# Central Limit Theorems and Statistical Inference for some Random Graph Models

Hanan Baaqeel

Thesis submitted to The University of Nottingham
for the degree of Doctor in Statistics and Probability

June 2015

# Abstract

Random graphs and networks are of great importance in many fields including mathematics, computer science, statistics, biology and sociology. This research aims to develop statistical theory and methods of statistical inference for random graphs in novel directions. A major strand of the research is the development of conditional goodness-of-fit tests for random graph models and for random block graph models. On the theoretical side, this entails proving a new conditional central limit theorem for a certain graph statistics, which are closely related to the number of two-stars and the number of triangles, and where the conditioning is on the number of edges in the graph. A second strand of the research is to develop composite likelihood methods for estimation of the parameters in exponential random graph models. Composite likelihood methods based on edge data have previously been widely used. A novel contribution of the thesis is the development of composite likelihood methods based on more complicated data structures. The goals of this PhD thesis also include testing the numerical performance of the novel methods in extensive simulation studies and through applications to real graphical data sets.

# Acknowledgements

While a completed thesis bears the single name of the student, the process that leads to its completion is always accomplished in combination with the dedicated work of other people. I wish to acknowledge my appreciation to certain people.

First of the all, I thank my God, for letting me through all the difficulties and allowing me to achieve my goal. You are the one who let me finish my degree.

I would like to express my special appreciation and thanks to my advisor Professor Andy Wood, you have been a tremendous mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist. Your advice on research has been invaluable. I am also grateful to Professor Philip O'Neill for his comments and his advice on the thesis. I would also like to thank two other peoples in the school of Mathematical Sciences, Anthony Hennessey and Dr. David Hodge, for their help at various times.

I would especially like to thank my my mother, Feryal Nouh, and my father, Mohamed Baaqeel, for all of the sacrifices that you have made on my behalf. Your prayers for me were what sustained me thus far.

I would also like to thank my beloved husband, Captain Ahmed Seit, for supporting me for everything, and especially I can not thank you enough for encouraging me throughout this experience. I could never have accomplished this thesis without his support and understanding.

I also wish to thank my wonderful daughters Ragad, Sara and Yara, also my gorgeous son Mohamed. I would like to express my thanks to them for doing their best to understand a mother who had to be confined to her study for such a long time.

I must express my gratitude towards my mother-in law, brothers, sisters, friends and all people who instilled in me confidence and the drive for pursuing my PhD.

Last but certainly not least, thanks to the government of Saudi Arabia, and King Abdulaziz University. I could not have gone through the doctoral program overseas without their financial support.

# Contents

Chapter 1

# Introduction

## 1.1 Background: Networks and Graphs

Networks are of great importance in the modern world, the internet and facebook being two prominent examples. Moreover, the understanding and modelling of networks is of major importance in many fields within the physical and social sciences including Biology, Computer Science, Economics, Geography and Sociology. The recent book by Kolaczyk (2010) gives an extensive account covering theory and many applications from a statistical perspective. In recent times this has been an active topic of research for statisticians; see, for example, the papers by Bickel et al. (2011), Chatterjee and Diaconis (2013), Caimo and Friel (2013) and Olhede and Wolfe (2014).

It is difficult to give a precise definition of a network which covers all cases of possible interest, but broadly speaking a network consists of collection of *units*, e.g. genes or people, plus information about the connections between them. In many applications of network modelling, randomness (or stochasticity) is a key feature of the network under consideration. In other situations, we may choose to model uncertainty using randomness, without necessarily forming a judgment as to whether the randomness is inherent in the network being studied. Ether way, we are led to consideration of random networks.

From a mathematical perspective, a network is often modelled as a graph. Graph theory is a well-established field of mathematics which is reviewed briefly in Chapter 2. In short, a graph consists of two aspects: a set of vertices (or nodes); and a set of edges which connect pairs of vertices. It would be an over-simplification to say that the study of networks is equivalent to the study of graphs because networks of interest in real-world applications often have further features or structure in addition to the structure represented by a graph. Nevertheless, the graphical structure of a network is often of primary interest and is an important object of study in its own right.

Just as it is important to consider random networks in many areas of application, so it is important to consider random graphs. In fact, the study of random graphs has been an active area of research within probability theory for over half a century, starting with work of Erdős and Rényi (1959, 1960, 1961) and Gilbert (1959). For up-to-date accounts of the theory of random graphs see the books by Bollobás (2001) and Durrett (2010); see also Kolaczyk (2010) for an accessible review of this large body of work.

Despite the progress in the theory of random graphs, it is fair to say that statistical theory for the analysis of random graph models is still rather under-developed, even though there have been some notable contributions in the past. No doubt this is partly due to the difficulty in developing an asymptotic theory, as the number of vertices of the random graph goes to infinity, for parameter estimators in most of the random graph models which have been considered to date. This difficulty arises because of the complex nature of the dependence structure in such models. However, even allowing for this, there are still a number of basic statistical questions which have not been addressed in the literature. The purpose of this thesis is to consider several such questions.

## 1.2 Main Contributions of the Thesis

The first question to be considered is the joint asymptotic behavior of the random graph statistics $u_1$, the number of edges, $u_2$, the number of 2-stars, and $u_3$, the number of triangles in the Erdös-Rényi-Gilbert random graph model in which each edge is present with probability $p$ and the edges are statistically independent. In this asymptotic framework, the number of vertices in the graph goes to infinity. It is proved that, suitably standardized, $u_1$, $u_2$ and $u_3$ are jointly asymptotic normal. This finding is not surprising but what does seem surprising is that the limiting covariance matrix has rank 1 rather than rank 3. Consequently, the limiting covariance matrix is degenerate. We suspect that this result may be known but, despite an extensive search, we have not been able to find it anywhere in the literature. The rank deficiency of the covariance matrix is a negative result from the point of view of statistical inference.

The second question to be considered is whether this degeneracy can be removed by conditioning $u_2$ and $u_3$ on $u_1$, the number of edges. It is proved that a conditional central limit theorem holds in this case. Moreover, the limiting covariance matrix has full rank 2. Thus, conditioning on $u_1$ removes the degeneracy. However, it turns that it is a major task to give a fully rigorous proof of this conditional central limit theorem. This proof is the most substantial component of the thesis.

The primary statistical motivation for considering these central limit theorems is to see whether they provide the basis for goodness-of-fit tests. Due to the degeneracy in the unconditional central limit theorem mentioned above, the unconditional approach is not useful from the point of view of goodness-of-fit tests. However, the conditional central limit theorem does lead to a potentially useful conditional goodness-of-fit tests, especially in the context of block graph models, considered later in the thesis.

The third question considered in the thesis is the use of novel composite likelihoods for parameter estimation in a widely-studied 3-parameter Exponential Random Graph Model *(ERGM)*. *ERGMs* and composite likelihood methods are reviewed briefly in Chapter 2. Theoretical asymptotic analysis of these new estimators does not seem possible using existing large-sample theory but their practical performance is investigated in a simulation study.

## 1.3 Structure of the Thesis

The outline of the thesis is as follows. Chapter 2 contains review material on random graphs, relevant results from probability and statistics and other miscellaneous mathematical results. Most of the material is standard and is included for convenience. However, the final section of the chapter reviews some relevant publications of more advance work on the probability and statistics of random graphs.

Chapter 3 contains a statement and proof of the joint central limit theorem for $u_1$, $u_2$ and $u_3$. The chapter also includes the statements and proofs of some elementary counting lemmas which are useful for calculating second moments of $u_2$, $u_3$. These lemmas are also used in Chapter 4.

The proof of the conditional central limit theorem, which as mentioned above is quite challenging, is spread over Chapter 4 and Chapter 5. In Chapter 4 a general conditional moment result is stated and proved. More specifically, the general conditional moment result gives a precise estimate of the order of the expectation of arbitrary products of centered identically distributed binary variables conditional on their total being fixed. This result, which we believe may be of independent interest, plays a crucial role in the proof of conditional central limit theorem which stated and proved in Chapter 5.

Chapter 5 contains some fairly complex counting approximation lemmas which also play a vital part in the proof of the conditional central limit theorem. This chapter also contains the statement and proof of the conditional central limit theorem.

In Chapter 6, three new composite likelihood estimators are suggested for use in a 3-parameter *ERGM* of interest. In Section 6.2 the composite likelihoods are derived and computational algorithms are presented for their calculation, and in Section 6.3, the results of a simulation study of these estimators is presented.

Finally, in Chapter 7, discussion, conclusions and possibilities for future research are described.

Chapter 2

# Review of Background and relevant Techniques

## 2.1 Introduction

In this chapter we present technical background which is relevant to later chapters in the thesis. In Section 2.2, random graph models are reviewed. Section 2.3 covers miscellaneous mathematical topics including the spectral decomposition theorem for symmetric matrices, equivalence relations and partitions. In Section 2.4, some important topics in probability and statistics are covered, including the projection method and the method of moments for proving central limit theorems, both of which are important later in the thesis. A review of composite likelihood methods is given in Section 2.5. Finally, a review of some more advanced literature on the statistics and probability of random graphs is given in Section 2.6.

## 2.2 Background on Random Graph Models

The term *network* refers to a collection of elements and their relations. For mathematical purposes, a *network* is represented as a *graph*, as defined in graph theory. Graph theory is a branch of mathematics which adds precision to this notion and provides a body of definitions, tools and results for examining graphs and their properties.

### 2.2.1 Basic Terminology in Graph Theory

A *graph* $\mathcal{G}(V, E)$ is a mathematical structure which consists of two sets, $V$ and $E$. $V$ is a nonempty finite set, whose elements are called *vertices* (or *nodes*). The set $E$ is a subset of $V \times V$, the Cartesian product of $V$ with itself, so that the elements of $E$ are pairs of vertices. We think of the elements of $E$ as edges. If $e = (u, v) \in E$ where $u, v \in V$, then we say that vertices $u$ and $v$ are adjacent. If the ordering of $u$ and $v$ does not matter, i.e. if $(u, v)$ is identified with $(v, u)$, then the edges are said to be *undirected*. If, on the other hand, the ordering of $u$ and $v$ in $(u, v)$ does matter, the graph is said to have directed edges. A directed graph is abbreviated as *digraph* with the directions of the edges indicated by arrows. Figure 2.1 (b), illustrates the form of digraph.

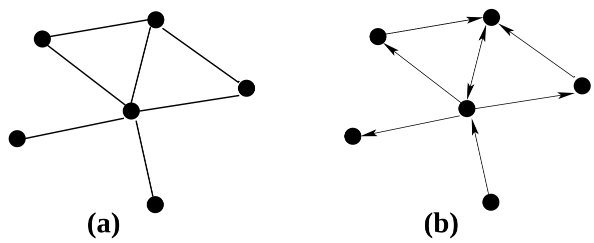


**Figure 2.1: (a)** A simple undirected graph. **(b)** A digraph

When an edge joins a vertex to itself, this is called *loop*, and if there is more than one edge connecting two (different) vertices that is called *multiple* (or *parallel*) edge. A graph without loops or parallel edges is called a *simple graph*. We will dealing with the simple undirected graphs throughout this thesis. Figure 2.1 (a) represents a simple undirected graph.

Throughout the thesis, the number of vertices, $|V|$, will denoted by $n$ and is called the *order* of the graph. The number of edges, $|E|$, is called the *size* of the graph and is denoted it by $m$.
The *degree* of a vertex is the number of edges with an end-point in that vertex. Also, a vertex $u \in V$ is *incident* on an edge $e \in E$ if $e = (u, v)$ or $e = (v, u)$.

A graph is said to be *connected* if and only if any vertex in the graph can reached from all other vertices in the graph by moving along edges, and the graph is said

to be *disconnected* otherwise. A disconnected graph splits into *components* where the vertices in each component are connected. Furthermore, a graph $\mathcal{H}(V_{\mathcal{H}}, E_{\mathcal{H}})$ is a *subgraph* of $\mathcal{G}(V_{\mathcal{G}}, E_{\mathcal{G}})$ if $V_{\mathcal{H}} \subseteq V_{\mathcal{G}}$ and $E_{\mathcal{H}} \subseteq E_{\mathcal{G}}$.
*Complete Graph:* In a simple graph, if each pair of vertices is connected, the graph is said to be a *complete graph*. A complete graph with $n$ vertices is denoted by $K_n$. Figure 2.2 represents the $K_4$ graph. Moreover, a complete subgraph is called a *clique*.

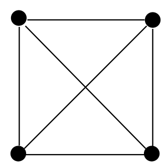

**Figure 2.2:** The complete graph $K_4$

*Bipartite Graph:* Consider a graph $\mathcal{G}$, and suppose $V$ can be partitioned into two disjoint sets, $A$ and $B$, in such a way that each edge in $\mathcal{G}$ links two vertices, with one vertex from $A$ and one from $B$. Then $\mathcal{G}$ is a *bipartite graph*, and in the sociology literature it is called a *two-mode network*; see Newman (2010). Figure 2.3 represents a bipartite graph:

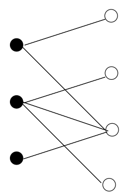

**Figure 2.3:** A bipartite graph

Bipartite graphs arise in Section 5.6 when we consider block models. Moreover, when each vertex in set $A$ is connected to all vertices in set $B$ where $A$ has $n_A$ vertices and $B$ has $n_B$ vertices, this graph is called a *complete bipartite graph* and denoted by $K_{n_A,n_B}$. Such a graph is represented in Figure 2.4.

In graph theory, there are several ways to represent a graph mathematically. A convenient representation of a graph for our purposes is the *adjacency matrix*. Consider an undirected graph with $n$ vertices with each vertex having a label $1, \cdots, n$. Consider $y_{ij}$ is a binary variable representing the presence or absence

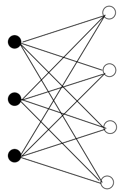

**Figure 2.4:** A complete bipartite graph $K_{3,4}$

of an edge between the vertices $i$ and $j$. Formally, for a simple undirected graph of $n$ nodes, consider the graph adjacency matrix $y = \{y_{ij}\}_{1\leq i<j\leq n}$, where $y_{ij}$ is a binary indicator for edge $\{i, j\}$:

$$y_{ij} = \begin{cases} 1 & \text{if there is an edge between } i \text{ and } j; \\ 0 & \text{otherwise.} \end{cases} \tag{2.2.1}$$

For a simple undirected graph, we may define the adjacency matrix to be symmetric, with $y_{ij} = y_{ji}$, for all $i \neq j$. Also since the graph is simple (i.e. no loops), we define $y_{ii} = 0$ for all $i$.
This thesis is concerned with *random* graph models. We briefly highlight some of the most common random graph models which are widely studied in the literature.

### 2.2.2 Random Graph Models

We now briefly describe some important random graph models.
**The Erdös-Rényi-Gilbert Models.** In a series of papers by Erdős and Rényi (1959, 1960, 1961) and Gilbert (1959), random graph models are introduced. The term *random graph* is used in this sense to refer to a model specifying a finite collection $\mathbb{G}$ of graphs and a uniform probability $\mathbb{P}(.)$ over $\mathbb{G}$. Precisely, for a given $n$ and $m$, the number of vertices and edges, or the order and the size of the graph, respectively, there is a collection $\mathbb{G}$ of all graphs $\mathcal{G}(V, E)$ with assign probability $\mathbb{P}(\mathcal{G}) = \binom{N}{m}^{-1}$, for each $\mathcal{G} \in \mathbb{G}$, where $N = \binom{n}{2}$ is the total number of distinct pairs of vertices. In other words, in this model we choose uniformly $m$ distinct pairs of vertices at random from all possible pairs and link them by an edge. To obtain a simple undirected graph, we restrict the vertices to be distinct to avoid loops and multiple edges. Therefore, the graph is created by choosing uniformly at random among the set of all simple graphs with exactly $n$ vertices

and $m$ edges.
Most commonly studied is the closely-related random graph proposed by Gilbert (1959), denoted $RG(n, p)$, in which every possible edge occurs independently with probability $0 < p < 1$. In the $RG(n, p)$ model, we fix the number of vertices and the probability of presence an edge between distinct pairs vertices, but the number of edges is not fixed. Then the definition of this random graph model, $RG(n, p)$, is the ensemble of graphs in which each simple undirected graph $\mathcal{G}$ appears with $n$ vertices and $m$ edges with probability

$$P(\mathcal{G}) = p^m(1-p)^{N-m},$$

where $N = n(n-1)/2$ is the number of possible edges. We focus here on homogeneous random graphs. However, there has been a large amount of attention in defining and studying inhomogeneous random graphs, such as Bollobás et al. (2007).

**Exponential Random Graph Models.** Exponential random graph models are a family of probability distributions for a class of random graphs which can be used for representing and analyzing data about social and other networks. See, for example Kolaczyk (2010) for further details. There are many techniques that measure properties of an observed graph which are useful for describing and understanding the observed graph. However, for a particular number of vertices, this observed graph represents one realization of a large number of possible graphs, as the outcome of some stochastic mechanism. Therefore, the principal goal is to estimate model parameter from data and then evaluate how well the model represents the data. In other words, the observed network is seen as one particular pattern of edges out of a large set of possible patterns. In general, we do not know what stochastic mechanism generated the observed network, and the goal in formulating a model is to suggest a reasonable and theoretically principled hypothesis for this process; see (Robins, Pattison, Kalish and Lusher (2007)). Fortunately, exponential families have a variety of common properties which makes this class of distributions mathematically convenient for purposes of inference and simulation.
Suppose $\mathcal{G}(V, E)$ is a random graph, and let $y = \{y_{ij}\}_{1\leq i<j\leq n}$ be the (random) adjacency matrix for $\mathcal{G}$. As before, $y_{ij} = y_{ji}$ is a binary random variable to indicate the present or absence of an edge $\{i, j\} \in E$, in the graph $\mathcal{G}$, where $i, j \in V$.

An exponential random graph model (*ERGM*) gives the joint distribution of the elements in $y$, and has the following general form:

$$P_\theta\{Y = y\} = \exp\left\{\sum_{\alpha=1}^{A} \theta_\alpha \sum_{\beta=1}^{B_\alpha} u_{\alpha\beta}(y) - \psi(\theta)\right\}, \tag{2.2.2}$$

where

$$u_{\alpha\beta}(y) = \prod_{\{i,j\}\in H_{\alpha\beta}} y_{ij}; \tag{2.2.3}$$

for each $\alpha = 1, \cdots, A$ and $\beta = 1, \cdots, B_\alpha$, $H_{\alpha\beta}$ consists of one or more pairs of vertices; $\theta = (\theta_1, \cdots, \theta_A)^T$ is the parameter vector; and the normalising constant $\psi(\theta)$ is given by

$$\psi(\theta) = \log\left(\sum_{y\in\{0,1\}^N} \exp\left\{\sum_{\alpha=1}^{A} \theta_\alpha \sum_{\beta=1}^{\beta_\alpha} u_{\alpha\beta}(y)\right\}\right), \tag{2.2.4}$$

where the outer sum in (2.2.4) is over all possible adjacency matrices, and $N = n(n-1)/2$. Different choices for $A, B_1, \cdots, B_A$ are considered in the models (2.2.5)-(2.2.7) and (2.2.9) below. The representation of (2.2.2) is equivalent to formula (6.24) in Kolaczyk (2010) but we find (2.2.2) more convenient because it makes the role of the inner summation (over $\beta$) in the expanent of (2.2.2) more explicit. The model (2.2.2) is an exponential family distribution with natural vector parameter $\theta$ and sufficient statistics $u_{\alpha\beta}(y)$, $\alpha = 1, \cdots, A$ and $\beta = 1, \cdots, B_\alpha$. Exponential families of random graphs are among the most extensively used. They represent flexible models for complex networks, particularly social networks. Exponential random graph models (*ERGMs*), also called $P^*$ models, are a family of probability distributions for a class of random graphs. *ERGMs* are used, for example, to represent structural of social network observed; see Snijders et al. (2006).

We now consider several examples of *ERGMs* which have been considered in the literature.

*Bernoulli random graphs.* In this model the presence or absence of any edge is independent of the presence or absence of all the other edges in a graph. So we assume that $y_{ij}$ is independent of $y_{i^*j^*}$, where $\{i,j\} \neq \{i^*, j^*\}$. This leads to the model

$$P_\theta\{Y = y\} = \exp\left\{ \sum_{1 \leqslant i < j \leqslant n} \theta_{ij} y_{ij} - \psi(\theta) \right\}. \tag{2.2.5}$$

In other words, every edge $\{i, j\}$ is present in the graph independently with probability $p_{ij} = \exp\{\theta_{ij}\}/(1 + exp\{\theta_{ij}\})$.

Model (2.2.5) is obtained from model (2.2.2) by putting $B_1 = \cdots = B_A = 1$, the number of elements in each $H_{\alpha\beta}$ is 1 and $A = N = n(n-1)/2$. Under the assumption of homogeneity, i.e. $\theta_{ij} = \theta$, Gilbert (1959) model is recovered, i.e.

$$P_\theta\{Y = y\} = \exp\left\{ \theta L(y) - \psi(\theta) \right\}, \tag{2.2.6}$$

where $L(y) = \sum_{i,j} y_{ij} = m$ is the number of edges in the graph, and the probability of an edge being present will be $p = \exp\{\theta\}/(1 + exp\{\theta\})$. In this case, $A = 1$ and $B_1 = n(n-1)/2$.

*Block models.* Block models can also be represented in terms of model (2.2.2). For example, when the vertex set $V$ splits into two sets $V_1$ and $V_2$, and there is homogeneity both within and between sets, this leads to a model of the form

$$P_\theta\{Y = y\} = \exp\left\{ \theta_{11} L_{11}(y) + \theta_{12} L_{12}(y) + \theta_{22} L_{22}(y) - \psi(\theta) \right\}, \tag{2.2.7}$$

where $L_{11}(y)$ and $L_{22}(y)$ are the number of edges connecting two elements of $V_1$ and connecting two elements of $V_2$, respectively, and $L_{12}$ is the number of edges connecting an element of $V_1$ withe an element of $V_2$. Clearly, (2.2.7) is of the form (2.2.2) with $A = 3$, $B_1 = n_1(n_1 - 1)/2$, $B_2 = n_2(n_2 - 1)/2$ and $B_3 = n_1 n_2$. We return to block models in Section 5.6.

*Markov random graphs.* As before, let $\mathcal{G}$ denote a random graph. To define the *Markov property*, let $i, j, k, l$ denote four distinct vertices. Then $\mathcal{G}$ satisfies the *Markov property* if

$$P(y_{ij}, y_{kl}|rest) = P(y_{ij}|rest)P(y_{kl}|rest), \tag{2.2.8}$$

where '*rest*' here consists of all the elements of $y$ apart from $y_{ij}$ and $y_{kl}$. In words, (2.2.8) says that $y_{ij}$ and $y_{kl}$ are conditionally independent given the rest of the elements of $y$.

Frank and Strauss (1986) characterize Markov graphs using the assumption that the distribution remains the same when the vertices are relabeled. They use the Hammersley-Clifford theorem (Besag, 1974) to prove a random graph is a Markov graph iff the probability distribution can be written as (2.2.9).

The concept of *Markov dependence* for a random graph model was introduced by Frank and Strauss (1986). This model defines that two potential edges are dependent whenever they share a vertex, conditional on all other potential edges. That is, the presence or absence of $\{i, j\}$ in the graph will depend on that of $\{i, k\}, \{j, l\}$, for all $k \neq i, j$ and $l \neq i, j$, such random graph called *Markov graph* and given by.

$$P_\theta\{Y = y\} = \exp\left\{\sum_{k=1}^{n-1} \theta_k S_k(y) + \theta_n T(y) - \psi(\theta)\right\}, \qquad (2.2.9)$$

where y is the adjacency matrix for a random graph, and $S_1(y) = m$ is the number of edges, $S_k$ is the number of k-stars, for $2 \leqslant k \leqslant (n-1)$, and $T(y)$ is the number of triangles. The statistics $S_k$ and $T$ are defined by

$$\begin{aligned} S_1(y) &= \sum_{1\le i<j\le n} y_{ij} && \text{number of edges,} \\ S_k(y) &= \sum_{1\le i\le n} \binom{y_{i+}}{k} && \text{number of } k\text{-stars } (k \ge 2), \\ T(y) &= \sum_{1\le i<j<h\le n} y_{ij}y_{ih}y_{jh} && \text{number of triangles,} \end{aligned}$$

where $y_{i+} = \sum_{j=1}^{n} y_{ij}$, *the degree of node i*, $\theta = (\theta_1, ..., \theta_n)^T$ is the parameter vector of the distribution, and $\psi(\theta)$ is a normalizing constant,

$$\psi(\theta) = log\left(\sum_{y\in\{0,1\}^{n(n-1)/2}} \exp\left\{\sum_{k=1}^{n-1} \theta_k S_k(y) + \theta_n T(y)\right\}\right)$$

which ensures that the sum of probabilities equals 1. It is obvious, when $k = 1$,

$S_1(y)$ is a 1-star representing the number of edges. Clearly (2.2.9) is of the form given in model (2.2.2) with $A = n$ and

$$\begin{aligned} B_1 &= \frac{n(n-1)}{2} \\ B_2 &= \frac{n(n-1)(n-2)}{2!} \\ B_j &= \frac{n(n-1)\cdots(n-j)}{j!}, 1 \leqslant j \leqslant n-1 \\ B_{n-1} &= n \\ B_n &= \frac{n(n-1)(n-2)}{3!}. \end{aligned}$$

The model will be more tractable when $\theta_2 = \ldots, \theta_{n-1} = \tau = 0$, in which case this distribution reduces to the Bernoulli graph model, where all edges occur with the same probability $e^{\theta_1}/(1+e^{\theta_1})$ independently. Frank and Strauss (1986) observed that parameter estimation is difficult when the probability distribution depends on the number of edges, number of 2-stars, and the number of triangles ($\theta_3 = \ldots = \theta_{n-1} = 0$). This is known as *triad model* and has the form

$$P_\theta\{Y = y\} = \exp\{\theta_1 u_1(y) + \theta_2 u_2(y) + \theta_3 u_3(y) - \psi(\theta)\}, \tag{2.2.10}$$

where the parameter vector to be estimated is $\theta = (\theta_1, \theta_2, \theta_3)^T$, and the sufficient statistics $u(y) = (u_1(y), u_2(y), u_3(y))$, considered important, is defined by

$$\begin{aligned} u_1(y) &= \sum_{1 \leq i < j \leq n} y_{ij} && \text{number of edges} \\ u_2(y) &= \sum_{1 \leq i < j \leq n} \sum_{k \neq i,j} y_{ik} y_{jk} && \text{number of 2-stars} \\ u_3(y) &= \sum_{1 \leq i < j < k \leq n} y_{ij} y_{ik} y_{jk} && \text{number of triangles,} \end{aligned}$$

and $\psi(\theta)$ is a normalizing constant. Thus, they assumed models of any one of the three parameters given that the other two are fixed at 0, and elaborated a simulation-based method to approximate the maximum likelihood estimation. Moreover, they suggested a type of conditional logistic regression method to estimate all parameters.

*Vertex degree models.* Another type of *ERGM* is based on vertex degrees. This models have the form

$$P_\theta\{Y = y\} = \exp\left\{\sum_{i=1}^{n} \theta_i y_{i+} - \psi(\theta)\right\}. \tag{2.2.11}$$

where $y_{i+} = \sum_{i \neq j} y_{ij}$, and $\theta = (\theta_1, \cdots, \theta_n)^T$ is the parameter vector and as before $\psi(\theta)$ is chosen to ensure the probabilities sum to 1.

## 2.3 Relevant Mathematical Techniques and Results

This section contains some miscellaneous mathematical concepts and results used later in the thesis.

### 2.3.1 Spectral Decomposition Theorem in Linear Algebra

A set of vectors $z_1, z_2, ..., z_r \in \mathbb{R}^m$ is said to be *linearly independent* if there exists no set of scalars $c_1, c_2, ..., c_r$, not all zero, such that $\sum_{i=1}^{r} c_i z_i = 0_m$, the $m$-vector of zeros. A $m \times n$ matrix $A$ is said to be of *rank* $r$ if the maximum number of linearly independent columns is $r$. Suppose now that $A$ $(n \times n)$ is a square matrix. Consider the quadratic form

$$x^T A x = \sum_{i,j=1}^{n} a_{ij} x_i x_j, \tag{2.3.1}$$

where $x = (x_1, ..., x_n)^T$ and $A = (a_{ij})$ is a symmetric matrix, i.e. $A^T = A$ where $T$ denotes transpose. This matrix $A$ and the quadratic form are called *positive semidefinite* if $x^T A x \geqslant 0$ for all $x \in \mathbb{R}^n$. If $x^T A x > 0$ for all $x \neq 0_n$, the $n \times 1$ vector of zeros, then $A$ and the quadratic form are called *positive definite*. Let $A$ be a square, $n \times n$ symmetric matrix. A real scalar $\lambda$ is said to be an eigenvalue of $A$ if there exist a non-zero vector $x$ in $\mathbb{R}^n$ such that

$$Ax = \lambda x \tag{2.3.2}$$

The vector $x$ is then referred to as an eigenvector associated with the eigenvalue $\lambda$. The eigenvalues of the matrix A are solution of the characteristic equation

$$det(\lambda I - A) = 0, \tag{2.3.3}$$

where the notation *det* refers to the determinant of a matrix. An important result of linear algebra, called the spectral decomposition theorem, states that for any symmetric matrix, there are exactly $n$ eigenvalues, and they are all real; further, that the associated eigenvectors can be chosen so as to form an orthonormal basis. The result offers a simple way to decompose the symmetric matrix as a product of simple transformations.

