LOCAL LIMIT THEOREM FOR A χ^2- TYPE STATISTIC IN INTERNET GRAPHS

I.A. Cheplyukova, Yu.L. Pavlov
Karelian Research Centre of the Russian Academy of Sciences
Petrozavodsk, RUSSIA
e-mail: chia@krc.karelia.ru, pavlov@krc.karelia.ru

Abstract

We consider random Internet graphs consisting of $N+1$ vertices under the condition that the sum of vertex degrees is equal to n. We get the limit distribution of χ^2- type statistic for these graphs such that $N, n \to \infty, 1 < C \leq n/N < \zeta(\tau), \tau > 0$, where $\zeta(\tau)$ is the Rimann's zeta-function.

The asymptotic behaviour of the corresponding statistic is well-known for the polynomial scheme connected with the goodness-of-fit χ^2 test. However, the need to obtain such result appears also when testing statistical hypotheses in many combinatorial problems. In the present paper we consider one of such problems. The object of the research is the known structure of random graph (see e.g. [2, 7]), which is a suitable model of the Internet. Therefore, such models are called Internet graphs. The model is defined as follows. Let the number of the vertices in the graph be equal to $N+1$ and vertices are numbered from 0 to N. The degrees of the vertices $\eta_1, \ldots, \eta_N$ are drawn independently from power-law distributions with positive exponents $\tau_1, \ldots, \tau_N$ such that

$$\mathbf{P}\{\eta_i \geq k\} = k^{-\tau_i}, \qquad i = 1, \ldots, N, \qquad k = 1, 2, \ldots \tag{1}$$

The vertex 0 is auxiliary in character and has the degree 0 if the sum of vertex degrees $\eta_1 + \ldots + \eta_N$ is even, otherwise the degree is 1. It is clear that we need to use the auxiliary vertex 0 because the sum of degrees of all graph vertices must be even.

To describe of the graph we will use the notion of a semiedge or stub, i.e. an edge incident to a concrete vertex but with the adjacent vertex not defined yet. All semiedges of vertices are numbered in arbitrary order. The graph is constructed by joining each stub to another equiprobably to form edges.

There are many papers where the results describing the limit behaviour of different characteristics of random graphs were obtained. In [2] it was found that in practice the typical values of parameters $\tau_1, \ldots, \tau_N$ are usually the same and belong to the interval $(1, 2)$. In [1, 6] it was for the first time suggested the generalized allocation scheme is used to study asymptotic behaviour of Internet graphs. This method has been introduced and supported by Kolchin V.F. (see e.g. [3]).

We consider the subset of Internet graphs under the conditions that $\tau > 0$ and $\eta_1 + \ldots + \eta_N = n$. Let the null hypothesis be that the parameters of distributions (1) are the same and equal to $\tau : \tau_1 = \tau_2 = \ldots = \tau_N = \tau$. To test this hypothesis we propose the statistic

$$\chi^2 = \sum_{i=1}^{N} \frac{(\eta_i - n/N)^2}{n/N}. \tag{2}$$

Let $\xi_1, \ldots \xi_N$ be independent identically distributed random variables such that

$$\mathbf{P}\{\xi_1 = k\} = \frac{\lambda^k}{1-(1-\lambda)\Phi(\lambda,\tau,1)}\left(\frac{1}{k^\tau} - \frac{1}{(k+1)^\tau}\right), \tag{3}$$

where $k = 1, 2, \ldots, \quad 0 < \lambda < 1$ and $\Phi(x, s, a)$ is the Lerch transcendent function:

$$\Phi(x, s, a) = \sum_{k=0}^{\infty} x^k/(k+a)^s, \qquad a > 0. \tag{4}$$

Let the parameter $\lambda = \lambda(n, N)$ of the distribution (3) be determined by the relation

$$\frac{\Phi(\lambda,\tau,1) - (1-\lambda)\Phi(\lambda,\tau-1,1)}{1-(1-\lambda)\Phi(\lambda,\tau,1)} = \frac{n}{N}. \tag{5}$$

Denote $\nu_1 = \xi_1(\xi_1 - 1)/2, a = \mathbf{E}\nu_1, \sigma^2 = \mathbf{D}\nu_1, \rho = cov(\xi_1, \nu_1)/\sqrt{\mathbf{D}\xi_1\mathbf{D}\nu_1}$. Below we will prove the local limit theorem for statistic (2) when $N, n \to \infty$ and $1 < C_1 \le n/N < \zeta(\tau)$, where $\zeta(\tau)$ is the Rimann's zeta-function. Note that judging by the results of paper [8] the distribution of statistic (2) weakly converges to the normal law.

