PLISKA

STUDIA MATHEMATICA BULGARICA

ПЛИСКА

БЪЛГАРСКИ МАТЕМАТИЧЕСКИ СТУДИИ

PROBABILITY DISTRIBUTIONS RELATED TO SOME EPIDEMIC MODELS ON RANDOM GRAPHS

LJUBEN R. MUTAFČIEV

Random graphs, corresponding to random self-mappings of the finite set $\{1, 2, \dots, n\}$, are considered. It is supposed that a bacterium is placed at one of the cyclic elements of the random graph. In this case the probability distribution of the infected area for two schemes of infection's spread is derived. More exactly, it is supposed that (i) the infection deliveres inversely to the directions of the arcs, and (ii) the infection deliveres in two sides; to the arc-direction and conversely. The density of the arcsine law as a limit version of the exact probability distribution in the cases (i) and (ii) is obtained as $n \to \infty$.

Consider the set $\mathbf{T}_n$ of all mappings of the finite set $X_n = \{1, 2, \dots, n\}$ into itself, which satisfy the condition: $Tx \neq x$ for each $T \in \mathbf{T}_n$ and each $x \in X_n$. There are $(n-1)^n$ different mappings in $\mathbf{T}_n$. Each mapping $T \in \mathbf{T}_n$ is a digraph G_T, whose points belong to the set X_n; an oriented arc goes from x to y iff $y = Tx$. G_T may consist of disjoined components and each component includes only one cycle. We classify the components of G_T corresponding to their size, i. e. to the number of points they consist of. Let an uniform probability distribution P_n on $\mathbf{T}_n$ be given (each mapping $T \in \mathbf{T}_n$ has probability $(n-1)^{-n}$). The random mappings just described are the second type mappings studied by Harris [1].

We shall consider random variables connected with two epidemic models on the random graphs G_T. These epidemic interpretations are introduced by Gertsbakh [2].

Define $T^k x$ to be the k-th iteration of $T \in \mathbf{T}_n$ on $x \in X_n$, where k is integer. In other words, $T^k x = T(T^{k-1} x)$, and $T^0 x = x$. If for some $k > 0$, $T^k x = y$, y is said to be a k-th successor of x in T. The set of all successors of x in T is $S_T(x) = \{x, Tx, \dots, T^{n-1} x\}$. If for some $k \leq 0$, $T^k x = y$, y is said to be a k-th predecessor of x in T. The set of all k-th predecessors of x in T is denoted by $T^{(k)}(x)$, and $P_T(x) = \bigcup_{k=-n}^{0} T^{(k)}(x)$ is the set of all predecessors of x in T.

Let one bacterium be placed at an arbitrary element $x_0 \in X_n$. An inverse epidemic process (IEP) is defined by the infection being delivered from the infected point x_0 to all its predecessors. The area which will be infected is the set $P_T(x_0)$.

Imagine that the arcs connecting vertices in G_T have lost their orientation, and the arc between two arbitrary vertices x and y carries infection in two directions: from x to y if x has been infected first, and conversely. In this way is defined the two-sided epidemic process (TEP): the infection is delivered from the point x_0 "backward" to all its predecessors $P_T(x_0)$, "forward" to all its successors $S_T(x_0)$ and again "backward" from each $x \in S_T(x_0)$ to all its predecessors. The infected area in this case will be $B_T(x_0) = P_T(x_0) \cup S_T(x_0) \cup R_T(x_0)$, where $R_T(x_0) = \cup_x P_T(x), x \in S_T(x_0)$.

Denote the number of distinct elements in the set $P_T(x_0)$ ($B_T(x_0)$) by $|P_T(x_0)|$ ($|B_T(x_0)|$), and let $v_1 = v_1(T) = |P_T(x_0)|$, $v_2 = v_2(T) = |B_T(x_0)|$. We shall study the random variables v_1 and v_2 on the probability space $(\mathbf{T}_n, \mathfrak{B}(\mathbf{T}_n), \mathsf{P}_n)$, where $\mathfrak{B}(\mathbf{T}_n)$ is the σ-algebra of the subsets of $\mathbf{T}_n$.

