

Chapter 1

Specification of ERGM Count

In this chapter, we will first define ERGM Count and its relevant shaping functions as proposed by Krivitsky (2012). Then, we will introduce the simplest form of ERGM Count which are modeled Poisson and Geometric Distribution. The various shapes of ERGM Count will then be explored and the interpretation of each the shapes will also be discussed accordingly.

1.1 Model Definition

Let the set of actors in the network be N and n represent the number of actors involved, such that $n \equiv |N|$. Let \mathbb{Y} be the set of all dyads where a dyad represent a unique pair of actors that may have relational interest. Since the later applied model is directed, $\mathbb{Y} \subseteq N \times N$.

Let the edge y_{ij} represents the value of relational interest of dyad (i, j) that originates from actor i and received by actor j . In Binary ERGM, each edge can only have binary values, $y_{ij} \in \{0, 1\}$. But in ERGM Count, an edge can take the value of any whole numbers, such that $y_{ij} \in \mathbb{N}_0$. Hence the set of possible network configuration, or the sample space is a set $\mathcal{Y} \subseteq \mathbb{N}_0^{\mathbb{Y}}$. Let the random network of count configuration be $\mathbf{Y} \in \mathcal{Y}$, with \mathbf{y} as its realization.

The unknown model parameter $\boldsymbol{\theta} \in \Theta$ where $\Theta \subseteq \mathbb{R}^q$ is then mapped to \mathbb{R}^p by the function $\eta : \Theta \rightarrow \mathbb{R}^p$. Given chosen sufficient statistics $g : \mathcal{Y} \rightarrow \mathbb{R}^p$ (which may also depend on external covariate \mathbf{x}), and reference measure $h : \mathcal{Y} \rightarrow [0, \infty)$. The ERGM Count of \mathbf{Y} has Probability Mass

Function (PMF):

$$Pr_{\boldsymbol{\theta};h,\eta,g}(\mathbf{Y} = \mathbf{y}) = \frac{h(\mathbf{y}) \exp(\eta(\boldsymbol{\theta}) \cdot g(\mathbf{y}))}{k_{h,\eta,g}(\boldsymbol{\theta})}, \quad (1.1)$$

where the denominator is a normalization constant defined as:

$$k_{h,\eta,g}(\boldsymbol{\theta}) = \sum_{\mathbf{y} \in \mathcal{Y}} h(\mathbf{y}) \exp(\eta(\boldsymbol{\theta}) \cdot g(\mathbf{y})). \quad (1.2)$$

There is an additional constraint for model parameter $\boldsymbol{\theta}$ to be an element from the set:

$$\Theta \subseteq \Theta_N = \{\boldsymbol{\theta}' \in \mathbb{R}^q : k_{h,\eta,g}(\boldsymbol{\theta}') < \infty\} \quad (1.3)$$

This constraint is not necessary in Binary ERGM as their sample space is finite (Krivitsky, 2012). In many occasions, the constraint for the parameters is not known explicitly.

To simplify the notation, the PMF can be defined up to normalization constant as:

$$Pr(\mathbf{Y} = \mathbf{y}|\boldsymbol{\theta}) \propto h(\mathbf{y}) \exp(\eta(\boldsymbol{\theta}) \cdot g(\mathbf{y})) \quad (1.4)$$

This assumes the constraint 1.3 is met.

1.2 Reference Measure and Baseline Distribution

In Binary ERGM, if the model has *dyadic independence* property, the model will reduce to edge-wise logistic regression. Krivitsky (2012) stated that ERGM of valued network should reduce into Generalized Linear Model given *dyadic independence*.

The specification of an appropriate reference measure enable such relationship to occur. Suppose the model only has one predictor θ , with sufficient statistics, **sum** : $g(\mathbf{y}) = \sum_{(i,j) \in \mathbb{Y}} y_{ij}$. If the reference measure, $h(\mathbf{y}) = 1$, the model's PMF will be:

$$Pr(\mathbf{Y} = \mathbf{y}|\theta) \propto \exp \left(\theta \sum_{(i,j) \in \mathbb{Y}} y_{ij} \right) = \prod_{(i,j) \in \mathbb{Y}} (\exp(\theta))^{y_{ij}} \quad (1.5)$$

