Degree, Poisson, and Negative Binomial Distributions

Barum Park

Department of Sociology
Cornell University
b.park@cornell.edu

March 8, 2021

In DiPrete and colleagues used the relationship between the Poisson and Negative Binomial distributions to set up their null model and measure of segregation. Here, we explore the relation between these distributions in a bit more detail. We start with the degree distribution of $G(n,p)$ graphs (sometimes call Erdos-Renyi or Poisson graphs) and then move on to the model of Diprete and colleagues.

$G(n,p)$ and Poisson Graphs

Consider a graph G with n nodes. Suppose, for now, that n remains fixed and that we place the (undirected) edges between these nodes uniformly at random with probability p. This is the simplest random graph model we might imagine and is referred to as the $G(n,p)$ model. Notice that there is also a $G(n,m)$ model in which a the probability distribution is not defined over edges but over the set of all possible graphs with n nodes and m edges. In practice, the $G(n,p)$ model and $G(n,m)$ model behave very similarly.

How would the degree distribution of a $G(n,p)$ look like? Each node i can be connected to any of the $n-1$ other nodes in the graph. Since these edges are formed independently and with probability p, we see that the degree of node i, D_i, will follow a Binomial$(n-1,p)$ distribution, i.e.,

$$P(D_i = k) = \binom{n-1}{k} p^k (1-p)^{n-k-1}.$$

Now, if we let the size of the network grow ($n \to \infty$) and the "success" probability shrink ($p \to 0$) in such a way that the product $(n-1)p$ converges to a constant λ, then D_i will converge in distribution to a random variable that has a Poisson distribution with parameter λ, i.e., for the PMF of D_i, we have

$$\binom{n-1}{k} p^k (1-p)^{n-k-1} \longrightarrow \frac{\lambda^k e^{-\lambda}}{k!}$$

as $n \to \infty$ and $p \to 0$ such that $(n-1)p \to \lambda$. (you can find a proof of this in any standard statistics textbook or here). Hence, in large networks, where edges are placed independently across the nodes, the degree distribution can be approximated with a Poisson distribution with parameter $\lambda = (n-1)p$.

Degree distributions in Randomly Sampled Ego-Networks

The distributions considered in the article of DiPrete et al. are a bit different from those above. The main reason is that they are studying randomly sampled ego-networks: a set of respondents is randomly sampled from the population and each respondent is asked about his/her ties to alters. So, we do not observe all the ties in the network (indeed, we observe only a negligible fraction). Assuming that there is no limit to the number of alters that the sampled respondent can enlist, the number of ties that the ego holds will accurately describe her degree. Further, the sample degree-distribution would be a consistent estimator of the population degree distribution.[1]

Suppose we use this method to collect information regarding the alters of the randomly sampled respondents. In particular, we ask them whether each of their alters belong to certain demographic groups/categories (e.g., race, education level, etc). If we have at our disposal the population proportion of each of these categories, we can create a reasonable null model of how the degree distribution should look like if the population mixes randomly.

Let Y_{ig} be the total number of ties that respondent i has to members of group g. Suppose that the true population-network from which we have sampled these respondents (and their ego-networks) is a $G(N,p)$ graph, i.e., all ties are generated uniformly at random between all individuals in society with probability p. As discussed above, we might model the degree of each sampled respondent, $D_i = \sum_{g=1}^{G} Y_{ig}$, with a Poisson distribution with parameter $(N-1)p$. Also, under random mixing, the number ties that each group receives (i.e., their in-degree) would be proportional to their population share. Hence, we have

$$\mathrm{E}[D_i] = (N-1)p \qquad \mathrm{E}[Y_{ig} \,|\, D_i] = D_i \pi_g$$

where π_g is the population share of group g. It follows that

$$\mathrm{E}[Y_{ig}] = \mathrm{E}_{D_i}[\mathrm{E}[Y_{ig} \,|\, D_i]] = (N-1)p\pi_g$$

and

$$Y_{ig} \sim \text{Poisson}(\mu_g), \qquad \mu_g = (N-1)p\pi_g.$$

The important point is that the the parameter of this distribution varies only by group g but not the individual: it assumes that the expected degree of each respondent is the same, which, of course is an unrealistic assumption.

To relax this assumption, we might regard $D_i = \alpha_i$ as a parameter itself, and write

$$Y_{ig} \sim \text{Poisson}(\mu_{ig}), \qquad \mu_{ig} = \alpha_i \pi_g$$

which is the null model used in DiPrete et al. (2011). Recall that this model still entails the assumption that the population mixes randomly. We have allowed for heterogeneity in the expected degrees but kept the assumption that the ties are distributed proportionally to group size (i.e., no group is favored by any of the respondents).

The last step to arrive at DiPrete et al.'s model is to add a parameter to the model that captures the relative propensity of individual i to have a tie of group g, by adding the parameter

$$\nu_{ig} = \frac{i\text{'s idiosyncratic rate of creating ties to group } g}{\text{Expected number of ties from } i \text{ to } g \text{ under random mixing}} = \frac{\lambda_{ig}}{\alpha_i \pi_g}$$

so that ν_{ig} might be interpreted as the "excess rate" of individual i to create ties to members of group g. Substituting this expression, one obtains

$$\mu_{ig} = \alpha_i \pi_g \nu_{ig}.$$

[1] Formally, this is saying that the empirical distribution function (EDF) of the degrees converges to the CDF of the underlying random variable. Intuitively, this can be understood as saying that, if our sample size would be infinitely large, the degree distribution of in the sample would look like the population degree-distribution.

