we essentially use the Stirling formula which allows us to reduce calculations to the evaluation of a multivariate Gaussian integral. We establish an explicit formula for the determinant of the matrix

of the quadratic form that defines the integrand. Finally, in Section 4 we prove that the ratio of the integral to the sum tends to 1. Here we use the general properties of the Riemann integral and uniformly distributed sequences. In conclusion we discuss possible generalizations and unsolved problems.

2. Equivalent Statements of the Main Theorem and the Pascal Pyramid

Let us first note that if $p_0 \to 0$, then $\gamma \to 1$. Reducing the nominator of fraction (6) by $p_0^{-\gamma}$, we write the following statement in this case:

Theorem 2 (the case of $\gamma = 1$). *Let $p_i > 0$ be the probability of the symbol ω_i, $i = 1, \ldots, n$, while $\sum_{i=1}^{n} p_i = 1$ (there is no stop symbol). Assume that at least one of the ratios $\log p_i / \log p_j$, $i, j \in \{1, \ldots, n\}$, is irrational. Let us consider all possible finite words (including the empty one) and sort them in the nonincreasing order of probabilities (we equate the probability of the empty word to 1 and calculate the probability of any other word as the product of probabilities of its letters). Let $p(r)$ be the probability of the rth word in the list (the word with the rank of r). Then the limit $\lim_{r\to\infty} p(r)/r^{-1}$ exists and equals $H^{-1}(\mathbf{p})$, where $H(\mathbf{p})$ is the entropy of the vector $\mathbf{p} = (p_1, \ldots, p_n)$; that is, $H(\mathbf{p}) = -\sum_{i=1}^{n} p_i \ln p_i$.*

In the statement of Theorem 2, as well as in Theorem 1, we use the bold font for the vector whose components are denoted by the same letter with the index ranging from 1 to n. In what follows we use the bold font for analogous denotations without mentioning this fact.

One can easily see that Theorem 2 is not just a particular case of Theorem 1, but these theorems are equivalent. Namely, the replacement of probabilities p_i^{γ} with new ones p_i turns the general case into the particular one. Therefore, in what follows we neglect p_0, assuming (without loss of generality) that $\sum_{i=1}^{n} p_i = 1$.

Fix some probability $q \in (0, 1]$ and denote by $Q(q)$ the rank of the last word whose probability is not less than q in the list of all words sorted in the nonincreasing order of their probabilities. Let us redefine the function $p(r)$ for noninteger r as $p(r) = p(\lfloor r \rfloor)$ (here $\lfloor \cdot \rfloor$ is the integer part of a number). Evidently, functions $q = p(r)$ and $r = Q(q)$ $(q \in (0, 1],\ r \geq 1)$ are inverse (more exactly, quasi-inverse); namely, the graph of one of the hyperbola-shaped, decreasing stepwise functions turns into another one when axes r and q switch roles (in the first case, q is the argument and r is the value and vice versa in the second case).

It can be clearly seen that $\lim_{r\to\infty} cp(r)/r^{-1} = 1$ is equivalent to

$$\frac{\lim_{q\to 0} c^{-1} Q(q)}{q^{-1}} = 1. \tag{7}$$

Therefore the equality in the assertion of Theorem 2 is equivalent to that

$$\frac{\lim_{q\to 0} Q(q)}{q^{-1}} = H^{-1}(\mathbf{p}). \tag{8}$$

Denote the logarithm of the denominator in the last fraction by $z = -\ln q$ (i.e., $q = e^{-z}$) and let $\widetilde{Q}(z) = Q(e^{-z})$. In view of considerations in the above paragraph the equality in the assertion of Theorem 2 is equivalent to that

$$\lim_{z\to\infty} \left(\ln \widetilde{Q}(z) - z\right) = -\ln H(\mathbf{p}). \tag{9}$$

Recall the proof of inequality (2) in [11]. It is reduced to the proof of the boundedness of the difference $\ln \widetilde{Q}(z) - z$ for the introduced function $\widetilde{Q}(z)$ with $z \geq 0$. Nonnegative values of z form the definition domain of the function $\widetilde{Q}(z)$ because $q \leq 1 \Leftrightarrow z \geq 0$. For convenience we redefine the function $\widetilde{Q}(z)$ by putting $\widetilde{Q}(z) = 0$ for $z < 0$.

Let $a_i = -\ln p_i$. Considering all possible variants of the last letters in words, whose quantity equals the value of the function $\widetilde{Q}$, we obtain the functional equation $\widetilde{Q}(z) = \widetilde{Q}(z - a_1) + \cdots + \widetilde{Q}(z - a_n) + \chi(z)$, where χ is the Heaviside step (i.e., the function that vanishes with negative values of the argument and equals 1 with nonnegative values). For $z \geq M = \max\{a_1, \ldots, a_n\}$ we get the following recurrent correlation:

$$Q_n(z) = Q_n(z - a_1) + \cdots + Q_n(z - a_n), \tag{10}$$

where $Q_n(z) = \widetilde{Q}(z) + 1/(n-1)$.

The equality $\sum_{i=1}^{n} p_i = 1$ implies that the function $\operatorname{const} \exp z$ satisfies (10). Since the function $Q_n(z)$ takes a finite number of positive values within $[0, M]$ interval, there exist positive c_1 and c_2 such that

$$c_1 \exp z < Q_n(z) < c_2 \exp z \tag{11}$$

for all $0 \leq z \leq M$.

Replacing terms in the right-hand side of the recurrent correlation (10) with their lower (upper) bounds, we extend the solution set of inequality (11) to the domain $0 \leq z \leq M + m$, where $m = \min\{a_1, \ldots, a_n\}$. Repeating this procedure several times, in a finite number of steps we prove that the inequality is valid for any arbitrarily large z. Performing the logarithmic transformation of the inequality, we conclude that $\ln Q_n(z) - z$ is bounded, and then so is the difference $\ln \widetilde{Q}(z) - z$.

Let us return to Theorem 2. As was mentioned above, Theorem 2 asserts (under certain assumptions) not only the boundedness of $\ln \widetilde{Q}(z) - z$ but also the validity of equality (9). Let us recall the combinatory sense of the function $\widetilde{Q}$; it is mentioned in [11]. Evidently, all words that contain k_1 letters of the 1st kind, k_2 letters of the 2nd kind, …, and k_n letters of the nth kind have one and the same probability of $\Pr(\mathbf{k}) = p_1^{k_1} \cdots p_n^{k_n}$ (i.e., $-\ln \Pr(\mathbf{k}) = \sum_{i=1}^{n} k_i a_i$); ranks of these words are consecutive. The quantity of such words is defined by the multinomial coefficient

$$M(\mathbf{k}) = \frac{(k_1 + \cdots + k_n)!}{k_1! \cdots k_n!}. \tag{12}$$

Considering the nonnegative part of the n-dimensional integer grid and associating the point $(k_1, \ldots, k_n)$ with the number $M(k_1, \ldots, k_n)$, we get one of the variants of the Pascal

pyramid. By the definition of the function $\widetilde{Q}$ the value $\widetilde{Q}(z)$ equals the sum of multinomial coefficients $M(\mathbf{k})$ over all integer vectors $\mathbf{k}$ that lie inside the n-dimensional simplex $S(z) = \{\mathbf{x} : \mathbf{x} \geq 0,\ \sum_{i=1}^{n} a_i x_i \leq z\}$:

$$\widetilde{Q}(z) = \sum_{\mathbf{k}\in S(z)} M(\mathbf{k}). \tag{13}$$

As a result, we obtain one more equivalent statement of the main theorem, which we are going to prove.