**Theorem 2.1 (see e.g. Mardia et al. (1980))**
*We can decompose any real symmetric $n \times n$ matrix $A$ with the spectral decomposition*

$$A = \sum_{i=1}^{n} \lambda_i q_i q_i^T = Q\Lambda Q^T, \quad \Lambda = \; diag(\lambda_1, \cdots, \lambda_n), \tag{2.3.4}$$

*where the $q_i$ are $n \times 1$ column vectors and the matrix $Q := [q_1, ..., q_n]$ is orthonormal (that is, $Q^T Q = Q\ Q^T = I_n$), where $I_n$ is the $n \times n$ identity matrix, and contains the unit eigenvectors of $A$, while the diagonal matrix $\Lambda$ contains the eigenvalues of $A$.*

### 2.3.2 Equivalence Relation and Partitions

An *equivalence relation* on a set $X$ is a relation $\sim$ on X such that the following properties are hold:

1. $x \sim x$ for all $x \in X$ . (The relation is reflexive.)
2. If $x \sim y$ , then $y \sim x$ . (The relation is symmetric.)
3. If $x \sim y$ and $y \sim z$ , then $x \sim z$ . (The relation is transitive.)

A *partition* of a set $X$ is a set $P = \{A_1, \cdots, A_n\}$ of blocks that are subsets of $X$ such that the following holds.

1. If $A_j \in P$ then $A_j \neq \emptyset$ for $j = 1, \cdots, n$, where $\emptyset$ is the empty set.
2. $\bigcup_{j=1}^{n} A_j = X$.

3. $A_j \bigcap A_k = \emptyset$ if $j \neq k$.

If $\sim$ is an equivalence relation on $X$, we define the equivalence class of $a \in X$ to be the set $[a] = \{b \in X | a \sim b\}$

**Result 1:** $[a] = [b]$ if and only if $a \sim b$.

**Result 2:** The set of all equivalence classes form a partition of $X$.

See, for example, Chapter 2 of Ayres (1965) for the above material.

A set $X$ will said to be *partially ordered* (the possibility of a total ordering is not excluded) by a binary relation $\mathcal{R}$ if for arbitrary $a, b, c \in X$,

1. $\mathcal{R}$ is reflexive, i.e. $a\mathcal{R}a$
2. $\mathcal{R}$ is anti-symmetric, i.e., $a\mathcal{R}b$ and $b\mathcal{R}a$ if and only if $a = b$
3. $\mathcal{R}$ is transitive, i.r., $a\mathcal{R}b$ and $b\mathcal{R}c$ implies $a\mathcal{R}c$.
   For more details see e.g Ayres (1965).

The type of partial ordering that is relevant in this thesis is that between the partitions of a fixed finite set $A$. Suppose we are given two partitions of $A$,

$$\Upsilon^{(1)} = \{\upsilon_1^{(1)}, \cdots, \upsilon_\alpha^{(1)}\} \text{ and } \Upsilon^{(2)} = \{\upsilon_1^{(2)}, \cdots, \upsilon_\beta^{(2)}\}.$$

That is, for $j = 1, 2$, $\upsilon_\gamma^{(j)} \subseteq A$, $\upsilon_\gamma^{(j)} \bigcap \upsilon_\delta^{(j)} = \emptyset$, the empty set, if $\gamma \neq \delta$, and $\bigcup_{\gamma=1}^{\alpha} \upsilon_\gamma^{(1)} = \bigcup_{\delta=1}^{\beta} \upsilon_\delta^{(2)} = A$. Then we say that

$$\Upsilon^{(1)} \leqslant \Upsilon^{(2)}$$

if for all $\gamma = 1, \cdots, \alpha$, $\upsilon_\gamma^{(1)} \subseteq \upsilon_\delta^{(2)}$ for some $\delta = 1, \cdots, \beta$.

We shell see in Chapter 5 that the diamond partition is 'larger' than tilde partition.

## 2.4 Background in Probability and Statistics

Here, briefly, we will lay some foundation of terminology, notation, and concepts of topics in probability and statistics, to be used in later chapters. An excellent introduction to many of the probability and statistics concepts is the book by van der Vaart (2000). Mardia et al. (1980) gives excellent coverage of multivariate analysis and related topics.

### 2.4.1 Univariate and Multivariate Normal Distributions

The normal distribution is most important and widely used distribution in statistics, because of the central limit theorem; see subsection 2.4.3. It is also called the Gaussian distribution after the mathematician Karl Friedrich Gauss. A normal distribution for a random variable $X$ with mean $\mu$ and variance $\sigma^2$ has probability density function

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{2.4.1}$$

Let $X = (X_1, \cdots, X_k)^T$ be a random vector with mean vector $\mu$ $(k \times 1)$ and covariance matrix $\Sigma$ $(k \times k)$

$$\mu = \begin{pmatrix} \mu_1 \\ \dots \\ \mu_k \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \dots & \sigma_{1k} \\ \vdots & \ddots & \vdots \\ \sigma_{k1} & \dots & \sigma_{kk} \end{pmatrix}.$$

A random vector $X = (X_1, \cdots, X_k)^T$, where $X_i \in \mathbb{R}$, is said to have the multivariate normal distribution with mean $\mu$ $(k \times 1)$ and covariance $\Sigma$ $(k \times k)$ defined by $\mu = E[X]$ and $\Sigma = \text{Cov}(X) = E[(X - \mu)(X - \mu)^T]$, if every fixed linear combination of its components $Y = a_1 X_1 + \cdots + a_k X_k$ is normally distributed. That is, for any constant vector $a = (a_1, ..., a_k)^T \in \mathbb{R}^k$, the random variable $Y = a^T X$ has a univariate normal distribution $N(\xi, \sigma^2)$ with density (2.4.1), with $\xi = \mu^T a$ and $\sigma^2 = a^T \Sigma a$.

The multivariate normal distribution is said to be "non-degenerate" if and only if the covariance matrix $\Sigma$ is positive definite or, equivalently, if $\Sigma$ has the full rank $k$. In this case the distribution has density on $\mathbb{R}^k$ given by

$$f_X(x_1, \dots, x_k) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right)$$

Also, $|\Sigma|$ is the determinant of $\Sigma$. The statistic $(x - \mu)^T \Sigma^{-1} (x - \mu)$ in the non-degenerate case is known as the (square) Mahalanobis distance, which represents the squared distance of the test point $x$ from the mean $\mu$ in the metric determined by $\Sigma^{-1}$.

### 2.4.2 Different Types of Stochastic Convergence

In this section, we provide a review of basic types of convergence of sequences of random variables and vectors, and we will explain three of them.

**Convergence in distribution.** Let $Y_1, Y_2, \cdots$ be a sequence of real random vectors in $\mathbb{R}^k$, and let $Y$ denote a random vector in $\mathbb{R}^k$. For each $n = 1, 2, \cdots$ suppose $F_n$ denote the distribution function of $Y_n$; that is

$$F_n(y) = P(Y_{n1} \leqslant y_1, \cdots, Y_{nk} \leqslant y_k),$$

where $y = (y_1, \cdots, y_k)^T$ and $Y_n = (Y_{n1}, \cdots, y_{nk})^T$. Then the sequence $Y_n$ is said to converge in distribution to the random vector $Y$ as $n \to \infty$ if

$$\lim_{n\to\infty} F_n(y) = F(y),$$

for every $y \in \mathbb{R}^k$ at which $F$ is continuous, where $F(y)$ is the distribution function of $Y$. This type of convergence also called *weak convergence* or *convergence in law*, and written as $Y_n \xrightarrow{d} Y$.

**Convergence in probability.** A sequence $Y_n \in \mathbb{R}^k$ is said to converge in probability to $Y$ if for every $\epsilon > 0$

$$\lim_{n\to\infty} P(\parallel Y_n - Y \parallel > \epsilon) = 0,$$

where $\parallel a \parallel$ is the Euclidean norm for $a \in \mathbb{R}^k$, i.e. $\parallel a \parallel = (a^T a)^{1/2} = (\sum_{i=1}^k a_i^2)^{1/2}$. This type of convergence is denoted by $Y_n \xrightarrow{p} Y$.

**Almost sure Convergence.** The sequence $Y_n$ is said to converge almost surely to $Y$ if

$$P(\omega : \lim_{n\to\infty} \parallel Y_n(\omega) - Y(\omega) \parallel = 0) = 1,$$

and written as $Y_n \xrightarrow{as} Y$, where $\omega$ is an element of the sample space $\Omega$. Note that, for each $n$, $Y_n$ and $Y$ must be defined on the same sample space in convergence almost surly and convergence in probability. However, this is not a requirement in the definition of the convergence in distribution.

Other notations are useful, to avoid messy details in asymptotic calculation in probability theory. We define the stochastic symbols $o_p(.)$ and $O_p(.)$. Let $\{X_n\}$ be a sequence of random variables. We say $X_n = o_p(1)$ means $X_n \xrightarrow{p} 0$ as $n \to \infty$, while $X_n = O_p(1)$ means $X_n$ is asymptotically bounded as $n \to \infty$, i.e. given $\epsilon > 0$, there exists an $A = A(\epsilon)$ such that

$$P(|X_n| \geqslant A) \leqslant \epsilon,$$

for $n$ sufficiently large. More generally, let $a_n$ is a sequence of positive numbers. If we say $X_n = o_p(a_n)$ then this means that $X_n/a_n = o_p(1)$, and likewise, $X_n = O(a_n)$ means $X_n/a_n = O_p(1)$.

The following result is known as Slutsky's lemma; see e.g. van der Vaart (2000), p.11.

**Lemma 2.1 (Slutsky.)**
*Let $X_n$, $X$ and $Y_n$ be random vectors or variables. If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$ for a constant $c$, then*

**(i)** $X_n + Y_n \xrightarrow{d} X + c$;

**(ii)** $Y_n X_n \xrightarrow{d} cX$;

**(iii)** $Y_n^{-1} X_n \xrightarrow{d} c^{-1} X$ *provided* $c \neq 0$.

This result is very useful for proving convergence in distribution and is used to prove Corollary 2.1, which is used in Chapter 3.

### 2.4.3 Central Limit Theorem

The most famous example of the convergence in distribution is the central limit theorem (CLT) case in the independent and identically distributed.

**Theorem 2.2 (Multivariate CLT)**
*Let $X_i$ be a sequence of iid $p$-dimensional random vectors with $E(X_1) = \mu$ and covariance matrix $Cov(X_1) = \Sigma$. Then, if $\overline{X} = n^{-1} \sum_{i=1}^{n} X_i$ is the sample mean*

*of* $X_i$,

$$\sqrt{n}(\overline{X} - \mu) \xrightarrow{d} N_p(0_n, \Sigma).$$

*where* $0_n$ *is a zero vector.*

### 2.4.4 Projections

To derive the limit distribution of a sequence of statistics $T_n$, we consider the projection method due originally to Hoeffding (1948); see also van der Vaart (2000). Projection is commonly used for showing that the sequence of statistics $T_n$ is asymptotically equivalent to a sequence $S_n$, where the limit behavior of the latter is known. Lemma 2.1 is the basis of this method, which shows that the sequence $T_n = T_n - S_n + S_n$ converges in distribution to $S$ if both $T_n - S_n \xrightarrow{p} 0$ and $S_n \xrightarrow{d} S$.

Now assume a sequence of statistics $T_n$ and linear space $\mathcal{S}$ is given. For each $n$, let $\widehat{S}_n$ be the projection of $T_n$ on $\mathcal{S}$; where $\widehat{S}_n$ is defined as the random variable in $\mathcal{S}$ which minimises $E(T - \widehat{S}_n)^2$; see van der Vaart (2000), Chapter 11. If the limit behavior of $\widehat{S}_n$ is known, then the limiting behavior of the sequence $T_n$ follows from that of $\widehat{S}_n$ provided the quotient $\text{Var}(T_n)/\text{Var}(\widehat{S}_n)$ converges to 1.

**Theorem 2.3 (van der Vaart (2000))**
*For each* $n$*, let* $\mathcal{S}_n$ *be linear space of random variables with finite second moments that contain the constants. Let* $T_n$ *be a random variable with projection* $\widehat{S}_n$ *onto* $\mathcal{S}_n$*. If* $\text{Var}(T_n)/\text{Var}(\widehat{S}_n) \longrightarrow 1$ *as* $n \to \infty$ *then*

$$\frac{T_n - E(T_n)}{sd(T_n)} - \frac{\widehat{S}_n - E(\widehat{S}_n)}{sd(\widehat{S}_n)} \xrightarrow{p} 0, \quad \text{as } n \to \infty, \tag{2.4.2}$$

*where* $sd(X)$ *is the standard deviation of a random variable* $X$*.*

**Corollary 2.1**
*Assume that the conclusion of Theorem 2.3 holds and suppose in addition that*

$$\frac{\widehat{S}_n - E(\widehat{S}_n)}{sd(\widehat{S}_n)} \xrightarrow{d} N(0, 1). \tag{2.4.3}$$

*Then*

$$\frac{T_n - E(T_n)}{sd(T_n)} \xrightarrow{d} N(0, 1).$$

**Proof:** Now

$$\frac{T_n - E(T_n)}{sd(T_n)} = \left\{ \frac{T_n - E(T_n)}{sd(T_n)} - \frac{\widehat{S}_n - E(\widehat{S}_n)}{sd(\widehat{S}_n)} \right\} + \frac{\widehat{S}_n - E(\widehat{S}_n)}{sd(\widehat{S}_n)}. \tag{2.4.4}$$

From (2.4.2), the first term on the RHS of (2.4.4) goes to 0 in probability, while the second term on the RHS of (2.4.4) is asymptotically standard normal by assumption (2.4.3). Therefore the result follows from Slutsky's Lemma. □

In Chapter 3 we will need to make use of the following multivariate version of Corollary 2.1.

**Corollary 2.2**

*Suppose that $T_n = (T_{n,1}, \cdots, T_{n,k})^T$ is a sequence of random vectors and for each $n$ let $\mathcal{S}_n$ denote a linear space of random variables. For $j = 1, \cdots, k$ and, for each $n$, suppose that the projection of $T_{n,j}$ onto $\mathcal{S}_n$ is written $\widehat{S}_{n,j}$. Assume that*

**(i)** *for $j = 1, \cdots, k$,*

$$\frac{T_{n,j} - E(T_{n,j})}{sd(T_{n,j})} - \frac{\widehat{S}_{n,j} - E(\widehat{S}_{n,j})}{sd(\widehat{S}_{n,j})} \xrightarrow{p} 0; \tag{2.4.5}$$

**(ii)** *and*

$$S_n^* \xrightarrow{d} N_k(0_k, V^*), \tag{2.4.6}$$

*where*

$$S_n^* = \left( \frac{\widehat{S}_{n,1} - E(\widehat{S}_{n,1})}{sd(\widehat{S}_{n,1})}, \cdots, \frac{\widehat{S}_{n,k} - E(\widehat{S}_{n,k})}{sd(\widehat{S}_{n,k})} \right)^T,$$

*and*

$$V^* = \lim_{n \to \infty} Cov(S_n^*) \text{ exists.}$$

*Then*

$$T_n^* = \left( \frac{T_{n,1} - E(T_{n,1})}{sd(T_{n,1})}, \cdots, \frac{T_{n,k} - E(T_{n,k})}{sd(T_{n,k})} \right) \xrightarrow{d} N_k(0_k, V^*).$$

**Proof.** We use the Cramér-Wold device; see p.16 of van der Vaart (2000). This states that if $(X_n)_{n \geqslant 1}$ and $X$ are k-dimensional random vectors, then $X_n \xrightarrow{d} X$ if and only if for each fixed vector $a$, $a^T X_n \xrightarrow{d} a^T X$, as $n \to \infty$. To prove Corollary

2.2, it is therefore sufficient to show that for each fixed $a = (a_1, \cdots, a_k)^T$,

$$a^T T_n^* \xrightarrow{d} N(0, a^T V^* a) \text{ as } n \to \infty.$$

Using the inequality

$$\left| \sum_{j=1}^{k} c_j \right| \leqslant \sum_{j=1}^{k} |c_j|,$$

we have

$$|a^T T_n^* - a^T \widehat{S}_n^*| \leqslant \sum_{j=1}^{k} |a_j| \left| \frac{T_{n,j} - E(T_{n,j})}{sd(T_{n,j})} - \frac{\widehat{S}_{n,j} - E(\widehat{S}_{n,j})}{sd(\widehat{S}_{n,j})} \right|. \tag{2.4.7}$$

From (2.4.5), each component on the right-hand side of (2.4.7) convergence to 0 in probability, and therefore

$$a^T T_n^* - a^T \widehat{S}_n^* \xrightarrow{p} 0.$$

Therefore we may apply Corollary 2.1 to obtain

$$\begin{aligned} a^T T_n^* &= a^T T_n^* - a^T \widehat{S}_n^* + a^T \widehat{S}_n^* \\ &= o_p(1) + a^T \widehat{S}_n^* \\ &\xrightarrow{d} N(0, a^T V^* a), \end{aligned}$$

using (2.4.6). Consequently after applying the Cramér-Wold device we may conclude that

$$T_n^* \xrightarrow{d} N_k(0_k, V^*).$$

□

*Projection onto Sums:*

Let $X_1, X_2, \cdots, X_N$ be independent random vectors, and let $\mathcal{S}$ be the set of all random variables of the form

$$\sum_{i=1}^{N} g_i(X_i),$$

where $g_i : \mathbb{R} \to \mathbb{R}$ is such that $E(g^2(X_i)) < \infty$. This class is of interest, because the convergence in distribution of the sums can be derived from the central limit theorem. The projection of a random variable onto this class is known as its

*Hájek projection.*

**Lemma 2.2 (van der Vaart (2000))**
*Let $X_1, X_2, \cdots, X_N$ be independent random vectors. Then the projection of an arbitrary random variable $T$ with finite second moment onto the class $\mathcal{S}$ is given by*

$$\widehat{S} = \sum_{i=1}^{N} E(T|X_i) - (n-1)E(T).$$

For more explanations, see van der Vaart (2000). We will use the projection method in Chapter 3 for prove the central limit theorem given there.

### 2.4.5 The Method of Moments

There are many methods for proving central limit theorems. One of these is the projection method discussed in Subsection 2.4.4. A second important technique is the method of moments. Here, weak convergence of the sequence of distributions is proved by establishing that the moments converge. This approach requires conditions under which a distribution is uniquely determined by its moments.

**Theorem 2.4 (Billingsley (2012),p.412)**
*Let $\mu$ be a probability measure on the line having finite moments $\alpha_k = \int_{-\infty}^{\infty} x^k \mu(dx)$ of all positive integer orders. If the power series $\Sigma_k \alpha_k r^k/k!$ has a positive radius of convergence, then $\mu$ is the only probability measure with the moments $\alpha_1, \alpha_2, \ldots$.*

**Theorem 2.5 (Billingsley (2012),p.414)**
*Suppose that the distribution of $X$ is determined by its moments, as in Theorem 2.4, that the $X_n$ have moments of all orders, and that $\lim_{n\to\infty} E[X_n^r] = E[X^r]$ for $r = 1, 2, \ldots$. Then $X_n \xrightarrow{d} X$.*

Theorem 2.5 will be used to prove the main central limit theorem in Chapter 5.

### 2.4.6 Conditional Expectation

For elementary properties of conditional expectation; see e.g. Billingsley (2012). Let $X, Y$ and $Z$ denote random vectors. In the context considered here, all expectations specified below exist and are finite. A particular result we shell use later is the following. For general functions $f$ and $g$, we have

$$E\left[f(X)g(Y)|Z\right] = E\left[f(X)E[g(Y)|X,Z]|Z\right]. \tag{2.4.8}$$

## 2.5 Composite Likelihood Methods

### 2.5.1 Types of Composite Likelihood

Many application areas use composite likelihood methods for statistical inference for parameters (Varin et al. (2011)). The advantage of composite likelihood is to reduce the computational complexity so that it is possible to deal with large datasets and very complex models, especially when the use of standard likelihood is not tractable. A composite likelihood function can be derived by multiplying togather a collection of likelihood components. Each component is a conditional or marginal probability density or probability mass function.

Suppose $f(y;\theta)$ is a probability function of $m$-dimensional random vector $y = (y_1, \cdots, y_m)^T$ and for some unknown $p$-dimensional parameter vector $\theta \in \Theta$. Let $\{A_1, ..., A_K\}$ a set of marginal or conditional events with associated likelihoods $\mathcal{L}_k(\theta; y) \propto f(y \in A_k; \theta)$. A composite likelihood based on these components is a weighted product

$$\mathcal{L}_C(\theta; y) = \prod_{k=1}^{K} \mathcal{L}_k(\theta; y)^{w_k},$$

with weights $w_k \geq 0$ to be chosen; see Varin et al. (2011) for a helpful and up-to-date review. In what follows, $f$ is used as a generic symbol for a probability density function or probability mass function.

***Composite Conditional Likelihoods***

Suppose that the observations $y_1, \cdots, y_m$ have a neighborhood structure in that $y_r$ has neighbors $y_s$, $s \in N_r$, where for $r = 1, \cdots, m$, $N_r \subseteq \{1, \cdots, m\}$, $r \notin N_r$, is the set of indices of the neighbors of $y_r$. One type of composite likelihood is the product of the conditional densities of a single observation given its neighbors,

$$\begin{aligned}\mathcal{L}_C(\theta; y) &= \prod_{r=1}^{m} f(y_r | \{y_s : y_s \text{ is a neighbor of } y_r\}; \theta) \\ &= \prod_{r=1}^{m} f(y_r | \{y_s, s \in N_r\}; \theta)\end{aligned}$$

Alternatively, composite likelihood can be construct by pooling pairwise conditional densities

$$\mathcal{L}_C(\theta; y) = \prod_{r=1}^{m-1} \prod_{s=r+1}^{m} f(y_r | y_s; \theta),$$

or by pooling full conditional densities

$$\mathcal{L}_C(\theta; y) = \prod_{r=1}^{m} f(y_r | y_{-r}; \theta),$$

where $y_{-r}$ denotes the vector of all the observations but excluding $y_r$. Which of the above is used will depend in part on the structure of the problem and in part on convenience.

***Composite Marginal Likelihoods***

Under independence assumptions, the most straightforward composite marginal likelihood is

$$\mathcal{L}_{ind}(\theta; y) = \prod_{r=1}^{m} f(y_r; \theta),$$

i.e. the observations $y_1, \cdots, y_m$ are treated as independent. The marginal likelihood allows inference only on marginal parameters. However, when parameters linked to dependence are of interest, it is necessary to model blocks of observations, as in the pairwise likelihood (Varin et al. (2011)),

$$\mathcal{L}_{ind}(\theta; y) = \prod_{r=1}^{m-1} \prod_{s=r+1}^{m} f(y_r, y_s; \theta).$$

### 2.5.2 Asymptotic behavior of Composite Likelihood estimators

Let $X_1, ..., X_n$ denote an IID sample from a population with distribution function $F$. Suppose that we wish to construct an estimator of a parameter vector $\theta \in$

$\Omega_\theta \subseteq \mathbb{R}^d$ based on the sample $X_1, ..., X_n$ and using an estimating function given by

$$G(\theta) = \sum_{i=1}^{n} G_i(\theta) \equiv \sum_{i=1}^{n} G(X_i, \theta),$$

where $G(X_i, \theta) \in \mathbb{R}^d$, i.e. $G$ has the same dimension as $\theta$. Assume $\theta_0$ is such that

$$E_F\{G(X_1, \theta_0)\} \equiv \int G(x, \theta_0) dF(x) = 0,$$

and consider the sequence of estimating equations for $\theta$ given by

$$G(\theta) \equiv \sum_{i=1}^{n} G_i(\theta) = 0, \qquad n = d, d+1, ... \tag{2.5.1}$$

**Theorem 2.6**

*Under mild technical conditions, (2.5.1) admits a sequence of solutions $(\widehat{\theta}_n)_{n=d}^{\infty}$ with the following properties: as $n \to \infty$,*

$$\widehat{\theta}_n \xrightarrow{p} \theta_0,$$

*i.e. $\widehat{\theta}_n$ is a consistent estimator of $\theta_0$; and*

$$n^{1/2}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N_d(0, H(\theta_0)V(\theta_0)H(\theta_0)^T),$$

*where*

$$V(\theta) = \text{Cov}\{G(X_1, \theta)\}, \;\; H(\theta) = [E_F\{\nabla_\theta^T G(X, \theta)\}]^{-1},$$

*and $\nabla_\theta$ is the gradient operator.*

The mild technical conditions referred to in the theorem are fairly complex to state and, as we shall not be using this theorem in the thesis, we just refer to van der Vaart (2000) for further details.

In the case where the composite likelihood is in fact a standard likelihood, it can be seen that $H(\theta) = \bar{i}(\theta)^{-1}$ where $\bar{i}(\theta)$ is the Fisher information matrix for a single observation, and $V(\theta) = \bar{i}(\theta)$. Therefore,

$$\begin{aligned} H(\theta_0)V(\theta_0)H(\theta_0)^T &= \bar{i}(\theta_0)^{-1}\bar{i}(\theta_0)\bar{i}(\theta_0)^{-1} \\ &= \bar{i}(\theta_0)^{-1} \end{aligned}$$

and the standard result for maximum likelihood estimators

$$n^{\frac{1}{2}}(\widehat{\theta} - \theta_0) \xrightarrow{d} N\{0, \bar{i}(\theta_0)^{-1}\}.$$

is recovered.

In Chapter 6 we will consider some novel parameter estimators in exponential random graph models based on new composite likelihoods. These new estimators are relatively easy to compute. However, it is unfortunately the case that the asymptotic theory presented in this subsection is not applicable due to the complex dependence structure.

## 2.6 More Advanced Topics

Ruciński (1988) gives necessary and sufficient conditions for asymptotic normality of counts of a fixed graph in a classical random graph, $RG(n, p)$, and also considers Poisson convergence under some restrictive conditions. At random time, Janson (1990) considers a random graph that evolves in time by adding new edges and Janson (1990) proves a functional limit theorem for a class of statistics of the random graph in this time-dependent setting.

Snijders (2002), considers estimation of the parameters of the *ERGM* triad model using Markov chain Monte Carlo (MCMC) methods and using stochastic approximation methods to approximate the solution to the likelihood equation.

Chatterjee and Diaconis (2013) provide a method for theoretical analysis of a 2-parameter submodel of the *ERGM* triad model which is will be discussed in Section 6.4.

A strong point of *ERGMs* is that a discrete exponential family is formed with commonly used graph statistics as sufficient statistics (see Robins, Snijders, Wang, Handcock and Pattison (2007)). However, the presence of the unknown normalizing constant, $\psi(\theta)$, makes parameter estimation in *ERGMs* extremely difficult to handle from a statistical point of view, because it requires evaluating a sum over a very large number of graphs and often too large to be feasible graphs. Geyer and Thompson (1992) provide a Monte Carlo algorithm that uses samples from a distribution with known parameters to approximate the full likelihood, which is then maximized to estimate the MLE. Bayesian approaches have been considered by Caimo and Friel (2013) for example.

Other types of models of interest for random graphs are degree-based models, where the focus is on the vertex degree distribution, see for example Olhede and Wolfe (2012). By considering empirical counts of certain motifs in a graph, Bickel et al. (2011) provide a general method of moments approach that can be used to fit a large class of probability models. Moreover, graphs with a given degree distribution ware studied by Britton et al. (2006) and Chatterjee et al. (2011). However, degree distributions have not considered in this research.

Janson et al. (2004) provide exponential bounds for the upper tail for subgraph counts, and Kunegis (2014) provides a software for computing the mean and variance of subgraph counts in random graphs. Picard et al. (2008) provide an analytical expression of the mean and variance of the motif count under any exchangeable random graph model, and they approximate the motif count distribution by a compound Poisson distribution. Janson and Nowicki (1991) prove results concerning the asymptotic behaviour of a class of statistics in various random graph models.

Bloznelis and Götze (2001) study orthogonal decomposition of general symmetric statistics based on samples drawn without replacement from finite populations. Furthermore, in terms of the Hoeffding decomposition, they provide bounds for the reminders of the approximations, see van der Vaart (2000).
This provides a powerful approach for deriving limit distribution for symmetric statistics based on data obtained by simple random sampling without replacement from a finite population. Unfortunately the random graph statistics considered in Chapters 4 and 5 are not symmetric statistics in the sense required, as we now show.
Let $T = t(X_1, \cdots, X_n)$ denote a statistic based on simple random sample $X_1, \cdots, X_n$ drawn without replacement from a finite population $\mathcal{X} = \{x_1, \cdots, x_N\}$ consisting of $N$ units. We say that $T$ is a symmetric statistic if $t(x_1, \cdots, x_n)$ is a symmetric function of $x_1, \cdots, x_n$. This means that for any permutation

$\sigma : \{1, \cdots, n\} \to \{1, \cdots, n\}$, then

$$t(x_{\sigma(1)}, \cdots, x_{\sigma(n)}) = t(x_1, \cdots, x_n).$$

Here we will present a counter example to the claim that statistics of the type considered in Chapters 4 and 5 are symmetric statistics; see Bloznelis and Götze (2001) and Bloznelis and Götze (2002).