We recognize this as an edge-wise distribution $Y_{ij} \stackrel{\text{iid}}{\sim} \text{Geometric}(p = 1 - \exp(\theta))$ where $\theta < 0$. This corresponds to the baseline distribution of *geometric-reference*. Whereas if the reference measure is $h(\mathbf{y}) = \prod_{(i,j) \in \mathbb{Y}} (y_{ij}!)^{-1}$, the model will have PMF:

$$\Pr(\mathbf{Y} = \mathbf{y} | \theta) \propto \frac{\exp\left(\theta \sum_{(i,j) \in \mathbb{Y}} y_{ij}\right)}{\prod_{(i,j) \in \mathbb{Y}} y_{ij}!} = \prod_{(i,j) \in \mathbb{Y}} \frac{(\exp(\theta))^{y_{ij}}}{y_{ij}!} \quad (1.6)$$

We recognize this as another edge-wise distribution $Y_{ij} \stackrel{\text{iid}}{\sim} \text{Poisson}(\lambda = \exp(\theta))$ where $\theta \in \mathbb{R}$. This corresponds to the baseline distribution of *poisson-reference*. Notice that the 2 models have different model parameter constraints. We illustrate the shape of Poisson and Geometric Distribution in Figure 1.1 below.

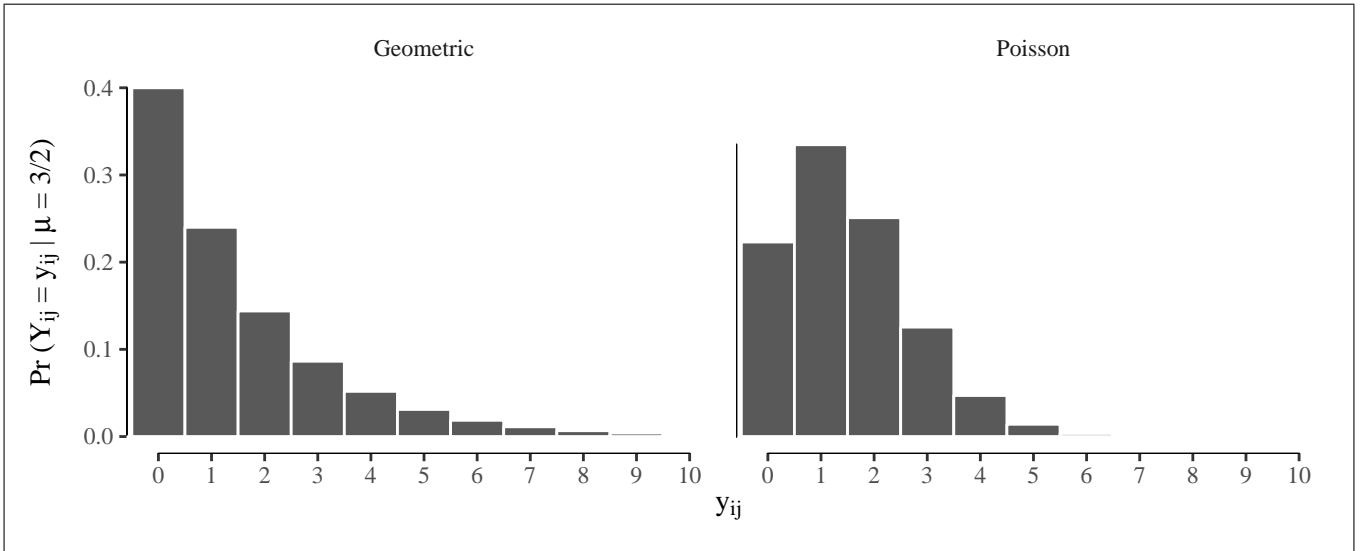


Figure 1.1: These plots show the shape of Geometric and Poisson Distribution. Both distributions have equal mean, $\mu = \frac{3}{2}$. Differs from Poisson, Geometric Distribution has a strictly decreasing probability as y_{ij} increased.

