International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:4, No:5, 2010

On the Use of Correlated Binary Model in Social Network Analysis

Elsayed A. Habib Elamir

***Abstract*—**In social network analysis the mean nodal degree and density of the graph can be considered as a measure of the activity of all actors in the network and this is an important property of a graph and for making comparisons among networks. Since subjects in a family or organization are subject to common environment factors, it is prime interest to study the association between responses. Therefore, we study the distribution of the mean nodal degree and density of the graph under correlated binary units. The cross product ratio is used to capture the intra-units association among subjects. Computer program and an application are given to show the benefits of the method.

***Keywords*—**Correlated Binary data, cross product ratio, density of the graph, multiplicative binomial distribution.

I. Introduction

GRAPHS have been widely used in social network analysis as a means of formally representing social relations and quantifying important social structural properties; see, for example [13] and [6]. The visual representation of data that a graph or sociogram offers often allow researchers to uncover patterns that might otherwise go undetected; see [2].

There are two main approaches in social network data collection- socio-centered and ego-centered approaches. Socio-centered approach is taking a census of ties in a population of actors – rather than a sample. Full network data allows for very powerful descriptions and analyses of social structures. Unfortunately, full network data can also be very expensive and difficult to collect. Obtaining data from every member of a population, and having every member rank or rate every other member can be very challenging tasks in any but the smallest groups. On the other hand, the ego-centered approach with alter connections is taking a sample from a population. This kind of approach can be quite effective for collecting a form of relational and attribute data. For example, we might take a simple random sample of male college students and ask them to report who are their close friends, and which of these friends know one another; see [4] and [8].

Density and mean nodal degree of the graph are informative in many applications. For example, if we observe children playing together, and represent children by nodes, and instances of pairs of children playing by lines in a graph, then a node with small degree would indicate a child who played with few others, and a node with a large degree would indicate a child who played with many others and it is informative to summarize the degree and density of all the actors in the networks and they can be considered as measures of the activity of all the actors in the network and this is an important property of a graph; see [14] and [10].

In this work we derive the distribution of the mean nodal degree and density of the graph under the correlated binary responses. The cross product ratio is used to capture the interaction among subjects in the network. The confidence interval of the density of the graph is obtained.

In Section 2 we summarize the concept of the density of the graph in social network analysis. The correlated binary distribution (multiplicative Bernoulli distribution) is summarized in Section 3. The estimation of multiplicative Bernoulli distribution is derived based on moment and maximum likelihood methods in Section 4. The distribution of the density of the graph is obtained in Section 5. An application is given in section 6.

II. Density of the Graph in Social Network Analysis

Graphs could use for representing non-directional and directional relations. Non-directional relations include such things as co-membership in formal organization or informal groups, such as "is married to" and "is a blood relative of"; see [3]. In a graph, nodes represent actors and lines represent ties between actors. A graph g consists of two sets of information: a set of nodes, $\aleph = n_1, .., n_g$ and a set of lines $\mathcal{L} = l_1, \ldots, l_L$ between pairs of nodes. There are g nodes and L lines. In a non-directional graph each line is an unordered pair of distinct nodes, n_i, n_j. Since lines are unordered pairs of nodes, the line between nodes n_i and n_j is identical to the line between n_j and n_i, there is no reflexive ties (loops). In graph of a social network with a single non-directional dichotomous relation, the nodes represent actor and lines represent the ties that exist between pairs of actors on the relation.

Figure 1 shows an example of non-directional graph where panel (a) shows empty graph with 5 nodes and zero lines, panel (b) shows complete graph with 5 nodes and 10 lines, and panels (c) and (d) show intermediate graph.

The density of the graph is a statistic that reports the proportional of possible lines that are actually present in the graph L to maximum possible lines 0.5g(g-1) since there are g nodes in the graph, and loops are excluded and its is value will be between 0 and 1.

Dr. Elsayed A. H. Elamir is with the University of Bahrain, College of Business, Department of Management & Marketing, P.O.Box 32038, Kingdom of Bahrain (corresponding author: +973-1743-8817; fax: +973-1744-9776; E-mail: shabib40@gmail.com).