**Example:** Consider a statistic

$$A = \sum_{i=1}^{n} \sum_{i \neq j < k \neq i} (y_{ij} - p)(y_{ik} - p),$$

which is scalar multiple of the statistic $\overline{C}_2$ defined in (4.6.1) and considered in Chapter 5. This statistic is related to the number of 2-stars in the graph. The $y_{ij}$ are the Bernoulli random variables defined in (2.2.1). Note that each term $(y_{ij} - p)(y_{ik} - p)$ has repeated vertex $i$.

For $n = 4$, we have

$$\begin{aligned} A(y) &= t_1(y_{12}, y_{13}, y_{14}, y_{23}, y_{24}, y_{3,4}) \\ &= t_1(X_1, X_2, X_3, X_4, X_5, X_6) \\ &= (y_{12} - p)(y_{13} - p) + (y_{12} - p)(y_{14} - p) + (y_{13} - p)(y_{14} - p) \\ &\quad +(y_{21} - p)(y_{23} - p) + (y_{21} - p)(y_{24} - p) + (y_{23} - p)(y_{24} - p) \\ &\quad +(y_{31} - p)(y_{32} - p) + (y_{31} - p)(y_{34} - p) + (y_{32} - p)(y_{34} - p), \end{aligned}$$

Now if we permute the random variables, we get some terms with no common vertex, i.e $(y_{ij} - p)(y_{uv} - p)$, for $i, j \neq u, v$; for instance, swapping $y_{12}$ with $y_{13}$, we get the following

$$\begin{aligned} A(y) &= t_2(X_2, X_1, X_3, X_4, X_5, X_6) \\ &= (y_{13} - p)(y_{12} - p) + (y_{13} - p)(y_{14} - p) + (y_{12} - p)(y_{14} - p) \\ &\quad +(y_{31} - p)(y_{23} - p) + (\mathbf{y_{31}} - \mathbf{p})(\mathbf{y_{24}} - \mathbf{p}) + (y_{23} - p)(y_{24} - p) \\ &\quad +(y_{21} - p)(y_{32} - p) + (y_{21} - p)(y_{34} - p) + (y_{32} - p)(y_{34} - p). \end{aligned}$$

As we notice, the fifth term in the summation, $(y_{31} - p)(y_{24} - p)$, has no common vertex, which mean the statistic, $C_2$ is not a symmetric statistic, since permutation of the arguments of the statistic changes the original statistic. Thus we are not able to use the results of Bloznelis and Götze (2001) in Chapter 5.

Chapter 3

# Central Limit Theorem: for some Random Graph Statistics

## 3.1 Introduction

The aim of this chapter is to state and prove a joint central limit theorem (CLT) for three random graph statistics in the Erdös-Rényi-Gilbert random graph model: the number of edges, the number of 2-stars and the number of triangles. This CLT is stated in Theorem 3.1. Although this CLT seems a very basic result to investigate, we have not been to find a statement of or reference to this result in the literature.

The most interesting and important, and perhaps surprising, aspect of this result is that it is degenerate in the following sense: the limiting covariance matrix of the centred and scaled trivariate statistic has rank 1 rather than rank 3, as is seen in Theorem 3.1. It is interesting to speculate that there is a connection between the degeneracy present in Theorem 3.1 and the approximate degeneracy of the 3-parameter exponential random graph model established by Chatterjee and Diaconis (2013).

The proof of Theorem 3.1 uses the projection method; see van der Vaart (2000) and Section 2.4.4. Also needed in the proof are several counting lemmas which are needed to calculate the variances and covariances of the three statistics.

The outline of this chapter is as follows. In Section 3.2, notation is introduced and the main result of the chapter, Theorem 3.1, is stated. In Section 3.3, the counting lemmas needed in the proof of Theorem 3.1, and also used later on in the thesis, are stated and proved. In Section 3.4, the variances and covariances of the three statistics under consideration are found using the lemmas in the previous section. Finally, Theorem 3.1 is proved in Section 3.5.

## 3.2 The Central Limit Theorem

All graphs we consider here are simple (no loops or parallel edges), have a finite number of vertices, and are undirected. Let

$$y_{ij} = \begin{cases} 1 & \text{if there is an edge connecting vertex } i \text{ to vertex } j; \\ 0 & \text{otherwise.} \end{cases} \tag{3.2.1}$$

Assume $V_n = \{1, 2, ..., n\}$ is a set of $n$ vertices. Then for $i \in V_n, j \in V_n$, $i \neq j$, we have $y_{ij} = y_{ji}$.

By $RG(n,p)$ we mean the following random graph: the $y_{ij}$ are independent and identically distributed with $P[y_{ij} = 1] = p$ and $P[y_{ij} = 0] = q = 1 - p$.

Let

$$\begin{aligned} u_1 &= \sum_{1 \leqslant i < j \leqslant n} y_{ij}, \\ u_2 &= \sum_{i=1}^{n} \sum_{i \neq j < k \neq i} y_{ij} y_{ik}, \\ u_3 &= \sum_{1 \leqslant i < j < k \leqslant n} y_{ij} y_{jk} y_{ki}. \end{aligned}$$

Note that $u_1$, $u_2$ and $u_3$ are, respectively, the number of edges, the number of 2-stars and the number of triangles. It will be convenient to work with centred and scaled versions of these statistics:

$$\overline{T}_1 = \frac{2}{n(n-1)} u_1 - p = \frac{2}{n(n-1)} \sum_{1 \leqslant i < j \leqslant n} (y_{ij} - p), \tag{3.2.2}$$

$$\overline{T}_2 = \frac{2}{n(n-1)(n-2)} u_2 - p^2 = \frac{2}{n(n-1)(n-2)} \sum_{i=1}^{n} \sum_{i \neq j < k \neq i} (y_{ij} y_{ik} - p^2), \tag{3.2.3}$$

$$\overline{T}_3 = \frac{6}{n(n-1)(n-2)} u_3 - p^3 = \frac{6}{n(n-1)(n-2)} \sum_{1 \leqslant i < j < k \leqslant n} (y_{ij} y_{jk} y_{ki} - p^3). \tag{3.2.4}$$

Note that $\overline{T}_1$, $\overline{T}_2$ and $\overline{T}_3$ are, respectively, the edge density, the 2-star density and the triangle density, each centred by its theoretical mean.
We are now in the position to state the main result of the chapter.

**Theorem 3.1**
*Let $\overline{T}_1$, $\overline{T}_2$ and $\overline{T}_3$ be as defined in (3.2.2), (3.2.3) and (3.2.4) respectively. Let $G_n(n,p)$, $n = 1,2,\cdots$ denote a sequence of random graphs from Erdös-Rényi-Gilbert model $RG(n,p)$, i.e. the number of vertices is $n$ and the probability of an edge being present is $p$. Then*

$$n\begin{pmatrix}\overline{T}_1\\ \overline{T}_2\\ \overline{T}_3\end{pmatrix} \xrightarrow{d} N_3(0_3, 2p(1-p)a\ a^T)$$

*as $n \to \infty$, i.e. the limiting distribution of $n(\overline{T}_1, \overline{T}_2, \overline{T}_3)^T$ is multivariate normal with mean the zero vector and covariance matrix $2p(1-p)a\ a^T$, where $a = (1, 2p, 3p^2)^T$.*

**Remark 3.1**
*Theorem 3.1 tells us that the covariance matrix has rank 1, and therefore the standardised joint limiting distribution of $\overline{T}_1$, $\overline{T}_2$ and $\overline{T}_3$ is degenerate.*

The proof of Theorem 3.1 is given in Section 3.5.

## 3.3 Some Counting Lemmas

In this section we present some counting lemmas which will be useful when calculating the variances and covariances of the statistics $\overline{T}_2$ and $\overline{T}_3$ defined in (3.2.3) and (3.2.4). The cases involving $\overline{T}_1$ are more elementary and do not require separate treatment. For example, when calculating the variance of $\overline{T}_2$, we need to evaluate

$$\begin{aligned} E[\overline{T}_2^2] &= \left\{\frac{2}{n(n-1)(n-2)}\right\}^2 \\ &\quad \times \sum_{i=1}^{n} \sum_{i\neq j<k\neq i} \sum_{\alpha=1}^{n} \sum_{\alpha\neq\beta<\gamma\neq\alpha} E\left\{(y_{ij}y_{ik}-p^2)(y_{\alpha\beta}y_{\alpha\gamma}-p^2)\right\}. \end{aligned} \quad (3.3.1)$$

When evaluating (3.3.1), it is helpful to know, for fixed $\alpha$, $\beta$ and $\gamma$, how many choices of $i$, $j$ and $k$ there are such that $\{\{i,j\},\{i,k\}\}$ has (i) zero, (ii) one or (iii) two elements in common with $\{\{\alpha,\beta\},\{\alpha,\gamma\}\}$. This is because the value of $E\{(y_{ij}y_{ik}-p^2)(y_{\alpha\beta}y_{\alpha\gamma}-p^2)\}$ depends on whether we are in case (i), case (ii) or case (iii). Results of this type are elementary but some care is needed. The required counting results are presented in Lemmas 3.1-3.3 and are used in Section 3.4. They are also used in the next chapter in Section 4.7 when calculating conditional moments.

The following lemma is useful when calculating $Var(\overline{T}_2)$.

**Lemma 3.1**

*Consider a set of vertices $V_n = \{1,...,n\}$ and fix $\alpha,\beta,\gamma \in V_n$, where $\alpha \neq \beta,\gamma$ and $\beta < \gamma$. Define $B_{\alpha\beta\gamma} = \{\{\alpha,\beta\},\{\alpha,\gamma\}\}$, i.e a set consisting of two pairs. Suppose now that we choose $i,j,k \in V_n$ with $i \neq j,k$ and $j < k$, and define $B_{ijk} = \{\{i,j\},\{i,k\}\}$. Then:*

**(i)** *The number of choices of the triple $\{i,j,k\}$ with $i \neq j,k$ and $j<k$ such that $B_{\alpha\beta\gamma}$ and $B_{ijk}$ have no pair in common, i.e, $B_{\alpha\beta\gamma}\bigcap B_{ijk} = \emptyset$, the empty set, is*

$$\frac{(n-3)(n^2-6)}{2}.$$

**(ii)** *The number of choices of $\{i,j,k\}$ with $i \neq j,k$ and $j<k$ such that $B_{\alpha\beta\gamma}$ and $B_{ijk}$ have precisely one pair in common, i.e. $|B_{\alpha\beta\gamma}\bigcap B_{ijk}| = 1$, is*

$$4n-10.$$

**(iii)** *The number of choices of $\{i,j,k\}$ with $i \neq j,k$ and $j<k$ such that $B_{\alpha\beta\gamma} = B_{ijk}$, the sets are the same, is 1.*

**Proof:** To prove Lemma 3.1 , we have to consider three cases.

**Case (i)** Here, $B_{ijk}$ has no pairs in common with $B_{\alpha\beta\gamma}$, i,e, $B_{ijk}\bigcap B_{\alpha\beta\gamma} = \emptyset$. There are precisely four ways in which Case (i) can arise.

1. ("$i$ in the middle"), $i=\alpha$, $j \neq \alpha,\beta,\gamma$ and $k \neq \alpha,\beta,\gamma,j$, leads to $(n-3)(n-4)$ instances for unordered pairs when $i \neq j,k$ and $j \neq k$, and therefore

$$\frac{(n-3)(n-4)}{2} \tag{3.3.2}$$

instances for ordered pairs when $j < k$.

2. ("$i$ in a terminal"), $i = \beta$, $j \neq \alpha, \beta$ and $k \neq \alpha, \beta, j$ leads to $(n-2)(n-3)$ instances for unordered pairs when $i \neq j, k$ and $j \neq k$, and therefore
$$\frac{(n-2)(n-3)}{2} \tag{3.3.3}$$
instances for ordered pairs when $j < k$.

3. ("$i$ in a terminal"), $i = \gamma$, $j \neq \alpha, \gamma$ and $k \neq \alpha, \gamma, j$, leads to $(n-2)(n-3)$ instances for unordered pairs when $i \neq j, k$ and $j \neq k$, and therefore
$$\frac{(n-2)(n-3)}{2} \tag{3.3.4}$$
instances for ordered pairs when $j < k$.

4. $i \neq \alpha, \beta, \gamma$, $j \neq i$ and $k \neq i, j$, $(n-3)(n-1)(n-2)$ instances for unordered pairs when $i \neq j, k$ and $j \neq k$, and therefore
$$\frac{(n-3)(n-1)(n-2)}{2} \tag{3.3.5}$$
instances for ordered pairs when $j < k$.

Consequently, the number of choices for ordered pairs when $j < k$ in Case (i) is

$$\begin{aligned}\frac{(n-3)(n-4)}{2} + (n-2)(n-3) + \frac{(n-1)(n-2)(n-3)}{2}& \\ &= \frac{(n-3)}{2}(n-4+2n-4+(n-1)(n-2)) \\ &= \frac{(n-3)(n^2-6)}{2}.\end{aligned} \tag{3.3.6}$$

**Case (ii)** Here, $B_{ijk}$ has one pair in common with $B_{\alpha\beta\gamma}$, i,e, $|B_{ijk} \bigcap B_{\alpha\beta\gamma}| = 1$. There are precisely eight ways in which Case (ii) can arise.

1. ("$i$ in the middle"), $i = \alpha, j = \beta, k \neq \alpha, \beta, \gamma$.
2. ("$i$ in the middle"), $i = \alpha, k = \beta, j \neq \alpha, \beta, \gamma$.

3. ("$i$ in the middle"), $i = \alpha, j = \gamma, k \neq \alpha, \beta, \gamma$.
4. ("$i$ in the middle"), $i = \alpha, k = \gamma, j \neq \alpha, \beta, \gamma$.

Each of the subcases above has $(n-3)$ instances. The remaining for cases are as follows.

5. ("$i$ in a terminal"), $i = \beta, j = \alpha, k \neq \alpha, \beta$.
6. ("$i$ in a terminal"), $i = \beta, k = \alpha, j \neq \alpha, \beta$.
7. ("$i$ in a terminal"), $i = \gamma, j = \alpha, k \neq \alpha, \gamma$.
8. ("$i$ in a terminal"), $i = \gamma, k = \alpha, j \neq \alpha, \gamma$.

In each of subcases 5-8 there are $(n-2)$ instances. Therefore, the number of choices for unordered pairs when $i \neq j, k$ and $j \neq k$ in Case (ii) is

$$4(n-3) + 4(n-2).$$

Therefore, the number of choices for ordered pairs when $j < k$ in Case (ii) is

$$\frac{1}{2!}[4(n-3) + 4(n-2)] = 4n - 10. \tag{3.3.7}$$

**Case (iii)** Finally, $B_{ijk}$ has precisely two elements in common with $B_{\alpha\beta\gamma}$, i,e, $|B_{ijk} \bigcap B_{\alpha\beta\gamma}| = 2$, which leads to either $i = \alpha$, $j = \beta$ and $k = \gamma$, or $i = \alpha$, $j = \gamma$ and $k = \beta$, (2 instances) for unordered pairs when $i \neq j, k$ and $j \neq k$. Therefore, there is just one instance for ordered pairs when $j < k$. ☐

**Remark 3.2**

*To check Lemma 3.1, note that the sum of the numbers in (3.3.6), (3.3.7) and Case (iii) is*

$$\frac{(n-3)(n^2-6)}{2} + 4n - 10 + 1 = \frac{n(n-1)(n-2)}{2}, \tag{3.3.8}$$

*which is the number of ways of selecting $i, j, k$ from $V_n$ with $i \neq j, k$ and $j < k$.*

The following lemma is useful when calculating $\mathrm{Var}(\overline{T}_3)$.

**Lemma 3.2**

*Consider a set of vertices $V_n = \{1, ..., n\}$ and fix $\alpha, \beta, \gamma \in V_n$, where $\alpha < \beta < \gamma$. Define $\widetilde{B}_{\alpha\beta\gamma} = \{\{\alpha, \beta\}, \{\beta, \gamma\}, \{\gamma, \alpha\}\}$, i.e a set consisting of three pairs.*

*Suppose now that we choose* $i, j, k \in V_n$ *with* $i < j < k$*, and define* $\widetilde{B}_{ijk} = \{\{i, j\}, \{j, k\}, \{k, i\}\}$*. Then:*

**(i)** *The number of choices of the triple* $\{i, j, k\}$ *with* $i < j < k$ *such that* $\widetilde{B}_{\alpha\beta\gamma}$ *and* $\widetilde{B}_{ijk}$ *have no pair in common, i.e,* $\widetilde{B}_{\alpha\beta\gamma} \bigcap \widetilde{B}_{ijk} = \emptyset$*, the empty set, is*

$$\frac{(n-3)(n^2-16)}{6}$$

**(ii)** *The number of choices of* $\{i, j, k\}$ *with* $i < j < k$ *such that* $\widetilde{B}_{\alpha\beta\gamma}$ *and* $\widetilde{B}_{ijk}$ *have precisely one pair in common, i.e.* $|\widetilde{B}_{\alpha\beta\gamma} \bigcap \widetilde{B}_{ijk}| = 1$*, is*

$$3(n-3).$$

**(iii)** *The number of choices of* $\{i, j, k\}$ *with* $i < j < k$ *such that* $\widetilde{B}_{\alpha\beta\gamma} = \widetilde{B}_{ijk}$*, the sets are the same, is 1.*

**Proof:** To prove Lemma 3.2, once again we have to consider three cases:

**Case (i)** Here, $\widetilde{B}_{ijk}$ has no pairs in common with $\widetilde{B}_{\alpha\beta\gamma}$, i,e, $\widetilde{B}_{ijk} \bigcap \widetilde{B}_{\alpha\beta\gamma} = \emptyset$. There are precisely two ways in which Case (i) can arise, when $i, j, k$ are unordered:

1. one of $i, j, k$ is equal one of $\alpha, \beta, \gamma$, therefore $9(n-3)(n-4)$ instances;
2. all are different, leading to $(n-3)(n-4)(n-5)$ instances.

Therefore, the number of instances in Case (i) of choices $i, j, k$ with $i \neq j \neq k \neq i$, (i.e. with no ordering imposed on $i, j, k$), is

$$9(n-3)(n-4) + (n-3)(n-4)(n-5).$$

Therefore, the number of instances of choices $i, j, k$ with $i < j < k$, (ordered elements), in Case (i) is

$$\begin{aligned} \frac{9(n-3)(n-4)}{3!} + \frac{(n-3)(n-4)(n-5)}{3!} &= \frac{(n-3)(n-4)(n-5+9)}{6}, \\ &= \frac{(n-3)(n^2-16)}{6}. \qquad (3.3.9) \end{aligned}$$

**Case (ii)** Here, $\widetilde{B}_{ijk}$ has one pair in common with $\widetilde{B}_{\alpha\beta\gamma}$, i,e, $|\widetilde{B}_{ijk} \bigcap \widetilde{B}_{\alpha\beta\gamma}| = 1$. The number of choices $i, j, k$ with $i \neq j \neq k \neq i$, (i.e. with no ordering imposed on $i, j, k$), in Case (ii) is

$$18(n-3).$$

Therefore, the number of choices of $i, j, k$ in Case (ii) with $i < j < k$, (ordered elements), is

$$\frac{18(n-3)}{3!} = 3(n-3) \tag{3.3.10}$$

**Case (iii)** Finally, $\widetilde{B}_{ijk}$ has precisely three pairs in common with $\widetilde{B}_{\alpha\beta\gamma}$, i,e, $\widetilde{B}_{ijk} = \widetilde{B}_{\alpha\beta\gamma}| = 3$, the number of choices of $i, j, k$ with $i < j < k$ in Case (iii) is 1. □

**Remark 3.3**

*As a check, that the sum of the numbers in (3.3.9), (3.3.10) and Case (iii) of Lemma3.2 is*

$$\frac{(n-3)(n^2-16)}{6} + 3(n-3) + 1 = \frac{n(n-1)(n-2)}{6}, \tag{3.3.11}$$

*the number of ways of selecting $i, j, k$ from $V_n$ with $i < j < k$.*

The following lemma is useful when calculating $\mathrm{Cov}(\overline{T}_2, \overline{T}_3)$

**Lemma 3.3**

*Consider a set of nodes $V_n = \{1, ..., n\}$ and fix $\alpha, \beta, \gamma \in V_n$, where $\alpha \neq \beta, \gamma$ and $\beta < \gamma$. Define $B_{\alpha\beta\gamma} = \{\{\alpha, \beta\}, \{\alpha, \gamma\}\}$, i.e a set consisting of two pairs. Suppose now that we choose $i, j, k \in V_n$ with $i < j < k$ , and define $\widetilde{B}_{ijk} = \{\{i, j\}, \{j, k\}, \{k, i\}\}$, i.e a set consisting of three pairs. Then:*

**(i)** *The number of choices of the triple $\{i, j, k\}$ with $i < j < k$ such that $B_{\alpha\beta\gamma}$ and $S_{ijk}$ have no pair in common, i.e, $B_{\alpha\beta\gamma} \bigcap \widetilde{B}_{ijk} = \emptyset$, the empty set, is*

$$\frac{(n-3)(n^2-10)}{6}.$$

**(ii)** *The number of choices of $(i, j, k)$ with $i < j < k$ such that $B_{\alpha\beta\gamma}$ and $\widetilde{B}_{ijk}$*

*have precisely one pair in common, i.e.* $|B_{\alpha\beta\gamma} \bigcap \widetilde{B}_{ijk}| = 1$, *is*

$$2(n-3).$$

**(iii)** *The number of choices of* $\{i, j, k\}$ *with* $i < j < k$ *such that* $B_{\alpha\beta\gamma}$ *and* $\widetilde{B}_{ijk}$ *have precisely two pair in common, i.e.* $|B_{\alpha\beta\gamma} \bigcap \widetilde{B}_{ijk}| = 2$, *is 1.*

**Proof :** We have to consider three cases to prove Lemma 3.3:

**Case (i)** First we consider the case in which $\widetilde{B}_{ijk}$ has no pairs in common with $B_{\alpha\beta\gamma}$, i,e, $\widetilde{B}_{ijk} \bigcap B_{\alpha\beta\gamma} = \emptyset$. There are precisely five ways in which Case (i) can arise, when $i, j, k$ are unordered:

1. none of $i, j, k$ equals $\alpha, \beta, \gamma$, [$(n-3)(n-4)(n-5)$ instances];
2. one of $i, j, k$ equals $\alpha$, [$3(n-3)(n-4)$ instances];
3. one of $i, j, k$ equals $\beta$ , [$3(n-3)(n-4)$ instances];
4. one of $i, j, k$ equals $\gamma$ , [$3(n-3)(n-4)$ instances];
5. two of $i, j, k$ equals $\beta$ and $\gamma$, [$6(n-3)$ instances].

Therefore, the number of choices $i, j, k$ with $i \neq j \neq k \neq i$ in Case (i) is

$$\begin{aligned} & (n-3)(n-4)(n-5) + 3(n-3)(n-4) + 3(n-3)(n-4) + \\ & 3(n-3)(n-4) + 6(n-3) \\ = \; & (n-3)(n^2-10) \end{aligned}$$

Therefore, the number of choices $i, j, k$ with $i < j < k$ in Case (i) is

$$\frac{(n-3)}{6}(n^2-10). \tag{3.3.12}$$

**Case (ii)** Now we consider the case when $\widetilde{B}_{ijk}$ has one pair in common with $B_{\alpha\beta\gamma}$, i,e, $|\widetilde{B}_{ijk} \bigcap B_{\alpha\beta\gamma}| = 1$. The number of choices $i, j, k$ with $i \neq j \neq k \neq i$ in Case (ii) is

$$12(n-3)$$

Therefore, the number of choices $i, j, k$ with $i < j < k$ in Case (ii) is

$$\frac{12(n-3)}{3!} = 2(n-3) \tag{3.3.13}$$

**Case (iii)** Finally, the case in which $\widetilde{B}_{ijk}$ has precisely two pairs in common with $B_{\alpha\beta\gamma}$, i,e, $|\widetilde{B}_{ijk} \bigcap B_{\alpha\beta\gamma}| = 2$, Therefore, the number of choices $i, j, k$ with $i < j < k$ in Case (iii) is 1. □

**Remark 3.4**

*Note that the sum of the numbers in (i), (ii) and (iii) of the Lemma 3.3 by summing (3.3.12), (3.3.13) and Case (iii)*

$$\frac{(n-3)(n^2-10)}{6} + 2(n-3) + 1 = \frac{n(n-1)(n-2)}{6}, \tag{3.3.14}$$

*the number of ways of selecting $i, j, k$ from $V_n$ with $i < j < k$.*

## 3.4 Calculation of Variances and Covariances

We now present the means , variances and covariances involving $\overline{T}_1$, $\overline{T}_2$ and $\overline{T}_3$.

**Proposition 3.1**

*Consider a random graph $RG(n,p)$ as defined in Subsection 2.2.2, and let $\overline{T}_1$, $\overline{T}_2$*

*and $\overline{T}_3$ denote the statistics defined in (3.2.2)-(3.2.4) respectively. Then*

$$E\left[\overline{T}_k\right] = 0, \quad k = 1, 2, 3; \tag{3.4.1}$$

$$\text{Var}\left[\overline{T}_1\right] = \frac{2p(1-p)}{n(n-1)}; \tag{3.4.2}$$

$$\text{Var}\left[\overline{T}_2\right] = \frac{4(2n-5)}{n(n-1)(n-2)}p^3(1-p) + \frac{2}{n(n-1)(n-2)}p^2(1-p^2); \tag{3.4.3}$$

$$\text{Var}\left[\overline{T}_3\right] = \frac{18(n-3)}{n(n-1)(n-2)}p^5(1-p) + \frac{6}{n(n-1)(n-2)}p^3(1-p^3); \tag{3.4.4}$$

$$Cov\left[\overline{T}_1, \overline{T}_2\right] = \frac{4}{n(n-1)}p^2(1-p); \tag{3.4.5}$$

$$Cov\left[\overline{T}_1, \overline{T}_3\right] = \frac{6}{n(n-1)}p^3(1-p); \tag{3.4.6}$$

$$Cov\left[\overline{T}_2, \overline{T}_3\right] = \frac{12(n-3)}{n(n-1)(n-2)}p^4(1-p) + \frac{6}{n(n-1)(n-2)}p^3(1-p^2). \tag{3.4.7}$$

The proofs of the expectations in (3.4.1) are immediate. We will present the proofs of the properties (3.4.2)-(3.4.7) in the following Lemmas 3.4 - 3.9 , respectively.

**Lemma 3.4**
*In the setting of Proposition 3.1,*

$$\text{Var}(\overline{T}_1) = \frac{2p(1-p)}{n(n-1)}.$$

**Proof:** Fix $\alpha, \beta \in V_n$, where $1 \leqslant \alpha < \beta \leqslant n$. Then,

$$\begin{aligned} E[\overline{T}_1^2] &= E\left[\frac{2}{n(n-1)}\sum_{\alpha<\beta}(y_{\alpha\beta}-p)\overline{T}_1\right] \\ &= \frac{2}{n(n-1)}\sum_{\alpha<\beta}E[(y_{\alpha\beta}-p)\overline{T}_1], \end{aligned} \tag{3.4.8}$$

and

$$\begin{aligned} E[(y_{\alpha\beta}-p)\overline{T}_1] &= E\left[(y_{\alpha\beta}-p)\frac{2}{n(n-1)}\sum_{i<j}(y_{ij}-p)\right] \\ &= \frac{2}{n(n-1)}E\left[(y_{\alpha\beta}-p)^2+(y_{\alpha\beta}-p)\sum_{\substack{i<j\\(i,j)\neq(\alpha,\beta)}}(y_{ij}-p)\right] \\ &= \frac{2}{n(n-1)}Var(y_{\alpha\beta})+0, \qquad (3.4.9) \\ &= \frac{2}{n(n-1)}p(1-p), \qquad (3.4.10) \end{aligned}$$

where in (3.4.9) the second expectation is zero because $y_{ij}$ and $y_{\alpha\beta}$ are independent if $\{\alpha,\beta\}\neq\{i,j\}$ and $y_{y_{ij}}$ are IID Bernoulli with probability $p$. Therefore, from (3.4.8) and (3.4.10) and using the fact that $E[\overline{T}_1]=0$, from (3.4.1) with $k=1$,

$$\begin{aligned} \mathrm{Var}(\overline{T}_1) &= E[\overline{T}_1^2]=\frac{2}{n(n-1)}\sum_{1\leqslant\alpha<\beta\leqslant n}E[(y_{\alpha\beta}-p)\overline{T}_1], \\ &= \frac{2}{n(n-1)}\sum_{1\leqslant\alpha<\beta\leqslant n}\frac{2}{n(n-1)}p(1-p), \\ &= \frac{2p(1-p)}{n(n-1)}, \end{aligned}$$

as required. □

**Lemma 3.5**

*In the setting of Proposition 3.1,*

$$\mathrm{Var}(\overline{T}_2)=\frac{4(2n-5)}{n(n-1)(n-2)}p^3(1-p)+\frac{2}{n(n-1)(n-2)}p^2(1-p^2).$$

**Proof:** Fix $\alpha,\beta,\gamma\in V_n$, where $\alpha\neq\beta,\alpha\neq\gamma$ and $\beta<\gamma$. The $n(n-1)(n-2)/2$ choices of $i,j,k\in V_n$ such that $i\neq j,k$ and $j<k$ split into three cases

**Case (i)** $\{\{i,j\},\{j,k\}\}$ has no elements in common with $\{\{\alpha,\beta\},\{\alpha,\gamma\}\}$. From Lemma 3.1 , there are $(n-3)(n^2-6)/2$ such instances).