Theorem 3 (the multinomial statement). *Let a_i, $i = 1, \ldots, n$, be arbitrary positive numbers such that at least one of the ratios a_i/a_j, $i, j \in \{1, \ldots, n\}$, be irrational and $\sum_{i=1}^{n} p_i = 1$, where $p_i = \exp(-a_i)$. Let a function $\widetilde{Q}$ obey formula (13). Then*

$$\lim_{z\to\infty} \frac{\widetilde{Q}(z)}{\exp(z)} = H^{-1}(\mathbf{p}), \tag{14}$$

where $H(\mathbf{p}) = \sum_{i=1}^{n} a_i p_i$.

3. The Proof of an Analog of Theorem 3 with Integration instead of Summation

3.1. Reduction of the Integration to the Calculation of a Gaussian Integral. The function $M(k_1, \ldots, k_n)$ is defined for integer nonnegative vectors $\mathbf{k}$. Let us redefine it for noninteger vectors by replacing (in this case) $x!$ in Definition (12) with $\Gamma(x+1)$. In what follows we use the denotation $M(x_1, \ldots, x_n)$ (or $M(\mathbf{x})$) for the corresponding function which is continuous for nonnegative x_i. Further we consider this function and study its properties only for such (nonnegative) x_i.

In this section we prove the following theorem.

Theorem 4 (on the integral). *Let a_i, $i = 1, \ldots, n$, be arbitrary positive numbers such that $\sum_{i=1}^{n} p_i = 1$, where $p_i = \exp(-a_i)$. Let a function $f(z)$ obey the formula $f(z) = \int_{\mathbf{x}\in S(z)} M(\mathbf{x}) d\mathbf{x}$, where $d\mathbf{x} = \prod_{i=1}^{n} dx_i$. Then*

$$\lim_{z\to\infty} \frac{f(z)}{\exp(z)} = H^{-1}(\mathbf{p}). \tag{15}$$

Proof. Let us first recall some evident properties of the integrand. Note that the existence of the (Riemann) integral of $f(z)$ over the compact set $S(z)$ evidently follows from the continuity of $M(\mathbf{x})$ in the domain under consideration.

If all components of the vector $(x_1, \ldots, x_n)$, possibly, except one component x_i, equal zero, then by definition we have $M(x_1, \ldots, x_n) \equiv 1$. Let us prove that otherwise the function $M(x_1, \ldots, x_n)$ is strictly increasing in x_i. Since the gamma function is positive definite, it suffices to prove that in this case the partial derivative of $\ln M(x_1, \ldots, x_n)$ with respect to x_i is positive. It equals

$$(\ln\Gamma)'(x_1 + \cdots + x_n + 1) - (\ln\Gamma)'(x_i + 1). \tag{16}$$

The positiveness of this difference follows from the fact that the function $(\ln\Gamma)'$ is increasing; this property, in turn, follows from the logarithmic convexity of the gamma function (it is well known [15] that $(\ln\Gamma)''(x) = \sum_{i=0}^{\infty} 1/(i+x)^2 > 0$ with $x > 0$).

The proved assertion implies that the function $M(\mathbf{x})$ attains its maximum in the domain $S(z)$ at the boundary $\langle \mathbf{a}, \mathbf{x}\rangle = z$, where $\langle \mathbf{a}, \mathbf{x}\rangle = \sum_{i=1}^{n} a_i x_i$. Let us calculate the exact asymptotics of the maximal value of the function $M(\mathbf{x})$ in the domain $S(z)$ with $z \to \infty$. For the vector $\mathbf{x}$ we denote by x the sum of its components and parameterize $\mathbf{x}$ by the value x and ratios $q_i = x_i/x$:

$$x_i = q_i x, \quad q_i \geq 0,\ i = 1, \ldots, n, \qquad \sum_{i=1}^{n} q_i = 1. \tag{17}$$

Let us use one simplest corollary of the Stirling formula [15], namely, the fact that with a nonnegative argument the value of the difference $\ln\Gamma(x+1) - (x\ln(x) - x + \ln(x+1)/2)$ is bounded. We obtain that, with any $x > 0$,

$$\ln M(x_1, \ldots, x_n) = xH(\mathbf{q}) + O(\ln(x+1)), \tag{18}$$

where $H(\mathbf{q}) = -\sum_{i=1}^{n} q_i \ln q_i$ (this correlation is closely connected with the so-called entropy inequality for multinomial coefficients).

We seek for the maximum of this function with $z \to \infty$ under one additional condition (namely, the requirement that the maximum is attained at the boundary) $\langle \mathbf{a}, \mathbf{x}\rangle = z$, where $a_i = -\ln p_i$, $0 < p_i < 1$, and $\sum_{i=1}^{n} p_i = 1$. Since $a_i > 0$, we get $O(\ln(x+1)) = O(\ln z)$. Moreover, the condition $\langle \mathbf{a}, \mathbf{x}\rangle = z$ with mentioned x_i gives the correlation

$$x = zH(\mathbf{q}; \mathbf{p})^{-1}, \tag{19}$$

where $H(\mathbf{q}; \mathbf{p}) = \sum_{i=1}^{n} a_i q_i = -\sum_{i=1}^{n} q_i \ln p_i$. Substituting this expression in (18), we conclude that the maximum of $\ln M$ (accurate to $O(\ln z)$) is attained at a vector $\mathbf{q}$ such that the fraction $H(\mathbf{q})/H(\mathbf{q}; \mathbf{p})$ takes on the maximal value. Recall that the difference $H(\mathbf{q}; \mathbf{p}) - H(\mathbf{q})$ takes on only nonnegative values and is called the Kullback–Leibler distance (divergence) $D(\mathbf{q} \mid \mathbf{p})$ between distributions $\mathbf{q}$ and $\mathbf{p}$ (see [16]). The minimum of this difference is attained at only one value of $\mathbf{q} = \mathbf{p}$; evidently, an analogous assertion is also true for $H(\mathbf{q}; \mathbf{p})/H(\mathbf{q})$: if $\mathbf{q} \neq \mathbf{p}$

$$\frac{H(\mathbf{q})}{H(\mathbf{q}; \mathbf{p})} < 1. \tag{20}$$

Consequently, the maximum of the function $\ln M(\mathbf{x})$ in the domain $S(z)$ is attained (accurate to $O(\ln z)$) at the intersection of the hyperplane $\langle \mathbf{a}, \mathbf{x}\rangle = z$ with the straight line $x_i = p_i x$, $i = 1, \ldots, n$, where it equals $z + O(\ln z)$.