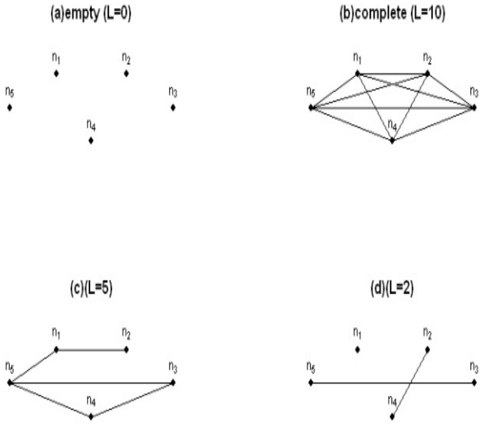


Fig.1 an example of non-directional graph in social network

The density of the graph is

$$\delta = \frac{\sum_{i=1}^{g-1}\sum_{j=i+1}^{g} Z_{ij}}{0.5g(g-1)} = \frac{L}{0.5g(g-1)} = \frac{L}{N}$$

where $N = 0.5g(g-1)$ and

$$Z_{ij} = \begin{cases} 1 & \text{if there is line between } (n_i, n_j) \\ 0 & \text{otherwise} \end{cases}$$

While degree of the graph is a concept that considers the number of lines incident with each node in a graph, $d(n_i)$. The mean nodal degree is a statistic that reports the average degree of the nodes in the graph and is defined as

$$\bar{d} = \frac{\sum_{i=1}^{g} d(n_i)}{g} = (g-1)\delta$$

See; [7] and [14].

Since subjects in a family or organization share common genetic traits or are subject to common environmental factors, we consider the binary random variables, Z_{ij}, are not independent. Therefore, we study the distribution of δ and $\bar{d}$ under dependence among Z_{ij}. Different models for this dependence provide a wider range of models than are provided by the Bernoulli distribution. Among these, [12] has derived the multivariate multiplicative Bernoulli distribution for the joint distribution of n binary-dependent variables introduced by [5] as an alternative to Altham's multiplicative distribution [1].

III. Multiplicative Bernoulli Distribution

[12] had derived the multivariate multiplicative Bernoulli distribution under dependent of the indicator random variables and the units are exchangeable as

$$P(\boldsymbol{Z} = \boldsymbol{z}) = \frac{\psi^{y}(1-\psi)^{n-y}\omega^{y(n-y)}}{\sum_{t=0}^{n}\binom{n}{t}\psi^{t}(1-\psi)^{n-t}\omega^{t(n-t)}},$$

$$\boldsymbol{z} = (0,\dots,0_n)_1,\dots,(1,\dots,1_n)_{2^n}$$

$y = \sum z_i$, $y = 0,1,\dots,n$, ψ and ω are the parameters and $\boldsymbol{Z} = (Z_1,\dots,Z_n)$ is a vector of indicator random variables Z_i, $z_i = 0$ or 1. There are 2^n possible outcomes of $\boldsymbol{Z}$, n is the sample size and there are $\binom{n}{0}$ ways for $y = 0$, $\binom{n}{1}$ ways for $y = 1$,…, $\binom{n}{n}$ ways for $y = n$. Also, the distribution could be rewritten as

$$P(\boldsymbol{Z} = \boldsymbol{z}) = \frac{\prod_{i=1}^{n}\psi^{z_i}(1-\psi)^{1-z_i}\omega^{nz_i-(\sum z_i)^2/n}}{\sum_{t=0}^{n}\binom{n}{t}\psi^{t}(1-\psi)^{n-t}\omega^{t(n-t)}}$$

The multivariate Bernoulli distribution is obtained for $\omega = 1$ as

$$P(\boldsymbol{Z} = z) = \prod_{i=1}^{n}\psi^{z_i}(1-\psi)^{1-z_i} = \tau^{y}(1-\tau)^{n-y}$$

with $E(Z_i) = \tau$, $V(Z_i) = \tau(1-\tau)$ and $Cov(Z_i, Z_j) = 0$ and $\psi = \tau = P(Z = 1)$; see [9]. The multiplicative Bernoulli distribution is provided wider range than are provided by the multivariate Bernoulli distribution; see, Figures 2 and 3.