On a side note, there is an error in Krivitsky (2012) on the normalization constant of the 2 distributions above. For equation 1.5, Krivitsky (2012) normalization constant is $\prod_{(i,j) \in \mathbb{Y}} \frac{1}{1 - \exp(\theta)}$ when it is supposed to be $\prod_{(i,j) \in \mathbb{Y}} 1 - \exp(\theta)$. Whereas for equation 1.6, the normalization constant is supposed to be $\prod_{(i,j) \in \mathbb{Y}} e^{-\exp(\theta)}$ instead of $\prod_{(i,j) \in \mathbb{Y}} \frac{1}{\exp(\theta)}$. However, these mentioned errors do not affect his conclusion on the respective models.

Besides the Poisson and Geometric Distribution, there are other *reference measures* which corresponds to other *baseline distribution*. Krivitsky and Butts (2016) introduced *binomial-reference* and *truncated geometric-reference* where it still focus on count edge's value, but there is a cap on the maximum attainable value. These constrained models will not be explored in this paper.

1.3 ERGM Count Modelling

Previously, the form of the ERGM Count was reduced to edge-wise Poisson and Geometric Distribution. These two models are rather simple and easily interpreted. However, both may not be the model that best fit the later application on Chapter 4. In addition, these two modes do not take into account the complex relationship of a network. We will explore more complex models and their interpretations later.

In this section, we will focus on *poisson-reference* ERGM Count without complex constraints: $\mathcal{Y} \in \mathcal{N}_0^{\mathcal{Y}}$ and $h(\mathbf{y}) = \prod_{(i,j) \in \mathbb{Y}} (y_{ij}!)^{-1}$.

Furthermore, the linear ERGM Count restriction will also be applied where $\eta(\boldsymbol{\theta}) = \boldsymbol{\theta}$ and thus $q = p$. This restriction makes the model comprehensible. In linear ERGM Count, if Θ_N is an open set, then $\theta'_k > \theta_k \implies E(g_k(\mathbf{Y})|\theta, \theta'_k) > E(g_k(\mathbf{Y})|\theta, \theta_k)$. In other words, if other model parameters $\theta \setminus \theta_k$ are fixed, the expected sufficient statistics g_k is strictly increasing in θ_k . Larger θ_k will result in a distribution of networks with more feature gauged by g_k .

1.3.1 Exogenous Covariates

Let the possibly predictive variables outside the network structure or exogenous covariate be x_{ijk} . A *poisson-reference* ERGM Count will reduce to Poisson Regression Model if we introduce an edge-independent sufficient statistics:

$$g_k(\mathbf{y}) = \sum_{(i,j) \in \mathbb{Y}} y_{ij} x_{ijk}. \quad (1.7)$$

Let the model sufficient statistics only be exogenous covariate and **sum** the PMF will reduce to:

$$\begin{aligned} Pr(\mathbf{Y} = \mathbf{y}|\theta) &\propto \frac{\exp\left(\theta_1 \sum_{(i,j) \in \mathbb{Y}} y_{ij} + \theta_k \sum_{(i,j) \in \mathbb{Y}} y_{ij} x_{ijk}\right)}{\prod_{(i,j) \in \mathbb{Y}} y_{ij}!} \\ &= \prod_{(i,j) \in \mathbb{Y}} \frac{(\exp(\theta_1 + \theta_k x_{ijk}))^{y_{ij}}}{y_{ij}!} \end{aligned}$$

This can be recognised as an edge-wise distribution $Y_{ij} \stackrel{\text{iid}}{\sim} \text{poisson}(\lambda = \exp(\theta_1 + \theta_k x_{ijk}))$. Notice that adding exogenous covariate will maintain the poisson relationship yet changing the parameter

λ for edge (i, j) by multiplication of factor $e^{\theta_k x_{ijk}}$. It is obvious that further addition of another exogenous covariate will give the same effect. Hence, this model is equivalent to applying Poisson regression model to the network data.

In current R implementation of ERGM Count, there can be 3 representation of exogenous covariates for directed network. These are **nodecov** which is the attribute of the actor where the edge originate x_i and **nodeicov** which is the attribute of receiving actor x_j . The last one is x_{ij} or **edgescov** which represent the external edge attribute.