Let us now immediately prove Theorem 4. Note first that by using the L'Hopital rule we can reduce the proof to that of the formula obtained by differentiating $f(z)/\exp(z)$ numerator and denominator with respect to z and to the proof of the equality

$$\lim_{z\to\infty} \frac{\widehat{f}(z)}{\exp(z)} = H^{-1}(\mathbf{p}), \tag{21}$$

where $\widehat{f}(z) = \int_{\mathbf{x}\geq 0} M(\mathbf{x})\delta(z - \langle \mathbf{a}, \mathbf{x}\rangle)d\mathbf{x}$ and $\delta(\cdot)$ is the delta function.

Let ε be a real arbitrarily small positive value. Denote by Λ_ε the sector consisting of points $\mathbf{x}$, $x_i = q_i x$, and $\sum_i q_i = 1$, such that

$$p_i - \varepsilon < q_i < p_i + \varepsilon, \quad i = 1, \ldots, n. \tag{22}$$

With fixed z on the hyperplane $\langle \mathbf{a}, \mathbf{x}\rangle = z$ correlations (18) and (19) take the form

$$\ln M(\mathbf{x}) = \frac{zH(\mathbf{q})}{H(\mathbf{q};\mathbf{p})} + O(\ln(z)). \tag{23}$$

Let us now strengthen inequality (20); namely, let us prove that if for $\mathbf{q}$ correlations (22) are violated, then

$$\frac{H(\mathbf{q})}{H(\mathbf{q};\mathbf{p})} < 1 - C_1(\mathbf{p})\varepsilon^2, \tag{24}$$

where $C_1(\mathbf{p})$ is a positive constant independent of $\mathbf{q}$.

Since $H(\mathbf{q};\mathbf{p})$ is a convex combination of $-\ln p_i$, it evidently is bounded:

$$0 < \min_i(-\ln p_i) \leq H(\mathbf{q};\mathbf{p}) \leq \max_i(-\ln p_i). \tag{25}$$

Consequently, formula (24) is equivalent to the inequality

$$H(\mathbf{q};\mathbf{p}) - H(\mathbf{q}) = D(\mathbf{q} \mid \mathbf{p}) > C_2(\mathbf{p})\varepsilon^2. \tag{26}$$

The latter correlation follows from the well-known property of the Kullback-Leibler divergence

$$D(\mathbf{q} \mid \mathbf{p}) \geq \frac{1}{4}\left(\sum_{i=1}^{n}|p_i - q_i|\right)^2 \tag{27}$$

(see, e.g., lemma 3.6.10 in [16]).

The proved inequality (24) (in view of formula (23)) implies that outside the domain Λ_ε the function $M(\mathbf{x})$ is exponentially small in comparison to the maximal value inside the domain which equals $\exp(z)$. More precisely, with $x \notin \Lambda_\varepsilon$ and $\langle \mathbf{a}, \mathbf{x}\rangle = z$, we get

$$M(\mathbf{x}) < \exp\left\{\left(1 - C\varepsilon^2\right)z\right\} \quad \text{for some } C > 0. \tag{28}$$

Note that the condition of the exponential smallness in comparison to $\exp z$ remains valid, even if ε depends on z and tends to 0 as z increases, though not too fast. In what follows we assume that

$$\varepsilon = \varepsilon(z) = z^{-1/2+\delta}, \quad \text{where } \delta > 0 \text{ is sufficiently small.} \tag{29}$$

One can easily see that the same exponential upper bound as in (28) also takes place not only for M function but also for its integral over the domain whose volume grows according to a power law:

$$\int_{\mathbf{x}\notin\Lambda_{\varepsilon(z)},\mathbf{x}\geq 0} M(\mathbf{x})\,\delta(z - \langle \mathbf{a}, \mathbf{x}\rangle)\,d\mathbf{x} < \exp\left\{\left(1 - C\varepsilon^2\right)z\right\} \tag{30}$$

with $z \to \infty$. Therefore in limit (21) we can treat $\widehat{f}(z)$ as the integral

$$\int_{\mathbf{x}\in\Lambda_{\varepsilon(z)}} M(\mathbf{x})\,\delta(z - \langle \mathbf{a}, \mathbf{x}\rangle)\,d\mathbf{x}. \tag{31}$$

Let us define the asymptotics (18) of the function $M(\mathbf{x})$ in the domain $\Lambda_{\varepsilon(z)}$ more precisely. Let us use the standard Stirling formula, namely, the fact that with $x \to \infty$ it holds that $\ln\Gamma(x+1) = x\ln(x) - x + \ln(x)/2 + \ln(2\pi)/2 + R(x)$, where $0 < R(x) < 1/(12x)$. We obtain that, in the domain $\Lambda_{\varepsilon(z)}$,

$$M(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^{n-1}}} \cdot \exp\left\{xH(\mathbf{q}) + \frac{\ln(x)}{2} - \frac{\sum_{i=1}^{n}\ln(x_i)}{2} + O\left(\frac{1}{z}\right)\right\}. \tag{32}$$

Here, as usual, $x = \sum_{i=1}^{n} x_i$; $q_i = x_i/x$. Therefore, we conclude that when considering the asymptotics of function (31) we can treat $M(\mathbf{x})$ as follows:

$$\widetilde{M}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^{n-1}}}\exp\left\{xH(\mathbf{q}) + \frac{\ln(x)}{2} - \frac{\sum_{i=1}^{n}\ln(x_i)}{2}\right\}. \tag{33}$$

In the latter formula we can write the exponent as

$$\left\{\right\} = x\ln x + \frac{\ln(x)}{2} - \sum_{i=1}^{n}\left(x_i\ln x_i + \frac{\ln(x_i)}{2}\right). \tag{34}$$

Let us write the Taylor expansion up to second-order terms near the maximum point in the plane $\langle \mathbf{a}, \mathbf{x}\rangle = z$, that is, near the point $\mathbf{x}' = \mathbf{p}z/H(\mathbf{p})$ (in what follows we denote by x_i' coordinates of the point $\mathbf{x}'$ and do by x' the sum of these coordinates which evidently equals $zH(\mathbf{p})^{-1}$).

First of all, note that

$$\widetilde{M}(\mathbf{x}') = \frac{1}{\sqrt{(2\pi x')^{n-1}\prod_{i=1}^{n}p_i}}\exp(z). \tag{35}$$

One can easily calculate second derivatives of expression (34):

$$\frac{\partial^2}{\partial x_i \partial x_j}\left\{\right\} = \begin{cases} x^{-1} - \left(2x^2\right)^{-1}, & \text{if } i \neq j, \\ x^{-1} - \left(2x^2\right)^{-1} - x_i^{-1} + \left(2x_i^2\right)^{-1}, & \text{else.} \end{cases} \tag{36}$$

(note that we do not use first derivatives in the Taylor expansion near the maximum point).