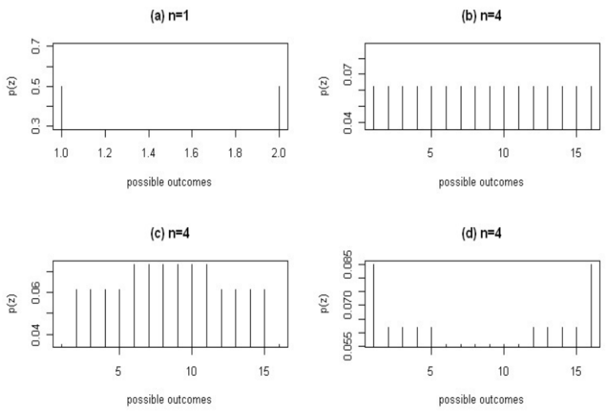


Fig. 2 multiplicative Bernoulli distribution (a) $\psi = 0.5$, $\omega = 1$ and $n = 1$, (b) $\psi = 0.5$, $\omega = 1$and $n = 4$, (c) $\psi = 0.5$, $\omega = 1.2$ and $n = 4$, and (d) $\psi = 0.5$, $\omega = 0.9$ and $n = 4$.

In Figure 2 panel (a) $n = 1$ and $\omega = 1$ we obtain the univariate Bernoulli distribution with equal probabilities at $\psi = \tau = 0.5$, panel (b) $n = 4$ and $\omega = 1$ we obtain the multivariate Bernoulli distribution for independent data with equal probabilities at $\psi = \tau = 0.5$, panel (c) $n = 4$ and $\omega = 1.2$ we obtain multiplicative Bernoulli for negative correlated binary data with higher probabilities in the middle and lower in the tails than independent data at $\psi = 0.5$, and panel (d) $n = 4$ and $\omega = 0.9$ we obtain the multiplicative Bernoulli for positive correlated binary data with lower probabilities in the middle and higher in the tails than

International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:4, No:5, 2010

independent data at $\psi = 0.5$.

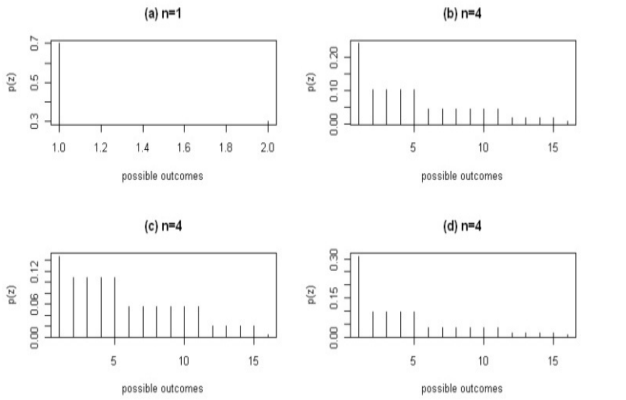


Fig. 3 (a) $\psi = 0.3$, $\omega = 1$, (b) $\psi = 0.3$, $\omega = 1$, (c) $\psi = 0.3$, $\omega = 1.2$ and (d) $\psi = 0.3$, $\omega = 0.9$.

Also in asymmetric distribution, Figure 3 shows the same pattern for negative correlated binary data the probability at the middle is higher and lower in tails than independent data while for positive correlated data the probability at the middle is lower and higher in the tails than independent data.

The expected value, variance, covariance and correlation of the multiplicative Bernoulli are given by

$$\mu = E(Z_i) = \tau,$$

$$\sigma^2 = \mathrm{V}(Z_i) = \tau[1-\tau],$$

$$E(Z_i Z_j) = \tau_2,$$

$$Cov(Z_i, Z_j) = \tau_2 - \tau^2$$

and

$$Cor(Z_i, Z_j) = \frac{\tau_2 - \tau^2}{\tau[1-\tau]}$$

where

$$\tau = \psi \frac{\kappa_{n-1}(\psi,\omega)}{\kappa_n(\psi,\omega)},$$

$$\tau_2 = \psi^2 \frac{\kappa_{n-2}(\psi,\omega)}{\kappa_n(\psi,\omega)}$$

and

$$\kappa_{n-a}(\psi,\omega) = \sum_{i=0}^{n-a} \binom{n-a}{i} \psi^i (1-\psi)^{(n-a-i)} \omega^{(n-a-i)(i+a)}$$

Table 1 shows values of μ, $E(Z_i Z_j) = E_{ij}$ and correlation (r) for some chosen values of ψ, ω and $n = 25$. We note that for $\omega < 1$, the correlation is positive, and for $\omega > 1$, the correlation is negative and for $\omega = 0$, the correlation is zero.