1.3.2 Zero Modification

The idea behind zero modification is to move away from the baseline distribution with new PMF that has more accurate representation of its $Pr(Y_{ij} = 0)$. The proposed statistics to represent zero modification is **nonzero**:

$$g(\mathbf{y}) = \sum_{(i,j) \in \mathbb{Y}} \mathbb{I}(y_{ij} > 0) \quad (1.8)$$

This statistics is simply counting the number of edges that has value bigger than 0. Suppose we model ERGM Count that only contain **sum** and **nonzero** statistics. Then the PMF will reduce to:

$$\begin{aligned} Pr(\mathbf{Y} = \mathbf{y} | \theta_1, \theta_2) &\propto \frac{\exp\left(\theta_1 \sum_{(i,j) \in \mathbb{Y}} y_{ij} + \theta_2 \sum_{(i,j) \in \mathbb{Y}} \mathbb{I}(y_{ij} > 0)\right)}{\prod_{(i,j) \in \mathbb{Y}} y_{ij}!} \\ &= \prod_{(i,j) \in \mathbb{Y}} \frac{\exp(\theta_1 y_{ij} + \theta_2 \mathbb{I}(y_{ij} > 0))}{y_{ij}!} \end{aligned}$$

This can be recognised as independent edge-wise Zero Modified Poisson, which is a Poisson Distribution with $Pr(Y_{ij} = 0)$ can be adjusted by θ_2 while $Pr(Y_{ij} = y_{ij}), y_{ij} = 1, 2, \dots$, maintain their relative proportion. The corresponding parameter θ_2 can take on any real values. If $\theta_2 < 0$, then the probability of an edge to take value of zero is higher than Poisson Distribution. Whereas if $\theta_2 > 0$, the number of non-zero edge has higher probability than Poisson Distribution, thus giving the name **nonzero**. Figure 1.2 will illustrate this relationship.