Suppose that in the case $n/N \to \zeta(\tau)$ one of the following conditions is valid:

1. $\tau > 6$;
2. $\tau = 6, \quad \sqrt{N}/|\ln(1-\lambda)| \to \infty$;
3. $4 < \tau < 6, \quad \sqrt{N}(1-\lambda)^{6-\tau} \to \infty$;
4. $\tau = 4, \quad \sqrt{N}(1-\lambda)^2|\ln^3(1-\lambda)| \to \infty$;
5. $0 < \tau < 4, \quad N(1-\lambda)^{-\tau} \to \infty$.

We have the following result.

Theorem 1 *Let $N, n \to \infty$ in such a way that $1 < C_1 \le n/N < \zeta(\tau)$. Then uniformly in the integer k such that $(k - Na)/\sigma\sqrt{N(1-\rho^2)}$ lies in any finite fixed interval*

$$\mathbf{P}\left\{\chi^2 = \frac{2Nk}{n} + N - n\right\} = \frac{1+o(1)}{\sigma\sqrt{2\pi N(1-\rho^2)}} \exp\left\{-\frac{(k-Na)^2}{2N\sigma^2(1-\rho^2)}\right\}.$$

To prove this theorem we will use some auxiliary statements (Lemmas 1-3).

We set $\zeta_N = \xi_1 + \ldots + \xi_N, \quad \mu_N = \sum\limits_{i=1}^{N} \xi_i(\xi_i - 1)/2$.

Lemma 1 *The equality*

$$\mathbf{P}\left\{\chi^2 = \frac{2Nk}{n} + N - n\right\} = \frac{\mathbf{P}\{\zeta_N = n, \mu_N = k\}}{\mathbf{P}\{\zeta_N = n\}} \tag{6}$$

holds.

Proof. Taking into account the equality $\eta_1 + \ldots + \eta_N = n$, it is easy to see from (2) that

$$\mathbf{P}\left\{\chi^2 = 2Nk/n + N - n\right\} = \mathbf{P}\left\{\sum_{i=1}^{N} \eta_i(\eta_i - 1)/2 = k\right\}. \tag{7}$$

It is not hard to show that the conditions of the generalized allocation scheme are valid (see [3]). From this and from (7) we get (6). Lemma 1 is proved.

It is clear that the sums ζ_N and μ_N from (6) form the array scheme. Therefore, to obtain the limit distribution of statistic (2) we have to prove local limit theorem for the random vector (ζ_N, μ_N). This assertion will be proved in Lemma 3. Let

$$\overline{X}_i = (\xi_i, \nu_i), i = 1, \ldots, N, S_N = (\overline{X}_1 + \ldots + \overline{X}_N) = (\zeta_N, \mu_N), \nu_i = \xi_i(\xi_i - 1)/2, \tag{8}$$

$$A_N = \mathbf{E}S_N, \sigma_1^2 = \mathbf{D}\xi_1, Q_N = \begin{pmatrix} 1/(\sigma_1\sqrt{N}) & 0 \\ 0 & 1/(\sigma\sqrt{N}) \end{pmatrix}, \Sigma = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}.$$

Using the properties and the asymptotic of the Lerch function the next lemma can be proved by the standard method of characteristic function.

Lemma 2 *Under the conditions of the Theorem 1, the sequence of distributions $(S_N - A_N)Q_N$ weakly converges to the normal law with the density*

$$g(\overline{x}) = (2\pi\sqrt{1-\rho^2})^{-1} \exp\left\{-\overline{x}\Sigma\overline{x}^T/2\right\}, \qquad \overline{x} = (x_1, x_2) \in R^2. \tag{9}$$

Lemma 2 shows that the distribution of statistic (2) weakly converges to the normal law. We will prove now that, in fact, local convergence takes place.

Lemma 3 *Under the conditions of the Theorem 1 we have*

$$\mathbf{P}\left\{S_N = \overline{z}\right\} = \det Q_N \left[g(\overline{z} - A_N)Q_N) + o(1)\right]$$

uniformly in $\overline{z} = (z_1, z_2)$ where $g(\overline{x})$ is given in (9).