The parameter ω is explained as a measure of intra-units association inversely related to the condition cross-product ratio (α)

$$\omega = 1/\sqrt{\alpha(k,h\,|rest)}$$

TABLE I
THE EXPECTED AND CORRELATION VALUES FOR MULTIPLICATIVE BERNOULLI DISTRIBUTION FOR CHOSEN VALUES OF ψ, ω AND $n = 25$

		ω						
		0.80	0.90	0.95	1	1.05	1.15	1.5
ψ	μ	0.002	0.041	0.158	0.30	0.369	0.422	0.465
0.30	E_{ij}	0	0.002	0.028	0.09	0.132	0.172	0.207
	r	0	0.025	0.022	0	-.013	-.025	-.036
	μ	0.5	0.5	0.5	0.5	0.5	0.5	0.5
0.50	E_{ij}	0.495	0.370	0.264	0.25	0.246	0.243	0.241
	r	0.98	0.48	0.056	0	-.016	-.024	-.036
	μ	0.999	0.991	0.966	0.90	0.811	0.698	0.59
0.90	E_{ij}	0.999	0.982	0.933	0.81	0.657	0.482	0.341
	r	0	0	0	0	-.013	-.024	-.033
	μ	1	0.999	0.997	0.99	0.97	0.880	0.689
0.99	E_{ij}	1	0.998	0.994	0.98	0.94	0.773	0.467
	r	0	0	0	0	0	-.01	-.03

Where the conditional cross-product ratio of any two units given all others is given by

$$\alpha(i,j|\text{rest}) = \frac{P(Z_i = 0, Z_j = 0|\text{rest})P(Z_i = 1, Z_j = 1|\text{rest})}{P(Z_i = 0, Z_j = 1|\text{rest})P(Z_i = 1, Z_j = 0|\text{rest})}$$

From [11] ω can be written in terms of Yule's measure of association Q as

$$Q = \frac{\alpha - 1}{\alpha + 1} = \frac{1-\omega^2}{1+\omega^2}$$

Q is the difference between the conditional probability that the scores are concordant (have like order) and the conditional probability that the scores are discordant (have unlike order), and Yule's measure of colligation "reasonable measure of association" Y as

$$\Upsilon = \frac{\sqrt{\alpha} - 1}{\sqrt{\alpha} + 1} = \frac{1-\omega}{1+\omega}$$

Υ is a measure of proportional reduction in predictive error for standardized 2×2 table; see [11].

From [12] ψ can be written as

$$\psi = P(Z_i = 1)\frac{\kappa_n(\psi,\omega)}{\kappa_{n-1}(\psi,\omega)} = \frac{e^{2v}}{1+e^{2v}}$$

$0 < \psi < 1$, and the probability of success is

$$P(Z_i = 1) = \tau = \psi \frac{\kappa_{n-1}(\psi,\omega)}{\kappa_n(\psi,\omega)},$$

Then ψ can be thought as the probability of a particular outcome in other words the weighted probability of success that would be governing the binary response of the n units. This weighted probability of success becomes the probability of success when the binary is independent, $\omega = 1$, or $\tau = 0.5$,

$\psi = P(Z_i = 1) = \tau$ where $\kappa_{n-a}(\psi, 1) = 1 \;\forall a, n$ and $\frac{\kappa_n(0.5,\omega)}{\kappa_{n-1}(0.5,\omega)} = 1 \;\forall \omega, n$.

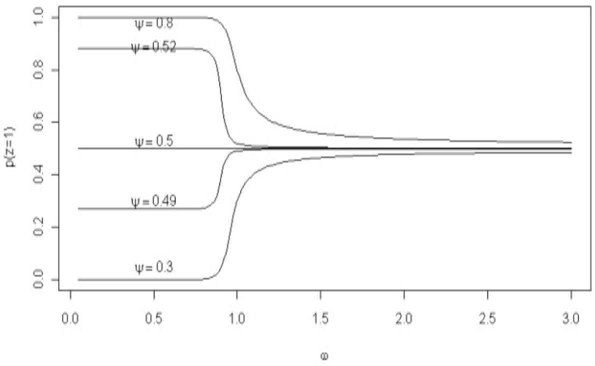


Fig. 4 relationship between ω, ψ and $P(Z_k = 1)$ for various values of ψ and $n = 25$.