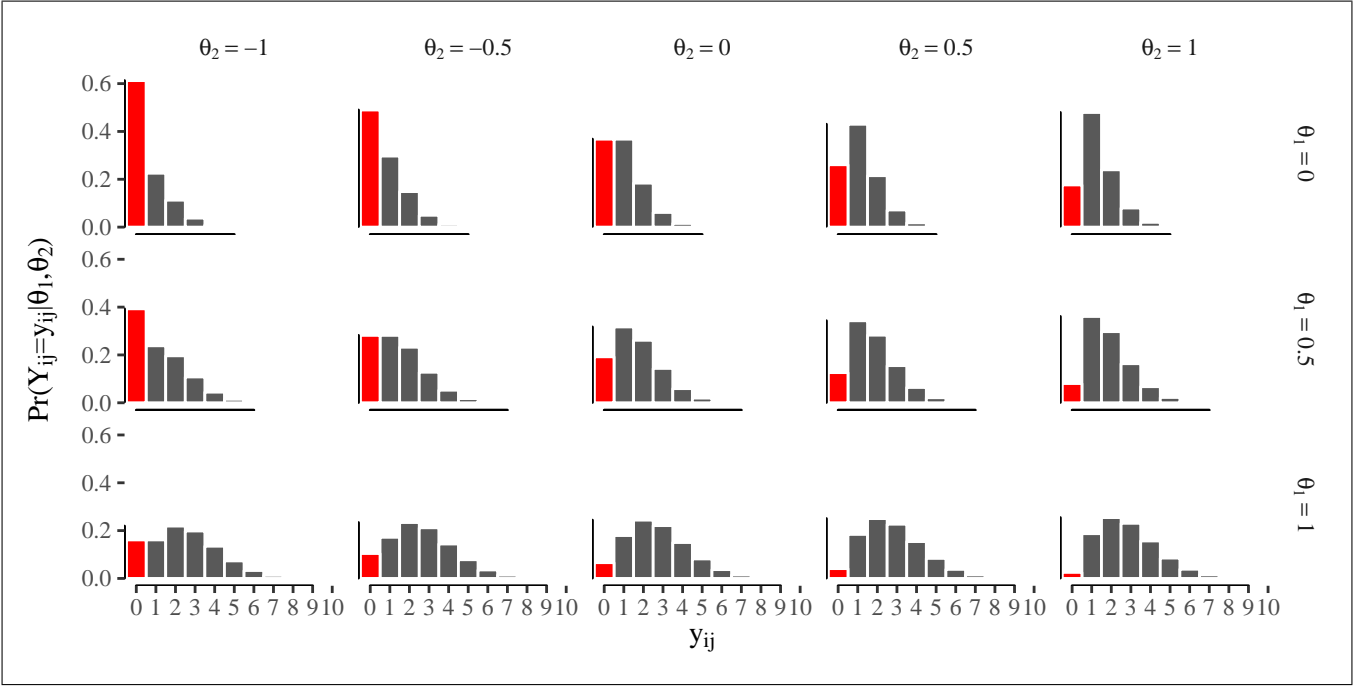


Figure 1.2: This figure shows the effect of adding on **nonzero** to the Poisson model. The top axis indicates parameter θ_2 , which affect the probability of zero edges. Notice that the effect diminished as parameter θ_1 (right axis) gets higher. Furthermore, e^{θ_1} is the mean parameter when $\theta_2 = 0$.

1.3.3 Dispersion Modelling

Poisson distribution has only 1 parameter that affect the distribution. This parameter found to affect the mean and the spread of the distribution. For one particular value of mean, there is only one particular spread poisson can attain. In other words, the spread is fixed and not adjustable. The same occurs with other one parameter distributions such as geometric.

One method to overcome this limitation is by adding sufficient statistic, **CMP**:

$$g(\mathbf{y}) = \sum_{(i,j) \in \mathbb{Y}} \log(y_{ij}!) \quad (1.9)$$

When only sufficient statistics use in the model is **sum** and **CMP** with poisson reference, the PMF:

$$\begin{aligned} Pr(\mathbf{Y} = \mathbf{y} | \theta_1, \theta_2) &\propto \frac{\exp\left(\theta_1 \sum_{(i,j) \in \mathbb{Y}} y_{ij} + \theta_2 \sum_{(i,j) \in \mathbb{Y}} \log(y_{ij}!)\right)}{\prod_{(i,j) \in \mathbb{Y}} y_{ij}!} \\ &\propto \prod_{(i,j) \in \mathbb{Y}} (\exp(\theta_1))^{y_{ij}} (y_{ij}!)^{\theta_2 - 1} \end{aligned}$$

Hence the model reduced to edge-wise *Conway-Maxwell-Poisson* **CMP** distribution $Y_{ij} \stackrel{\text{iid}}{\sim} \text{CMP}(\lambda =$

$\exp(\theta_1), \nu = 1 - \theta_2)$ where $\theta_2 \leq 1$. When θ_2 goes from 0 to $-\infty$, the distribution become less and less dispersed compare to Poisson. When $\theta_2 \rightarrow -\infty$, it is reduced to Bernoulli Distribution. Whereas if as θ_2 from 0 to 1, the distribution is more dispersed compare to Poisson. It is reduced to Geometric Distribution at $\theta_2 = 1$. Figure ... will illustrate the various shapes of CMP distribution.