Proof. Let $\langle\alpha\rangle$ be the distance from $\alpha \in \mathbf{R}$ to the nearest integer. We also set

$$\Omega(1/2, 1/4) = \{\overline{d} = (d_1, d_2) \in \mathbf{R}^2 : |d_1|, |d_2| \le 1/2, |\overline{d}| > 1/4\}; H(\overline{X}, \overline{d}) = \mathbf{E}\langle(\overline{X}^*, \overline{d})\rangle^2,$$

where $(\overline{X}^*, \overline{d})$ is the scalar product of the random vector $\overline{X}^*$ and vector $\overline{d}$, random variable $\overline{X}^*$ is obtained from the random vector $\overline{X}$ by symmetrization. Let also

$$H_N(\overline{d}) = \sum_{i=1}^{N} H(\overline{X}_i, \overline{d}) = N\mathbf{E}\langle(\overline{X}^*, \overline{d})\rangle^2, \qquad H_N = \inf_{\overline{d}\in\Omega(1/2,1/4)} H_N(\overline{d}).$$

For all $\overline{d} \in \Omega(1/2, 1/4)$ it is valid that

$$H_N(\overline{d}) \ge C_{11}N[\langle(d_1 + d_2)\rangle^2 + \langle(2d_1 + 3d_2)\rangle^2]. \tag{10}$$

Using Lemma 2, (3),(10) and the asymptotic of $\Phi(\lambda, \tau, 1)$, we get that the condition 1 of the theorem 4 from [4] is valid in the case $0 < C_1 \le n/N \le C_6 < \zeta(\tau)$ and

$n/N \to \zeta(\tau), \tau > 4$. This means that there is $\alpha > 0$ such that $H_N(\overline{d}) \geq \alpha|Q_N^{-1}d|^2$, wherefore Lemma 3 is proved by theorem 4 [4].

Let $n/N \to \zeta(\tau), 0 < \tau \leq 4$. Consider the condition 2 of theorem 4 [4]. In other words we have to show that there are $\alpha > 0, \delta \in (0, 2], M > 0, \beta > 0, \nu \in (0, 1/2)$ such that $H_N \to \infty$ and

$$B_N^2(\overline{\theta}, u) \geq \alpha u^{2-\delta}|Q_N^{-1}\overline{\theta}|^{\delta}, \tag{11}$$

where $B_N^2(\overline{\theta}, u) = N \sum_{(\overline{\theta}, \overline{X_1}) \leq u} (\overline{\theta}, \overline{X}_1)\mathbf{P}^*, \overline{\theta} = (\theta_1, \theta_2), |\overline{\theta}| = 1$, $\mathbf{P}^*$ is symmetrization distribution of the random vector $\overline{X_1}$ and $u \in [MH_N^{\nu}, \beta|Q_N^{-1}\overline{\theta}|]$. It is not hard to see that for $\nu > 0$

$$B_N^2(\overline{\theta}, u) \geq C_{12}N \sum_{u/2 \leq k \leq u} k^{3-\tau} \geq C_{12}Nu^{4-\tau}.$$

From this and the conditions of Lemma 3 we find that inequality (11) is valid. Lemma 3 is proved.

In [6] and [5] were demonstrated that under the conditions of the Theorem

$$\mathbf{P}\{\zeta_N = n\} = (\sigma_1\sqrt{2\pi N})^{-1}(1 + o(1)).$$

Then from Lemma 1 and Lemma 3 it follows the assertion of the Theorem holds.

References

[1] Cheplyukova I.A., Pavlov Yu.L. (2007). Limit distributions of vertex degree in conditional power-law random graphs. In:*Transactios of the XXVI international seminar on stability problems for stochastic models* . Karniel, pp. 52-59.

[2] Faloutsos C., Faloutsos P., Faloutsos M. (1999). On power–law relationships of the Internet topology. *Computer communications Rev.*. Vol. **29**, pp. 251-252.

[3] Kolchin V.F. (1986). *Random mappings*. Springer, New York.

[4] Mukhin A.B. (1991). Local limit theorems for latice random variables. *Theory of probability and its aplications*. Vol. **36**, 4, pp. 698-713.

[5] Pavlov Yu.L., (2009). On the limit distributions of the vertex degree of conditional Internet graphs. *Discrete mathematics and applications*. Vol. **19**, 4, pp. 337-360.

[6] Pavlov Yu.L., Cheplyukova I.A. (2008). Random graphs of Internet type and the generalized allocations scheme. *Discrete mathematics and applications*. Vol. **18**, 5, pp. 447-464.

[7] Reittu H., Norros I. (2004). On the power–law random graph model of massive data networs. *Performance evalution*. Vol. **55**, pp. 3-23.

[8] Ronzhin A.F. (1988). Goodness-of-fit tests for generalized schemes based on separable statistics. *Theory of probability and its aplications*. Vol. **33**, 1, pp. 86-95.