**Case (ii)** $\{\{i,j\},\{i,k\}\}$ has precisely one element in common with $\{\{\alpha,\beta\},\{\alpha,\gamma\}\}$. From Lemma 3.1 , there are $(4n-10)$ such instances.

**Case (iii)** $\{\{i,j\},\{i,k\}\} = \{\{\alpha,\beta\},\{\alpha,\gamma\}\}$, i.e, $\alpha = i$, $\beta = j$ and $\gamma = k$. There is one such instance.

Now

$$\begin{aligned}
E\{(y_{\alpha\beta}y_{\alpha\gamma} - p^2)(y_{ij}y_{ik} - p^2)\} &= E\{y_{\alpha\beta}y_{\alpha\gamma}y_{ij}y_{ik} - p^2 y_{\alpha\beta}y_{\alpha\gamma} - p^2 y_{ij}y_{ik} + p^4\}, \\
&= E\{y_{\alpha\beta}y_{\alpha\gamma}y_{ij}y_{ik}\} - p^4, \\
&= \begin{cases} p^4 - p^4 = 0 & \text{in Case (i);} \\ p^3 - p^4 = p^3(1-p) & \text{in Case (ii);} \\ p^2 - p^4 = p^2(1-p^2) & \text{in Case (iii).} \end{cases}
\end{aligned}$$

Therefore, using Lemma 3.1,

$$\begin{aligned}
E\{(y_{\alpha\beta}y_{\alpha\gamma} - p^2)\overline{T}_2\} &= \frac{2}{n(n-1)(n-2)} \sum_{i=1}^{n} \sum_{i\neq j<k\neq i} E\left\{(y_{\alpha\beta}y_{\alpha\gamma} - p^2)(y_{ij}y_{ik} - p^2)\right\} \\
&= \frac{2}{n(n-1)(n-2)} \left\{ \frac{(n-3)(n^2-6)}{2}.0 + (4n-10)p^3(1-p) \right. \\
&\quad \left. +1.p^2(1-p^2) \right\}, \\
&= \frac{2}{n(n-1)(n-2)} \left\{(4n-10)p^3(1-p) + p^2(1-p^2)\right\}.
\end{aligned} \tag{3.4.11}$$

Consequently, since $E[\overline{T}_2] = 0$, and using (3.4.11),

$$\begin{aligned}
\mathrm{Var}(\overline{T}_2) &= E[\overline{T}_2^2], \\
&= \frac{2}{n(n-1)(n-2)} \sum_{\alpha=1}^{n} \sum_{\alpha\neq\beta<\gamma\neq\alpha} E\{(y_{\alpha\beta}y_{\alpha\gamma} - p^2)\overline{T}_2\}, \\
&= \left\{\frac{2}{n(n-1)(n-2)}\right\}^2 \sum_{\alpha=1}^{n} \sum_{\alpha\neq\beta<\gamma\neq\alpha} \{(4n-10)p^3(1-p) \\
&\quad +p^2(1-p^2)\}, \\
&= \frac{2}{n(n-1)(n-2)}\{(4n-10)p^3(1-p) + p^2(1-p^2)\},
\end{aligned}$$

as required. □

**Lemma 3.6**

*In the setting of Proposition 3.1,*

$$\mathrm{Var}(\overline{T}_3) = \frac{18(n-3)}{n(n-1)(n-2)}p^5(1-p) + \frac{6}{n(n-1)(n-2)}p^3(1-p^3).$$

**Proof:** Fix $\alpha, \beta, \gamma \in V_n$, where $\alpha < \beta < \gamma$. The $\frac{n(n-1)(n-2)}{6}$ choices of $i, j, k \in V_n$, with $i < j < k$ split into three cases.

**Case (i)** $\{\{i,j\},\{j,k\},\{k,i\}\}$ has no elements in common with $\{\{\alpha,\beta\},\{\beta,\gamma\},\{\gamma,\alpha\}\}$. From Lemma 3.2 there are $(n-3)(n^2-16)/6$ such instances;

**Case (ii)** $\{\{i,j\},\{j,k\},\{k,i\}\}$ has precisely one element in common with $\{\{\alpha,\beta\},\{\beta,\gamma\},\{\gamma,\alpha\}\}$. From Lemma 3.2 there are $3(n-3)$ such instances;

**Case (iii)** $\{\{i,j\},\{j,k\},\{k,i\}\}$ is equal to $\{\{\alpha,\beta\},\{\beta,\gamma\},\{\gamma,\alpha\}\}$, i.e. $\alpha = i$, $\beta = j$ and $\gamma = k$. There is one such instance.

Now

$$\begin{aligned}
& E\{(y_{\alpha\beta}y_{\beta\gamma}y_{\gamma\alpha} - p^3)(y_{ij}y_{jk}y_{ki} - p^3)\} \\
= \; & E\{y_{\alpha\beta}y_{\beta\gamma}y_{\gamma\alpha}y_{ij}y_{jk}y_{ki} - p^3 y_{\alpha\beta}y_{\beta\gamma}y_{\gamma\alpha} - p^3 y_{ij}y_{jk}y_{ki} + p^6\} \\
= \; & E\{y_{\alpha\beta}y_{\beta\gamma}y_{\gamma\alpha}y_{ij}y_{jk}y_{ki}\} - p^6, \\
= \; & \begin{cases} p^6 - p^6 = 0 & \text{in Case (i);} \\ p^5 - p^6 = p^5(1-p) & \text{in Case (ii);} \\ p^3 - p^6 = p^3(1-p^3) & \text{in Case (iii).} \end{cases}
\end{aligned}$$

Using Lemma 3.2,

$$\begin{aligned}
& E\{(y_{\alpha\beta}y_{\beta\gamma}y_{\gamma\alpha} - p^3)\overline{T}_3\} \\
= \; & \frac{6}{n(n-1)(n-2)} \sum_{1 \leqslant i<j<k \leqslant n} E\left\{(y_{\alpha\beta}y_{\beta\gamma}y_{\gamma\alpha} - p^3)(y_{ij}y_{jk}y_{ki} - p^3)\right\} \\
= \; & \frac{6}{n(n-1)(n-2)} \left\{ \frac{(n-3)(n^2-16)}{6}.0 + 3(n-3)p^5(1-p) + p^3(1-p^3) \right\}, \\
= \; & \frac{6}{n(n-1)(n-2)} \left\{ 3(n-3)p^5(1-p) + p^3(1-p^3) \right\}. \qquad (3.4.12)
\end{aligned}$$

Consequently, since $E[\overline{T}_3] = 0$ and using (3.4.12),

$$\begin{aligned}
\text{Var}(\overline{T}_3) &= E\{\overline{T}_3^2\}, \\
&= \frac{6}{n(n-1)(n-2)} \sum_{1\leqslant\alpha<\beta<\gamma\leqslant n} E\{(y_{\alpha\beta}y_{\beta\gamma}y_{\gamma\alpha} - p^3)\overline{T}_3\}, \\
&= \left\{\frac{6}{n(n-1)(n-2)}\right\}^2 \sum_{1\leqslant\alpha<\beta<\gamma\leqslant n} \left\{3(n-3)p^5(1-p) + p^3(1-p^3)\right\}, \\
&= \frac{6}{n(n-1)(n-2)} \left\{3(n-3)p^5(1-p) + p^3(1-p^3)\right\},
\end{aligned}$$

as required. □

**Lemma 3.7**

*In the setting of Proposition 3.1,*

$$Cov(\overline{T}_1, \overline{T}_2) = \frac{4}{n(n-1)} p^2(1-p).$$

**Proof:** Fix $\alpha, \beta \in V_n$, where $\alpha < \beta$. The $n(n-1)(n-2)/2$ choices of $i, j, k \in V_n$ with $i \neq j, k$ $j < k$ splits into two cases.

**Case (i)** Either $\{i, j\}$ or $\{i, k\}$ is equal to $\{\alpha, \beta\}$. The number of such instances is $2(n-2)$.

**Case (ii)** Neither $\{i, j\}$ or $\{i, k\}$ is equal to $\{\alpha, \beta\}$. By substraction, the number of such cases is

$$\begin{aligned}
\frac{n(n-1)(n-2)}{2} - 2(n-2) &= \frac{(n^2-n)(n-2) - 4(n-2)}{2} \\
&= \frac{(n-2)(n^2-n-4)}{2}
\end{aligned}$$

Now

$$\begin{aligned}
E[(y_{\alpha\beta} - p)(y_{ij}y_{ik} - p^2)] &= E[y_{\alpha\beta}y_{ij}y_{ik} - py_{ij}y_{ik} - p^2y_{\alpha\beta} + p^3] \\
&= E[y_{\alpha\beta}y_{ij}y_{ik}] - p^3 \\
&= \begin{cases} p^2 - p^3 = p^2(1-p) & \text{in Case (i);} \\ p^3 - p^3 = 0 & \text{in Case (ii).} \end{cases}
\end{aligned}$$

Therefore,

$$\begin{aligned} E[(y_{\alpha\beta}-p)\overline{T}_2] &= \frac{2}{n(n-1)(n-2)}\sum_{i=1}^{n}\sum_{i\neq j<k\neq i} E\Big\{(y_{\alpha\beta}-p)(y_{ij}y_{ik}-p^2)\Big\} \\ &= \frac{2}{n(n-1)(n-2)}\left\{2(n-2)p^2(1-p)+\frac{(n-2)(n^2-n-4)}{2}.0\right\}, \\ &= \frac{4}{n(n-1)}p^2(1-p). \end{aligned}$$

Finally, since $E[\overline{T}_1]=E[\overline{T}_2]=0$,

$$\begin{aligned} \mathrm{Cov}(\overline{T}_1,\overline{T}_2) &= E[\overline{T}_1\overline{T}_2], \\ &= \frac{2}{n(n-1)}\sum_{1\leqslant\alpha<\beta\leqslant n} E[(y_{\alpha\beta}-p)\overline{T}_2] \\ &= \frac{4}{n(n-1)}p^2(1-p), \end{aligned}$$

as required. □

**Lemma 3.8**
*In the setting of Proposition 3.1,*

$$Cov(\overline{T}_1,\overline{T}_3)=\frac{6}{n(n-1)}p^3(1-p).$$

**Proof:** Fix $\alpha,\beta\in V_n$, where $\alpha<\beta$. The $n(n-1)(n-2)/6$ choices of $i,j,k\in V_n$ with $i<j<k$ splits into two cases.

**Case (i)** One of $\{i,j\}$, $\{i,k\}$ and $\{j,k\}$ is equal to $\{\alpha,\beta\}$. The number of such instances is $(n-2)$.

**Case (ii)** None of $\{i,j\}$, $\{i,k\}$ or $\{j,k\}$ are equal to $\{\alpha,\beta\}$. By substraction, the number of such cases is

$$\begin{aligned} \frac{n(n-1)(n-2)}{2}-(n-2) &= \frac{(n-2)(n^2-n-6)}{6} \\ &= \frac{(n-2)(n+2)(n-3)}{6} \\ &= \frac{(n^2-4)(n-3)}{6} \end{aligned}$$

Now

$$\begin{aligned} E[(y_{\alpha\beta}-p)(y_{ij}y_{jk}y_{ki}-p^3)] &= E[y_{\alpha\beta}y_{ij}y_{jk}y_{ki} - py_{ij}y_{jk}y_{ki} - p^3 y_{\alpha\beta} + p^4] \\ &= E[y_{\alpha\beta}y_{ij}y_{jk}y_{ki}] - p^4 \\ &= \begin{cases} p^3 - p^4 = p^3(1-p) & \text{in Case (i);} \\ p^4 - p^4 = 0 & \text{in Case (ii).} \end{cases} \end{aligned}$$

Therefore,

$$\begin{aligned} E[(y_{\alpha\beta}-p)\overline{T}_3] &= \frac{6}{n(n-1)(n-2)} \sum_{1\leqslant i<j<k\leqslant n} E\left\{(y_{\alpha\beta}-p)(y_{ij}y_{jk}y_{ki}-p^3)\right\} \\ &= \frac{6}{n(n-1)(n-2)}\left\{(n-2)p^3(1-p) + \frac{(n^2-4)(n-3)}{6}.0\right\}, \\ &= \frac{6}{n(n-1)}p^3(1-p). \end{aligned}$$

Finally, since $E[\overline{T}_1] = E[\overline{T}_3] = 0$,

$$\begin{aligned} \mathrm{Cov}(\overline{T}_1, \overline{T}_3) &= E[\overline{T}_1\overline{T}_3], \\ &= \frac{2}{n(n-1)} \sum_{1\leqslant\alpha<\beta\leqslant n} E[(y_{\alpha\beta}-p)\overline{T}_3] \\ &= \frac{6}{n(n-1)}p^3(1-p), \end{aligned}$$

as required. □

**Lemma 3.9**

*In the setting of Proposition 3.1,*

$$Cov(\overline{T}_2, \overline{T}_3) = \frac{12(n-3)}{n(n-1)(n-2)}p^4(1-p) + \frac{6}{n(n-1)(n-2)}p^3(1-p^2).$$

**Proof:** Fix $\alpha, \beta, \gamma \in V_n$, where $\alpha \neq \beta, \gamma$ and $\beta < \gamma$. The $n(n-1)(n-2)/6$ choices of $i, j, k \in V_n$ with $i < j < k$ splits into three cases.

**Case (i)** $\{i,j\}$, $\{i,k\}$ and $\{j,k\}$ has no elements in common with $\{\{\alpha,\beta\},\{\alpha,\gamma\}\}$. From Lemma 3.3, there are $\frac{(n-3)(n^2-10)}{6}$ such instances.

**Case (ii)** $\{i,j\}$, $\{i,k\}$ or $\{j,k\}$ has precisely two elements in common with

$\{\{\alpha,\beta\},\{\alpha,\gamma\}\}$. From Lemma 3.3, there are $2(n-3)$ such instances.

**Case (iii)** $\{i,j\}$, $\{i,k\}$ or $\{j,k\}$ has precisely one element in common with $\{\{\alpha,\beta\},\{\alpha,\gamma\}\}$. From Lemma 3.3, there are one such instances.

Now

$$\begin{aligned} E[(y_{\alpha\beta}y_{\alpha\gamma}-p^2)(y_{ij}y_{jk}y_{ki}-p^3)] &= E[y_{\alpha\beta}y_{\alpha\gamma}y_{ij}y_{jk}y_{ki}-p^2y_{ij}y_{jk}y_{ki}-p^3y_{\alpha\beta}y_{\alpha\gamma}+p^5] \\ &= E[y_{\alpha\beta}y_{\alpha\gamma}y_{ij}y_{jk}y_{ki}]-p^5 \\ &= \begin{cases} p^5-p^5=0 & \text{in Case (i);} \\ p^4-p^5=p^4(1-p) & \text{in Case (ii);} \\ p^3-p^5=p^3(1-p^2) & \text{in Case (iii).} \end{cases} \end{aligned}$$

Therefore,

$$\begin{aligned} E[(y_{\alpha\beta}-p^2)\overline{T}_3] &= \frac{6}{n(n-1)(n-2)}\sum_{1\leqslant i<j<k\leqslant n} E\left\{(y_{\alpha\beta}y_{\alpha\gamma}-p^2)(y_{ij}y_{jk}y_{ki}-p^3)\right\}, \\ &= \frac{6}{n(n-1)(n-2)}\left\{\frac{(n^2-10)(n-3)}{6}.0+2(n-3)p^4(1-p)+\right\}, \\ &= \frac{6}{n(n-1)(n-2)}\{2(n-3)p^4(1-p)+p^3(1-p^2)\}. \end{aligned} \tag{3.4.13}$$

Consequently, from (3.4.13), and since $E[\overline{T}_2]=E[\overline{T}_3]=0$,

$$\begin{aligned} \mathrm{Cov}(\overline{T}_2,\overline{T}_3) &= E[\overline{T}_2\overline{T}_3], \\ &= \frac{2}{n(n-1)(n-2)}\sum_{\alpha=1}^{n}\sum_{\alpha\neq\beta<\gamma\neq\alpha} E[(y_{\alpha\beta}y_{\alpha\gamma}-p^2)\overline{T}_3] \\ &= \frac{2}{n(n-1)(n-2)}\sum_{\alpha=1}^{n}\sum_{\alpha\neq\beta<\gamma\neq\alpha}\frac{6}{n(n-1)(n-2)}\{2(n-3)p^4(1-p)+p^3(1-p^2)\} \\ &= \frac{6}{n(n-1)(n-2)}\{2(n-3)p^4(1-p)+p^3(1-p^2)\}, \end{aligned}$$

as required. □

**Remark 3.5**

*We can summarize a general expectation formula for the statistics $\overline{T}_1$, $\overline{T}_2$ and $\overline{T}_3$, as following:*

$$E[\overline{T}_r\overline{T}_s]\sim\frac{2rs}{n^2}p^{r+s-1}(1-p), \tag{3.4.14}$$

*for* $r, s = 1, 2, 3$. *Consequently, since* $E[\overline{T}_k] = 0$ *for* $k = 1, 2, 3$,

$$\begin{aligned}
Var(\overline{T}_1) = E[\overline{T}_1\overline{T}_1] &\sim \frac{2}{n^2}p(1-p),\\
Var(\overline{T}_2) = E[\overline{T}_2\overline{T}_2] &\sim \frac{8}{n^2}p^3(1-p),\\
Var(\overline{T}_3) = E[\overline{T}_3\overline{T}_3] &\sim \frac{18}{n^2}p^5(1-p),\\
Cov(\overline{T}_1, \overline{T}_2) = E[\overline{T}_1\overline{T}_2] &\sim \frac{4}{n^2}p^2(1-p),\\
Cov(\overline{T}_1, \overline{T}_3) = E[\overline{T}_1\overline{T}_3] &\sim \frac{6}{n^2}p^3(1-p),\\
Cov(\overline{T}_2, \overline{T}_3) = E[\overline{T}_2\overline{T}_3] &\sim \frac{12}{n^2}p^4(1-p).
\end{aligned}$$

## 3.5 Proof of Theorem 3.1

We first note that $\overline{T}_1$ satisfies a central limit theorem and then apply the projection method to $\overline{T}_2$ and $\overline{T}_3$. Since

$$\overline{T}_1 = \frac{2}{n(n-1)} \sum_{1 \leqslant i < j \leqslant n} (y_{ij} - p)$$

is the sample mean of centred independent and identically distributed Bernoulli($p$) random variables, $\overline{T}_1$ satisfies a central limit theorem:

$$\frac{\overline{T}_1 - E[\overline{T}_1]}{sd(\overline{T}_1)} = \sqrt{\frac{n(n-1)}{2p(1-p)}}\,\overline{T}_1 \sim \frac{n}{\sqrt{2p(1-p)}}\overline{T}_1 \xrightarrow{d} N(0, 1).$$

Note for future reference that

$$\mathrm{Var}(\overline{T}_1) = \frac{2p(1-p)}{n(n-1)}.$$

**Projection of $\overline{T}_2$:**

Consider

$$\overline{T}_2 = \frac{2}{n(n-1)(n-2)} \sum_{i=1}^{n} \sum_{i \neq j < k \neq i} (y_{ij}y_{jk} - p^2).$$

Then

$$\begin{aligned} E[\overline{T}_2|y_{\alpha\beta}] &= \frac{2}{n(n-1)(n-2)} E\Big[y_{\alpha\beta}\Big(\sum_{k\neq\alpha\beta} y_{\alpha k} + \sum_{k\neq\alpha\beta} y_{k\beta}\Big) - 2(n-2)p^2|y_{\alpha\beta}\Big] \\ &= \frac{2}{n(n-1)(n-2)}[y_{\alpha\beta}2(n-2)p - 2(n-2)p^2], \\ &= \frac{4p}{n(n-1)}(y_{\alpha\beta}-p), \end{aligned}$$

because of independence of the $y_{ij}$. Consequently, from Lemma 2.2, the projection, say $\widehat{S}_2$, of $\overline{T}_2$ onto the IID random variables $\{y_{ij}\}_{1\leqslant i<j\leqslant n}$ is

$$\begin{aligned} \widehat{S}_2 &= \sum_{1\leqslant\alpha<\beta\leqslant n} E[\overline{T}_2|y_{\alpha\beta}] - (n-1)E[\overline{T}_2], \\ &= \frac{4p}{n(n-1)} \sum_{1\leqslant\alpha<\beta\leqslant n} (y_{\alpha\beta}-p), \\ &= 2p\overline{T}_1. \end{aligned} \tag{3.5.1}$$

The variance of the projection $\widehat{S}_2$ is the following:

$$\begin{aligned} \mathrm{Var}(\widehat{S}_2) &= \mathrm{Var}(2p\overline{T}_1), \\ &= 4p^2\mathrm{Var}(\overline{T}_1), \\ &= \frac{8p^3(1-p)}{n(n-1)}, \\ &\sim \frac{8p^3(1-p)}{n^2}, \end{aligned}$$

which from Lemma 3.5 is equivalent to $\mathrm{Var}(\overline{T}_2)$, as $n\to\infty$. Consequently, due to Theorem 2.3,

$$\frac{\overline{T}_2 - E[\overline{T}_2]}{sd(\overline{T}_2)} - \frac{\widehat{S}_2 - E[\widehat{S}_2]}{sd(\widehat{S}_2)} \xrightarrow{p} 0,$$

and in view of (3.5.1), (2.4.2) holds and so we may apply Corollary 2.1 to conclude that

$$\frac{\overline{T}_2 - E[\overline{T}_2]}{sd(\overline{T}_2)} \xrightarrow{d} N(0,1), \quad \text{as} \quad n\to\infty.$$

**Projection of $\overline{T}_3$:**

Consider

$$\overline{T}_3 = \frac{6}{n(n-1)(n-2)} \sum_{1\leqslant i<j<k\leqslant n} (y_{ij}y_{jk}y_{ki} - p^3).$$

Then

$$\begin{aligned} E[\overline{T}_3|y_{\alpha\beta}] &= \frac{3}{n(n-1)(n-2)} E\left[ y_{\alpha\beta}(\sum_{k\neq\alpha\beta} y_{\beta k}y_{k\alpha}) - (n-2)p^3 \right] \\ &= \frac{6}{n(n-1)(n-2)} \left[ y_{\alpha\beta}(n-2)p^2 - (n-2)p^3 \right] \\ &= \frac{6p^2}{n(n-1)}(y_{\alpha\beta} - p). \end{aligned}$$

Consequently, from Lemma 2.2 the projection, say $\widehat{S}_3$, of $\overline{T}_3$ onto the IID random variables $\{y_{ij}\}_{1\leqslant i<j\leqslant n}$ is

$$\begin{aligned} \widehat{S}_3 &= \sum_{1\leqslant\alpha<\beta\leqslant n} E[\overline{T}_3|y_{\alpha\beta}] - (n-1)E[\overline{T}_3] \\ &= \frac{6p^2}{n(n-1)} \sum_{1\leqslant\alpha<\beta\leqslant n} (y_{\alpha\beta} - p), \\ &= 3p^2\overline{T}_1. \end{aligned} \tag{3.5.2}$$

The variance of the projection $\widehat{S}_3$ is the following:

$$\begin{aligned} \mathrm{Var}(\widehat{S}_3) &= \mathrm{Var}(3p^2\overline{T}_1), \\ &= 9p^4\mathrm{Var}(\overline{T}_1), \\ &= \frac{18p^5(1-p)}{n(n-1)}, \\ &\sim \frac{18p^5(1-p)}{n^2}, \end{aligned}$$

which from Lemma 3.6 is equivalent to $\mathrm{Var}(\overline{T}_3)$, as $n \to \infty$. Consequently, due to Theorem 2.3,

$$\frac{\overline{T}_3 - E[\overline{T}_3]}{sd(\overline{T}_3)} - \frac{\widehat{S}_3 - E[\widehat{S}_3]}{sd(\widehat{S}_3)} \xrightarrow{p} 0,$$

and in view of (3.5.2), (2.4.2) holds and so we may apply Corollray 2.1 to conclude that

$$\frac{\overline{T}_3 - E[\overline{T}_3]}{sd(\overline{T}_3)} \xrightarrow{d} N(0,1), \quad \text{as} \quad n \to \infty.$$

To complete the proof, we apply Corollary 2.2.
The covariance matrix of $n\widehat{S} = (\widehat{S}_1, \widehat{S}_2, \widehat{S}_3)^T = n\overline{T}_1(1, 2p, 3p)^T)$ is

$$n^2\mathrm{Var}(\overline{T}_1)\begin{pmatrix} 1 \\ 2p \\ 3p^2 \end{pmatrix}\begin{pmatrix} 1 & 2p & 3p^2 \end{pmatrix} \sim 2p(1-p)\begin{pmatrix} 1 \\ 2p \\ 3p^2 \end{pmatrix}\begin{pmatrix} 1 & 2p & 3p^2 \end{pmatrix}.$$

Because of the rank of covariance matrix is 1 as opposed to 3, the joint probability distribution of $\overline{T}_1$, $\overline{T}_2$ and $\overline{T}_3$ is degenerate. □

## 3.6 Summary

In this chapter we proved a joint central limit theorem (CLT) for the following three statistics assuming the Erdös-Rényi-Gilbert random graph model: the number of edges, $u_1$, the number of 2-stars, $u_2$, and the number of triangles, $u_3$. The standardised version of these statistics is jointly trivariate normal in the limit as the number of vertices, $n$, goes to infinity. However, the most interesting finding is that the limiting covariance of the standardised variables has rank 1 as opposed to rank 3.
In the following chapter we present various moment results, including the important Theorem 4.1, which are used in Chapter 5 to prove Theorem 5.1, a conditional central limit theorem for $u_2$ and $u_3$ conditional on $u_1$, the number of edges.

Chapter 4

# Conditional Moment Results for Random Graph Models

## 4.1 Introduction

In the previous chapter it was proved in Theorem 3.1 that the joint distribution of $(\overline{T}_1, \overline{T}_2, \overline{T}_3)^T$, suitably standardized, is asymptotically trivarate normal as $n \to \infty$, but is degenerate in the sense that the limiting covariance matrix has rank 1 rather than 3. The main goal of this chapter and the next is to prove a conditional central limit Theorem for $(\overline{T}_2, \overline{T}_3)^T$ given $T_1$, and show that the limiting bivariate normal distribution is non-degenerate in the sense that its covariance matrix has full rank 2. A key part of proving the conditional central limit theorem in Chapter 5 is precisely describing the behavior of the conditional expectation of general products of the form

$$E\left[\prod_{u=1}^{q}(y_{i_u,j_u} - p)\middle| \sum_{1\leqslant i<j\leqslant n} y_{ij} = m\right], \tag{4.1.1}$$

where $m = Np$ and $N = n(n-1)/2$. Note that in (4.1.1) the expectation is conditional on the event $\sum_{1\leqslant i<j\leqslant n} y_{ij} = m$.

**Remark 4.1**

*Here we mention an important point concerning notation. In Chapter 3, $p$ denoted the probability of an edge being present in the homogeneous Bernoulli random graph model. In contrast, in this chapter and the next, $p$ is always defined by*

*$p = m/N$, where $m$, the number of edges in the graph, is the variable we condition on.*

The main result of this chapter is Theorem 4.1, which describes the behavior of conditional expectations of the form (4.1.1) as $n \to \infty$. A second goal is to find all first and second conditional moments of $(\overline{C}_2, \overline{C}_3)$, conditional on the event $\sum_{1\leqslant i<j\leqslant n} y_{ij} = m$, where $\overline{C}_2$ and $\overline{C}_3$ are defined in (4.6.1) and (4.6.2) respectively. On this conditioning event, the quantities $\overline{C}_2$ and $\overline{C}_3$ are related to $\overline{T}_2$ and $\overline{T}_3$ by a simple linear transformation; see Section 4.6.
The outline of this chapter is as follows. In Section 4.2, the main results of the chapter, Theorem 4.1, is stated. In Section 4.3, some useful expressions concerning sampling without replacement from a finite population of zero-one variables are presented. A selection of lemmas needed in the proof of Theorem 4.1 are stated and proved in Section 4.4, and in Section 4.5 Theorem 4.1 is proved. Finally, in Section 4.6, the variables $\overline{C}_2$ and $\overline{C}_3$ are introduced, and their first and second moments are calculated in Section 4.7.

## 4.2 General Conditional Moments Theorem

The theorem below describes the asymptotic behavior of conditional moments of arbitrarily high order when $n$, the number of vertices in a random Erdös-Rényi-Gilbert graph, goes to infinity, and where the conditioning is on $m = m^{(n)}$, the number of edges present in the random graph with $n$ vertices. Equivalently, $m^{(n)}$ is the number of $y_{ij}^{(n)}$ equal to 1 in the sample of size $N^{(n)}$ where $N^{(n)} = n(n-1)/2$ is the maximum number of possible edges.

The main theorem of this chapter is now stated.