If $x \in \Lambda_\varepsilon$, then by formula (19) we have $x - x' = z(H(\mathbf{q};\mathbf{p})^{-1} - H(\mathbf{p})^{-1}) = zO(\varepsilon)$ (in the latter inequality we use the continuity of the function $H(\mathbf{q};\mathbf{p})^{-1}$). Consequently,

$$\begin{aligned} x_i - x_i' &= xq_i - x'p_i \\ &= \left(x' + zO(\varepsilon)\right)\left(p_i + O(\varepsilon)\right) - x'p_i = zO(\varepsilon). \end{aligned} \tag{37}$$

In particular, with chosen $\varepsilon = \varepsilon(z)$ we have $|x_i - x_i'| = O(z^{1/2+\delta})$. We obtain that, in the domain $\Lambda_{\varepsilon(z)}$,

$$\widetilde{M}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi x')^{n-1}\prod_{i=1}^{n} p_i}} \exp(z) \times \exp\left\{ \frac{\sum_{i,j=1}^{n}(x_i - x_i')(x_j - x_j')}{2x'} - \frac{\sum_{i=1}^{n}(x_i - x_i')^2}{2p_i x'} + O\left(z^{-1/2+3\delta}\right) \right\}. \tag{38}$$

Here the term $O(z^{-1/2+3\delta})$ contains both the remainder of terms of the series whose order exceeds 2 and the value of $O(z^{-1+2\delta})$ added by some omitted second-order terms. With $z \to \infty$ we can neglect the term of $O(z^{-1/2+3\delta})$. Therefore, in integral (31) in place of $M(\mathbf{x})$ we should substitute the function $\widehat{M}(\mathbf{x})$ which differs from $\widetilde{M}(\mathbf{x})$ in the fact that its exponent does not contain the term of $O(z^{-1/2+3\delta})$.

Let us change variables in the integral as follows: $y_i = (x_i - x_i')/\sqrt{x'}$. Since the degree of homogeneity of the delta-function equals -1, we obtain that limit (21) coincides with

$$\frac{1}{\sqrt{(2\pi)^{n-1}\prod_{i=1}^{n} p_i}} \int_{\mathbb{R}^n} \delta(\langle \mathbf{a}, \mathbf{y}\rangle) \exp\left\{ -\frac{\langle \mathscr{B}\mathbf{y}, \mathbf{y}\rangle}{2} \right\} d\mathbf{y}, \tag{39}$$

where $\mathscr{B}$ is $n \times n$ matrix, whose all elements equal -1, except diagonal components which are greater by $1/p_i$.

3.2. Calculation of the Determinant

Lemma 5. *Let $n \geq 2$. Consider $n \times n$ matrix B, where all nondiagonal elements equal 1, while $b_{ii} = 1 + k_i$. Then*

(1) *the determinant of this matrix equals*

$$\prod_{i=1}^{n} k_i \left(1 + \sum_{j=1}^{n} \frac{1}{k_j} \right); \tag{40}$$

(2) *the algebraic complement of the element with indices (i, j), $i \neq j$, equals*

$$-\prod_{\ell \in [n]\setminus\{i,j\}} k_\ell, \quad \text{where } [n] = \{1, \ldots, n\}. \tag{41}$$

Corollary 6. *The matrix $\mathscr{B}$ in formula (39) is degenerate.*

Proof of Lemma 5. Note that the first item of Lemma 5 defines the value of the algebraic complement of the diagonal element of such a matrix. Let us prove the theorem by induction.

With $n = 2$ in the formula in item (2) we get the product over the empty set; it is accepted that this product equals 1. The formula in item (1) remains valid with $n = 1$. In the induction step we assume that the formula in item (1) is proved for all dimensions less than n and has to be proved for the case when the dimension equals n, while the formula in item (2) is proved for all dimensions not greater than n and has to be proved for $(n+1) \times (n+1)$ matrix.

For proving item (1) we can use the expansion by the last row. Multiplying the algebraic complement by the diagonal element $k_n + 1$, we get the sum

$$\begin{aligned} &\prod_{i=1}^{n} k_i \left(1 + \sum_{j=1}^{n-1} \frac{1}{k_j} \right) + \prod_{i=1}^{n-1} k_i \left(1 + \sum_{j=1}^{n-1} \frac{1}{k_j} \right) \\ &\quad = \prod_{i=1}^{n} k_i \left(1 + \sum_{j=1}^{n-1} \frac{1}{k_j} \right) + \prod_{i=1}^{n-1} k_i + \sum_{j=1}^{n-1} \prod_{\ell \in [n]\setminus\{n,j\}} k_\ell. \end{aligned} \tag{42}$$

The expansion by the entire last row, taking into account the induction hypothesis for item (2), make the third part in row (42) vanish. First two terms in formula (42) together give the desired sum.

In order to prove item (2), let us expand the determinant considered in this item (algebraic complement of the element with (i, j) indices of the matrix B with $(n+1) \times (n+1)$ dimension) by the row whose number in the initial matrix of B was equal to j. Generally speaking, for clarity, we use the same indices as in the numeration of the initial matrix. Since the algebraic complement considered in this item and the occurring algebraic complement for the element with indices (j, i) (obtained by the expansion by a row of the determinant under consideration) have opposite signs, the value added by the element with indices (j, i) equals

$$-\prod_{r \in [n+1]\setminus\{i,j\}} k_r \left(1 + \sum_{\ell \in [n+1]\setminus\{i,j\}} \frac{1}{k_\ell} \right) \tag{43}$$

(here we have used the induction hypothesis for item (1)). The difference from the desired formula consists in the last term which equals (taking into account the first multiplier)

$$-\sum_{\ell \in [n+1]\setminus\{i,j\}} \prod_{r \in [n+1]\setminus\{i,j,\ell\}} k_r. \tag{44}$$

It vanishes, when taking into account the contribution of the remaining $n - 1$ elements in the jth row of the considered matrix. □

Lemma 7. *Let B_1 be the matrix mentioned in Lemma 5 (its dimension is $n \times n$, $n \geq 2$). Assume that $b_{ii} = 1 - 1/p_i$, $i = 1, \ldots, n$, where p_i are arbitrary nonzero numbers. Denote by B_2 a matrix of the same dimension in the form $a^T a$, where $a = (a_1, \ldots, a_n)$ is an arbitrary numeric row and T is the transposition sign. Let s be an arbitrary real number. Then*

$$\begin{aligned} &\det(sB_2 - B_1) \\ &\quad = \det(-B_1) \\ &\qquad + \frac{s\left(\left(\sum_{i=1}^{n} a_i p_i\right)^2 - \left(\sum_{j=1}^{n} p_j - 1\right)\sum_{i=1}^{n} a_i^2 p_i\right)}{\prod_{\ell=1}^{n} p_\ell}. \end{aligned} \tag{45}$$

Corollary 8. *Let a vector* $\mathbf{p} = (p_1, \ldots, p_n)$ *satisfy additional constraints* $p_i > 0$, $\sum_{i=1}^n p_i = 1$ *(i.e.,* $-B_1 = \mathscr{B}$*), while* $a_i = -\ln p_i$. *Then*

$$\sqrt{s \det\left(s^{-1}B_2 - B_1\right)} = \frac{H(\mathbf{p})}{\sqrt{\prod_{\ell=1}^n p_\ell}}. \tag{46}$$

Proof of Lemma 7. By the differentiation rule for determinants, the derivative of the determinant of $n \times n$ matrix equals the sum of determinants of n matrices such that in the ith one all elements of the ith row are replaced with their derivatives. We obtain that $\partial^2 \det(sB_2 - B_1)/\partial s^2$ is the sum of determinants of matrices each one of which contains either the zero row or two various rows of the matrix B_2. Since rank $B_2 = 1$, we get $\partial^2 \det(sB_2 - B_1)/\partial s^2 = 0$.