At the beginning we may notice as in [2] that two different points of view are acceptable: (i) the bacterium is placed always at the same point x_0 for each mapping $T \in \mathbf{T}_n$, and (ii) the bacterium is randomly "thrown" on the set of n vertices X_n, with each of the n different occupations being eqally probable. It is easy to see that both approaches are equivalent. Actually, if $P_n^{(i)}(\Pi)$ and $P_n^{(ii)}(\Pi)$ are the probabilities that randomly chosen mapping $T \in \mathbf{T}_n$ has a fixed property Π in the cases (i) and (ii) respectively, and $N(\Pi)$ is the number of these mappings, then obviously $P_n^{(i)}(\Pi) = N(\Pi)(n-1)^{-n}$. In the case (ii), according to the formula of the total probability,

$$\begin{aligned} P_n^{(ii)}(\Pi) &= \sum_{x_0 \in X_n} \mathsf{P}_n\{\Pi \mid \text{the bacterium is placed at } x_0\} \\ &\quad \times \mathsf{P}_n\{\text{the bacterium is placed at } x_0\} \qquad (1) \\ &= \sum_{x_0 \in X_n} N(\Pi)(n-1)^{-n}\frac{1}{n} = N(\Pi)(n-1)^{-n} = P_n^{(i)}(\Pi) \end{aligned}$$

which yields the desired equivalence. Thus we can study the random variables v_1 and v_2 only in the case (i) accepting that the bacterium is placed at the point 1 ($x_0 = 1$) for each mapping $T \in \mathbf{T}_n$.

First, we shall consider the random variable v_1. Its probability distribution is known [1], and follows from easy combinatorial arguments:

$$\mathsf{P}_n\{v_1 = j\} = \binom{n-1}{j-1} j^{j-2}(n-j-1)^{n-j}(n-1)^{-n+1}, \; j = 1, \ldots, n. \qquad (2)$$

From this formula some limit expressions can be obtained. For example, if $j = tn$, $0 < t < 1$, and $n \to \infty$, applying Stirling's formula, we find

$$\mathsf{P}_n\{v_1 = tn\} = \frac{1}{nt\sqrt{2\pi(1-t)}} + o\left(\frac{1}{n}\right).$$

For an arbitrary fixed natural number j, using also Stirling's formula, we obtain another limit version of (2):

$$\lim_{n\to\infty} \mathsf{P}_n\{v_1 = j\} = \frac{e^{-j} j^{j-1}}{j!}.$$

For the probability distribution of the random variable v_2 we get

$$\mathsf{P}_n\{v_2 = j\} = \binom{n-1}{j-1}(j-1)! \sum_{k=0}^{j-2} \frac{j^k}{k!}(n-j-1)^{n-j}(n-1)^{-n}, \; j = 2, \ldots, n. \qquad (3)$$

The relation (3) follows from the fact that the number of all indecomposable mappings in the set $\mathbf{T}_j$, $j = 2, 3, \ldots$, is $B_j = (j-1)! \sum_{k=0}^{j-2} j^k/k!$ (the mapping

$T \in \mathbf{T}_j$ is called indecomposable iff it generates only one cycle). It is known [2] that the number $A_{j,k}$ of indecomposable mappings in $\mathbf{T}_j$ with k cyclic elements is $A_{j,k}=\binom{j}{k} k!\, j^{j-k-1}$, $j \geqq k \geqq 2$. Then, actually,

$$B_j=\sum_{k=2}^{j} A_{j,k}=(j-1)!\sum_{k=0}^{j-2} \frac{j^k}{k!}.$$

The limit value of the sum $\Sigma_{k=0}^{j-2} j^k/k!$ can be obtained using the normal approximation of the Poisson's distribution [3, Chapter 3, § 18]

$$\lim_{\lambda\to\infty} \sum_{k<\lambda+x\sqrt{\lambda}} \frac{\lambda^k e^{-\lambda}}{k!}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-u^2/2}\,du.$$

Putting in this relation $\lambda=j$, and $x=0$, we get

$$\lim_{j\to\infty} \sum_{k=0}^{j-1} \frac{j^k e^{-j}}{k!}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} e^{-u^2/2}\,du=\frac{1}{2}.$$