Figures 4 shows the relationship between, ω, ψ and $P(Z_k = 1)$ at $n = 25$. We notice that when $\psi < 0.5$, $P(Z = 1) < 0.5$, and for $\psi > 0.5$, $P(Z = 1) > 0.5$, also for the values of $\omega < 1$, ψ tends quickly to 0 or 1 and for the values of $\omega > 1$, ψ tends quickly to 0.5.

IV. Estimation

A. Method of moments

We may use the sample expected value of Z_i and correlation between Z_i and Z_j to find moment estimates for ψ and ω as follows. The sample moments mean and correlation are

$$m_1 = \sum_{i=1}^{n} z_i / n$$

and

$$r = corr(Z_i, Z_j)$$

Equating these sample moments to the corresponding population moments we get the estimates of ψ and ω by solving the two equations

$$m_1 = \hat{\psi} \frac{\kappa_{n-1}(\hat{\psi}, \hat{\omega})}{\kappa_n(\hat{\psi}, \hat{\omega})},$$

and

$$r = \frac{\hat{\tau}_2 - \hat{\tau}^2}{\hat{\tau}[1 - \hat{\tau}]}$$

By solving these two equations we obtain moment estimate $\hat{\psi}$ and $\hat{\omega}$.

B. Maximum likelihood Estimation

By noticing that in a vector of binary responses z there are $n(n-1)/2$ pairs of responses and if the order is irrelevant three type of pairs are distinguishable: there are $(n-y)(n-y-1)/2$ pairs of $(z_i = 0, z_j = 0)$, $y(y-1)/2$ pairs of $(z_i = 1, z_j = 1)$, and $(n-y)y$ pairs of $(z_i = 0, z_j = 1)$, or $(z_i = 1, z_j = 0)$, for $y = 0, \dots, n$ and $y = \sum_{i=1}^{n} z_i$. In view of exchangeability and absence of second and higher-order interactions the estimate of ω using cross product ratio is

$$\hat{\omega} = 1/\sqrt{\hat{\alpha}}$$

and estimate of cross product ratio is

$$\hat{\alpha} = \frac{0.25y(y-1)(n-y)(n-y-1)}{\neq (z_i = 0, z_j = 1)\,\#(z_j = 1, z_i = 0)}$$

To find maximum likelihood estimate for ψ we could use the maximum likelihood method for $P(\boldsymbol{Z})$ as

$$L(\psi | n, \boldsymbol{z}, \omega) = \frac{\psi^y (1-\psi)^{n-y} \omega^{y(n-y)}}{\sum_{t=0}^{n} \binom{n}{t} \psi^t (1-\psi)^{n-t} \omega^{t(r-t)}} = \frac{q_0}{q_1}$$

$0 < \psi < 1$. The estimate $\hat{\psi}$ of ψ can be obtained by maximize L in the range (0,1). This value is the solution of the function

$$\frac{d}{d\psi} \log L(\psi | r, y, \omega) = \frac{dl}{d\psi} = \frac{q_1 \acute{q}_{0\psi} - q_0 \acute{q}_{1\psi}}{q_1^2}$$

where

$$\acute{q}_{0\psi} = \psi^y (1-\psi)^{(n-y)} \omega^{y(n-y)} \left[\frac{y}{\psi} - \frac{n-y}{1-\psi}\right]$$

and

$$\acute{q}_{1\psi} = \sum_{t=0}^{n} \binom{n}{t} \psi^t (1-\psi)^{(n-t)} \omega^{t(n-t)} \left[\frac{t}{\psi} - \frac{n-t}{1-\psi}\right]$$

Substituting by these values in $\frac{dl}{d\psi}$ we obtain

$$\frac{dl}{d\psi} = \frac{y}{\psi} - \frac{(n-y)}{1-\psi} - \frac{\acute{q}_{1\psi}}{q_1} = 0$$

This equation can be solved using numerical methods. We can find the solution using the function *uniroot* in R-software; see Appendix A.