**Theorem 4.1**

*Consider a sequence of finite populations of $N^{(n)} = n(n-1)/2$ binary variables (i.e. zero-one variables) $y_{ij}^{(n)}$, $1 \leqslant i < j \leqslant n$, and suppose that for each $n$, $\sum_{1\leqslant i<j\leqslant n} y_{ij}^{(n)} = m^{(n)}$. Suppose also that for each $n$ we sample $q$ observations*

$$y_{i_1,j_1}, \cdots, y_{i_q,j_q}$$

*randomly, without replacement, from the full set $(y_{ij})_{1\leqslant i<j\leqslant n}$ of binary variables. Let $r_1, \cdots, r_q$ denote any fixed positive integers. Finally, assume that for $n$*

*sufficiently large,* $p^{(n)} = m^{(n)}/N^{(n)} \in (a,b)$ *for some constant* $a$ *and* $b$ *satisfying* $0 < a < b < 1$. *Then*

$$E\left[\prod_{u=1}^{q}(y_{i_u,j_u} - p^{(n)})^{r_u}\middle| m^{(n)}\right] = O(N^{-\lfloor (t+1)/2 \rfloor}), \tag{4.2.1}$$

*where* $N = N^{(n)}$, $\lfloor x \rfloor$ *denotes the largest integer not greater than x, i.e,*

$$\lfloor x \rfloor = max\left\{h \in \mathbb{Z}\middle| h \leqslant x\right\},$$

*and*

$$t = \sum_{u=1}^{q} I_{\{r_u=1\}}, \tag{4.2.2}$$

*where*

$$I_{\{A\}} = \begin{cases} 1 & \textit{if } A \textit{ is true;} \\ 0 & \textit{if } A \textit{ is false.} \end{cases}$$

Theorem 4.1 plays a crucial role in the proof of the conditional central limit theorem stated in Theorem 5.1 in Chapter 5, but we also believe it is of independent interest.

**Remark 4.2**

**(i)** *Note that the expectation in (4.2.1) is with respect to simple random sampling from a finite population.*

**(ii)** *In the formulation of the theorem, we have taken the population size to be* $N^{(n)} = n(n-1)/2$ *and we have used two indices,* $i_u$ *and* $j_u$. *These choices are purely to make the link with random graphs clear. We could have used a single index* $i_u$ *in the statement of the theorem if we had wanted to.*

## 4.3 Sampling without Replacement from a Finite Binary Population

Suppose that a population consists of $N$ elements, each of which is either a success or a failure. Suppose that the proportion of successes is $p$, and the number of successes in the population is then $m = Np$. To make inferences about $p$ we take a simple random sample of size $s$ without replacement. Interest centers here on the joint distribution of $Y_1, \cdots, Y_s$ drawn from the population without replacement, where $Y_i$ is the number of successes on the ith draw (trial). The $Y_i$ are binary valued since on each draw we get either a success $Y_i = 1$ or we get a failure $Y_i = 0$. The $Y_i$ are not independent so the trials are not Bernoulli trials. Clearly $P(Y_1 = 1) = Np/N = p$, the proportion of successes in the population, and $P(Y_1 = 0) = (N - Np)/N = 1 - p$. Now consider samples of size 2 (*i.e.* $s = 2$), selected without replacement. Then

$$
\begin{aligned}
P(Y_1 = 1, Y_2 = 1) &= \frac{Np}{N}\frac{Np-1}{N-1} = \frac{N}{N-1}p^2\left(1 - \frac{1}{Np}\right),\\
P(Y_1 = 0, Y_2 = 1) &= \frac{N-Np}{N}\frac{Np}{N-1} = \frac{N}{N-1}p(1-p),\\
P(Y_1 = 1, Y_2 = 0) &= \frac{Np}{N}\frac{N-Np}{N-1} = \frac{N}{N-1}p(1-p),\\
P(Y_1 = 0, Y_2 = 0) &= \frac{N-Np}{N}\frac{N-Np-1}{N-1} = \frac{N}{N-1}(1-p)^2\left(1 - \frac{1}{N(1-p)}\right).
\end{aligned}
\tag{4.3.1}
$$

Define

$$
\pi_{r,s-r} = P\left(\sum_{i=1}^{s} Y_i = r, \sum_{i=1}^{s}(1 - Y_i) = s - r\right). \tag{4.3.2}
$$

Note that, $\pi_{r,s-r}$ is the probability of selecting $r$ successes and $s - r$ failures in a simple random sample of size $s$ selected from the population without replacement. Then

$$
\begin{aligned}
\pi_{2,0} &= P(Y_1 = 1, Y_2 = 1) = \binom{2}{2}\frac{N}{N-1}p^2\left(1 - \frac{1}{Np}\right),\\
\pi_{0,2} &= P(Y_1 = 0, Y_2 = 0) = \binom{2}{0}\frac{N}{N-1}(1-p)^2\left(1 - \frac{1}{N(1-p)}\right),\\
\pi_{1,1} &= P(Y_1 = 1, Y_2 = 0) + P(Y_1 = 0, Y_2 = 1),\\
&= \binom{2}{1}\frac{N}{N-1}p(1-p).
\end{aligned}
\tag{4.3.3}
$$

Now suppose we are selecting a simple random sample of size $s = 3$ without replacement from the finite binary population of size $N$, and let $r$ denote number of ones. Then $3-r$ is the number of zeros, where the possible values of $r$ are $0,1,2,3$. From (4.3.2) the probability of having $r$ ones and $3-r$ zeros is $\pi_{r,3-r}$. Then

$$
\begin{aligned}
\pi_{0,3} &= \binom{3}{0}\left(\frac{N-Np}{N}\right)\left(\frac{N-Np-1}{N-1}\right)\left(\frac{N-Np-2}{N-2}\right),\\
&= \frac{N^3}{N(N-1)(N-2)}(1-p)^3\left(1-\frac{1}{N(1-p)}\right)\left(1-\frac{2}{N(1-p)}\right),\\
\pi_{3,0} &= \binom{3}{3}\left(\frac{Np}{N}\right)\left(\frac{Np-1}{N-1}\right)\left(\frac{Np-2}{N-2}\right),\\
&= \frac{N^3}{N(N-1)(N-2)}p^3\left(1-\frac{1}{Np}\right)\left(1-\frac{2}{Np}\right),\\
\pi_{1,2} &= \binom{3}{1}\left(\frac{N-Np}{N}\right)\left(\frac{Np}{N-1}\right)\left(\frac{N-Np-1}{N-2}\right),\\
&= \frac{3N^3}{N(N-1)(N-2)}p(1-p)^2\left(1-\frac{1}{N(1-p)}\right),\\
\pi_{2,1} &= \binom{3}{2}\left(\frac{Np}{N}\right)\left(\frac{Np-1}{N-1}\right)\left(\frac{N-Np}{N-2}\right)\\
&= \frac{3N^3}{N(N-1)(N-2)}p^2(1-p)\left(1-\frac{1}{Np}\right).
\end{aligned}
\tag{4.3.4}
$$

For the general case and under simple random sampling without replacement, the probability of selecting a particular sequence of length $s$ with $r$ ones and $s-r$ zeros is given by

$$
\frac{m_{(r)}(N-m)_{(s-r)}}{N_{(s)}} = \frac{m(m-1)\cdots(m-r+1)(N-m)(N-m-1)\cdots(N-m-s+r+1)}{N(N-1)\cdots(N-s+1)},
$$

where e.g. $m_{(r)} = m(m-1)\cdots(m-r+1)$ for $r \geqslant 1$ and $m_{(0)} = 1$. Therefore, the probability of selecting a sample of size $s$ without replacement having $r$ ones and $s-r$ zeros is given by

$$\begin{aligned}
\pi_{r,s-r} &= \binom{s}{r}\frac{m_{(r)}(N-m)_{(s-r)}}{N_{(s)}}, \\
&= \binom{s}{r}\frac{m^r(N-m)^{s-r}}{N^s}\frac{N^s}{N_{(s)}}\left[\frac{m_{(r)}(N-m)_{(s-r)}}{m^r(N-m)^{s-r}}\right], \\
&= \binom{s}{r}p^r(1-p)^{s-r}\frac{N^s}{N_{(s)}}\left[\frac{m\cdots(m-r+1)(N-m)\cdots(N-m-s+r+1)}{m^r(N-m)^{s-r}}\right], \\
&= \binom{s}{r}p^r(1-p)^{s-r}\frac{N^s}{N_{(s)}}\left[\left(1-\frac{1}{m}\right)\cdots\left(1-\frac{r+1}{m}\right).\left(1-\frac{1}{N-m}\right)\cdots\right. \\
&\quad \left.\left(1-\frac{(s+r+1)}{N-m}\right)\right], \\
&= \binom{s}{r}p^r(1-p)^{s-r}\frac{N^s}{N_{(s)}}\left[\prod_{j=0}^{r-1}\left(1-\frac{j}{m}\right)\right]\left[\prod_{k=0}^{s-r-1}\left(1-\frac{k}{N-m}\right)\right], \\
&= \binom{s}{r}p^r(1-p)^{s-r}.\frac{N^s}{N_{(s)}}\left[\prod_{j=0}^{r-1}\left(1-\frac{j}{Np}\right)\right]\left[\prod_{k=0}^{s-r-1}\left(1-\frac{k}{N(1-p)}\right)\right], \qquad (4.3.5)
\end{aligned}$$

using the fact that $m = Np$ and $N-m = N(1-p)$. Also, we define $\prod_{j=0}^{r-1}(1-\frac{j}{Np})$ to be 1 when $r=0$, and $\prod_{k=0}^{s-r-1}(1-\frac{k}{N(1-p)})$ to be 1 when $r=s$.

## 4.4 Some preliminary lemmas

The following lemmas will be used in the proof of Theorem 4.1. We will start with the following definition.

**Definition 4.1**

*Let $A_k = \{1,...,k\}$ and $\underline{\alpha}_k = (\alpha_1,...,\alpha_k)$, and consider a function $f_r(\underline{\alpha}_k)$, where $r \leqslant k$, which calculates the summation of the product of all subsets of components of $\underline{\alpha}_k$ of size $r$. Define*

$$f_r(\underline{\alpha}_k) = \sum_{C\subset A_k:|C|=r}\ \prod_{j\in C}\alpha_j, \qquad (4.4.1)$$

*where the sum is over the $\binom{k}{r}$ distinct subsets of $A_k$ with precisely $r$ elements. In the case $r=0$ it is natural and convenient to define*

$$f_0(\underline{\alpha}_k) = 1. \tag{4.4.2}$$

Moreover, from Definition 4.1, two results will be presented in Lemma 4.1.

**Lemma 4.1**

*Let $A_k = \{1, ..., k\}$ and $\underline{\alpha}_k = (\alpha_1, ..., \alpha_k)$, and consider the function $f_r(\underline{\alpha}_k)$ defined in (4.4.1) and (4.4.2). Then for $r \geqslant 1$,*

$$f_r(\underline{\alpha}_{r+1}) = \left(\prod_{i=1}^{r} \alpha_i\right) + \alpha_{r+1} f_{r-1}(\underline{\alpha}_r), \tag{4.4.3}$$

*and for $i = 1, \cdots, r$,*

$$f_{r-i+1}(\underline{\alpha}_{r+1}) = \alpha_{r+1} f_{r-i}(\underline{\alpha}_r) + f_{r-i+1}(\underline{\alpha}_r). \tag{4.4.4}$$

**Proof:** From (4.4.1) we have

$$\begin{aligned}
f_r(\underline{\alpha}_{r+1}) &= \sum_{C \subset A_{r+1}:|C|=r} \prod_{j \in C} \alpha_j, \\
&= \text{term not involving } \alpha_{r+1} + \text{terms involving } \alpha_{r+1}, \\
&= \left(\prod_{i=1}^{r} \alpha_i\right) + \alpha_{r+1} \sum_{C \subset A_r:|C|=r-1} \prod_{j \in C} \alpha_j, \\
&= \left(\prod_{i=1}^{r} \alpha_i\right) + \alpha_{r+1} f_{r-1}(\underline{\alpha}_r),
\end{aligned}$$

from the definition of $f_{r-1}(\underline{\alpha}_r)$. This establishes (4.4.3).

To prove (4.4.4), we see that

$$\begin{aligned}
f_{r-i+1}(\underline{\alpha}_{r+1}) &= \sum_{C \subset A_{r+1}:|C|=r-i+1} \prod_{j \in C} \alpha_j, \\
&= \text{term involving } \alpha_{r+1} + \text{terms not involving } \alpha_{r+1}, \\
&= \alpha_{r+1} \sum_{C \subset A_r:|C|=r-i} \prod_{j \in C} \alpha_j + \sum_{C \subset A_r:|C|=r-i+1} \prod_{j \in C} \alpha_j, \\
&= \alpha_{r+1} f_{r-i}(\underline{\alpha}_r) + f_{r-i+1}(\underline{\alpha}_r),
\end{aligned}$$

from the definition of $f_{r-i}(\underline{\alpha}_r)$ and $f_{r-i+1}(\underline{\alpha}_r)$; see (4.4.1). □

The following lemma, which builds on Lemma 4.1, will be useful in Part I of the proof the Theorem 4.1 in Section 4.5.

**Lemma 4.2**

*With $A_t = \{1, ..., t\}$ and $\underline{\alpha}_t = (\alpha_1, ..., \alpha_t)$ defined as before, and $z$ any real number, we have*

$$\prod_{i=1}^{t}(\alpha_i - z) = \prod_{i=1}^{t}\alpha_i + \sum_{r=1}^{t}(-1)^r z^r f_{t-r}(\underline{\alpha}_t), \tag{4.4.5}$$

*where $f_{t-r}(\underline{\alpha}_k)$ is defined in (4.4.1).*

**Proof:** To proof the formula (4.4.5), we use the induction technique. The lemma certainly holds when $k = t = 1$ provided we define $f_0(\alpha_t) = 1$. When $k = t = 2$,

$$\begin{aligned}
\prod_{i=1}^{2}(\alpha_i - z) &= (\alpha_1 - z)(\alpha_2 - z), \\
&= \alpha_1\alpha_2 - z(\alpha_1 + \alpha_2) + z^2, \\
&= \prod_{i=1}^{2}\alpha_i - z f_1(\underline{\alpha}_2) + z^2 f_0(\underline{\alpha}_2) \\
&= \prod_{i=1}^{2}\alpha_i + \sum_{r=1}^{2}(-1)^r z^r f_{2-r}(\underline{\alpha}_2),
\end{aligned}$$

as required. So, (4.4.5) holds for $k = 2$. Assume now that (4.4.5) holds for $k = t$, and consider $k = t + 1$. Then

$$\begin{aligned}
\prod_{i=1}^{t+1}(\alpha_i - z) &= (\alpha_{t+1} - z)\prod_{i=1}^{t}(\alpha_i - z), \\
&= (\alpha_{t+1} - z)\left[\prod_{i=1}^{t}\alpha_i + \sum_{i=1}^{t}(-1)^i z^i f_{t-i}(\underline{\alpha}_t)\right],
\end{aligned} \tag{4.4.6}$$

using the induction assumption. Expanding the RHS of (4.4.6),

$$
\begin{aligned}
RHS &= \alpha_{t+1}\prod_{i=1}^{t}\alpha_i - z\prod_{i=1}^{t}\alpha_i \\
&\quad +\alpha_{t+1}\sum_{i=1}^{t}(-1)^i z^i f_{t-i}(\underline{\alpha}_t) - z\sum_{i=1}^{t}(-1)^i z^i f_{t-i}(\underline{\alpha}_t) \\
&= \prod_{i=1}^{t+1}\alpha_i - z\Big\{\prod_{i=1}^{t}\alpha_i + \alpha_{t+1}f_{t-1}(\underline{\alpha}_t)\Big\} + \\
&\quad \sum_{i=2}^{t}(-1)^i z^i\Big\{f_{t-i}(\underline{\alpha}_t)\alpha_{t+1} + f_{t-i+1}(\underline{\alpha}_t)\Big\} + (-1)^{t+1}z^{t+1} \qquad (4.4.7)\\
&= \prod_{i=1}^{t+1}\alpha_i - z f_t(\underline{\alpha}_{t+1}) \\
&\quad +\sum_{i=2}^{t}(-1)^i z^i f_{t-i+1}(\underline{\alpha}_{t+1}) + (-1)^{t+1}z^{t+1}f_0(\underline{\alpha}_{t+1}), \qquad (4.4.8)\\
&= \prod_{i=1}^{t+1}\alpha_i + \sum_{i=1}^{t+1}(-1)^i z^i f_{t-i+1}(\underline{\alpha}_{t+1}),
\end{aligned}
$$

where in moving from (4.4.7) to (4.4.8), we have made use of (4.4.3) and made multiple use of (4.4.4). Then, (4.4.5) holds for $k = t+1$, and by the principle of induction it therefore holds for all positive integers $k$. □

The following result is well-known but for convenience we provide a statement and proof.

**Lemma 4.3**

*Let $t$ and $\alpha$ be integers such that $t > 0$ and $0 \leqslant \alpha < t$. Then*

$$\sum_{r=0}^{t}(-1)^r\binom{t}{r}r^{\alpha} = 0. \tag{4.4.9}$$

**Proof:** We prove the formula (4.4.9) using the induction technique, with the beginning step when $t = 2$ and $\alpha = 0, 1, 2$. For $\alpha = 0$, the LHS of (4.4.9) will be

$$
\begin{aligned}
\sum_{r=0}^{2}(-1)^r\binom{2}{r} &= (-1)^0\binom{2}{0} + (-1)^1\binom{2}{1} + (-1)^2\binom{2}{2} \\
&= 1 - 2 + 1 \\
&= 0.
\end{aligned}
$$

For $\alpha = 1$, the LHS of (4.4.9) will be

$$\begin{aligned}\sum_{r=0}^{2}(-1)^r\binom{2}{r}r &= (-1)^0\binom{2}{0}0+(-1)^1\binom{2}{1}+(-1)^2\binom{2}{2}2\\ &= 0-2+2\\ &= 0,\end{aligned}$$

and for $\alpha = 2$, the LHS of (4.4.9) will be

$$\begin{aligned}\sum_{r=0}^{2}(-1)^r\binom{2}{r}r^2 &= (-1)^0\binom{2}{0}0^2+(-1)^1\binom{2}{1}1^2+(-1)^2\binom{2}{2}2^2\\ &= 1-2+4\\ &\neq 0.\end{aligned}$$

So, (4.4.9) holds for $t = 2$ and $\alpha = 0$ and 1.
Assume, for $t = k$ and $\alpha = 0, ..., k-1$, (4.4.9) holds. Now let $t = k+1$ and consider $\alpha = 0, ..., k$. Then for $\alpha = 0$, the LHS of (4.4.9) will be

$$\begin{aligned}\sum_{r=0}^{k+1}(-1)^r\binom{k+1}{r} &= (1-1)^{k+1},\\ &= 0,\end{aligned}$$

from the binomial theorem. Fix $\alpha \in \{1, ..., k\}$. The LHS of (4.4.9) will be

$$\begin{aligned}\sum_{s=0}^{k+1}(-1)^s\binom{k+1}{s}s^\alpha &= \sum_{s=1}^{k+1}(-1)^s\binom{k+1}{s}s^\alpha\\ &= \sum_{s=1}^{k+1}(-1)^s\frac{(k+1)!}{s!(k+1-s)!}s.s^{\alpha-1},\\ &= \sum_{s=1}^{k+1}(-1)^s\frac{(k+1)!}{(s-1)!(k+1-s)!}s^{\alpha-1}.\end{aligned}$$

Let $r = s - 1$. Then

$$\begin{aligned}
\sum_{r=0}^{k}(-1)^{r+1}\binom{k+1}{r+1}(r+1)^{\alpha} &= (k+1)\sum_{r=0}^{k}(-1)^{r+1}\frac{(k)!}{(r)!(k-r)!}(r+1)^{\alpha-1}, \\
&= -(k+1)\sum_{r=0}^{k}(-1)^{r}\binom{k}{r}(r+1)^{\alpha-1}, \\
&= -(k+1)\sum_{r=0}^{k}(-1)^{r}\binom{k}{r}\sum_{j=0}^{\alpha-1}r^{j}1^{(\alpha-1-j)}\binom{\alpha-1}{j}, \\
&= -(k+1)\sum_{j=0}^{\alpha-1}\binom{\alpha-1}{j}\sum_{r=0}^{k}(-1)^{r}\binom{k}{r}r^{j} \\
&= 0,
\end{aligned}$$

because $\sum_{r=0}^{k}(-1)^{r}\binom{k}{r}r^{j} = 0$ for $j = 1, 2, ..., k-1$, by using the inductive assumption. Thus, (4.4.9) holds for $t = k+1$. Therefore, (4.4.9) is true for $0 \leqslant \alpha < t$ and $t = 1, 2, ....$ □

Define the quantity $T_\alpha(r)$ as follows

$$T_\alpha(r) = \begin{cases} 1 & \alpha = 0,\ r \geqslant 0; \\ 0 & \alpha > r; \\ \sum_{1 \leqslant i_1 < \cdots < i_\alpha \leqslant r} i_1 \cdots i_\alpha & 1 \leqslant \alpha \leqslant r. \end{cases} \tag{4.4.10}$$

The quantity $T_\alpha(r)$ will play an important role in the proof of Theorem 4.1. The key property of $T_\alpha(r)$ we need is stated and proved in Lemma 4.4

**Lemma 4.4**

*Let $T_\alpha(r)$ be as defined in (4.4.10). Then for all integers satisfying $0 < \alpha \leqslant r$, $T_\alpha(r)$ is a polynomial in $r$ of degree $2\alpha$.*

**Remark 4.3**

*The quantity $T_\alpha(r)$ can be written as $T_\alpha(r) = f_\alpha(\underline{\delta}_r)$, and $\underline{\delta}_r = (1, ..., r)$, where $f_\alpha$ is defined in (4.4.1). However, it is more convenient below to use separate notation.*

**Proof:** First of all, we calculate $T_\alpha(r)$ in the cases $\alpha = 1$ and $\alpha = 2$, assuming $r \geqslant \alpha$; otherwise, $T_\alpha(r)$=0, from definition (4.4.10).

*Case*: $\alpha = 1$, $r \geqslant \alpha$. Here

$$T_1(r) = \sum_{i=1}^{r} i = \frac{1}{2}r(r+1),$$

which is a polynomial in $r$ of degree $2\alpha = 2$.

*Case*: $\alpha = 2$, $r \geqslant \alpha$. In this case, using the standard result

$$\sum_{i=1}^{r} i^2 = \frac{1}{6}r(r+1)(2r+1), \tag{4.4.11}$$

we have

$$\begin{aligned} T_2(r) &= \sum_{1\leqslant i<j\leqslant r} ij, \\ &= \frac{1}{2}\sum_{1\leqslant i\neq j\leqslant r} ij, & (4.4.12)\\ &= \frac{1}{2}\Bigg[\sum_{1\leqslant i,j\leqslant r} ij - \sum_{i=1}^{r} i^2\Bigg], & (4.4.13)\\ &= \frac{1}{2}\Bigg[\bigg\{\frac{r(r+1)}{2}\bigg\}^2 - \frac{1}{6}r(r+1)(2r+1)\Bigg], \\ &= \frac{r(r+1)}{24}\Bigg[3r^2+3r-4r-2\Bigg], \\ &= \frac{r(r+1)}{24}\Bigg[(r-1)(3r+2)\Bigg], \end{aligned}$$

which is a polynomial in $r$ of degree $2\alpha = 4$. In (4.4.12), the summation is over all $i \neq j$, including $i > j$ and $j > i$; and the first summation in (4.4.13) is over all $i, j$ with no restriction.

We now consider the following inductive hypothesis:
$\mathcal{P}_\alpha$: $T_\alpha(r)$ is polynomial in $r$ of degree $2\alpha$ when $r \geqslant \alpha$.

We already know from above that $\mathcal{P}_1$ and $\mathcal{P}_2$ are true. We shall now show that if $\alpha \geqslant 0$ and $\mathcal{P}_\alpha$ is true then $\mathcal{P}_{\alpha+1}$ is also true, i.e.
$\mathcal{P}_{\alpha+1}$: $T_{\alpha+1}(r)$ is a polynomial in $r$ of degree $2(\alpha+1)$ when $r \geqslant \alpha+1$.

Consider the general case

$$T_\alpha(r) = \sum_{1 \leqslant i_1 < \ldots < i_\alpha \leqslant r} i_1 \cdots i_\alpha.$$

There are two types of terms in the sum: those for which $i_\alpha = r$, and those for which $i_\alpha < r$. The sum of all the terms of the first type is given by

$$\begin{aligned} \sum_{1 \leqslant i_1 < \ldots < i_\alpha = r} i_1 \cdots i_{\alpha-1} r = r & \sum_{1 \leqslant i_1 < \ldots < i_{\alpha-1} \leqslant r-1} i_1 \cdots i_{\alpha-1}, \\ = r & \;\; T_{\alpha-1}(r-1), \end{aligned} \tag{4.4.14}$$

by definition of $T_{\alpha-1}(r-1)$; see (4.4.10).

The sum of all the terms of the second type, with $i_\alpha < r$, is given by

$$\sum_{1 \leqslant i_1 < \ldots < i_\alpha \leqslant r-1} i_1 \cdots i_\alpha = T_\alpha(r-1), \tag{4.4.15}$$

again by definition. Therefore combining (4.4.14) and (4.4.15), we have the identity

$$T_\alpha(r) = T_\alpha(r-1) + rT_{\alpha-1}(r-1). \tag{4.4.16}$$

It will be slightly more convenient to work with the identity

$$T_{\alpha+1}(r+1) = T_{\alpha+1}(r) + (r+1)T_\alpha(r), \tag{4.4.17}$$

obtained by replacing $\alpha$ and $r$ in (4.4.16) by $\alpha + 1$ and $r + 1$, respectively.

Now

$$T_{\alpha+1}(r+1) = \sum_{j=\alpha}^{r} \Big\{ T_{\alpha+1}(j+1) - T_{\alpha+1}(j) \Big\}$$

due to cancelation and the fact that $T_{\alpha+1}(\alpha) = 0$ from (4.4.10). Therefore, using (4.4.17),

$$T_{\alpha+1}(r+1) = \sum_{j=\alpha}^{r} (j+1) T_\alpha(j). \tag{4.4.18}$$

We now use the inductive hypothesis $\mathcal{P}_\alpha$ and write

$$T_\alpha(j) = \sum_{k=0}^{2\alpha} A_k^{[\alpha]} \; j^k,$$

i.e. we may write $T_\alpha(j)$ as a polynomial in $j$ of degree $2\alpha$, where the coefficients $A_k^{[\alpha]}$ depend on $\alpha$ but not on $j$.

Therefore, equating coefficients of powers of $j$,

$$\begin{aligned} T_{\alpha+1}(r+1) &= \sum_{j=\alpha}^{r} (j+1) T_\alpha(j), \\ &= \sum_{j=\alpha}^{r} (j+1) \sum_{k=0}^{2\alpha} A_k^{[\alpha]} \; j^k, \\ &= \sum_{k=0}^{2\alpha+1} B_k^{[\alpha]} \sum_{j=\alpha}^{r} \; j^k, \end{aligned} \tag{4.4.19}$$

where

$$B_k^{[\alpha]} = \begin{cases} A_0^{[\alpha]} & k = 0; \\ A_k^{[\alpha]} + A_{k-1}^{[\alpha]} & 1 \leqslant k \leqslant 2\alpha; \\ A_{2\alpha}^{[\alpha]} & k = 2\alpha + 1. \end{cases}$$

Now

$$\begin{aligned} \sum_{j=\alpha}^{r} j^k &= \sum_{j=1}^{r} j^k - \sum_{j=1}^{\alpha-1} j^k, \\ &= \sum_{j=1}^{r} j^k - g_k(\alpha), \end{aligned}$$

where $g_k(\alpha)$ depends only on $k$ and $\alpha$, and not $j$.

It is well known that $\sum_{j=1}^{r} j^k$ is a polynomial in $r$ of degree $k+1$. This follows from the result that

$$\sum_{j=1}^{r} j^k = \frac{1}{k+1} r^{k+1} + \frac{1}{2} r^k + \frac{1}{k+1} \sum_{i=2}^{k} \binom{k+1}{i} B_i r^{k+1-i} \tag{4.4.20}$$

where $B_i$ are the Bernoulli numbers; see Conway and Guy (1996), page 106.

Since the largest power of $j$ in (4.4.19) is $(2\alpha+1)$, it follows from (4.4.20) that (4.4.19) is a polynomial in $r$ of degree $2\alpha+1+1 = 2(\alpha+1)$.

Finally, we want to show $T_{\alpha+1}(r+1)$ is a polynomial of degree $2(\alpha+1)$ in $r+1$, not $r$. To see this is the case, write

$$\begin{aligned} T_{\alpha+1}(r+1) &= \sum_{k=0}^{2(\alpha+1)} C_k^{[\alpha+1]} r^k, \\ &= \sum_{k=0}^{2(\alpha+1)} C_k^{[\alpha+1]} (r+1-1)^k. \end{aligned} \tag{4.4.21}$$

Expanding and using the binomial theorem, we obtain

$$(r+1-1)^k = \sum_{l=0}^{k} \binom{k}{l} (r+1)^l (-1)^{k-l},$$

and after substitution in (4.4.21), we get

$$\begin{aligned} T_{\alpha+1}(r+1) &= \sum_{k=0}^{2(\alpha+1)} C_k^{[\alpha+1]} \sum_{l=0}^{k} \binom{k}{l} (r+1)^l (-1)^{k-l} \\ &= \sum_{l=0}^{2(\alpha+1)} (r+1)^l \sum_{k=l}^{2(\alpha+1)} (-1)^{k-l} C_k^{[\alpha+1]} \binom{k}{l}, \\ &= \sum_{k=0}^{2(\alpha+1)} D_k^{[\alpha+1]} (r+1)^k. \end{aligned}$$

Then, by replacing $r+1$ by $r$, we have

$$T_{\alpha+1}(r) = \sum_{k=0}^{2(\alpha+1)} D_k^{[\alpha+1]} r^k,$$

for all integers $r \geqslant \alpha+1$. Therefore, $\mathcal{P}_{\alpha+1}$ is implied by $\mathcal{P}_\alpha$ and the lemma is proved.