Thus, $\det(sB_2 - B_1)$ is a linear function of s, whose free term evidently equals $\det(-B_1)$. It is clear that for calculating the coefficient $\det(sB_2 - B_1)$ at s it suffices to summate products of each element of the matrix B_2 by the algebraic complement of the corresponding element of the matrix $-B_1$. If an element has indices (i, j), $i \neq j$, then by item (2) of Lemma 5 this product equals $a_i a_j p_i p_j / \prod_{\ell=1}^n p_\ell$.

Let us explain the positive sign in the last formula. We calculate an algebraic complement of the $-B_1$ matrix element. The matrix has $n \times n$ dimension, and therefore the found algebraic complement differs from the algebraic complement of the corresponding B_1 matrix element for $(-1)^{n-1}$ times. According to item (2) of Lemma 5, the algebraic complement of the corresponding B_1 matrix element is a "minus" product of $n-2$ multipliers k_i. In the given case each of k_i factors is negative (equals $-1/p_i$) which results in positive sign of the last formula in the above paragraph.

Assume that this formula is valid for all (i, j). Then we get the sum

$$\frac{\sum_{i,j=1}^n a_i a_j p_i p_j}{\prod_{\ell=1}^n p_\ell} = \frac{\left(\sum_{i=1}^n a_i p_i\right)^2}{\prod_{\ell=1}^n p_\ell}. \tag{47}$$

However by item (1) of Lemma 5 the algebraic complement of the diagonal element b_{ii} of the matrix $-B_1$ equals

$$(-1)^{n-1}\left(\prod_{j:j\neq i}\left(\frac{-1}{p_j}\right) + \sum_{j:j\neq i}\prod_{\ell\notin\{i,j\}}\left(\frac{-1}{p_\ell}\right)\right) \tag{48}$$

(here and below we omit the evident requirement that values of all indices belong to the set $[n]$).

Multiplying the first term in parentheses, that is, $\prod_{j:j\neq i}(-1/p_j)$, by $(-1)^{n-1}a_i^2$ and summing over all i, we get $\sum_{i=1}^n a_i^2 p_i / \prod_{\ell=1}^n p_\ell$. Let us multiply the resting term in parentheses (48) by $(-1)^{n-1}a_i^2$, sum over all i, and subtract the value

$$\frac{\sum_{i=1}^n a_i^2 p_i^2}{\prod_{\ell=1}^n p_\ell} \tag{49}$$

from the obtained result (note that the subtrahend was "illegally" included in formula (47)). It gives the overall contribution of the second term in formula (48), which equals

$$\frac{-\sum_{j=1}^n p_j \sum_{i=1}^n a_i^2 p_i}{\prod_{\ell=1}^n p_\ell}. \tag{50}$$

Taking into account all the calculation elements of the determinant $\det(sB_2 - B_1)$ allows completing the proof of Lemma 7. □

For completing the proof of Theorem 4 let us use Corollary 8. Let us replace the δ-function in integral (39) (as was proved earlier, this integral equals the limit considered in Theorem 4): $\delta(t) = \lim_{\sigma\to 0}(1/\sqrt{2\pi}\sigma)\exp\{-t^2/2\sigma^2\}$. Treating the limit multiplied by the coefficient at the exponent as a multiplier in the integral, we come to the limit of the Gaussian integral

$$\lim_{\sigma\to 0}\frac{\int_{\mathbb{R}^n}\exp\left\{-\left\langle\left(\sigma^{-2}B_2 + \mathscr{B}\right)\mathbf{y}, \mathbf{y}\right\rangle/2\right\}d\mathbf{y}}{\sigma\sqrt{(2\pi)^n\prod_{i=1}^n p_i}}, \tag{51}$$

that is,

$$\lim_{\sigma\to 0}\frac{1}{\sigma\sqrt{\prod_{i=1}^n p_i \det\left(\sigma^{-2}B_2 + \mathscr{B}\right)}}; \tag{52}$$

Immediately applying Corollary 8, we get desired $H^{-1}(\mathbf{p})$. This completes the proof.

4. The Ratio between the Sum and the Integral

What remains is to prove that, under assumptions of Theorem 3, the ratio of the integral of the function M calculated over the domain $S(z)$ to the sum of values of this function at integer points of this domain tends to 1 as $z \to \infty$. For comparing the integral of the function and the sum of its values in the same domain one usually applies the Koksma-Hlawka inequality (see [17]). Note that usually one considers the integral over a fixed domain (as a rule, the cube $[0, 1]^n$), whereas the domain in the case under consideration is varying. However, we intend only to prove the convergence of the fraction to 1 and do not need to estimate the asymptotic difference between the integral and the sum, which simplifies the task.

Evidently, it suffices to calculate the limit of the ratio for an arbitrary infinite increasing sequence $z_1, z_2, \ldots$, such that $z_i \to \infty$.

Theorem 9. *Let* $\Omega_1, \Omega_2, \ldots$ *be a sequence of Jordan measurable sets such that* $\Omega_i \subset \Omega_{i+1}$ *for all* $i = 1, 2, \ldots$. *Assume that* $f(x)$, $x \in \Omega$, *where* $\Omega = \bigcup_i \Omega_i$, *is an integrable and bounded on each of the domains* Ω_i *function such that* $f(x) \geq 0$ *and* $\int f(x)d\Omega_i \to \infty$ *as* $i \to \infty$. *Assume also that* K *is a countable set of points from* Ω *such that each of the sets* $K_i = K \cap \Omega_i$ *is finite. Then if for any sufficiently small* $\alpha > 0$ *there exists a partition of* Ω *onto a countable number of Jordan measurable*

sets $X_j = X_j(\alpha)$, $j = 1, 2, \ldots$, such that $\Omega_i = \bigcup_{j=1}^{n_i} X_j$ for some $n_i = n_i(\alpha)$, while

$$\frac{\sup_{x\in X_j} f(x)}{\inf_{x\in X_j} f(x)} < 1+\alpha \quad \text{starting with some } j, \tag{53}$$

$$\frac{|K\cap X_j|}{\mu X_j} \longrightarrow 1 \quad \text{as } j \longrightarrow \infty, \tag{54}$$

then in this case there exists the limit

$$\lim_{i\to\infty} \frac{\int f(x)\, d\Omega_i}{\sum_{x\in K_i} f(x)} = 1. \tag{55}$$

Proof. Evidently,

$$\mu X_j \inf_{x\in X_j} f(x) \le \int f(x)\, dX_j \le \mu X_j \sup_{x\in X_j} f(x). \tag{56}$$

Therefore, in view of (53) we conclude that starting with some j it holds that

$$\frac{\int f(x)\, dX_j/\mu X_j}{\sum_{x\in K\cap X_j} f(x)/|K\cap X_j|} \in (1-\alpha, 1+\alpha) \tag{57}$$

with $\alpha < 1$. In accordance with (54) we conclude that $\mu X_j/|K\cap X_j| \in (1-\alpha, 1+\alpha)$ for all j, except a finite number of values of the index. Therefore, there exists ℓ such that, with all $j > n_\ell$,

$$\frac{\int f(x)\, dX_j}{\sum_{x\in K\cap X_j} f(x)} \in \left((1-\alpha)^2, (1+\alpha)^2\right). \tag{58}$$

Representing this correlation as a double inequality and summing it over all j from $n_\ell + 1$ to n_i, we obtain

$$\frac{\int f(x)\, d(\Omega_i \setminus \Omega_\ell)}{\sum_{x\in K_i\setminus K_\ell} f(x)} \in \left((1-\alpha)^2, (1+\alpha)^2\right) \tag{59}$$

with $i > \ell$.