Thus, taking into account that $j^{j-1} e^{-j}/(j-1)!\to 0$, as $j\to\infty$, we have

$$\sum_{k=0}^{j-2} \frac{j^k}{k!}=\frac{1}{2} e^j(1+o(1)). \tag{4}$$

Now, applying (4) and Stirling's formula in (3), it is easy to check that for $j=tn$, $0<t<1$, and $n\to\infty$, the limit relation

$$\mathsf{P}_n\{v_2=tn\}=\frac{1}{2n\sqrt{1-t}}+o\left(\frac{1}{n}\right) \tag{5}$$

holds. The limit expression (5) shows that under the same assumptions the sequence of radom variables $\{v_2/n\}_{n=1}^{\infty}$ converges weakly to beta-distributed random variable with parameters 1 and 1/2.

Remark. The events $\{v_i=tn\}$, $i=1, 2$, will take place in our main asymptotic result too. They denote that the bacterium, placed at the point 1 will infect t-part $(0<t<1)$ of whole population X_n.

Remembering the epidemic mechanism, it is not difficult to show that the relation $B_T(x)=P_T(x)$ holds iff x is one of the cyclic elements of the mapping $T \in \mathbf{T}_n$. According to the equivalence obtained in (1), we can write $v_1(T)=v_2$ for each $T \in \mathbf{T}_n$ containing the element 1 as cyclical. Let C_n be the set of all mappings $T \in \mathbf{T}_n$ in which the element 1 is one of the cyclical points. Further we shall obtain the conditional probability distribution of the random variables v_i, $i=1, 2$ with the condition C_n. These two variables have same values with equal probabilities on this set. As a limit version of the exact distribution we shall derive the density of the arcsine law. Consequently, if the bacterium is placed at a cyclic point of the random mapping, then $n/2$ is the least probable area of the population X_n which will be infected $(n\to\infty)$. In order to prove our main result we need two preliminary lemmas.

Let $\mathbf{T}_{n,0}$ be the set of all indecomposable mappings of $\mathbf{T}_n$, and $T \in \mathbf{T}_{n,0}$. The first orbit of the cycle of T consists of those vertices which have as their direct successors the cyclic elements. The predecessors of the elements on the first orbit constitute the second one, and so on. The following combinatorial lemma is based on some arguments contained in [4].

Lemma 1. $|C_n \cap \mathbf{T}_{n,0}| = (n-1)n^{n-2}$, $n=2, 3, \ldots$.

Proof. Let us denote by $E_{k,1}^{(n_1,\ldots,n_p)}$ the set of all indecomposable mappings which have the following three properties: (i) they have 1 as a cyclical element, (ii) they have exactly k cyclical elements, (iii) they have exactly n_l elements on the l-th orbit, $l=1, \ldots, p$.

It is clear that the elements $\{2, 3, \ldots, n\}$ can be partitioned into $p+1$ groups of $k-1$, $n_1, \ldots, n_p$ elements in $(n-1)!/(k-1)!\,n_1!\ldots n_p!$ different ways, and there are $k!/k=(k-1)!$ ways to arrange the cycle of k elements (the elements of the first group and the element 1). The places of the elements on the l-th orbit can be chosen in $n_{l-1}^{n_l}$ ways, $l=1,\ldots, p$, where $n_0=k$. So we find the formula

$$|E_{k,1}^{(n_1,\ldots,n_p)}| = \frac{(n-1)!}{n_1!\ldots n_p!}\, k^{n_1} n_1^{n_2} \ldots n_{p-1}^{n_p}.$$

If $[n-k]_p$ stands for the collecti o n of all non-zero p-part partitions $(n_1, \ldots, n_p$ of $n-k$ elements, then (see [4])

$$\sum_p \sum_{[n-k]_p} \frac{k^{n_1} n_1^{n_2} \ldots n_{p-1}^{n_p}}{k n_1! \ldots n_p!} = \frac{n^{n-k-1}}{(n-k)!}.$$