## 4.5 Proof of the General Conditional Moments Theorem

The proof is a quite lengthy, so we split it into two parts, Part I and Part II. In Part I, it will be shown that proving the results in the general case reduces to proving the result in the case

$$r_1 = ... = r_q = 1, \; q = t, \tag{4.5.1}$$

where $t$ is defined in (4.2.2). In other words, $r_u = 1$ for all $u = 1, \cdots, q$, and for no $u$ is $r_u > 1$. Then, in Part II, the result will be proved under condition (4.5.1).

**Part I:**

Define $A = \{u : r_u = 1\}$ and $B = \{u : r_u > 1\}$. Then $A \bigcap B = \phi$, the empty set, and $A \bigcup B = \{1, ..., q\}$. Also, define $y_A^{(n)} = \{y_{i_u,j_u} : u \in A\}$ and $y_B^{(n)} = \{y_{i_u,j_u} : u \in B\}$. In this notation,

$$\prod_{u=1}^{q} (y_{i_u,j_u} - p^{(n)})^{r_u} = \left[\prod_{u \in A} (y_{i_u,j_u} - p^{(n)})\right] \left[\prod_{u \in B} (y_{i_u,j_u} - p^{(n)})^{r_u}\right]. \tag{4.5.2}$$

The situation considered here relates to formula (2.4.8) when the following substitution are made:

$$\begin{aligned} X =& y_B^{(n)}, \quad Y = y_A^{(n)}, \quad Z = m^{(n)}, \\ f(X) =& \prod_{u\in B}(y_{i_u,j_u} - p^{(n)})^{r_u}, \\ g(Y) =& \prod_{u\in A}(y_{i_u,j_u} - p^{(n)}). \end{aligned} \tag{4.5.3}$$

Since the sampling here is equivalent to finite population sampling, all the expectations indicated in (2.4.8) and (4.5.3) can be expressed as finite sums. Moreover, since $|f(X)|$ in (4.5.3) is bounded above by1, the theorem will follow if it can be proved that

$$E\left[\prod_{u\in A}(y_{i_u,j_u} - p^{(n)})\middle| m^{(n)}, y_B^{(n)}\right] = O(N^{-\lfloor\frac{t+1}{2}\rfloor}), \tag{4.5.4}$$

for all possible $y_B^{(n)}$, where $N = N^{(n)}$. There are only a finite number of possible outcomes of the vector $y_B^{(n)}$. In particular, for $n$ sufficiently large there are $2^{q-t}$ possible values. A further point to note is that, due to the nature of the sampling, i.e. simple random sampling without replacement, conditioning on $m^{(n)}$ and $y_B^{(n)}$ in (4.5.4) is equivalent to conditioning on $\widetilde{m}^{(n)} = m^{(n)} - \sum_{u\in B} y_{i_u,j_u}^{(n)}$. This simplification is used below.

Consideration of the LHS of (4.5.4) leads to the following conclusions. The conditional distribution of $y_A^{(n)}$ given $y_B^{(n)}$ and $m^{(n)}$ corresponds to simple random sampling without replacement with sample size reduced from $N^{(n)}$ to $\widetilde{N}^{(n)} = N^{(n)} - (q-t)$; the number of ones in the finite population is reduced from $m^{(n)}$ to $\widetilde{m}^{(n)} = m - \sum_{u\in B} y_{i_u,j_u}^{(n)}$; and the number of zeros in the finite population is reduced from $N^{(n)} - m^{(n)}$ to $\widetilde{N}^{(n)} - \widetilde{m}^{(n)}$.

Let us now suppose that the theorem holds in all cases in which condition (4.5.1) is satisfied. Then, in view of comments in the previous paragraph, it can be concluded that

$$E\left[\prod_{u\in A}(y_{i_u,j_u} - \widetilde{p}^{(n)})\middle| \widetilde{m}^{(n)}\right] = O(\widetilde{N}^{-\lfloor\frac{t+1}{2}\rfloor}), \tag{4.5.5}$$

where $\widetilde{p}^{(n)} = \frac{\widetilde{m}^{(n)}}{\widetilde{N}^{(n)}}$ and $\widetilde{N} = \widetilde{N}^{(n)}$. Now $\widetilde{N}^{(n)}/N^{(n)}$ converges to 1 as $n \to \infty$ because $\widetilde{N}^{(n)} = N^{(n)} - (q-t)$ and $q$ and $t$ are fixed, so we can replace the RHS of (4.5.5) by $O(N^{-\lfloor\frac{t+1}{2}\rfloor})$.

So (4.5.4) will follow from (4.5.5) if it can be shown that

$$E\left[\prod_{u\in A}(y^{(n)}_{i_u,j_u} - \widetilde{p}^{(n)})\middle|\widetilde{m}^{(n)}\right] = E\left[\prod_{u\in A}(y^{(n)}_{i_u,j_u} - p^{(n)})\middle|\widetilde{m}^{(n)}\right] + O(\widetilde{N}^{-\lfloor\frac{t+1}{2}\rfloor}).$$

But $0 \leqslant \sum_{u\in B} y^{(n)}_{i_u,j_u} \leqslant q-t$, so

$$\begin{aligned}
\widetilde{p}^{(n)} &= \frac{\widetilde{m}^{(n)}}{\widetilde{N}^{(n)}} \\
&= \frac{m^{(n)} - \sum_{u\in B} y_{i_u,j_u}}{N^{(n)} - (q-t)} \\
&= \frac{\frac{m^{(n)}}{N^{(n)}} - \frac{\sum_{u\in B} y_{i_u,j_u}}{N^{(n)}}}{1 - \frac{q-t}{N^{(n)}}} \\
&= \frac{p^{(n)} - O(\frac{1}{N^{(n)}})}{1 + O(\frac{1}{N^{(n)}})} \\
&= p^{(n)} + O(N^{-1})
\end{aligned}$$

uniformly over $y^{(n)}_B$. Moreover, using Lemma (4.2),

$$\begin{aligned}
\prod_{u\in A}(y^{(n)}_{i_u,j_u} - \widetilde{p}^{(n)}) &= \prod_{u\in A}\{(y^{(n)}_{i_u,j_u} - p^{(n)}) - (\widetilde{p}^{(n)} - p^{(n)})\} \\
&= \prod_{u\in A}(y^{(n)}_{i_u,j_u} - p^{(n)}) \\
&\quad + \sum_{r=1}^{t}(-1)^r(\widetilde{p}^{(n)} - p^{(n)})^r \sum_{C\subset A:|C|=t-r}\ \prod_{u\in C}(y^{(n)}_{i_u,j_u} - p^{(n)}),
\end{aligned} \tag{4.5.6}$$

where the sum in the final line is over all subsets $C$ of $A$ with $t-r$ elements. Therefore, taking expectation in (4.5.6), conditional on $\widetilde{m}^{(n)}$, we find that

$$
\begin{aligned}
& E\left[\prod_{u\in A}(y^{(n)}_{i_u,j_u}-\widetilde{p}^{(n)})\middle|\widetilde{m}^{(n)}\right] \\
= \; & E\left[\prod_{u\in A}(y^{(n)}_{i_u,j_u}-p^{(n)})\middle|\widetilde{m}^{(n)}\right]+\sum_{r=1}^{t}(-1)^r(\widetilde{p}^{(n)}-p^{(n)})^r \\
& \times E\left[\sum_{C\subset A:|C|=t-r}\;\prod_{u\in C}(y^{(n)}_{i_u,j_u}-p^{(n)})\middle|\widetilde{m}^{(n)}\right] \\
= \; & E\left[\prod_{u\in A}(y^{(n)}_{i_u,j_u}-p^{(n)})|\widetilde{m}^{(n)}\right]+ \\
& \sum_{r=1}^{t}O(N^{-r})\sum_{C\subset A:|C|=t-r.}E\left[\prod_{u\in C}(y^{(n)}_{i_u,j_u}-p^{(n)})\middle|\widetilde{m}^{(n)}\right] \\
= \; & E\left[\prod_{u\in A}(y^{(n)}_{i_u,j_u}-p^{(n)})\middle|\widetilde{m}^{(n)}\right]+\sum_{r=1}^{t}O(N^{-r})O(N^{-\lfloor\frac{t-r+1}{2}\rfloor}),
\end{aligned}
\tag{4.5.7}
$$

assuming in the final line that the theorem holds in all cases satisfying (4.5.1). But for all integers satisfying $1\leqslant r\leqslant t$,

$$
r+\left\lfloor\frac{t-r+1}{2}\right\rfloor\geqslant\left\lfloor\frac{t+1}{2}\right\rfloor \tag{4.5.8}
$$

Consequently, Part I follows from (4.5.7).

**Part II:**

For the remainder of the proof it is assumed that condition (4.5.1) holds, i.e. $r_1=\cdots=r_q=1$ and $q=t$.

Therefore, using (4.3.5), the probability of selecting a sequence of length $t$ having $r$ ones and $t-r$ zeros, $\pi_{r,t-r}$, we get

$$\begin{aligned}
& E\left[\prod_{u=1}^{t}(y_{i_u,j_u}-p)\middle| m\right] \\
= & \sum_{r=0}^{t}\pi_{r,t-r}(1-p)^r(-p)^{t-r}, \\
= & \sum_{r=0}^{t}\binom{t}{r}\frac{N^t}{N_{(t)}}p^r(1-p)^{t-r}\left[\prod_{j=0}^{r-1}\left(1-\frac{j}{Np}\right)\right]\times\left[\prod_{k=0}^{t-r-1}\left(1-\frac{k}{N(1-p)}\right)\right] \\
& (1-p)^r p^{t-r}(-1)^{t-r}, \\
= & (-1)^t\frac{N^t}{N_{(t)}}p^t(1-p)^t\times \\
& \sum_{r=0}^{t}(-1)^r\binom{t}{r}\left[\prod_{j=0}^{r-1}\left(1-\frac{j}{Np}\right)\right]\left[\prod_{k=0}^{t-r-1}\left(1-\frac{k}{N(1-p)}\right)\right]. \qquad (4.5.9)
\end{aligned}$$

Then, using (4.4.10),

$$\prod_{j=0}^{r-1}\left(1-\frac{j}{Np}\right)=\sum_{\alpha=0}^{r-1}(-1)^{\alpha}\frac{1}{(Np)^{\alpha}}T_{\alpha}(r-1), \qquad (4.5.10)$$

and

$$\prod_{k=0}^{t-r-1}\left(1-\frac{k}{N(1-p)}\right)=\sum_{\beta=0}^{t-r-1}(-1)^{\beta}\frac{1}{(N(1-p))^{\beta}}T_{\beta}(t-r-1), \qquad (4.5.11)$$

where $T_{\alpha}(r)$ is defined in (4.4.10). Consequently, using (4.5.10) and (4.5.11),

$$
\begin{aligned}
&\left[\prod_{j=0}^{r-1}\left(1-\frac{j}{Np}\right)\right]\left[\prod_{k=0}^{t-r-1}\left(1-\frac{k}{N(1-p)}\right)\right] \\
&= \left[\sum_{\alpha=0}^{r-1}(-1)^{\alpha}\frac{1}{(Np)^{\alpha}}T_{\alpha}(r-1)\right]\left[\sum_{\beta=0}^{t-r-1}(-1)^{\beta}\frac{1}{(N(1-p))^{\beta}}T_{\beta}(t-r-1)\right] \\
&= \sum_{\alpha=0}^{r-1}\sum_{\beta=0}^{t-r-1}(-1)^{\alpha+\beta}\frac{1}{(Np)^{\alpha}(N(1-p))^{\beta}}T_{\alpha}(r-1)T_{\beta}(t-r-1) \\
&= \sum_{\gamma=0}^{t-2}(-1)^{\gamma}\frac{1}{N^{\gamma}}\sum_{\alpha=max(0,\gamma-t+r+1)}^{\min(r-1,\gamma)}\frac{1}{p^{\alpha}}\frac{1}{(1-p)^{\gamma-\alpha}}T_{\alpha}(r-1)T_{\gamma-\alpha}(t-r-1) \qquad (4.5.12) \\
&= \sum_{\gamma=0}^{t-2}(-1)^{\gamma}\frac{1}{N^{\gamma}}\sum_{\alpha=0}^{\gamma}\frac{1}{p^{\alpha}}\frac{1}{(1-p)^{\gamma-\alpha}}T_{\alpha}(r-1)T_{\gamma-\alpha}(t-r-1) \qquad (4.5.13)
\end{aligned}
$$

We can change $\min(r-1,\gamma)$, the upper limit of summation in (4.5.12) to $\gamma$, the upper limit of summation in (4.5.13), because $T_{\alpha}(r-1)=0$ if $(r-1)<\alpha\leqslant\gamma$, from the definition of $T_{\alpha}(r)$ in (4.4.10). Also, We can change $max(0,\gamma-t+r+1)$, the lower limit of summation in (4.5.12) to 0, the lower limit of summation in (4.5.13). Then by substituting (4.5.13) into (4.5.9) and writing

$$C_N(p,t)=(-1)^t\frac{N^t}{N_{(t)}}p^t(1-p)^t,$$

we obtain

$$
\begin{aligned}
E\left[\prod_{u=1}^{t}(y_{i_u,j_u}-p)\middle|m\right] &= C_N(p,t)\sum_{r=0}^{t}(-1)^r\binom{t}{r}\sum_{\gamma=0}^{t-2}(-1)^{\gamma}\frac{1}{N^{\gamma}} \\
&\quad \sum_{\alpha=0}^{\gamma}\frac{1}{p^{\alpha}}\frac{1}{(1-p)^{\gamma-\alpha}}T_{\alpha}(r-1)T_{\gamma-\alpha}(t-r-1), \\
&= C_N(p,t)\sum_{\gamma=0}^{t-2}(-1)^{\gamma}\frac{1}{N^{\gamma}}\times \\
&\quad \sum_{\alpha=0}^{\gamma}\frac{1}{p^{\alpha}}\frac{1}{(1-p)^{\gamma-\alpha}}\sum_{r=0}^{t}(-1)^r\binom{t}{r}T_{\alpha}(r-1)T_{\gamma-\alpha}(t-r-1).
\end{aligned} \qquad (4.5.14)
$$

Using Lemma 4.4, for fixed $t$ and variable $r$,

$$T_\alpha(r-1)T_{\gamma-\alpha}(t-r-1)$$

is a polynomial of degree $2\alpha + 2(\gamma-\alpha) = 2\gamma$. Therefore, from Lemma 4.3,

$$\sum_{r=0}^{t}(-1)^r \binom{t}{r} T_\alpha(r-1)T_{\gamma-\alpha}(t-r-1) = 0,$$

for all $\gamma$ such that $2\gamma < t$. However, for $\gamma \geqslant t/2$, these equations are not equal 0. Therefore, the leading term in (4.5.14) is $\frac{1}{N^{\gamma^*}}$ where $\gamma^*$ is the smallest $\gamma$ such that $2\gamma^* \geqslant t$. So $\gamma^* = \frac{t}{2}$ if $t$ is even and $\gamma^* = \frac{t+1}{2}$ if $t$ is odd; i.e. $\gamma^* = \lfloor \frac{t+1}{2} \rfloor$, the integer part of $\frac{t+1}{2}$. The proof is now complete. □

## 4.6 Preliminaries for $\overline{C}_2$ and $\overline{C}_3$

Rather than work with $\overline{T}_2$ and $\overline{T}_3$, the 2-star and triangle densities, respectively, it will be more convenience to work with the equivalent variables $\overline{C}_2$ and $\overline{C}_3$ defined below:

$$\overline{C}_2 = \frac{2}{n(n-1)(n-2)} \sum_{i=1}^{n} \sum_{i \neq j < k \neq i} (y_{ij}-p)(y_{ik}-p) \tag{4.6.1}$$

$$\overline{C}_3 = \frac{6}{n(n-1)(n-2)} \sum_{1 \leqslant i<j<k \leqslant n} (y_{ij}-p)(y_{jk}-p)(y_{ki}-p), \tag{4.6.2}$$

where $p = N^{-1}\sum_{1\leqslant i<j\leqslant n} y_{ij}$.

**Lemma 4.5**
*If $p = N^{-1}\sum_{1\leqslant i<j\leqslant n} y_{ij}$ then $\overline{C}_2 = \overline{T}_2$, where $\overline{C}_2$ is defined in (4.6.1) and $\overline{T}_2$ is defined in (3.2.3).*

**Proof:** Starting from the definition of $\overline{C}_2$ in (4.6.1), we have

$$\begin{aligned}\overline{C}_2 &= \frac{2}{n(n-1)(n-2)} \sum_{i=1}^{n} \sum_{i \neq j < k \neq i} (y_{ij}-p)(y_{ik}-p) \\ &= \frac{2}{n(n-1)(n-2)} \sum_{i=1}^{n} \sum_{i \neq j < k \neq i} (y_{ij}y_{ik} + p^2 - p(y_{ij}+y_{ik}))\end{aligned} \tag{4.6.3}$$

Now

$$\begin{aligned}\frac{2}{n(n-1)(n-2)}\sum_{i=1}^{n}\sum_{i\neq j<k\neq i}(y_{ij}+y_{ik}) &= \frac{1}{2}\frac{2}{n(n-1)(n-2)}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i,j}(y_{ij}+y_{ik}) \\ &= \frac{2}{n(n-1)(n-2)}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i,j}y_{ij} \\ &= \frac{2}{n(n-1)(n-2)}\sum_{i=1}^{n}\sum_{j\neq i}(n-2)y_{ij} \\ &= \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}y_{ij} \\ &= 2.\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{i<j}y_{ij} \\ &= 2p.\end{aligned} \tag{4.6.4}$$

Consequently, using (4.6.3) and (4.6.4), we have

$$\begin{aligned}\overline{C}_2 &= \left(\frac{2}{n(n-1)(n-2)}\sum_{i=1}^{n}\sum_{i\neq j<k\neq i}(y_{ij}y_{ik}+p^2)\right)-2p^2 \\ &= \frac{2}{n(n-1)(n-2)}\sum_{i=1}^{n}\sum_{i\neq j<k\neq i}(y_{ij}y_{ik}+p^2-2p^2) \\ &= \frac{2}{n(n-1)(n-2)}\sum_{i=1}^{n}\sum_{i\neq j<k\neq i}(y_{ij}y_{ik}-p^2) \\ &= \overline{T}_2,\end{aligned}$$

as required. □

**Lemma 4.6**
*When $p = N^{-1}\sum_{1\leqslant i<j\leqslant n} y_{ij}$ then $\overline{C}_3 = \overline{T}_3 - 3p\overline{T}_2$, where $\overline{C}_3$, $\overline{T}_2$ and $\overline{T}_3$ are defined in (4.6.2), (3.2.3) and (3.2.4) respectively.*

**Proof:** Starting from the definition of $\overline{C}_3$ in (4.6.2), we have

$$\begin{aligned}
\overline{C}_3 &= \frac{6}{n(n-1)(n-2)} \sum_{1\leqslant i<j<k\leqslant n} (y_{ij}-p)(y_{jk}-p)(y_{ki}-p) \\
&= \frac{6}{n(n-1)(n-2)} \\
&\quad \sum_{1\leqslant i<j<k\leqslant n} \left\{ y_{ij}y_{jk}y_{ki} - p^3 - p(y_{ij}y_{ik}+y_{ij}y_{kj}+y_{jk}y_{ki}) + p^2(y_{ij}+y_{jk}+y_{ki}) \right\} \\
&= \frac{6}{n(n-1)(n-2)} \sum_{1\leqslant i<j<k\leqslant n} (y_{ij}y_{jk}y_{ki} - p^3) \\
&\quad - \frac{p}{n(n-1)(n-2)} \sum_{i=1}^{n} \sum_{j\neq i} \sum_{k\neq i,j} (y_{ij}y_{ik}+y_{ij}y_{kj}+y_{jk}y_{ki}) \\
&\quad + \frac{p^2}{n(n-1)(n-2)} \sum_{i=1}^{n} \sum_{j\neq i} \sum_{k\neq i,j} (y_{ij}+y_{jk}+y_{ki}). \qquad (4.6.5)
\end{aligned}$$

In the two lines above we used the fact that

$$\sum_{1\leqslant i<j<k\leqslant n} (y_{ij}y_{ik}+y_{ij}y_{kj}+y_{jk}y_{ki}) = \frac{1}{6} \sum_{i=1}^{n} \sum_{j\neq i} \sum_{k\neq i,j} (y_{ij}y_{ik}+y_{ij}y_{kj}+y_{jk}y_{ki}),$$

and

$$\sum_{1\leqslant i<j<k\leqslant n} (y_{ij}+y_{jk}+y_{ki}) = \frac{1}{6} \sum_{i=1}^{n} \sum_{j\neq i} \sum_{k\neq i,j} (y_{ij}+y_{jk}+y_{ki}).$$

Since the adjacency matrix is symmetric, then (4.6.5) becomes

$$\begin{aligned}
\overline{C}_3 &= \frac{6}{n(n-1)(n-2)} \sum_{1\leqslant i<j<k\leqslant n} (y_{ij}y_{jk}y_{ki} - p^3) \\
&\quad - \frac{3p}{n(n-1)(n-2)} \sum_{i=1}^{n} \sum_{j\neq i} \sum_{k\neq i,j} y_{ij}y_{ik} \\
&\quad + \frac{3p^2}{n(n-1)(n-2)} \sum_{i=1}^{n} \sum_{j\neq i} (n-2) y_{ij} \\
&= \overline{T}_3 - \frac{6p}{n(n-1)(n-2)} \sum_{i=1}^{n} \sum_{i\neq j<k\neq i} y_{ij}y_{ik} + \frac{3p^2}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq i} y_{ij}
\end{aligned}$$

since $\sum_{i=1}^{n}\sum_{j\neq i} y_{ij} = n(n-1)p$, twice the number of edges. Then

$$\begin{aligned}
\overline{C}_3 &= \overline{T}_3 - \left(\frac{6p}{n(n-1)(n-2)}\sum_{i=1}^{n}\sum_{i\neq j<k\neq i} y_{ij}y_{ik}\right) + 3p^3 \\
&= \overline{T}_3 - 3p\left(\frac{2}{n(n-1)(n-2)}\sum_{i=1}^{n}\sum_{i\neq j<k\neq i} (y_{ij}y_{ik} - p^2)\right) \\
&= \overline{T}_3 - 3p\overline{T}_2,
\end{aligned}$$

as required. □

In the next section we focus on finding conditional means, variances and covariances of $\overline{C}_2$ and $\overline{C}_3$.

## 4.7 First and Second Moments of $\overline{C}_2$ and $\overline{C}_3$

Here we will calculate all the conditional first and second moment of our statistics, $\overline{C}_2$ and $\overline{C}_3$ in the conditional case. Specifically, we find $E[\overline{C}_2|m]$, $\text{Var}[\overline{C}_2|m]$, $E[\overline{C}_3|m]$, $\text{Var}[\overline{C}_3|m]$, and $\text{Cov}[\overline{C}_2, \overline{C}_3|m]$.

For an arbitrary random graph $RG(n, p)$, $n$ is the number of vertices, and $p$ is the sample mean of the number of edges in the graph. The number of possible edges in the graph is

$$N = \frac{n(n-1)}{2}.$$

Therefore, the number of edges present is $Np$, and the number of absent edges is $N - Np$.

**Proposition 4.1**
*Let $(y_{ij})_{1\leqslant i<j\leqslant n}$ denote the adjacency matrix of Erdös-Rényi-Gilbert random graph. Let $\overline{C}_2$ and $\overline{C}_3$ be the statistics defined in (4.6.1) and (4.6.2) with $p = N^{-1}\sum_{1\leqslant i<j\leqslant n} y_{ij}$. Then, conditional on the event $\sum_{1\leqslant i<j\leqslant n} y_{ij} = m$, the following results hold.*

$$E\left[\overline{C}_2 \middle| m\right] = -\frac{p(1-p)}{N-1}; \quad (4.7.1)$$

$$E\left[\overline{C}_3 \middle| m\right] = \frac{2p(1-p)(1-2p)}{(N-1)(N-2)}; \quad (4.7.2)$$

$$Var\left[\overline{C}_2 \middle| m\right] = \frac{2p^2(1-p)^2}{n(n-1)(n-2)} + O(n^{-4}); \quad (4.7.3)$$

$$Var\left[\overline{C}_3 \middle| m\right] = \frac{6p^3(1-p)^3}{n(n-1)(n-2)} + O(n^{-6}); \quad (4.7.4)$$

$$Cov\left[\overline{C}_2, \overline{C}_3 \middle| m\right] = O(n^{-4}). \quad (4.7.5)$$

Statements (4.7.1) - (4.7.5) are proved in Subsections 4.7.1-4.7.6, respectively.

### 4.7.1 First Conditional Moment of $\overline{C}_2$

**Lemma 4.7**

*In the notation used above, the conditional covariance of* $y_{ij}$ *and* $y_{ik}$*,* $j \neq k$*, given the number of edges is*

$$E\left[(y_{ij}-p)(y_{ik}-p) \middle| m\right] = -\frac{p(1-p)}{N-1}.$$

**Proof:** In a population consists of binary variables, zeros and ones, in (4.3.2)we defined $\pi_{r,s-r}$ as the probability of choosing $r$ ones and $s-r$ zeros when sampling without replacement. In this case, the sample size is 2 (i.e. $s = 2$ and $r = 0, 1, 2$). The conditional covariance of $y_{ij}$ and $y_{ik}$, $j \neq k$, given $m$, the number of edges, is

$$E\left[(y_{ij}-p)(y_{ik}-p) \middle| m\right] = \sum_{r=0}^{2} \pi_{r,2-r}(1-p)^r(-p)^{2-r}, \quad (4.7.6)$$

where all the results of $\pi_{r,2-r}$ where $r = 0, 1, 2$ are presented in (4.3.3).

Then by substitution in (4.7.6) using the equations in (4.3.3), we get

$$\begin{aligned} E\left[(y_{ij}-p)(y_{ik}-p)\middle|m\right] &= \frac{Np^2(1-p)^2}{N-1}\left[(1-\frac{1}{Np})-2+(1-\frac{1}{N(1-p)})\right] \\ &= -\frac{Np^2(1-p)^2}{N-1}\left[\frac{(1-p+p)}{Np(1-p)}\right], \\ &= -\frac{p(1-p)}{N-1}. \end{aligned}$$

□

To find the first conditional moment of $\overline{C}_2$, $E[\overline{C}_2|p]$, we need to use Lemma 4.7. Thus the conditional expectation of $\overline{C}_2$ is

$$\begin{aligned} E\left[\overline{C}_2\middle|m\right] &= E\left[\frac{2}{n(n-1)(n-2)}\sum_{i=1}^{n}\sum_{i\neq j<k\neq i}(y_{ij}-p)(y_{ik}-p)\middle|m\right] \\ &= \frac{2}{n(n-1)(n-2)}\sum_{i=1}^{n}\sum_{i\neq j<k\neq i}E\left[(y_{ij}-p)(y_{ik}-p)\middle|m\right] \\ &= \frac{2}{n(n-1)(n-2)}\sum_{i=1}^{n}\sum_{i\neq j<k\neq i}\left(-\frac{p(1-p)}{N-1}\right) \\ &= -\frac{p(1-p)}{N-1}. \end{aligned}$$

### 4.7.2 First Conditional Moment of $\overline{C}_3$

**Lemma 4.8**

*In the notation of context, the conditional expectation of* $y_{ij}$, $y_{jk}$ and $y_{ki}$ *given the number of edges is*

$$E\left[(y_{ij}-p)(y_{jk}-p)(y_{ki}-p)\middle|m\right] = \frac{2p(1-p)(1-2p)}{(N-1)(N-2)}.$$

**Proof:** Conditional on the number of edges $m$,

$$
\begin{aligned}
&E\left[(y_{ij}-p)(y_{jk}-p)(y_{ik}-p)\Big|m\right] \\
&= \sum_{r=0}^{3} \pi_{r,3-r}(1-p)^r(-p)^{3-r}, \\
&= \pi_{0,3}(-p)^3 + \pi_{1,2}(1-p)(-p)^2 + \pi_{2,1}(1-p)^2(-p) + \pi_{3,0}(1-p)^3,
\end{aligned} \tag{4.7.7}
$$

where, $\pi_{r,3-r}$; $r = 0, \cdots, 3$, is defined in (4.3.4). Then by substitution in (4.7.7), we get

$$
\begin{aligned}
&E\left[(y_{ij}-p)(y_{jk}-p)(y_{ki}-p)\Big|m\right] \\
= \;& \frac{N^2}{(N-1)(N-2)}p^3(1-p)^3 \\
& \left[\left(1-\frac{1}{Np}\right)\left(1-\frac{2}{Np}\right)\right. \\
& \left.-\left(1-\frac{1}{N(1-p)}\right)\left(1-\frac{2}{N(1-p)}\right) - 3\left(1-\frac{1}{Np}\right) + 3\left(1-\frac{1}{N(1-p)}\right)\right], \\
= \;& \frac{N^2}{(N-1)(N-2)}p^3(1-p)^3\left[\frac{2}{N^2p^2} - \frac{2}{N^2(1-p)^2}\right], \\
= \;& \frac{N^2}{(N-1)(N-2)}p^3(1-p)^3\left[\frac{2}{N^2p^2(1-p)^2}\left((1-p)^2-p^2\right)\right], \\
= \;& \frac{2p(1-p)(1-2p)}{(N-1)(N-2)}.
\end{aligned}
$$

□

To find the first conditional moment of $\overline{C}_3$, $E[\overline{C}_3|m]$, we need Lemma 4.8. Thus the conditional expectation of triangles statistic is

$$\begin{aligned} E\left[\overline{C}_3\middle| m\right] &= E\left[\frac{6}{n(n-1)(n-2)}\sum_{1\leqslant i<j<k\leqslant n}(y_{ij}-p)(y_{jk}-p)(y_{ki}-p)\middle| m\right] \\ &= \frac{6}{n(n-1)(n-2)}\sum_{1\leqslant i<j<k\leqslant n} E\left[(y_{ij}-p)(y_{jk}-p)(y_{ki}-p)\middle| m\right] \\ &= \frac{6}{n(n-1)(n-2)}\sum_{1\leqslant i<j<k\leqslant n}\left(\frac{2p(1-p)(1-2p)}{(N-1)(N-2)}\right) \\ &= \frac{2p(1-p)(1-2p)}{(N-1)(N-2)}. \end{aligned}$$

### 4.7.3 Second Conditional Moment of $\overline{C}_2$

To derive the second conditional moment of the centered 2-stars density, $\overline{C}_2$, we should derive the following conditional expectation,

$$E\left[(y_{ij}-p)(y_{ik}-p)(y_{i^*j^*}-p)(y_{i^*k^*}-p)\middle| p\right],$$

here we have three cases:

**Case (i)** Two edge in common, i.e. $i=i^*$, $j=j^*$ and $k=k^*$;

**Case (ii)** One edge in common, i.e. $i=i^*$, $j=j^*$ and $k\neq k^*$;

**Case (iii)** No edge in common, i.e. $\{i,j\},\{i,k\},\{i^*,j^*\},\{j^*,k^*\}$, all are different.

Now, we will investigate each case in detail.