Note that by condition the numerator in the latter fraction (different from the integral $\int f(x) d\Omega_i$ by a constant value) tends to infinity. Then the same is true for the denominator. Note that the denominator differs from $\sum_{x\in K_i} f(x)$ by a constant value.

Therefore we conclude that all limit points of the sequence $\int f(x) d\Omega_i / \sum_{x\in K_i} f(x)$ lie inside the interval $((1-\alpha)^2, (1+\alpha)^2)$. Due to the arbitrariness of the choice of positive α Theorem 9 is proved. □

Corollary 10 (completion of the proof of Theorem 1). *Let $f(z)$ be the function mentioned in assumptions of Theorem 4 and let $\widetilde{Q}(z)$ obey formula (13). Then if at least one of the ratios a_i/a_k, $i, k \in \{1, \ldots, n\}$, $i \ne k$, is irrational, then*

$$\lim_{z\to\infty} \frac{f(z)}{\widetilde{Q}(z)} = 1. \tag{60}$$

Proof. For clarity we denote by z the parameter that defines the boundary of the considered domain, and do by ζ the corresponding parameter of the hyperplane that contains a certain interior point $\mathbf{x}$ of this domain; that is, $\zeta(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x}\rangle$.

First of all, note that considerations in Section 3.1 imply that both in the sum and in the integral we can replace $S(z)$ with the domain

$$\widehat{\Lambda}(z) = S(z) \cap \Lambda_{\varepsilon(\zeta)}, \quad \text{where } \varepsilon(\zeta) = \zeta^{-1/2+\delta}, \tag{61}$$

and replace the function $M(\mathbf{x})$ with $\widetilde{M}(\mathbf{x})$ defined by formula (34). Therefore, we need to prove that

$$\frac{\int_{\mathbf{x}\in\widehat{\Lambda}(z)} \widetilde{M}(\mathbf{x})\, d\mathbf{x}}{\sum_{\mathbf{k}\in\widehat{\Lambda}(z)} \widetilde{M}(\mathbf{k})} \longrightarrow 1 \tag{62}$$

(or that the difference of logarithms of the numerator and denominator tends to zero).

In view of Theorem 4 the logarithm of the numerator in the latter fraction is a uniformly continuous function of z, while the logarithm of the denominator evidently is a nondecreasing function. Therefore for proving the existence of the limit with $z \to \infty$ it suffices to prove the existence of the limit for a sequence in the form $z_n = \kappa n$, $n = 1, 2, \ldots$, where κ is an arbitrarily small positive value (as the difference between the numerator and denominator of the logarithms in an arbitrary point slightly differs from the value of difference in the nearest points z_n in this sequence). Namely, just for this fixed sequence we consider the ratio from the right-hand side of (62).

In order to apply Theorem 9, for an arbitrary sufficiently small positive α we construct a partition of $\Lambda_{\varepsilon(\zeta)}$ onto domains X_j satisfying assumptions of the theorem. Namely, we construct this partition by dividing of an infinite quantity of "flapjacks" located between neighboring hyperplanes in the forms $\zeta(\mathbf{x}) = c_r$ and $\zeta(\mathbf{x}) = c_{r+1}$, $r = 1, 2, \ldots$, where $c_{r+1} = c_r + \text{Const}$, onto a finite number of domains X_j.

Evidently, for any $\alpha \le 2\kappa$ we can choose a sequence c_r such that

$$c_{r+1} - c_r = \text{Const} < \frac{\alpha}{2}; \quad \text{for any } n\ \exists r : z_n = c_r. \tag{63}$$

To this end, it suffices to put $c_r = \text{Const}\, r$, where $\text{Const} = \kappa/\lceil 2\kappa/\alpha\rceil$ (here $\lceil\cdot\rceil$ is an upward rounding to the nearest integer).

Let $C_r = \{\mathbf{x} : c_r \le \zeta(\mathbf{x}) < c_{r+1}\}$. Denote by F_r the rth "flapjack" $C_r \cap \Lambda_{\varepsilon(\zeta)}$. We are going to "cut" F_r onto a finite number of domains X_j. We numerate the countable number of domains X_j, $j = 1, 2, \ldots$, so as to make domains X_j obtained by "cutting" F_r with the least r have lesser numbers, while the order of numbering inside the partition of F_r plays no role.

Since $\varepsilon(\zeta) = o(\zeta)$, with $\mathbf{x}, \mathbf{y} \in F_r$, it holds that $y_i = x_i + o(x_i)$ (cf. with (37)). Consequently, with $r \to \infty$ we get $\ln y_i - \ln x_i \to 0$ and $\ln y - \ln x \to 0$.

By formula (33),

$$\ln \widetilde{M}(\mathbf{x}) = \text{const} + g(\mathbf{x}) + \frac{\ln(x)}{2} - \frac{\sum_{i=1}^{n} \ln(x_i)}{2}, \tag{64}$$

where

$$g(\mathbf{x}) = xH(\mathbf{q}) = \sum_{i=1}^{n} x_i \ln x_i - \left(\sum_{i=1}^{n} x_i\right) \ln\left(\sum_{j=1}^{n} x_i\right). \quad (65)$$

We get grad $g = \ln \mathbf{q} = (\ln q_1, \ldots, \ln q_n)$ and

$$\frac{\partial^2 g}{\partial x_i \partial x_j} = O\left(\frac{1}{x}\right), \quad i, j \in \{1, \ldots, n\}. \quad (66)$$

Using expansion in a series $\ln \widetilde{M}$ with evaluation of the second-order terms and considerations of the previous paragraph we obtain the following important observation. If $\mathbf{x}, \mathbf{y} \in F_r$ and

$$|x_i - y_i| = o(\sqrt{x}) = o(\sqrt{c_r}), \quad i = 1, \ldots, n, \quad (67)$$

then with sufficiently large r it holds that

$$\ln \widetilde{M}(\mathbf{x}) - \ln \widetilde{M}(\mathbf{y}) < |\langle \ln \mathbf{q}, \mathbf{x} - \mathbf{y} \rangle| + \frac{\alpha}{10}. \quad (68)$$

Since $\mathbf{q} \to \mathbf{p}$ as $r \to \infty$, with sufficiently large r it holds that

$$|\langle \ln \mathbf{q}, \mathbf{x} - \mathbf{y} \rangle| < |\langle \mathbf{a}, \mathbf{x} - \mathbf{y} \rangle| + \frac{\alpha}{20} < (c_{r+1} - c_r) + \frac{\alpha}{20}. \quad (69)$$

As a result, we obtain that with sufficiently small α, starting with some r, it holds that

$$\frac{\widetilde{M}(\mathbf{x})}{\widetilde{M}(\mathbf{y})} < 1 + \alpha. \quad (70)$$

Therefore, dividing F_r onto domains X_j so as to fulfill correlation (67) for all points $\mathbf{x}, \mathbf{y}$ that belong to one domain, we guarantee the validity of assumption (53) in Theorem 9. Note that it suffices to fulfill condition (67) for all indices i except one, because the validity of this condition for the remaining index follows from the fact that $\mathbf{x}, \mathbf{y} \in C_r$.