Using this relation, we get

$$|C_n \cap \mathbf{T}_{n,0}| = \sum_{k=2}^{n} \sum_p \sum_{[n-k]_p} |E_{k,1}^{(n_1,\ldots,n_p)}| = (n-1)! \sum_{k=2}^{n} k\, \frac{n^{n-k-1}}{(n-k)!}$$

$$= (n-1)! \sum_{k=0}^{n-2} (n-k) \frac{n^{k-1}}{k!} = (n-1)!\, \Big[\sum_{k=0}^{n-2} \frac{n^k}{k!} - \sum_{k=1}^{n-2} \frac{n^{k-1}}{(k-1)!}\Big]$$

$$= \frac{(n-1)!\, n^{n-2}}{(n-2)!} = (n-1)\, n^{n-2}.$$

The following combinatorial identity is a corrolary of well-known identities of Abelian type [5, Chapter 6.6]. It is not essential for the proof of our basic result but it gives a possibility to write our relations in a simpler form.

Lemma 2. *For each integer $n \geq 2$ the following identity*

$$\sum_{k=2}^{n} \binom{n}{k} k^{k-1} (k-1)(n-k-1)^{n-k} = n! \sum_{k=0}^{n-2} \frac{(n-1)^k}{k!}$$

holds.

Proof. Let $S_n = \Sigma_{k=2}^{n} \binom{n}{k} k^{k-1} (k-1)(n-k-1)^{n-k}$. This sum can be represented as follows:

$$(6)\quad S_n = S_n^{(1)} - S_n^{(2)},\ S_n^{(1)} = \sum_{k=2}^{n} \binom{n}{k} k^k (n-k-1)^{n-k},\ S_n^{(2)} = \sum_{k=2}^{n} \binom{n}{k} k^{k-1} (n-k-1)^{n-k}.$$

For simplifying the sum $S_n^{(1)}$ we shall use the following Cauchy's [5, Section 6.6.1] identity:

$$\sum_{k=0}^{n}\binom{n}{k}(x+ka)^k(y-ka)^{n-k}=n!\sum_{k=0}^{n}\frac{(x+y)^k}{k!}a^{n-k}.$$

Putting here $x=0$, $a=1$, and $y=n-1$, we obtain

$$S_n^{(1)}=n!\sum_{k=0}^{n-1}\frac{(n-1)^k}{k!}-n(n-2)^{n-1}. \tag{7}$$

For the sum $S_n^{(2)}$ we shall apply another well-known identity (also [5, Section 6.6.1]):

$$\sum_{k=0}^{n-1}\binom{n}{k}(\alpha+n-k)^{n-k-1}(x+k)^k=\alpha^{-1}[(x+\alpha+n)^n-(x+n)^n].$$

Substituting $n-k=l$, $x=-1$, and taking the limit of both sides as $\alpha\to 0$, we find

$$S_n^{(2)}=n(n-1)^{n-1}-n(n-2)^{n-1}. \tag{8}$$

Relations (6)—(8) yield Lemma 2.

Now we shall formulate the following statement.

Theorem. *If* $j\in[2,n]$ *is an arbitrary integer, and* $0<t<1$, *then for* $i=1,2$

$$\mathsf{P}_n\{v_i=j\mid C_n\}=\binom{n}{j}j^{j-1}(j-1)(n-j-1)^{n-j}\Big[n!\sum_{k=0}^{n-2}\frac{(n-1)^k}{k!}\Big]^{-1},\ n\to\infty, \tag{9}$$

$$\mathsf{P}_n\{v_i=tn\mid C_n\}=\frac{1}{\pi n\sqrt{t(1-t)}}+o\Big(\frac{1}{n}\Big),\ n\to\infty. \tag{10}$$

Proof. First we shall prove that

$$\mathsf{P}_n(\{v_i=j\}\cap C_n)=\binom{n-1}{j-1}j^{j-2}(j-1)(n-j-1)^{n-j}(n-1)^{-n}. \tag{11}$$

Actually, the cyclic element 1, in which is placed the bacterium, will infect area with size j iff the element 1 belongs to a component with size j. There are $\binom{n-1}{j-1}$ different ways to choose the other $j-1$ elements of this component. Lemma 1 gives the number of ways in which this component will be constructed. This number is $(j-1)j^{j-2}$. Finally, the remaining $n-j$ vertices can be connected in a graph G_T, $T\in\mathbf{T}_{n-j}$, in $(n-j-1)^{n-j}$ different ways, which proves relation (11). Using (11), and the identity of Lemma 2, we obtain that

$$n\,\mathsf{P}_n(C_n)=n\sum_{k=2}^{n}\mathsf{P}_n(\{v_i=k\}\cap C_n) \tag{12}$$

$$=(n-1)^{-n}\sum_{k=2}^{n}\binom{n}{k}k^{k-1}(k-1)(n-k-1)^{n-k}=n!(n-1)^{-n}\sum_{k=0}^{n-2}\frac{(n-1)^k}{k!}.$$

Dividing (11) by (12) we get (9).