1) Second derivative and variance

The second derivative $\frac{d^2}{d\psi^2} \log L(\psi | n, y, \omega)$ at $\hat{\psi}$ must be negative

$$\frac{d^2 l}{d\psi^2} = -\frac{y}{\psi^2} - \frac{(n-y)}{(1-\psi)^2} - \left[\frac{q_1 \acute{\acute{q}}_{1\psi} - \acute{q}_{1\psi}^2}{q_1^2}\right] < 0$$

where

$$\acute{\acute{q}}_{1\psi} = \sum_{t=0}^{n} \binom{n}{t} \psi^t (1-\psi)^{(n-t)} \omega^{t(n-t)} \left[\frac{t(t-1)}{\psi^2} - \frac{2t(n-t)}{\psi(1-\psi)} + \frac{(n-t)(n-t-1)}{(1-\psi)^2}\right]$$

The expected value of $\frac{d^2 l}{d\psi^2}$ is

$$E\left(\frac{d^2 l}{d\psi^2}\right) = -\frac{E(y)}{\psi^2} - \frac{E(n-y)}{(1-\psi)^2} - \left[\frac{q_1 \acute{\acute{q}}_{1\psi} - \acute{q}_{1\psi}^2}{q_1^2}\right]$$

where $E(y) = \mu$ and $E(n-y) = (n-\mu)$ we obtain

$$E\left(\frac{d^2 l}{d\psi^2}\right) = -\frac{\mu}{\psi^2} - \frac{(n-\mu)}{(1-\psi)^2} - \left[\frac{q_1 \acute{\acute{q}}_{1\psi} - \acute{q}_{1\psi}^2}{q_1^2}\right]$$

The asymptotic variance of $\hat{\psi}$ is

$$V(\hat{\psi}) \approx \frac{-1}{E\left(\frac{d^2 l}{d\psi^2}\right)} = \frac{1}{\frac{\mu}{\psi^2} + \frac{(n-\mu)}{(1-\psi)^2} + \left[\frac{q_1 \acute{\acute{q}}_{1\psi} - \acute{q}_{1\psi}^2}{q_1^2}\right]}$$

This can be estimated as

$$\hat{V}(\hat{\psi}) \approx \frac{1}{\frac{\hat{\mu}}{\hat{\psi}^2} + \frac{(n-\hat{\mu})}{\left(1-\hat{\psi}\right)^2} + \left[\frac{\hat{q}_1 \acute{\acute{q}}_{1\hat{\psi}} - \acute{q}_{1\hat{\psi}}^2}{\hat{q}_1^2}\right]}$$

We may obtain the approximate variance for $\widehat{\omega}$ as

$$\widehat{\omega} = g(\hat{\alpha}) = 1/\sqrt{\hat{\alpha}}$$

Using Taylor's approximation where

$$V(\widehat{\omega}) \approx V(\hat{\alpha})[\acute{g}(\mu_{\hat{\alpha}})]^2$$

The variance of cross product ratio $(\hat{\alpha})$ can be obtained from [11] as

$$V(\hat{\alpha}) \approx \frac{\alpha^2}{n}\left[\frac{1}{P_{11}} + \frac{1}{P_{12}} + \frac{1}{P_{21}} + \frac{1}{P_{22}}\right]$$

hence,

$$[\acute{g}(\mu_{\hat{\alpha}})]^2 = [-0.5(\mu_{\hat{\alpha}})^{-1.5}]^2 = 0.25\alpha^{-3}$$

By noting that $\alpha^{-1} = \omega^2$ we obtain

$$V(\widehat{\omega}) \approx \frac{\alpha^{-1}}{4n}\left[\frac{1}{P_{11}} + \frac{1}{P_{12}} + \frac{1}{P_{21}} + \frac{1}{P_{22}}\right] = \frac{\omega^2}{4n}\left[\frac{1}{P_{11}} + \frac{1}{P_{12}} + \frac{1}{P_{21}} + \frac{1}{P_{22}}\right]$$

This can be estimated as

$$\hat{V}(\widehat{\omega}) \approx \frac{\hat{\alpha}^{-1}}{4}\left[\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}\right] = \frac{\widehat{\omega}^2}{4}\left[\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}\right]$$

where

$$n_{11} = 0.5(n-y)(n-y-1), \quad n_{12} = \neq\left(z_i = 0, z_j = 1\right), \quad n_{21} = \neq\left(z_i = 1, z_j = 0\right), \quad n_{22} = 0.5y(y-1)$$