**Case (i):**

In this case when we have two edge in common, i.e. $i=i^*$, $j=j^*$ and $k=k^*$. The number of instance in this case is 1, see Lemma 3.1. This case reduces the expectation to

$$
\begin{aligned}
E\left[(y_{ij}-p)^2(y_{ik}-p)^2\middle|m\right] &= \sum_{r=0}^{2}\pi_{r,2-r}[(1-p)^2]^r[(-p)^2]^{(2-r)},\\
&= \pi_{0,2}p^4+\pi_{1,1}(1-p)^2p^2+\pi_{2,0}(1-p)^4,\\
&= \frac{N}{N-1}p^2(1-p)^2\left[p^2(1-\frac{1}{N(1-p)}+\right.\\
&\quad \left. 2p(1-p)+(1-p)^2(1-\frac{1}{Np})\right]\\
&= \frac{Np^2(1-p)^2}{N-1}[1+O(N^{-1})].
\end{aligned}
$$

where $\pi_{r,2-r}$, $r=0,1,2$, are defined in (4.3.3).

**Case (ii):**

In this case when we have one edge in common, i.e. $i=i^*$, $j=j^*$ and $k\neq k^*$. The number of instance in this case is $(4n-10)$, see (3.3.7) in Lemma 3.1. This case reduces the conditional expectation to

$$
E\left[(y_{ij}-p)^2(y_{ik}-p)(y_{ik^*}-p)\middle|m\right].
$$

Let $r$ = number of 1's, and $2-r$ = number of 0's, where $r=0,1,2$. Consider the probability of having 0 in the first term then $r$ ones in the rest and $2-r$ zeros in the rest is $\tau_{0;r,2-r}$, and the probability of having 1 in the first term then $r$ ones in the rest and $2-r$ zeros in the rest is $\tau_{1;r,2-r}$. The conditional expectation as following

$$
\begin{aligned}
&E\left[(y_{ij}-p)^2(y_{ik}-p)(y_{ik^*}-p)\middle|m\right]\\
&=\sum_{r=0}^{2}\tau_{0;r,2-r}(1-p)^r(-p)^{4-r}+\sum_{r=0}^{2}\tau_{1;r,2-r}(1-p)^{2+r}(-p)^{2-r}
\end{aligned}
\tag{4.7.8}
$$

where

$$
\tau_{0;r,2-r}=\frac{N-Np}{N-2}\pi_{r,2-r},
$$

and

$$
\tau_{1;r,2-r}=\frac{Np}{N-2}\pi_{r,2-r},
$$

where $\pi_{r,2-r}$, $r=0,1,2$, are defined in (4.3.3).

However, we can apply simply Theorem 4.1, and we note that $t = 2$, therefore the leading term in (4.7.8) is $O(N^{-1})$.

**Case (iii):**

In this case there is no edge in common between 2-stars, i.e. $\{i, j\}, \{i, k\}, \{i^*, j^*\}$, $\{j^*, k^*\}$, all are different. The number of instance in this case is $\frac{n-3}{2}(n^2 - 6)$, see (3.3.6) in Lemma 3.1. Let $r$ = number of 1's, and $4 - r$ = number of 0's, where $r = 0, 1, 2, 3, 4$. Consider the probability of having $r$ ones and $4-r$ zeros is $\pi_{r,4-r}$. The conditional expectation of Case(iii) is

$$E\left[(y_{ij} - p)(y_{ik} - p)(y_{i^*j^*} - p)(y_{i^*k^*} - p)\middle| m\right] = \sum_{r=0}^{4} \pi_{r,4-r}(1 - p)^r(-p)^{4-r}, \tag{4.7.9}$$

The corresponding probabilities, $\pi_{r,4-r}$, again obtained by considering sampling without replacement from a finite binary population, and using formulas in (4.3.5). Simply, we can apply Theorem 4.1, and we note that $t = 4$, thus the leading term in (4.7.9) is $O(N^{-2})$.

To calculate the conditional expectation, $E[\overline{C}_2^2|p]$, for the three cases, we have to multiply each case by its number of instances, which already found it in Lemma 3.1 in Subsection 3.3. The number of instances in each case are:
1, $4n - 10$ and $\frac{n-3}{2}(n^2 - 6)$ in cases (i), (ii) and (iii) respectively.

Finally,

$$
\begin{aligned}
& E\left[\overline{C}_2^2\middle|m\right] \\
= & \left(\frac{2}{n(n-1)(n-2)}\right)^2 \sum_{i=1}^{n} \sum_{i\neq j<k\neq i} \sum_{i^*=1}^{n} \sum_{i^*\neq j^*<k^*\neq i^*} \\
& E\left[(y_{ij}-p)(y_{ik}-p)(y_{i^*j^*}-p)(y_{i^*k^*}-p)\middle|m\right] \\
= & \frac{2}{n(n-1)(n-2)} \sum_{i=1}^{n} \sum_{i\neq j<k\neq i} \frac{2}{n(n-1)(n-2)} \\
& \sum_{i^*=1}^{n} \sum_{i^*\neq j^*<k^*\neq i^*} E\left[(y_{ij}-p)(y_{ik}-p)(y_{i^*j^*}-p)(y_{i^*k^*}-p)\middle|m\right] \\
= & \frac{2}{n(n-1)(n-2)} \sum_{i=1}^{n} \sum_{i\neq j<k\neq i} \frac{2}{n(n-1)(n-2)} \\
& \left[p^2(1-p)^2 + (4n-10)O(N^{-1}) + \frac{n-3}{2}(n^2-6)O(N^{-2})\right], \\
= & \frac{2}{n(n-1)(n-2)} \left[p^2(1-p)^2 + O(n^{-1}) + O(n^{-1})\right], \\
= & \frac{2p^2(1-p)^2}{n(n-1)(n-2)} + O(n^{-4}).
\end{aligned}
$$

### 4.7.4 Second Conditional Moment of $\overline{C}_3$

To drive the second conditional moment of the centered triangles density, $\overline{C}_3$, we should drive the following conditional expectation,

$$
E\left[(y_{ij}-p)(y_{jk}-p)(y_{ki}-p)(y_{i^*j^*}-p)(y_{j^*k^*}-p)(y_{k^*i^*}-p)\middle|m\right],
$$

here we have also three cases, as the statistic number of 2-stars:

**Case (i)** Two edge in common, i.e. $i = i^*$, $j = j^*$ and $k = k^*$;

**Case (ii)** One edge in common, i.e. $i = i^*$, $j = j^*$ and $k \neq k^*$;

**Case (iii)** No edge in common, i.e. $\{i, j\}, \{j, k\}, \{k, i\}, \{i^*, j^*\}, \{j^*, k^*\}, \{k^*, i^*\}$, all are different.

Now, we will investigate each case in detail.

**Case (i):**

In this case when we have two edge in common, i.e. $i = i^*$, $j = j^*$ and $k = k^*$. The number of instance in this case is 1, see Lemma 3.2. This case reduces the expectation to

$$E\left[(y_{ij}-p)^2(y_{jk}-p)^2(y_{ki}-p)^2\bigg|m\right] = \sum_{r=0}^{3}\pi_{r,3-r}\left[(1-p)^2\right]^r\left[(-p)^2\right]^{3-r},$$

and by substitution in (4.3.4), we get

$$\begin{aligned}
& E\left[(y_{ij}-p)^2(y_{jk}-p)^2(y_{ki}-p)^2\bigg|m\right] \\
= & \frac{N^3}{N_{(3)}}p^3(1-p)^3 \\
& \left[p^3\left(1-\frac{1}{N(1-p)}\right)\left(1-\frac{2}{N(1-p)}\right)+3p^2(1-p)\left(1-\frac{1}{N(1-p)}\right)\right. \\
& \left.+3p(1-p)^2\left(1-\frac{1}{Np}\right)+(1-p)^3\left(1-\frac{1}{Np}\right)\left(1-\frac{2}{Np}\right)\right], \\
= & \frac{N^3}{N_{(3)}}p^3(1-p)^3 \\
& \left[1-\frac{3p^3}{N(1-p)}-\frac{3p^2}{N}-\frac{3(1-p)^2}{N}-\frac{3(1-p)^3}{Np}+\frac{2p^3}{N^2(1-p)^2}+\frac{2(1-p)^3}{N^2p^2}\right], \\
= & \frac{N^3}{N_{(3)}}p^3(1-p)^3 \\
& \left[1-\frac{3}{N}\left(\frac{p^3}{(1-p)}+\frac{(1-p)^3}{p}+p^2+(1-p)^2\right)+\frac{2}{N^2}\left(\frac{p^3}{(1-p)^2}+\frac{(1-p)^3}{p^2}\right)\right], \\
= & \frac{N^2}{(N-1)(N-2)}p^3(1-p)^3\left[1-O(\frac{1}{N})+O(\frac{1}{N^2})\right], \\
\simeq & \; p^3(1-p)^3.
\end{aligned}$$

**Case (ii):**

In this case when we have one edge in common, i.e. $i = i^*$, $j = j^*$ and $k \neq k^*$. The number of instance in this case is $3(n-3)$, see (3.3.10) in Lemma 3.2. This case reduces the conditional expectation to

$$E\left[(y_{ij}-p)^2(y_{jk}-p)(y_{ki}-p)(y_{j^*k^*}-p)(y_{k^*ß^*}-p)\bigg|m\right].$$

Let $r$ is the number of 1's, and $4-r$ is the number of 0's, where $r = 0, 1, 2, 3, 4$. Consider the probability of having 0 in the first term then $r$ ones in the rest and $4-r$ zeros in the rest is $\tau_{0;r,4-r}$, and the probability of having 1 in the first term then $r$ ones in the rest and $4-r$ zeros in the rest is $\tau_{1;r,4-r}$. The conditional expectation as following

$$E\left[(y_{ij}-p)^2(y_{jk}-p)(y_{ki}-p)(y_{j^*k^*}-p)(y_{k^*i^*}-p)\Big|m\right] = \sum_{r=0}^{4}\tau_{0;r,4-r}(1-p)^r(-p)^{6-r} + \sum_{r=0}^{4}\tau_{1;r,4-r}(1-p)^{2+r}(-p)^{4-r} \tag{4.7.10}$$

where

$$\tau_{0;r,4-r} = \frac{N-Np}{N-4}\pi_{r,4-r},$$

and

$$\tau_{1;r,4-r} = \frac{Np}{N-4}\pi_{r,4-r},$$

where $\pi_{r,4-r}$, $r = 0, 1, 2, 3, 4$, are defined as a general case in (4.3.5).

However, we can apply simply Theorem 4.1, and we note that $t = 4$, therefore the leading term in (4.7.10) is $O(N^{-2})$.

**Case (iii):**

In this case there is no edge in common between triangles, i.e. $\{i, j\}, \{j, k\}, \{k, i\}$, $\{i^*, j^*\}, \{j^*, k^*\}, \{k^*, i^*\}$, all are different. The number of instance in this case is $\frac{(n-3)(n^2-16)}{6}$, see (3.3.9) in Lemma 3.2. Let $r$ = number of 1's, and $(6-r)$ = number of 0's, where $r = 0, 1, ..., 6$. Consider the probability of having $r$ ones and $6-r$ zeros is $\pi_{r,6-r}$. The conditional expectation of $C_3^2$ given $m$, Case(iii), is

$$E\left[(y_{ij}-p)(y_{jk}-p)(y_{ki}-p)(y_{i^*j^*}-p)(y_{j^*k^*}-p)(y_{k^*i^*}-p)\Big|m\right] = \sum_{r=0}^{6}\pi_{r,6-r}(1-p)^r(-p)^{6-r}, \tag{4.7.11}$$

where

$$\pi_{r,6-r} = \binom{6}{r}\frac{N^6}{N_{(6)}}p^r(1-p)^{6-r}\prod_{\alpha=0}^{r-1}\left(1-\frac{\alpha}{Np}\right)\prod_{\beta=0}^{6-r-1}\left(1-\frac{\beta}{N(1-p)}\right),$$

and

$$N_{(6)} = N(N-1)(N-2)(N-3)(N-4)(N-5)$$

However, we can simply apply Theorem 4.1. We note that $t = 6$, therefore the leading term of the conditional expectation of $\overline{C}_3^2$ given $m$, Case(iii) in 4.7.11 is $O(N^{-3})$

To calculate the conditional expectation, $E[\overline{C}_3^2|p]$, for the three cases, we have to multiply each case by its number of instances, which already found it in Lemma 3.2 in subsection 3.3. The number of instances in each case are:
1, $3(n-3)$ and $\frac{(n-3)(n^2-16)}{6}$ in cases (i), (ii) and (iii) respectively.
Finally,

$$\begin{aligned}
& E\left[\overline{C}_3^2\middle|m\right] \\
= & \left(\frac{6}{n(n-1)(n-2)}\right)^2 \sum_{1\leqslant i<j<k\leqslant n}\ \sum_{1\leqslant i^*<j^*<k^*\leqslant n} \\
& E\left[(y_{ij}-p)(y_{jk}-p)(y_{ki}-p)(y_{i^*j^*}-p)(y_{j^*k^*}-p)(y_{k^*i^*}-p)\middle|m\right] \\
= & \frac{6}{n(n-1)(n-2)}\sum_{1\leqslant i<j<k\leqslant n}\frac{6}{n(n-1)(n-2)}\sum_{1\leqslant i^*<j^*<k^*\leqslant n} \\
& E\left[(y_{ij}-p)(y_{jk}-p)(y_{ki}-p)(y_{i^*j^*}-p)(y_{j^*k^*}-p)(y_{k^*i^*}-p)\middle|m\right] \\
= & \frac{6}{n(n-1)(n-2)}\sum_{1\leqslant i<j<k\leqslant n}\frac{6}{n(n-1)(n-2)} \\
& \left[p^3(1-p)^3+3(n-3)O(N^{-2})+\frac{(n-3)(n^2-16)}{6}(O(N^{-3}))\right], \\
= & \frac{6}{n(n-1)(n-2)}\left[p^3(1-p)^3+O(n^{-3})+O(n^{-3})\right], \\
= & \frac{6p^3(1-p)^3}{n(n-1)(n-2)}.
\end{aligned}$$

### 4.7.5 Limiting Conditional Variances of $\overline{C}_2$ and $\overline{C}_3$

We examine the limiting variance of standardized version of $\overline{C}_2$.

$$\begin{aligned}
&\text{Var} \quad \left[\sqrt{\frac{n(n-1)(n-2)}{2p^2(1-p)^2}}\left(\overline{C}_2 - E(\overline{C}_2|m)\right)\right] \\
&= \quad \frac{n(n-1)(n-2)}{2p^2(1-p)^2}\left[E(\overline{C}_2^2|m) - E^2(\overline{C}_2|m)\right], \\
&= \quad \frac{n(n-1)(n-2)}{2p^2(1-p)^2}\left[\frac{2p^2(1-p)^2}{n(n-1)(n-2)} - \frac{p^2(1-p)^2}{(N-1)^2}\right], \\
&= \quad \frac{n(n-1)(n-2)}{2p^2(1-p)^2}\left[\frac{2p^2(1-p)^2}{n(n-1)(n-2)} - O(\frac{1}{N^2})\right], \\
&= \quad 1 + O(n^{-1}).
\end{aligned}$$

Also, the limiting variance of standardized version of $\overline{C}_3$.

$$\begin{aligned}
&\text{Var} \quad \left[\sqrt{\frac{n(n-1)(n-2)}{6p^3(1-p)^3}}\left(\overline{C}_3 - E(\overline{C}_3|m)\right)\right] \\
&= \quad \frac{n(n-1)(n-2)}{6p^3(1-p)^3}\left[E(\overline{C}_3^2|m) - E^2(\overline{C}_3|m)\right], \\
&= \quad \frac{n(n-1)(n-2)}{6p^3(1-p)^3}\left[\frac{6p^3(1-p)^3}{n(n-1)(n-2)} - \frac{4p^2(1-p)^2}{(N-1)^2(N-2)^2}(1-2p)^2\right], \\
&= \quad \frac{n(n-1)(n-2)}{6p^3(1-p)^3}\left[\frac{6p^3(1-p)^3}{n(n-1)(n-2)} - O(\frac{1}{N^4})\right], \\
&= \quad 1.
\end{aligned}$$

### 4.7.6 Limiting Conditional Covariance of $\overline{C}_2$ and $\overline{C}_3$

We examine the covariance of standardized $C_2$ and $C_3$ to proof it is leading to 0.

$$\begin{aligned}
&\text{Cov} \quad \left[\sqrt{\frac{n(n-1)(n-2)}{2p^2(1-p)^2}}\left(\overline{C}_2 - E(\overline{C}_2|m)\right), \sqrt{\frac{n(n-1)(n-2)}{6p^3(1-p)^3}}\left(\overline{C}_3 - E(\overline{C}_3|m)\right)\right] \\
&= \quad \frac{n(n-1)(n-2)}{\sqrt{12p^5(1-p)^5}}\left[E(\overline{C}_2\overline{C}_3|m) - E(\overline{C}_2|m)E(\overline{C}_3|m)\right]. \qquad (4.7.12)
\end{aligned}$$

So,

$$\begin{aligned}\frac{n(n-1)(n-2)}{\sqrt{12p^5(1-p)^5}}E(\overline{C}_2|m)E(\overline{C}_3|m) &= O(n^3)O(n^{-2})O(n^{-4}) \\ &= O(n^{-3})\end{aligned}$$

where $N = O(n^2)$, then

$$\begin{aligned}&\text{Cov}\quad\left[\sqrt{\frac{n(n-1)(n-2)}{2p^2(1-p)^2}}\left(\overline{C}_2 - E(\overline{C}_2|m)\right), \sqrt{\frac{n(n-1)(n-2)}{6p^3(1-p)^3}}\left(\overline{C}_3 - E(\overline{C}_3|m)\right)\right] \\ &= \frac{n(n-1)(n-2)}{\sqrt{12p^5(1-p)^5}}E(\overline{C}_2\overline{C}_3|m). \qquad (4.7.13)\end{aligned}$$

To find $E(\overline{C}_2\overline{C}_3|m)$, we have three cases:

**Case (i):** No edge in common, where here $t = 5$, and the number of instances is $\frac{(n-3)(n^2-10)}{6}$;

**Case (ii):** One edge in common, where here $t = 3$, and the number of instances is $2(n-3)$;

**Case (iii):** Two edge in common, where here $t = 1$, and the number of instances is 1.

See Lemma 3.3. Therefore

$$
\begin{aligned}
&E(\overline{C}_2\overline{C}_3|m)\\
&\quad= E\left[\frac{2}{n(n-1)(n-2)}\sum_{\alpha=1}^{n}\sum_{\beta<\gamma}(y_{\alpha\beta}-p)(y_{\alpha\gamma}-p).\frac{6}{n(n-1)(n-2)}\right.\\
&\quad\left.\sum_{1\leqslant i<j<k\leqslant n}(y_{ij}-p)(y_{jk}-p)(y_{ki}-p)\right]\\
&\quad= \frac{2}{n(n-1)(n-2)}\sum_{\alpha=1}^{n}\sum_{\beta<\gamma}\frac{6}{n(n-1)(n-2)} \qquad (4.7.14)\\
&\quad\sum_{1\leqslant i<j<k\leqslant n}E\left[(y_{\alpha\beta}-p)(y_{\alpha\gamma}-p)(y_{ij}-p)(y_{jk}-p)(y_{ki}-p)\right]\\
&\quad= \frac{6}{n(n-1)(n-2)}\left[\frac{(n-3)(n^2-10)}{6}O(N^{-3})+2(n-3)O(N^{-2})+O(N^{-1})\right]\\
&\quad= \frac{6}{n(n-1)(n-2)}\left[O(n^{-3})+O(n^{-3})+O(n^{-2})\right]\\
&\quad= O(n^{-6})+O(n^{-6})+O(n^{-5}).
\end{aligned}
$$

Finally, by substitution (4.7.14) in (4.7.13), we have

$$
\begin{aligned}
&\mathrm{Cov}\quad\left[\sqrt{\frac{n(n-1)(n-2)}{2p^2(1-p)^2}}\left(\overline{C}_2-E(\overline{C}_2|m)\right),\sqrt{\frac{n(n-1)(n-2)}{6p^3(1-p)^3}}\left(\overline{C}_3-E(\overline{C}_3|m)\right)\right]\\
&=\quad \frac{n(n-1)(n-2)}{\sqrt{12p^5(1-p)^5}}E(\overline{C}_2\overline{C}_3|m)\\
&=\quad O(n^3)\left[O(n^{-6})+O(n^{-6})+O(n^{-5})\right]\\
&=\quad O(n^{-2})
\end{aligned}
$$

## 4.8 Summary

The main result of this chapter, Theorem 4.1, describes the asymptotic behavior as the number of vertices, $n$, goes to infinity, of a family of conditional expectations under the Erdös-Rényi-Gilbert random graph model, where the conditioning is on $m = m^{(n)}$, the number of edges in the graph. A second goal of this chapter was to find all first and second conditional moments of ($\overline{C}_2$ and $\overline{C}_3$), and the covariance between them, where ($\overline{C}_2$ and $\overline{C}_3$) are defined in (4.6.1) and (4.6.2) respectively. The results in this chapter play a vital role in the proof of Theorem 5.1, the conditional central limit theorem in the next chapter.

Chapter 5

# Central Limit Theorem: Conditional Case

## 5.1 Introduction

The statistics $\overline{T}_1$, $\overline{T}_2$ and $\overline{T}_3$, defined in (3.2.2) - (3.2.4) denote, respectively, the edge density, the 2-stars density and the triangle density in a random graph. In Theorem 3.1 we proved that, under the Erdös-Rényi-Gilbert random graph model, the vector $(\overline{T}_1, \overline{T}_2, \overline{T}_3)^T$, suitably standardized, satisfies a central limit theorem. What is noteworthy about this results, however, is that the limiting $3 \times 3$ covariance matrix has rank 1, so that the limiting multivariate Gaussian distribution is degenerate. This result seems some what surprising. We have not been able to find this result in the literature.

In this chapter, our main aim is to prove that this degeneracy is removed when we condition on the edge density $\overline{T}_1^u = \overline{T}_1 + p = m/N = p$, where $m$ is the number of edges present and $N = n(n-1)/2$ as before. The result we prove in Theorem 5.1 is equivalent to the following: Conditional on $\overline{T}_1^u$, $(\overline{T}_2, \overline{T}_3)^T$ suitably standardized satisfies a conditional central limit theorem with a limiting covariance matrix which has full rank 2. However, in Theorem 5.1 it turns out to be more convenient to work with the statistics $\overline{C}_2$ and $\overline{C}_3$ defined in (4.6.1) and (4.6.2) respectively. In fact, $\overline{C}_2$ and $\overline{C}_3$ are closely related to $\overline{T}_2$ and $\overline{T}_3$, on the event $\overline{T}_1^u$: Lemma 4.5 tells us that $\overline{C}_2 = \overline{T}_2$, and Lemma 4.6 tells us that $\overline{C}_3 = \overline{T}_3 - 3p\overline{T}_2$. Therefore, working with $\overline{T}_2$ and $\overline{T}_3$ conditional on $\overline{T}_1^u = p$ is equivalent to working with $\overline{C}_2$ and $\overline{C}_3$ conditional on $\overline{T}_1^u = p$ after a linear transformation.

The proof of Theorem 5.1, the main results of the chapter, uses the method of moments (see Section 2.4.5 of Chapter 2) and is split into three components: The proofs of Propositions 5.1, 5.2 and 5.3. The key results needed in these proofs are Theorem 4.1 and some counting lemmas, Lemma 5.1-5.4.
The outline of this chapter is as follows. In Section 5.2 we state the conditional central limit theorem in Theorem 5.1 and we also state the component results, Propositions 5.1-5.3. In Section 5.3 we state and prove the counting lemmas, Lemma 5.1-5.4, and in Section 5.4 we prove the Propositions 5.1-5.3, thereby completing the proof of Theorem 5.1. Finally, we present some numerical results in Section 5.5 which explore how good the conditional Gaussian approximation for $(\overline{C}_2, \overline{C}_3)^T$ is for various choices of $p$ and $n$, and in Section 5.6 we investigate real-world network data.

## 5.2 The Conditional Central Limit Theorem

Recall that $p = N^{-1}\sum_{1\leqslant i<j\leqslant n} y_{ij}$. Define standardized versions of $\overline{C}_2$ and $\overline{C}_3$ as follows:

$$\begin{aligned}
C_2 &= \sqrt{\frac{n(n-1)(n-2)}{2}}\ \overline{C}_2, \\
&= \sqrt{\frac{2}{n(n-1)(n-2)}}\ \sum_{i=1}^{n}\ \sum_{i\neq j<k\neq i} (y_{ij}-p)(y_{ik}-p), && (5.2.1)\\
&= \frac{1}{2}\left\{\frac{2}{n(n-1)(n-2)}\right\}^{\frac{1}{2}} \sum_{i=1}^{n}\ \sum_{j\neq i}\ \sum_{k\neq i,j} (y_{ij}-p)(y_{ik}-p), \\
&= \left\{\frac{2^{-1}}{n(n-1)(n-2)}\right\}^{\frac{1}{2}} \sum_{i\neq j\neq k\neq i}^{n} (y_{ij}-p)(y_{ik}-p), && (5.2.2)
\end{aligned}$$

where, by definition

$$\sum_{i\neq j\neq k\neq i}^{n} (y_{ij}-p)(y_{ik}-p) \equiv \sum_{i=1}^{n}\ \sum_{j\neq i}\ \sum_{k\neq i,j} (y_{ij}-p)(y_{ik}-p), \qquad (5.2.3)$$

and

$$
\begin{aligned}
C_3 &= \sqrt{\frac{n(n-1)(n-2)}{6}}\,\overline{C}_3, \\
&= \sqrt{\frac{6}{n(n-1)(n-2)}} \sum_{1\leqslant i<j<k\leqslant n} (y_{ij}-p)(y_{jk}-p)(y_{ki}-p), \qquad (5.2.4) \\
&= \frac{1}{6}\left\{\frac{6}{n(n-1)(n-2)}\right\}^{\frac{1}{2}} \sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i,j}(y_{ij}-p)(y_{jk}-p)(y_{ki}-p), \\
&= \left\{\frac{6^{-1}}{n(n-1)(n-2)}\right\}^{\frac{1}{2}} \sum_{i\neq j\neq k\neq i}^{n} (y_{ij}-p)(y_{jk}-p)(y_{ki}-p), \qquad (5.2.5)
\end{aligned}
$$

again using the equivalence in (5.2.3). Although equivalent, in subsequent calculations it will be more convenient to use (5.2.2), where $j$ and $k$ are not ordered, than (5.2.1), where $j$ and $k$ are ordered; and it will be more convenient to use (5.2.5), where $i$, $j$ and $k$ are not ordered, than (5.2.4), where $i$, $j$ and $k$ are ordered.

The following conditional central limit theorem for the two statistics $C_2$ and $C_3$ is conditional on $m$, the number of edges present in the graph.