Finally, let us use the irrationality of a_{i^*}/a_{k^*} for some $i^* \neq k^*$. Let us denote by I_{k^*} the set $\{1, \ldots, n\} \setminus \{k^*\}$ and do by $I_{i^*k^*}$ the set $\{1, \ldots, n\} \setminus \{i^*, k^*\}$. We are going to prove that, defining domains X_j by inequalities

$$l_{ji} \le x_i < L_{ji}, \quad i \in I_{k^*}, \text{ where } L_{ji} - l_{ji} > \text{const}\, c_r^{1/2-\delta}, \quad (71)$$

we fulfill condition (54) (with $K = \mathbb{Z}^n$). Here, as usual, δ is a sufficiently small real positive value, though in this case we can choose δ as any number in the interval $(0, 1/2)$ (roughly speaking, it is sufficient that the radius of the pieces X_j used to divide "flapjacks" F_r tends to infinity at $r \to \infty$).

Evidently, we can divide "almost all" F_r onto domains X_j so as to simultaneously fulfill inequalities (67) and conditions (71) on l and L (the remaining "cuttings" on the edges of the domain F_r which occur due to the inconsistency between the inequality $l_{ji} \le x_i < L_{ji}$, $i \in I_{k^*}$ and the definition of the boundary of the domain $\Lambda_{\varepsilon(z)}$ are asymptotically small).

Evidently, $\mu X_j = \prod_{i \in I_{k^*}} (L_{ji} - l_{ji}) \times (c_{r+1} - c_r)/a_{k^*}$. Since the difference $(L_{ji} - l_{ji})$ grows as $j \to \infty$, *the asymptotics* of the number of ways for choosing integer x_i such that $l_{ji} \le x_i < L_{ji}$ for $i \in I_{i^*k^*}$ coincide with $\prod_{i \in I_{i^*k^*}} (L_{ji} - l_{ji})$. Here and below we understand the asymptotics as a function of j such that the ratio of the considered quantity to this function tends to 1 as $j \to \infty$. In order to complete the proof of Corollary 10, what remains is to prove the following lemma.

Lemma 11. *Let the ratio a_{i^*}/a_{k^*} be irrational and $(L_{ji^*} - l_{ji^*}) \to \infty$. Assume also that the ratio $(c_{r+1} - c_r)/a_{k^*}$ equals a constant value lesser than 1 which is independent of r. Then for fixed x_i, $i \in I_{i^*k^*}$, the asymptotics of the number of ways to choose integer x_i, $i \in \{i^*, k^*\}$, such that $l_{ji^*} \le x_{i^*} < L_{ji^*}$ and $c_r \le \zeta(\mathbf{x}) < c_{r+1}$ simultaneously, equal $(L_{ji} - l_{ji}) \times (c_{r+1} - c_r)/a_{k^*}$.*

Proof of Lemma 11. In what follows we need standard denotations for the fractional part $\{\cdot\}$, floor $\lfloor \cdot \rfloor$, and ceil $\lceil \cdot \rceil$ of a number.

Let $c' = \sum_{i \in I_{i^*k^*}} a_i x_i$, $d_r = (c_r - c')/a_{k^*}$, $D_r = (c_{r+1} - c')/a_{k^*}$, and $\theta = a_{i^*}/a_{k^*}$. The condition $c_r \le \zeta(\mathbf{x}) < c_{r+1}$ is equivalent to the condition

$$\theta x_{i^*} + x_{k^*} \in [d_r, D_r). \quad (72)$$

If the difference $D_r - d_r$ (it equals $(c_{r+1} - c_r)/a_{k^*}$) is less than 1 (this inequality obviously holds for sufficiently small α) then with fixed x_{i^*} the integer value x_{k^*} satisfying condition (72) is defined uniquely, provided that it exists. Therefore, we need to estimate the quantity of values x_{i^*} in the interval $[l_{ji^*}, L_{ji^*})$ such that $\{\theta x_{i^*}\} \in [\{d_r\}, \{D_r\})$; here the latter correlation is understood in the sense of an interval on the unit circle, and the length of the considered interval is independent of r.

Recall the definition of a well-distributed sequence [17, section 1.5]. □

> Let (y_n) $n = 1, 2, \ldots$, be a sequence of real numbers. For integers $N \ge 1$ and $k \ge 0$ and a subset E of $[0, 1)$, let $A(E, N, k)$ be the number of terms among $\{y_{k+1}\}, \{y_{k+2}\}, \ldots, \{y_{k+N}\}$ that are lying in E.
>
> The sequence (y_n) $n = 1, 2, \ldots$, is said to be *well-distributed mod 1* if for all pairs a, b of real numbers with $0 \le a < b \le 1$ we have
>
> $$\lim_{N \to \infty} \frac{A([a, b); N, k)}{N} = b - a \quad \text{uniformly in } k = 0, 1, 2, \ldots. \quad (73)$$
>
> *Example.* The sequence $(n\theta)$ $n = 1, 2, \ldots$, with θ irrational is well-distributed mod 1.

The latter fact would have proved Lemma 11, if the interval of the unit circle $[\{d_r\}, \{D_r\})$ was independent of r. Let us clarify this property in the case of the inequality $\{D_r\} > \{d_r\}$. In what follows we always assume that this inequality is valid (evidently, as in the definition of the well-distribution

property, this leads to no loss of generality). Really, if k equals $\lceil l_{ji^*} \rceil - 1$ and N does the difference $\lfloor L_{ji^*} \rfloor - \lceil l_{ji^*} \rceil + 1$, then we obtain the uniform (in j) convergence

$$\lim_{N\to\infty} \frac{A\left([\{d\},\{D\});N,\left\lceil l_{ji^*}\right\rceil - 1\right)}{N} = D - d, \tag{74}$$

which is equivalent to the assertion of the lemma with the fixed value of $\{d_r\}$, $\{D_r\}$.

Note that if with fixed j equality (74) is valid for any subinterval in $[0,1)$, then we say that the corresponding sequence is *uniformly distributed modulo 1*. This property follows from the property of the *well-distribution modulo 1*. It is well known that (see [17, section 2.1]) for any sequence uniformly distributed modulo 1 the convergence is uniform with respect to all subintervals in $[0,1)$. Consequently, we get the uniform (in j) convergence

$$\lim_{N\to\infty} \frac{A\left([\{d_r\},\{D_r\});N,\left\lceil l_{ji^*}\right\rceil - 1\right)}{N} = \Delta, \tag{75}$$

where the *constant* Δ equals $D_r - d_r$. Therefore,

$$\lim_{j\to\infty} \frac{A\left([\{d_r\},\{D_r\});\left\lfloor L_{ji^*}\right\rfloor - \left\lceil l_{ji^*}\right\rceil + 1,\left\lceil l_{ji^*}\right\rceil - 1\right)}{\left(L_{ji^*} - l_{ji^*}\right)\Delta} = 1, \tag{76}$$

which coincides with the lemma assertion in a general case. This completes the proof.