Now, this creates the problem that the model is no longer identified. The respondent-to-group matrix will have $n \times G$ entries, where n is the sample size, but now we have $n + G + (n \times G)$ parameters. The solution is to assign a distribution to ν_{ig}, governed by a few parameters, and which is *independently* distributed from Y_{ig}. For example, if the distribution would have only two parameters, the number of parameters to be estimated from the model (after we integrate ν_{ig} out) would be $n + G + 2$ which will be strictly less than $n \times G$ in any reasonable dataset.

One convenient choice is the the Gamma distribution. The Gamma distribution has two parameters (α, β) and PDF

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \qquad x > 0, \alpha > 0, \beta > 0,$$

where $\Gamma(v) = \int_0^\infty x^{v-1} e^{-x} \mathrm{d}x$ is the Gamma function and can be understood as a generalized factorial function, since $\Gamma(n) = (n-1)!$ if n is an integer.

Assume $\nu_{ig} \sim \text{Gamma}(\phi, \phi)$, with ϕ being an integer. Then, maintaining the assumption of independence, the joint density of Y_{ig} and ν_{ig} is

$$p(y_{ig}, \nu_{ig}) = p(y_{ig} \mid \mu_{ig}) p(\nu_{ig}) = \underbrace{\left(\frac{\mu_{ig}^{y_{ig}}}{y_{ig}!} e^{-\mu_{ig}} \right)}_{\text{Poisson}(y_{ig} \mid \mu_{ig})} \underbrace{\left(\frac{\phi^\phi}{\Gamma(\phi)} \nu_{ig}^{\phi-1} e^{-\phi \nu_{ig}} \right)}_{\text{Gamma}(\nu_{ig} \mid \phi, \phi)}$$

and integrating ν_{ig} out, we obtain

$$\begin{aligned}
p(y_{ig}) &= \int_0^\infty p(y_{ig}, \nu_{ig}) \mathrm{d}\nu_{ig} \\
&= \int_0^\infty \left(\frac{\mu_{ig}^{y_{ig}}}{y_{ig}!} e^{-\mu_{ig}} \right) \left(\frac{\phi^\phi}{\Gamma(\phi)} \nu_{ig}^{\phi-1} e^{-\phi \nu_{ig}} \right) \mathrm{d}\nu_{ig} \\
&= \int_0^\infty \left(\frac{(\alpha_i \pi_g \nu_{ig})^{y_{ig}}}{y_{ig}!} e^{-\alpha_i \pi_g \nu_{ig}} \right) \left(\frac{\phi^\phi}{\Gamma(\phi)} \nu_{ig}^{\phi-1} e^{-\phi \nu_{ig}} \right) \mathrm{d}\nu_{ig} \\
&= \frac{(\alpha_i \pi_g)^{y_{ij}} \phi^\phi}{y_{ig}! \Gamma(\phi)} \int_0^\infty \underbrace{\nu_{ig}^{y_{ig}+\phi-1} e^{-(\alpha_i \pi_g + \phi) \nu_{ig}}}_{\text{This is the kernel of a Gamma}(y_{ig}+\phi,\, \alpha_i\pi_g+\phi)\text{ density!}} \mathrm{d}\nu_{ig} \\
&= \frac{(\alpha_i \pi_g)^{y_{ij}} \phi^\phi}{y_{ig}! \Gamma(\phi)} \frac{\Gamma(y_{ig} + \phi)}{(\alpha_i \pi_g + \phi)^{y_{ig}+\phi}} \underbrace{\int_0^\infty \frac{(\alpha_i \pi_g + \phi)^{y_{ig}+\phi}}{\Gamma(y_{ig} + \phi)} \nu_{ig}^{y_{ig}+\phi-1} e^{-(\alpha_i \pi_g + \phi) \nu_{ig}} \mathrm{d}\nu_{ig}}_{\text{This integrates to 1, since it's a PDF}} \\
&= \frac{(\alpha_i \pi_g)^{y_{ij}} \phi^\phi}{y_{ig}! \Gamma(\phi)} \frac{\Gamma(y_{ig} + \phi)}{(\alpha_i \pi_g + \phi)^{y_{ig}+\phi}} \\
&= \frac{(\alpha_i \pi_g)^{y_{ij}} \phi^\phi}{y_{ig}! (\phi - 1)!} \frac{(y_{ig} + \phi - 1)!}{(\alpha_i \pi_g + \phi)^{y_{ig}+\phi} (\alpha_i \pi_g + \phi)^\phi} \\
&= \binom{y_{ig} + \phi - 1}{y_{ig}} \left(\frac{\alpha_i \pi_g}{\alpha_i \pi_g + \phi} \right)^{y_{ig}} \left(\frac{\phi}{\alpha_i \pi_g + \phi} \right)^\phi \\
&= \binom{y_{ig} + \phi - 1}{y_{ig}} \eta_{ig}^{y_{ig}} (1 - \eta_{ig})^\phi
\end{aligned}$$

with $\eta_{ig} = \alpha_i \pi_g / (\alpha_i \pi_g + \phi)$, which we identify as a Negative-Binomial$(y_{ig} | \phi, \eta_{ig})$ distribution, where η is the "success probability" and ϕ the number of failures before the (hypothetical) experiment stops. Notice that this distribution has mean and variance equal to

$$\mathrm{E}[Y_{ig}] = \frac{\eta_{ig} \phi}{1 - \eta_{ig}} = \alpha_i \pi_g \qquad \text{and} \qquad \mathrm{Var}[Y_{ig}] = \frac{\eta_{ig} \phi}{(1 - \eta_{ig})^2} = \mathrm{E}[Y_{ig}] (\alpha_i \pi_g + \phi) / \phi.$$

DiPrete et al. (2011) assigned each group g a unique ϕ parameter and took $\omega_g = \left(\frac{\alpha_i \pi_g + \phi_g}{\phi_g}\right)$ as the "overdispersion" parameter.