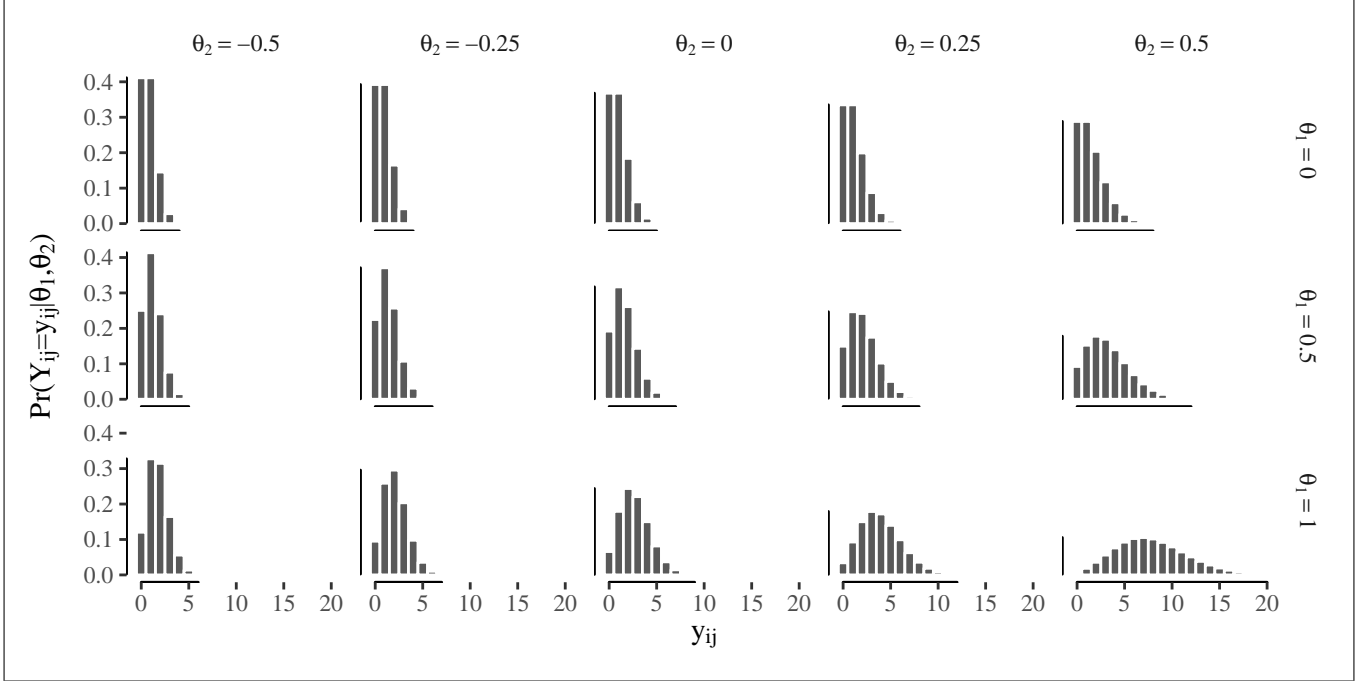


Figure 1.3: This image shows the various shapes of Conway-Maxwell-Poisson Distribution (CMP). The right axis corresponds to parameter θ_1 . Whereas the top axis indicates parameter θ_2 , which dominates the dispersion effect of the distribution. When $\theta_2 = 0$, CMP reduce to Poisson with mean e^{θ_1} . When $\theta_2 > 0$ the distribution become more dispersed, then it gets less dispersed when $\theta_2 < 0$.

1.3.4 Mutuality

The previous sections discussed the various models that has independent edge-wise assumption. Although, normalization constants tend to be intractable, they can be easily recognised as one particular distribution. However, fitting network data to independent edge-wise model will undermine the inherent relational characteristic of network properties.

In directed network, the strength of mutuality is one of the network structure that one might consider. In friendship network, mutuality might be defined as when one person claimed to be friend with someone, the opposite party will reciprocate and claim the same. It can also be defined when one declare to be an enemy, the other party also declare the referral to be an enemy. Surely, the tendency to be mutual differs on different network. Hence being mutual in binary network can be defined as $y_{ij} = y_{ji} = 1$ or sometimes $y_{ij} = y_{ji}$.

Addition of mutuality will make the ERGM Count model difficult to interpret. Krivitsky suggest that we can understand the model through conditional distribution: $Pr(Y_{ij} = y_{ij} | \mathbf{Y} \in \mathcal{Y}_{ij}(y))$ where $\mathcal{Y}_{ij}(y)$ is a set where only edge y_{ij} can change it's value where others assumed to be constant. This means that we will know which value $y_{ij} \in \mathbb{N}_0$ will random variable Y_{ij} likely to attain, given one particular network values.

However, we will proof that addition of mutuality will reduce to model with independent pair-wise distribution. Where a pair represents 2 edges where the origin and destination nodes are opposite of each other. The proposed statistics to represent mutuality are:

- **Minimum**

One of the proposed statistics to represent mutuality is minimum value or **mutual(min)**:

$$g(\mathbf{y}) = \sum_{(i,j) \in \mathbb{Y}, i < j} \min(y_{ij}, y_{ji}) \quad (1.10)$$