V. DISTRIBUTION OF THE DENSITY OF THE GRAPH

If Z_{ij} has a multiplicative Bernoulli distribution, the δ has multiplicative binomial with expected value and variances

$$E(\delta) = \frac{\sum_{i=1}^{g-1}\sum_{j=i+1}^{g} E(Z_{ij})}{N} = \frac{N\tau}{N} = \tau$$

and

$$V(\delta) = \frac{\sum_{i=1}^{g-1}\sum_{j=i+1}^{g}\sum_{k=1}^{g-1}\sum_{h=k+1}^{g} Cov(Z_{ij}, Z_{kh})}{N^2}$$

If $ij = kh$, we have $NV(Z)$ and for $ij \neq kh$ we have $N(N-1)Cov(Z_i, Z_j)$ then

$$V(\delta) = \frac{N\tau(1-\tau) + N(N-1)(\tau_2 - \tau^2)}{N^2} = \frac{\tau(1-N\tau) + (N-1)\tau_2}{N}$$

The estimated values are

$$\hat{E}(\delta) = \hat{\tau}$$

and

$$\hat{V}(\delta) = \frac{\hat{\tau}(1-N\hat{\tau}) + (N-1)\hat{\tau}_2}{N}$$

The variance of the mean nodal degree is

$$V(\bar{d}) = (g-1)^2 V(\delta)$$

and its estimated is

$$\hat{V}(\bar{d}) = (g-1)^2 \hat{V}(\delta)$$

For large sample size the confidence interval for the population density is

$$\delta \mp Z_{1-\alpha/2}\sqrt{\hat{V}(\delta)}$$

The confidence interval for mean nodal degree is

$$\bar{d} \mp Z_{1-\alpha/2}\sqrt{\hat{V}(\bar{d})}$$

$Z_{1-\alpha/2}$ is the standard normal percentile.

VI. APPLICATION

The network in Figure 5 shows the lives near relation for six children; see [14]. The relation of "lives near" is non-directional because the statement "a child i is "lives near" a child j" is equivalent to " a child j is "lives near" a child i".

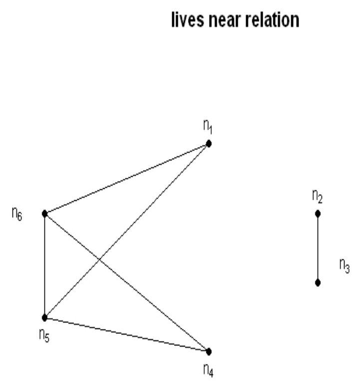


Fig. 5 lives near relation for six children

The number of nodes is $g = 6$, and the values of $z = c(0,0,0,1,1,1,0,0,0,0,0,0,1,1,1)$,
$N = 15$, $L = 6$ and the estimated density is

$$\delta = \frac{6}{0.5(6)(6-1)} = 0.4$$

Moment estimates: The sample moments are

$$m_1 = 0.4 \quad \text{and} \quad r = -0.042$$

The moment estimates are obtained as

$$\hat{\psi} = 0.267 \quad \text{and} \quad \hat{\omega} = 1.229$$

Maximum likelihood estimates:

The estimated values are

$$\hat{\psi} = 0.339 \quad \text{and} \quad \hat{\omega} = 1.095$$

See Appendix A for details. The Yule's measures of association are

$$\hat{Q} = \frac{1-\sqrt{1.095}}{1+\sqrt{1.095}} = -0.023 \quad \text{and} \quad \hat{\Upsilon} = \frac{1-1.095}{1+1.095} = -.045$$

The estimated variance of density using maximum likelihood is

$$\hat{V}(\delta) = 0.0098$$

If we use the Bernoulli independent model the variance is

$$\hat{V}(\delta_{ind}) = \frac{0.4(1-.4)}{15} = 0.016$$

and the efficiency with respect the dependent model is

$$eff = \frac{0.016}{0.0098} = 1.63$$

The 95% confidence interval is

$$0.40 \mp 1.96\sqrt{0.0098} = 0.21 \ldots .0.59$$

Also, the variance of the mean nodal degree ($\bar{d} = (5)(.4) = 2$) is

$$\hat{V}(\bar{d}) = (25)0.0098 = 0.2459$$

Its 95% confidence interval is

$$2 \mp 1.96\sqrt{0.2459} = 1.03 \ldots .2.97$$

VII. Conclusion

We have studied a new method for finding the distribution of mean nodal and density of the graph. The main advantage of the new method its generalization where it can be applied with dependent and independent binary variables. It is more efficient than classical method in case of units negatively associated. Moreover, we estimated the parameters of multiplicative Bernoulli distribution using method of moments and maximum likelihood and obtained the asymptotic variances of the estimates.