5. Conclusion

We have proved that in the monkey model the probability of words in the sorted list has the *exact* power asymptotics, provided that the ratio of logarithms of probabilities of certain letters is irrational.

Note that this condition is not only sufficient but also necessary. Really, otherwise logarithms of probabilities $a_i = -\ln p_i$, $i = 1,\ldots,n$, allow the representation $a_i = m_i v$, where m_i are natural numbers and v is independent of i. In this case formula (10) defines a linear recurrent correlation on a grid with the step of v. This does not affect the initial constancy of the function $\widetilde{Q}$ in cells of the grid with the mentioned step with any value of the argument. Such constancy piecewise of the function $\widetilde{Q}$ contradicts the existence of a finite limit for the ratio of $p(r)/r^{-\gamma}$.

It should be noted that using the expression for terms of linear recurring sequences via the corresponding powers of roots of the characteristic equation allows clear analysis of rate of convergence to the power law of the function $\widetilde{Q}$ (with a step of v on the grid) in this degenerate case. It would be more interesting to conduct such studies for more general case to which the main theorem of this paper is devoted.

A generalization of results obtained in this paper to the case of the Markov dependence is of even more interest. In this case an analog of the vector $\mathbf{p}^\gamma$ is a substochastic matrix of transition probabilities where the row and column that correspond to the absorbing state are deleted, and all elements of this matrix are raised to a power of γ such that its spectral radius equals 1. Denote this matrix by $\mathbf{P}^\gamma$. In the case considered above all rows of the matrix $\mathbf{P}^\gamma$ coincide with $\mathbf{p}^\gamma$. In a typical case, when the strong power law takes place (see Introduction), the matrix $\mathbf{P}$ is irreducible and the matrix transposed with respect to $\mathbf{P}^\gamma$ has a positive eigenvector that corresponds to the unit eigenvalue. Let us norm this vector so as to make the sum of its components equal 1 and denote the result by $\mathbf{w}$. In the case of the Markov chain with the transition probability matrix $\mathbf{P}^\gamma$ this vector defines an ergodic distribution.

If all rows of the considered matrix coincide with $\mathbf{p}^\gamma$, then one can easily see that $\mathbf{w}$ coincides with $\mathbf{p}^\gamma$. It is possible that, in the case of the Markov dependence with the irreducible matrix $\mathbf{P}$, an analog of Theorem 1 takes place. The role of the entropy of the vector $\mathbf{p}^\gamma$ in this case plays the conditional entropy of the matrix $\mathbf{P}^\gamma$ rows with the weights equal to the corresponding components of the vector $\mathbf{w}$. We will discuss this fact in another publication.

It should be noted that mean values defined by other type recurrent relations occur in the process of analyzing the digital trees (cf. formula (10) and the recurrent relation for A_N, where N is a natural number, in [18, p. 404]). Nevertheless, the results for these values almost wholly coincide with Theorem 2 (see [19]). These results were obtained using the Mellin transform. The Mellin transform may also be useful for our case but it is a discussible problem.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Russian Foundation for Basic Research, Grant no. 15-06-07402. The research of the first author was supported by the Russian Government Program of Competitive Growth of Kazan Federal University.

References

[1] R. Durrett, *Random Graph Dynamics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, UK, 2007.

[2] M. Mitzenmacher, "A brief history of generative models for power law and lognormal distributions," *Internet Mathematics*, vol. 1, no. 2, pp. 226–251, 2004.

[3] R. H. Baayen, *Word Frequency Distributions*, vol. 18 of *Text, Speech and Language Technology*, Kluwer Academic, Dordrecht, Netherlands, 2001.

[4] G. K. Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*, Addison-Wesley, Cambridge, Mass, USA, 1949.

[5] B. Mandelbrot, "An informational theory of the statistical structure of languages," in *Communication Theory*, W. B. Jackson, Ed., pp. 486–502, 1953.

[6] B. Mandelbrot, "On recurrent noise limiting coding, in Proceedings of the symposium on information networks," in *Proceedings of the Symposium on Information Networks*, pp. 205–221, Polytechnic Institute of Brooklyn, New York, NY, USA, April 1954.

[7] G. A. Miller, "Some effects of intermittent silence," *The American Journal of Psychology*, vol. 70, no. 2, pp. 311–314, 1957.

[8] W. Li, "Random texts exhibit Zipf's-law-like word frequency distribution," *IEEE Transactions on Information Theory*, vol. 38, no. 6, pp. 1842–1845, 1992.

[9] R. Perline and R. Perline, "Two universality properties associated with the monkey model of Zipf's law," *Entropy*, vol. 18, no. 3, article 89, 2016.

[10] B. Conrad and M. Mitzenmacher, "Power laws for monkeys typing randomly: the case of unequal probabilities," *IEEE Transactions on Information Theory*, vol. 50, no. 7, pp. 1403–1414, 2004.

[11] V. V. Bochkarev and E. Yu. Lerner, "The Zipf law for random texts with unequal letter probabilities and the Pascal pyramid," *Izvestiya Vysshikh Uchebnykh Zavedenii. Matematika*, vol. 56, no. 12, pp. 30–33, 2012.

[12] R. Edwards, E. Foxall, and T. J. Perkins, "Scaling properties of paths on graphs," *Electronic Journal of Linear Algebra*, vol. 23, pp. 966–988, 2012.

[13] V. V. Bochkarev and E. Yu. Lerner, "Strong power and subexponential laws for an ordered list of trajectories of a Markov chain," *Electronic Journal of Linear Algebra*, vol. 27, pp. 534–556, 2014.

[14] V. V. Bochkarev and E. Yu. Lerner, "Zipf and non-Zipf laws for homogeneous Markov chain," https://arxiv.org/abs/1207.1872.

[15] E. Artin, *The Gamma Function*, Translated by Michael Butler. Athena Series: Selected Topics in Mathematics, Holt, Rinehart and Winston, New York, NY, USA, 1964.

[16] Yu. Suhov and M. Kelbert, *Probability and statistics by example. II*, Markov Chains: A Primer in Random Processes and Their Applications, Cambridge University Press, Cambridge, UK, 2008.

[17] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, Pure and Applied Mathematics, Wiley-Interscience [John Wiley & Sons], New York, NY, USA, 1974.

[18] R. Sedgewick and P. Flajolet, *An Introduction to the Analysis of Algorithms*, Addison-Wesley, Boston, Mass, USA, 1995.

[19] P. Flajolet, M. Roux, and B. Vallee, "Digital trees and memoryless sources: from arithmetics to analysis," in *Proceedings of the 21st International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms (AofA '10)*, Vienna, Austria, 2010.