This is simply taking minimum values between 2 opposite edges of a dyad, and sum it for every possible pair. When a model only use the **sum** and **mutual(min)**, the PMF:

$$\begin{aligned} Pr(\mathbf{Y} = \mathbf{y} | \theta_1, \theta_2) &\propto \frac{\exp\left(\theta_1 \sum_{(i,j) \in \mathbb{Y}} y_{ij} + \theta_2 \sum_{(i,j) \in \mathbb{Y}, i < j} \min(y_{ij}, y_{ji})\right)}{\prod_{(i,j) \in \mathbb{Y}} y_{ij}!} \\ &= \frac{\exp\left(\theta_1 \sum_{(i,j) \in \mathbb{Y}, i < j} (y_{ij} + y_{ji}) + \theta_2 \sum_{(i,j) \in \mathbb{Y}, i < j} \min(y_{ij}, y_{ji})\right)}{\prod_{(i,j) \in \mathbb{Y}, i < j} y_{ij}! y_{ji}!} \\ &= \prod_{(i,j) \in \mathbb{Y}, i < j} \frac{\exp(\theta_1 (y_{ij} + y_{ji}) + \theta_2 \min(y_{ij}, y_{ji}))}{y_{ij}! y_{ji}!} \end{aligned}$$

Note that this cant be reduced into edge-wise distribution as per previous sections. Yet it has indendent pair-wise distribution where $\forall (i, j) \in \mathbb{Y}, i < j$, the joint distribution of (Y_{ij}, Y_{ji}) has PMF:

$$P(Y_{ij} = y_{ij}, Y_{ji} = y_{ji} | \theta_1, \theta_2) \propto \frac{\exp(\theta_1 (y_{ij} + y_{ji}) + \theta_2 \min(y_{ij}, y_{ji}))}{y_{ij}! y_{ji}!}$$

- **Negative Absolute Difference**

Another proposed statistics is negative absolute difference or **mutual(nabsdiff)** which has

the form:

$$g(\mathbf{y}) = \sum_{(i,j) \in \mathbb{Y}, i < j} -|y_{ij} - y_{ji}| \quad (1.11)$$

Note that the statistics g 's maximum value is 0. It is achieved when $\forall (i, j) \in \mathbb{Y}, i < j (y_{ij} = y_{ji})$ or every pair is equal. The negation of the statistics is necessary as the higher the statistics, the higher strength of mutuality. With the same method as above, the pairwise distribution (Y_{ij}, Y_{ji}) has PMF:

$$Pr(Y_{ij} = y_{ij}, Y_{ji} = y_{ji} | \theta_1, \theta_2) \propto \frac{\exp(\theta_1(y_{ij} + y_{ji}) - \theta_2|y_{ij} - y_{ji}|)}{y_{ij}!y_{ji}!}$$

- *Geometric Mean*

The last proposed statistics is geometric mean or ***mutual(geomean)*** which has the form:

$$g(\mathbf{y}) = \sum_{(i,j) \in \mathbb{Y}, i < j} \sqrt{y_{ij}}\sqrt{y_{ji}} \quad (1.12)$$

Similarly the model with statistics ***sum*** and ***mutual(geomean)*** will have independent pairwise ditribution (Y_{ij}, Y_{ji}) with PMF:

$$Pr(Y_{ij} = y_{ij}, Y_{ji} = y_{ji} | \theta_1, \theta_2) \propto \frac{\exp(\theta_1(y_{ij} + y_{ji}) + \theta_2\sqrt{y_{ij}}\sqrt{y_{ji}})}{y_{ij}!y_{ji}!}$$

Before we discuss the differences of each parameter, it is best to look at the shapes of the joint distribution.