The limit version (10) can be found in some different ways. For example, one can use Stirling's formula and approximation's formula for the sum $\Sigma_{k=0}^{n-2}(n-1)^k/k!$ as in (4). Here we shall give another proof, based on the properties of the probability distribution of the number of the cyclic points in random mappings [1].

Let us consider the following combinatorial problem. In how many different ways will a randomly chosen vertex be cyclic element for a randomly chosen mapping $T\in\mathbf{T}_n$?

The proof of (12) shows us that this number is equal to $n!\ \Sigma_{k=0}^{n-2}(n-1)^k/k!$. Now we shall count in a different manner. Suppose that the vertex which we shall choose belong to a mapping with k cyclical vertices. Harris [1] has found that the number or all mappings from $\mathbf{T}_n$ with k cyclical vertices is $n^{n-k}D_k\binom{n-1}{k-1}$ for $k\geqq 2$, where D_k is the nearest integer to $k!\,e^{-1}$. So Harris's proposition shows that the sum $\Sigma_{k=2}^{n}kD_k\binom{n-1}{k-1}n^{n-k}$ gives also the desired number of ways. Thus we obtain the following identity:

$$\sum_{k=2}^{n}kD_k\binom{n-1}{k-1}n^{n-k}=n!\sum_{k=0}^{n-2}\frac{(n-1)^k}{k!}. \tag{13}$$

But the left-hand side of (13) multiplied by $(n-1)^{-n}$ gives the mathematical expectation of the number of the cyclic points in a random mapping. The asymptotic form of this expectation as $n\to\infty$ is known [1]:

$$(n-1)^{-n}\sum_{k=2}^{n}kD_k\binom{n-1}{k-1}n^{n-k}=\sqrt{\frac{\pi n}{2}}\,(1+o(1)). \tag{14}$$

More exactly, in [1] the limit distribution of the number of the cyclic elements in a random mapping is obtained, which leads trivially to (14). Combining (12) and (14), we obtain

$$n\,\mathsf{P}_n(C_n)=\sqrt{\frac{\pi n}{2}}\,(1+o(1)). \tag{15}$$

But by virtue of (9), we have

$$\mathsf{P}_n\{v_i=j\mid C_n\}=\binom{n}{j}j^{j-1}(j-1)(n-j-1)^{n-j}[n\,\mathsf{P}_n(C_n)]^{-1}. \tag{16}$$

Now the asymptotic formula (10) can be easily verified putting $j=nt$ in (16) and applying Stirling's formula and (15). So the theorem is proved.

We shall notice that beside the proof of (10) we obtain also a curious combinatorial identity — (13).

The results (3), (5), (9), (10), and (13), were presented at the 14th European Meeting of Statisticians (see [6]) without proofs.

The author wishes to express his thanks to J. Jaworski for the helpful discussion about these results.

REFERENCES

1. B. H a r r i s. Probability distributions related to random mappings. *Ann. Math. Statist.*, **31**, 1960, 1045—1062.
2. I. B. G e r t s b a k h. Epidemic process on a random graph: some preliminary results. *J. Appl. Probab.*, **14**, 1977, 427—438.
3. A. R e n y i. Probability theory. Budapest, 1970.
4. L. K a t z. Probability of indecomposability of a random mapping function. *Ann. Math. Statist.*, **26**, 1955, 512—517.
5. J. K a u c k y. Kombinatoricke identity. Bratislava, 1975.
6. L. M u t a f c h i e v. Probabilities and asymptotical estimates for epidemic models on random graphs. Abstracts of 14th European Meeting of Statisticians, Wroclaw, 1981, 233—234.

Centre for Mathematics and Mechanics
1090 Sofia P. O. Box 373

Received 8. 3. 1982