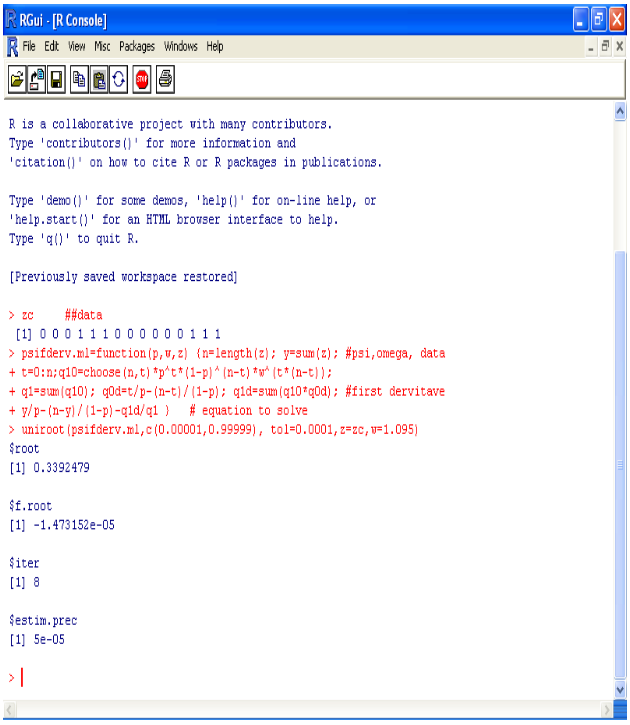
APPENDIX

Function to compute ML estimate for psi based on R-software:

The function is called psifderv.ml

- psifderv.ml=function(p,w,z) {n=length(z); y=sum(z); #psi,omega, data t=0:n;q10=choose(n,t)*p^t*(1-p)^(n-t)*w^(t*(n-t)); q1=sum(q10); q0d=t/p-(n-t)/(1-p); q1d=sum(q10*q0d); #first dervitave y/p-(n-y)/(1-p)-q1d/q1 } # equation to solve
- uniroot(psifderv.ml,c(0.00001,0.99999), tol=0.0001, z=zc, w=1.095) # uniroot function to find ML estimate for psi

The solution is 0.399 (given below $root)



References

[1] P. Altham, " Two generalizations of the binomial distribution. *Applied Statistics*, 27, 1978, pp. 162-167.

[2] B.S. Anderson, Butts, C. and Carley, K." The interaction of size and density with graph-level indices. *Social Networks*, 3, 1999, pp.239-267.

[3] P.J. Carrington, P.J., Scott, J. and Wasserman, S. "*Models and Methods in Social Network Analysis*" 1st Edition, Cambridge University Press, 2005.

[4] K.K.S. Chung, Hossain L. and Davis J. "Exploring sociocentric and egocentric approaches for social network analysis", http://kamp2005.vuw.ac.nz/papers/,2005

[5] D.R. Cox, D.R. "The analysis of multivariate binary data" *Applied Statistics*, 21, 1972,pp. 113-120.

[6] K. Faust "Comparing social networks: size, density, and local structure" *Advances in Methodology and Statistics*, 3, 2006, pp.185-216.

[7] K. Faust "Triadic configuration in limited choice sociometric networks: Empirical and theoretical results" *Social Networks*, 30, 2008, pp. 273-283.

[8] R. A. Hanneman "*Introduction to Social Network Methods*" http://faculty.ucr.edu/~hanneman, 2001.

[9] N.L.*Johnson, Kotz, S. and Kemp, A.W. "Univariate Discrete Distributions" 2nd Edition. Wiley, New York. 1992.*

[10] D. Knoke and Yang, S. "*Social Network Analysis*" 2nd Edition, Sage Publications, 2008.

[11] A.M. Liebetrau "*Measures of Association*" 1st Edition, Sage Publications, 1983.

[12] G. Lovison "An alternative representation of Altham's multiplicative-binomial distribution" *Statistics & Probability Letters*, 36,1998, 415-420.

[13] J. Scott "*Social Network Analysis: A handbook*" 2nd Edition, Sage Publcations, 2000.

[14] S. Wasserman and Faust, K. "*Social Network Analysis: Methods and Applications*" Cambridge University Press, New York.Yor, 1994.