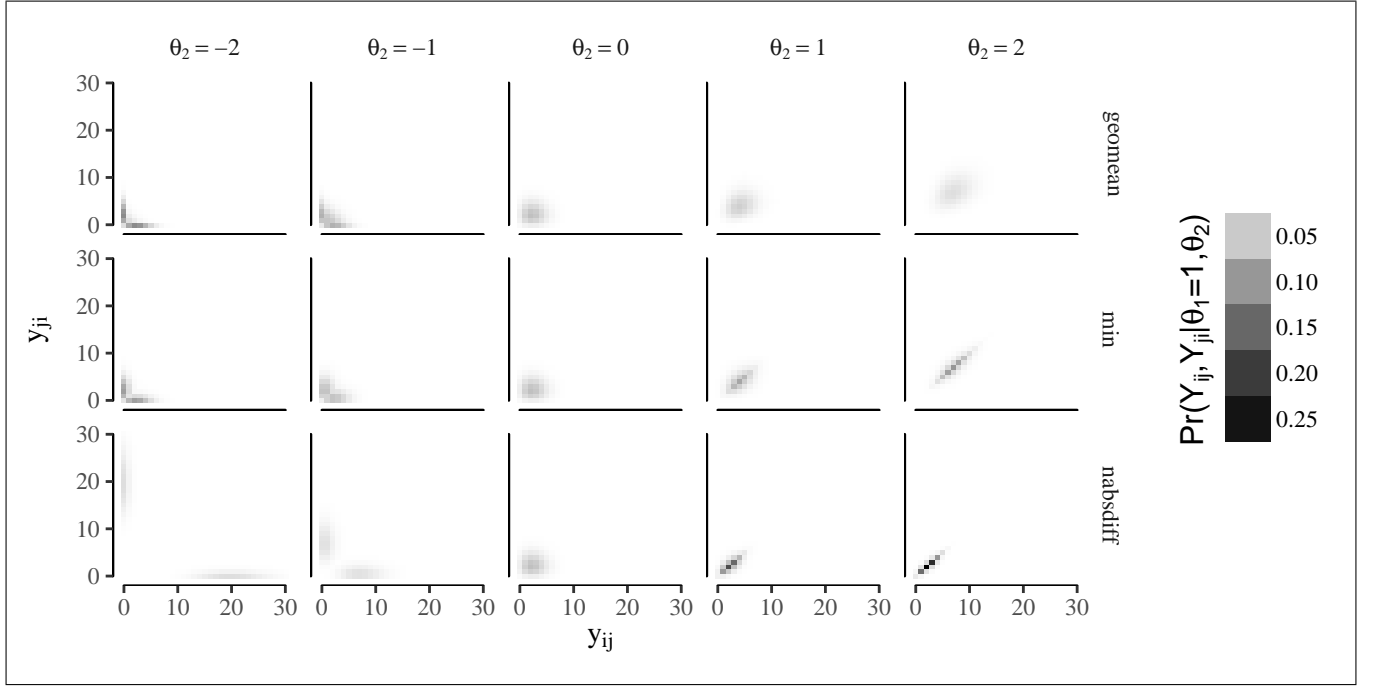


Figure 1.4: This image shows the various PMF of the joint distribution with $\theta_1 = 1$. The different shades of black indicate the region that has higher probability. Here the top axis represent θ_2 which coressponds to the parameter of *mutual* statistics. When $\theta_2 = 0$, the joint distribution follows 2 independent Poisson with $\lambda = e^{\theta_1}$

The differences of the joint distributions lies on the dependency statistics $\min(y_{ij}, y_{ji})$, $-|y_{ij} - y_{ji}|$, and $\sqrt{y_{ij}}\sqrt{y_{ji}}$ for *min*, *nabsdiff* and *geommean* respectively. When $\theta_2 > 0$ for *min* and *geommean*, the probability will increase as y_{ij} and y_{ji} increase. This shifted the distribution to higher pair values. For *nabsdiff* the level of y_{ij} and y_{ji} does not affect the reduction of probability. However, the reduction occurs less when disparity between y_{ij} and y_{ji} is low and 0 reduction at $y_{ij} = y_{ji}$. Hence the distribution is "pulled-in" to the mutual values. The same occurs in *min*, such that the highest increase occurs when $y_{ij} = y_{ji}$. Hence the distribution is "pulled-up" to mutual values.

When $\theta_2 < 0$, *min* and *geommean* will experience higher reduction at higher values of y_{ij} and y_{ji} . The reduction stop as y_{ij} or y_{ji} equals to zero. Hence, it creates bimodal distribution that has low values with opposing edge has zero value. The *min* shows less abrupt clouds as the reduction varies. The effect differs significantly from *nabsdiff*. There is increase in probability especially when the disparity is high, which creates the bimodal distribution. Since now the level of y_{ij} or y_{ji} affect the increase, the distribution cloud clutter on the higher values with opposinge edge is close to 0.

1.3.5 Transitivity

not done

Bibliography

- Krivitsky, P. N. (2012). Exponential-family random graph models for valued networks. *Electron. J. Statist.*, 6:1100–1128.
- Krivitsky, P. N. and Butts, C. T. (2016). Modeling valued networks with statnet.