**Theorem 5.1**

*Consider a sequences of random graphs with $n$ vertices and adjacency matrices $\left(y_{ij}^{(n)}\right)_{1\leqslant i<j\leqslant n}$ and define $N^{(n)} = n(n-1)/2$, $n = 1, 2, \cdots$. Suppose that, for each $n$,*

**(i)** *conditional on $\sum_{1\leqslant i<j\leqslant n} y_{ij}^{(n)} = m^{(n)} = N^{(n)}p^{(n)}$, the zero-one variables $y_{ij}^{(n)}(1 \leqslant i < j \leqslant n)$, are identically distributed.*

**(ii)** *As $n \to \infty$, $p^{(n)} = m^{(n)}/N^{(n)} \xrightarrow{p} p_0 \in (0, 1)$.*

*Then, conditional on $m^{(n)}$, i.e. conditional on the event $\sum_{1\leqslant i<j\leqslant n} y_{ij}^{(n)} = m^{(n)}$, we have*

$$
\left(\begin{array}{c} C_2 \\ C_3 \end{array}\right)\Bigg|\; m \xrightarrow{d} N_2\left\{\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} p_0^2(1-p_0)^2 & 0 \\ 0 & p_0^3(1-p_0)^3 \end{array}\right)\right\} \qquad (5.2.6)
$$

*as $n \to \infty$, i.e. the limiting distribution of $(C_2, C_3)^T|m^{(n)}$ is bivariate normal* *with mean the zero vector and covariance matrix,* $\text{diag}\{p_0^2(1-p_0)^2, p_0^3(1-p_0)^3\}$, *and we notice that* $C_2$ *and* $C_3$ *are independent.*

**Remark 5.1**

*In the follows we will simplify notation by writing* $y_{ij}^{(n)} = y_{ij}$, $m^{(n)} = m$, $p^{(n)} = p$ *and* $N^{(n)} = N$. *Note that the* $p$ *used in the definition of* $C_2$ *in (5.2.2) and* $C_3$ *in (5.2.5) is actually* $p^{(n)} = m^{(n)}/N^{(n)}$.

**Preliminaries Comments on the Proof:** The proof is based on the method of moments; see e.g. Billingsley (2012), Chung (2001) and Chapter 2, Section 2.4.5. A key role is played by Theorem 4.1, formula (4.2.1). It has already been shown that $E[\overline{C}_2|m] = O(n^{-2})$ and $E[\overline{C}_3|m] = O(n^{-4})$ in (4.7.1) and (4.7.2) respectively. It follows that

$$E\left[C_2\Big|m\right] = \sqrt{\frac{n(n-1)(n-2)}{2}}O(n^{-2}) = O(n^{-\frac{1}{2}}) \tag{5.2.7}$$

$$E\left[C_3\Big|m\right] = \sqrt{\frac{n(n-1)(n-2)}{6}}O(n^{-4}) = O(n^{-\frac{5}{2}}) \tag{5.2.8}$$

Also, from the results for $\text{Var}(\overline{C}_2|m)$, $\text{Var}(\overline{C}_3|m)$ and $\text{Cov}(\overline{C}_2, \overline{C}_3|m)$ obtained in Proposition 4.1, it follows immediately that

$$\text{Var}(C_2|m) \to p_0^2(1-p_0)^2, \quad \text{Var}(C_3|m) \to p_0^3(1-p_0)^3, \quad \text{Cov}(C_2, C_3|m) \to 0, \tag{5.2.9}$$

as $n \to \infty$. Consequently, the first and second moments for $C_2$ and $C_3$ already obtained are consistent with what is stated in (5.2.6).

It remains to show that all the higher order moments of $C_2$ and $C_3$ (conditional on $m$) converge to the corresponding moments of the bivariate normal given in (5.2.6). Specifically, to apply the method of moments approach, we need to show that for each pair of non-negative integers $r$ and $s$,

$$E[C_2^r C_3^s|m] \to E[Z_1^r Z_2^s] = E[Z_1^r]E[Z_2^s] \tag{5.2.10}$$

as $n \to \infty$, where $(Z_1, Z_2)^T$ is bivariate normal with mean $(0,0)^T$ and covariance matrix $\text{diag}\{p_0^2(1-p_0)^2, p_0^3(1-p_0)^3\}$. Note that the equality is valid in (5.2.10) because, due to the covariance matrix of $(Z_1, Z_2)^T$ being diagonal, and because

$(Z_1, Z_2)^T$ is bivariate normal, $Z_1$ and $Z_2$ are independent.

It has already been show in (5.2.7)-(5.2.8) that (5.2.10) holds for all integers $r$ and $s$ satisfying $0 \leqslant r + s \leqslant 2$. It remains to show that (5.2.10) holds for all non-negative integers $r$ and $s$ such that $r + s \geqslant 3$.
It will be helpful to split the proof into three cases:

**Case I** the marginal conditional moments of $C_2$ ($r \geqslant 3$, $s = 0$).

**Case II** the marginal conditional moments of $C_3$ ($r = 0$, $s \geqslant 3$).

**Case III** the joint conditional moments of $C_2$ and $C_3$ ($r \geqslant 1, s \geqslant 1, r + s \geqslant 3$).

Case I, II and III are covered below in Propositions 5.1, 5.2 and 5.3, respectively. The proofs of these propositions are given in Section 5.4. □

**Proposition 5.1**
*Assume that the conditions Theorem 5.1 hold. Then for each fixed $r \geqslant 1$,*

$$E[C_2^r|m] \overset{n\to\infty}{\longrightarrow} \left\{ \begin{array}{ll} 0 & \text{if } r \text{ is odd;} \\ (r-1)!!p_0^r(1-p_0)^r & \text{if } r \text{ is even} \end{array} \right\}$$

*where $(r-1)!! = (r-1)(r-3)...5.3.1$ for even $r$. Consequently, since the limiting conditional moments of $C_2$ are the same as those of $Z \sim N(0, p_0^2(1-p_0)^2)$, we may conclude that $C_2 \overset{p}{\longrightarrow} N(0, p_0^2(1-p_0)^2)$.*

**Proposition 5.2**
*Assume that the conditions Theorem 5.1 hold. Then for each fixed $s \geqslant 1$,*

$$E[C_3^s|m] \overset{n\to\infty}{\longrightarrow} \left\{ \begin{array}{ll} 0 & \text{if } s \text{ is odd;} \\ (s-1)!!p_0^{3s/2}(1-p_0)^{3s/2} & \text{if } s \text{ is even,} \end{array} \right\}$$

*Consequently, since the limiting conditional moments of $C_3$ are the same as those of $Z \sim N(0, p_0^3(1-p_0)^3)$, we may conclude that $C_3 \overset{p}{\longrightarrow} N(0, p_0^3(1-p_0)^3)$.*

**Proposition 5.3**
*Assume that the conditions Theorem 5.1 hold. Then for $r, s \geqslant 1$,*

$$E[C_2^r C_3^s|m] \stackrel{n\to\infty}{\longrightarrow} \left\{ \begin{array}{ll} (r-1)!!(s-1)!!p_0^r(1-p_0)^r p_0^{3s/2}(1-p_0)^{3s/2} & \textit{if r,s even;} \\ 0 & \textit{otherwise} \end{array} \right\}.$$

*Consequently, since the limiting joint conditional moments of $C_2$ and $C_3$ are the same as those of the corresponding joint moments of $Z_1 \sim N(0, p_0^2(1-p_0)^2)$ and $Z_2 \sim N(0, p_0^3(1-p_0)^3)$, where $Z_1$ and $Z_2$ are independent and normal, it follows that $C_2$ and $C_3$ are asymptotically independent and normal.*

Before proving Propositions 5.1-5.3, we provide some preliminary results. These results depend on some counting arguments which are quite complicated.

## 5.3 Some Counting Lemmas

In this section we make use of two types of equivalence relation: Tilde equivalence relations, defined in subsection 5.3.1; and diamond equivalence relations, defined in subsection 5.3.2. Tilde equivalence relations are relevant when we want to count the number of tilde singletons, i.e. the number of tilde equivalence classes with one element. The number of tilde singletons will play the role of $t$ in Theorem 4.1. In contrast, diamond equivalence relations are more convenient to use when we counting indices as in the key Lemma 5.1. As will be seen, there is actually a relationship between the two types of equivalence relation considered: A given tilde equivalence relation uniquely determines a diamond equivalence relation. However, the uniqueness does not go the other way. There are many tilde equivalence relations which determine the same diamond equivalence relation.
Consider $C_2^r$, which is given by

$$C_2^r = \{2n(n-1)(n-2)\}^{-r/2} \sum_{i_1\neq j_1\neq k_1\neq i_1}^{n} \cdots \sum_{i_r\neq j_r\neq k_r\neq i_r}^{n} \prod_{u=1}^{r} (y_{i_u j_u} - p)(y_{i_u k_u} - p), \tag{5.3.1}$$

and $C_3^r$, which is given by

$$\begin{aligned} C_3^r &= \{6n(n-1)(n-2)\}^{-r/2} \\ &\times \sum_{i_1 \neq j_1 \neq k_1 \neq i_1}^{n} \cdots \sum_{i_r \neq j_r \neq k_r \neq i_r}^{n} \prod_{u=1}^{r} (y_{i_u j_u} - p)(y_{j_u k_u} - p)(y_{k_u i_u} - p), \end{aligned} \tag{5.3.2}$$

where (5.2.3) has been used repeatedly. First, we give some definitions. Define

$$\mathcal{N}_n = \{1, ...n\}, \tag{5.3.3}$$

$$\mathcal{A}_{r,n} = \{(i_u, j_u, k_u : u = 1, ..., r) \in \mathcal{N}_n^{3r} | i_u \neq j_u \neq k_u \neq i_u\} \tag{5.3.4}$$

and

$$T(y, A) = \prod_{u=1}^{r} (y_{i_u j_u} - p)(y_{i_u k_u} - p), \qquad A \in \mathcal{A}_{r,n}. \tag{5.3.5}$$

Since there is a one-to-one relationship between $\mathcal{A}_{r,n}$ and the set of those $(i_u, j_u, k_u : u = 1, ..., r)$ which appear on the RHS of (5.3.1), it follows that we have the identity

$$C_2^r = \{2n(n-1)(n-2)\}^{-r/2} \sum_{A \in \mathcal{A}_{r,n}} T(y, A), \tag{5.3.6}$$

where the sum in (5.3.6) is over all elements $A$ of $A_{r,n}$.

### 5.3.1 Definition of Tilde Equivalence Relation

Now consider a general $A \in \mathcal{A}_{r,n}$. It is seen that $T(y, A)$ depends on (and only on) those $y_{ij}$ with index sets

$$L_1 = \{i_1, j_1\}, \; L_2 = \{i_1, k_1\}, \cdots, \; L_{2r-1} = \{i_r, j_r\}, \; L_{2r} = \{i_r, k_r\} \tag{5.3.7}$$

Using the given $A \in \mathcal{A}_{r,n}$, define an equivalence relation $\sim$ on the set $\{L_1, L_2, ..., L_{2r}\}$ as follows:

$$\text{for } L_u, L_v \in \{L_1, L_2, ..., L_{2r}\}, \; L_u \sim L_v \; \text{ if and only if } \; L_u = L_v. \tag{5.3.8}$$

**Remark 5.2**

*When assessing equality in (5.3.8), the $L_u$ and $L_v$ in (5.3.7) are treated as sets* *with two elements rather than as ordered pairs. Note that* $y_{ij} = y_{ji}$ *due to the fact that we are considering graphs with undirected edges.*

The relation $\sim$ is easily seen to be:

**(i)** reflexive ($L_u \sim L_u$), because $L_u = L_u$;

**(ii)** symmetric ($L_u \sim L_v$ if and only if $L_v \sim L_u$), because $L_u = L_v$ if and only if $L_v = L_u$;

**(iii)** transitive ($L_u \sim L_v$ and $L_v \sim L_w$ implies $L_u \sim L_w$), because $L_u = L_v$ and $L_v = L_w$ then $L_u = L_w$.

Therefore it is an equivalence relation, which determines a partition of the set $\{L_1, L_2, ..., L_{2r}\}$ corresponding to equivalence classes. Let us write $\Upsilon^{(2r)} = \{\upsilon_1, ...\upsilon_\beta\}$ for a typical partition of $\{L_1, ..., L_{2r}\}$, where $\bigcup_{\gamma=1}^{\beta} \upsilon_\gamma = \{L_1, ..., L_{2r}\}$ and $\upsilon_\gamma \cap \upsilon_\delta = \emptyset$, the empty set, unless $\delta = \gamma$. The $\upsilon_\gamma$ are called the blocks of the partitions.

Let us now define

$$\mathcal{A}_{r,n}[\Upsilon^{(2r)}] = \{A \in \mathcal{A}_{r,n} | A \text{ determines partition } \Upsilon^{(2r)}\}. \tag{5.3.9}$$

Since each $A \in \mathcal{A}_{r,n}$ determines one and only one partition of $\{L_1, ..., L_{2r}\}$, we have the identity

$$C_2^r = \{2n(n-1)(n-2)\}^{-r/2} \sum_{0 \leqslant \Upsilon^{(2r)} \leqslant 1} \quad \sum_{A \in \mathcal{A}_{r,n}[\Upsilon^{(2r)}]} T(y, A), \tag{5.3.10}$$

where 0 is an abbreviation for the minimal partition $\{\{L_1\}, ..., \{L_{2r}\}\}$, 1 is an abbreviation for the maximal partition $\{L_1, L_2, ..., L_{2r}\}$, and

$$\mathcal{A}_{r,n} = \bigcup_{0 \leqslant \Upsilon^{(2r)} \leqslant 1} \mathcal{A}_{r,n}[\Upsilon^{(2r)}], \tag{5.3.11}$$

and the ordering is with respect to the partial ordering of set partitions; see Chapter 2. Note that the union in (5.3.11) is over all distinct partitions of the set $\{L_1, L_2, ..., L_{2r}\}$.

Finally, we define the number of singletons in partition $\Upsilon^{(2r)} = \{\upsilon_1, ..., \upsilon_\beta\}$ as follows:

$$t = \mathcal{S}(\Upsilon^{(2r)}) = \text{card}\{\gamma \in \{1, ..., \beta\} : |\upsilon_\gamma| = 1\},$$

where card is that for cardinality, i.e. number of elements, and $|\upsilon_\gamma| = \text{card}\{\upsilon_\gamma\}$.

**Example:** Note that, for $C_2^3$, where $r = 3$, we have to configure three potential 2-stars as in Figure 5.1. We use the notation for $L_u$ in (5.3.7) with $r = 3$.

There are several cases for these three 2-stars. To illustrate, we will explain

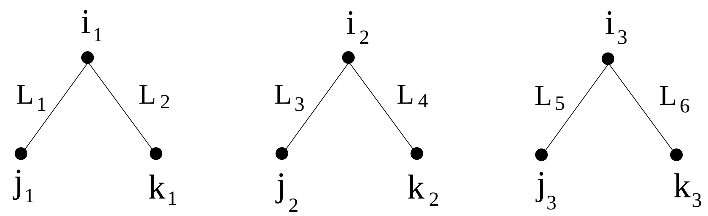


**Figure 5.1:** Three potential 2-Stars in Case 1. Here, none of the $L_u$ are equal, i.e. $L_u \neq L_v$ if $u \neq v$.

three cases in detail.

*Case 1.* Suppose there are no equalities among the $L_u$. Here we have $L_1, \cdots, L_6$; all are different. The resulting partition is a set of 6 blocks, one block for each of $L_u$, since there are no equalities in this case. So $\Upsilon^{(6)} = \{\upsilon_1, \cdots, \upsilon_6\}$, where $\upsilon_u = \{L_u\}$, $u = 1, \cdots, 6$. Therefore, we have 6 blocks each of them has size 1, i.e. $|\upsilon_u| = 1$. Consequently, the number of singletons in this case is $t = 6$.

*Case 2.* Two distinct equalities among the $L_u$; see Figure 5.2. Suppose we have, $L_1 = L_3$; $L_2 = L_4$; $L_5 \neq L_6$; $L_1$, $L_2$, $L_5$, $L_6$ all differint. In this case we have 4 blocks, $\Upsilon_2^{(6)} = \{\upsilon_1, \upsilon_2, \upsilon_3, \upsilon_4\}$, where $\upsilon_1 = \{L_1, L_3\}$; $\upsilon_2 = \{L_2, L_4\}$; $\upsilon_3 = \{L_5\}$; $\upsilon_4 = \{L_6\}$.

Therefore we have 2 blocks of length 1. Consequently, the number of singletons in this case is $t = 2$.

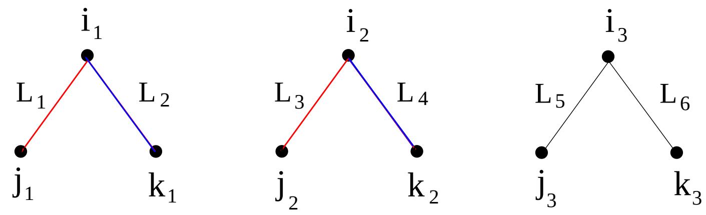


**Figure 5.2:** Three potential 2-Stars in Case 2. Here, $L_1 = L_3$ (red)and $L_2 = L_4$ (blue), and all other $L_u$ are distinct.

*Case 3.* Equalities among two distinct triples; see for instance Figure 5.3, and let we have

$$L_1 = L_3 = L_5 \neq L_2 = L_4 = L_6.$$

In this case we have 2 blocks, $\Upsilon_3^{(6)} = \{\upsilon_1, \upsilon_2\}$, where $\upsilon_1 = \{L_1, L_3, L_5\}$; $\upsilon_2 =$

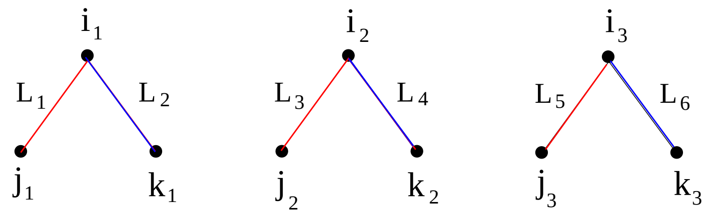


**Figure 5.3:** Three potential 2-Stars in Case 3. Here, $L_1 = L_3 = L_5$ (red) and $L_2 = L_4 = L_6$ (blue) and $L_1 \neq L_2$.

$\{L_2, L_4, L_6\}$. Because of none blocks of length 1, the number of singletons in this case is $t = 0$.

### 5.3.2 Definition of Diamond Equivalence Relation

Let us consider an equivalent relation $'\Diamond'$, this time on the set $\{\overline{L}_1, ..., \overline{L}_r\}$ rather than $\{L_1, ..., L_{2r}\}$, where here

$$\overline{L}_1 = \{i_1; j_1, k_1\}, \cdots, \overline{L}_r = \{i_r; j_r, k_r\}.$$

In the definition of $\overline{L}_1, \cdots, \overline{L}_r$, note that there is a semi-colon after the indices $i_1, \cdots, i_r$, respectively. This is to ensure that for a given $\overline{L}_u$, there is a unique way to construct two pairs of indices $\{i_u, j_u\}$ and $\{i_u, k_u\}$ from $\overline{L}_u$. As will be seen

below, this semi-colon will enable us to relate the diamond equivalence classes to the tilde equivalence classes. The diamond equivalent relation $'\Diamond'$ is defined as follows:
For $\overline{L}_u, \overline{L}_v \in \{\overline{L}_1, ..., \overline{L}_r\}$, we say that $\overline{L}_u \Diamond \overline{L}_v$ if at least one of the set equalities $\{i_u, j_u\} = \{i_v, j_v\}$ or $\{i_u, j_u\} = \{i_v, k_v\}$ or $\{i_u, k_u\} = \{i_v, j_v\}$ or $\{i_u, k_u\} = \{i_v, k_v\}$ holds.
Clearly from the definition $'\Diamond'$ is seen to be reflexive, symmetric and transitive and it therefore defines an equivalence relation. Let us use a bar to denote the corresponding partition of $\{\overline{L}_1, ..., \overline{L}_r\}$, i.e. $\overline{\Upsilon}^{(r)} = \{\overline{\upsilon}_1, ..., \overline{\upsilon}_\beta\}$, with blocks $\overline{\upsilon}_\gamma \subseteq \{\overline{L}_1, ..., \overline{L}_r\}$. Let $b_\gamma = |\overline{\upsilon}_\gamma|$, the number of elements in block $\overline{\upsilon}_\gamma$. The blocks $\overline{\upsilon}_\gamma$ are of two types:

**(i)** Those consisting of a single element , $\overline{L}_u$ say, of $\{\overline{L}_1, ..., \overline{L}_r\}$, corresponding to a pair $\{i_u, j_u\}$ and $\{i_u, k_u\}$; blocks of this type will be called diamond singletons.

**(ii)** Those blocks consisting of two or more elements of $\{\overline{L}_1, ..., \overline{L}_r\}$.

The blocks in (ii) contain elements (i.e. pairs) which satisfy equality constraints. Let $t_0$ denote the number of blocks of $\overline{\upsilon}_\gamma$ of type (i), i.e. the number of diamond singletons. Let $\alpha$ denote the number of blocks of type (ii) of size $b_1, ..., b_\alpha$ respectively, and suppose that these blocks have, respectively, $t_1, ..., t_\alpha$ singletons with respect to the tilde equivalence relation.
Note: to calculate the number of tilde singletons $t_\gamma$ in a diamond block $\overline{\upsilon}_\gamma$ of type (ii), we proceed as follows. Suppose $\overline{\upsilon}_\gamma = \{\overline{L}_{u_\delta} : \delta = 1, \cdots, b_\gamma\}$. Then for each $\overline{L}_{u_\delta}$ in the block, construct the two pairs $\{i_{u_\delta}, j_{u_\delta}\}$ and $\{i_{u_\delta}, k_{u_\delta}\}$, noting the semi-colon in the definition of the $\overline{L}_u = \{i_u; j_u, k_u\}$. Finally, count the number of tilde singletons, $t_\gamma$, among the $2b_\gamma$ pairs $\{i_{u_\delta}, j_{u_\delta}\}$ and $\{i_{u_\delta}, k_{u_\delta}\}$, $\delta = 1, \cdots, b_\gamma$.

### 5.3.3 Counting Indices over a Diamond Block

**Lemma 5.1**
*Let $\overline{\upsilon}_\gamma$ denote any block of $\overline{\Upsilon}^{(r)}$. Suppose that $|\overline{\upsilon}_\gamma| = b_\gamma \geqslant 2$, and let $t_\gamma$ denote the number of tilde singletons in $\overline{\upsilon}_\gamma$. Write $\mathfrak{f}_\gamma$ for the number of ways of choosing the indices*

$$\{i_{u_\delta}; j_{u_\delta}, k_{u_\delta}\}, \quad \delta = 1, \cdots, b_\gamma.$$

*Then*

$$\mathfrak{f}_\gamma = O\{n^{1+b_\gamma+I(t_\gamma>1)}\} \quad \textit{as } n \to \infty, \tag{5.3.12}$$

*where*

$$I(t_\gamma > 1) = \begin{cases} 1 & \textit{if } t_\gamma > 1, \\ 0 & \textit{otherwise}, \end{cases}$$

*is an indicator function.*

**Proof:** Due to the connectedness of the set $\overline{v}_\gamma$ with respect to the diamond equivalence relation $\Diamond$, there exist sequences $u_1, \cdots, u_{b_\gamma} \in \{1, \cdots, r\}$ and $v_1, \cdots, v_{b_\gamma-2} \in \{1, \cdots, r\}$ such that $\overline{v}_\gamma = \{\overline{L}_{u_1}, \cdots, \overline{L}_{u_{b_\gamma}}\}$, and the diamond relations

$$\overline{L}_{u_1} \quad \Diamond \quad \overline{L}_{u_2} \tag{5.3.13}$$

$$\overline{L}_{v_1} \quad \Diamond \quad \overline{L}_{u_3} \quad \text{for some } \ v_1 \in \{u_1, u_2\} \tag{5.3.14}$$

$$\overline{L}_{v_2} \quad \Diamond \quad \overline{L}_{u_4} \quad \text{for some } \ v_2 \in \{u_1, u_2, u_3\} \tag{5.3.15}$$

$$\vdots$$

$$\overline{L}_{v_{b_\gamma-2}} \quad \Diamond \quad \overline{L}_{u_{b_\gamma}} \quad \text{for some } \ v_{b_\gamma-2} \in \{u_1, \cdots, u_{b_\gamma-1}\} \tag{5.3.16}$$

The cases $t_\gamma = 0$, $t_\gamma = 1$ and $t_\gamma > 1$ are considered separately. When convenient we shall write $b(\gamma)$ by $b_\gamma$.

*For $t_\gamma = 0$.* First we focus on $\overline{L}_{u_1}$, and consider the number of choices for the indices $\{i_{u_1}, j_{u_1}, k_{u_1}\}$. As there are no constraints apart from $i_{u_1} \neq j_{u_1} \neq k_{u_1} \neq i_{u_1}$, there are $O(n^3)$ choices for the indices $i_{u_1}$, $j_{u_1}$ and $k_{u_1}$. Now we consider $\overline{L}_{u_2}$. Due to the relation $\overline{L}_{u_1} \Diamond \overline{L}_{u_2}$ in (5.3.13), there are at most $O(n)$ choices which have not already been fixed by the relation (5.3.13). This is because at least one of $\{i_{u_2}, j_{u_2}\}$ or $\{i_{u_2}, k_{u_2}\}$ is equal to at least one of $\{i_{u_1}, j_{u_1}\}$ or $\{i_{u_1}, k_{u_1}\}$. Similar arguments show that the number of choices of indices for each of

$$\{i_{u_\delta}, j_{u_\delta}, k_{u_\delta}\}, \quad \delta = 3, \cdots, b_\gamma - 1,$$

is at most $O(n)$. However, because $t_\gamma = 0$, neither of $\{i_{b(\gamma)}, j_{b(\gamma)}\}$ or $\{i_{b(\gamma)}, k_{b(\gamma)}\}$

can be a tilde singleton with respect to the tilde equivalence relation. Therefore all three of the indices $i_{b(\gamma)}, j_{b(\gamma)}, k_{b(\gamma)}$ must have already been determined by the earlier $i_{u_\delta}, j_{u_\delta}, k_{u_\delta}$, $\delta = 1, \cdots, b_\gamma - 2$, through the tilde equivalence relation, see Figure 5.4.

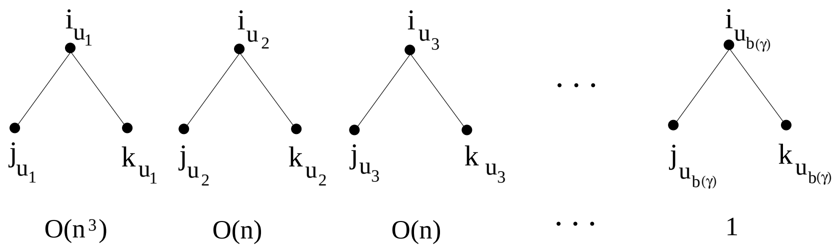


**Figure 5.4:** The potential 2-Stars in the block $\overline{\upsilon}_\gamma$.

Therefore

$$\mathfrak{f}_\gamma = O(n^3).O(n^{b_\gamma - 2}) = O(n^{1+b_\gamma}). \tag{5.3.17}$$

*For* $t_\gamma = 1$. Without loss of generality, we can arrange for the unique tilde singleton in $\overline{\upsilon}_\gamma$ to belong to $\overline{L}_{u_1}$. As in the case $t_\gamma = 0$, there are $O(n^3)$ ways of choosing the indices in $\overline{L}_{u_1}$ and at most $O(n)$ ways of choosing the indices in each of $\overline{L}_{u_2}, \cdots, \overline{L}_{u_\gamma - 1}$.

Finally, because there is only one tilde singleton in this case, which has already appeared in $\overline{L}_{u_1}$, it follows that the indices $i_{u_{b(\gamma)}}, j_{u_{b(\gamma)}}, k_{u_{b(\gamma)}}$ have already been determined by the tilde equivalence relation. Therefore, as in case $t_\gamma = 0$,

$$\mathfrak{f}_\gamma = O(n^3).O(n^{b_\gamma - 2}) = O(n^{1+b_\gamma}). \tag{5.3.18}$$

*For* $t_\gamma > 1$. In this case we cannot rule out the possibility that $\overline{L}_{u_{b_\gamma}}$ contains a tilde singleton, so there are potentially $O(n)$ ways of choosing the indices $i_{u_{b(\gamma)}}, j_{u_{b(\gamma)}}, k_{u_{b(\gamma)}}$ while still responding the tilde equivalence relation. Therefore in this case

$$\mathfrak{f}_\gamma = O(n^3).O(n^{b_\gamma - 1}) = O(n^{2+b_\gamma}). \tag{5.3.19}$$

Putting (5.3.17), (5.3.18) and (5.3.19) together, we obtain (5.3.12). □

The next Lemma plays a crucial role in the proof of Proposition 5.1 below. It turns out that, with minor modifications, essentially the same result can be used to prove Proposition 5.2 and Proposition 5.3, as is discussed below.

### 5.3.4 Counting Lemma for Proposition 5.1

The following Lemma is for counting indices over a tilde partition, $\Upsilon^{(2r)}$.

**Lemma 5.2**
*Let $t = 2t_0 + \sum_{\gamma=1}^{\alpha} t_\gamma$ denote the number of tilde singletons of $\Upsilon^{(2r)}$. Define*

$$E_{max} = \frac{3}{2}r + \frac{3}{2}\left\lfloor \frac{t}{2} \right\rfloor. \tag{5.3.20}$$

*Then $a_{r,n}[\Upsilon^{(2r)}]$, the cardinality of the set $A_{r,n}[\Upsilon^{(2r)}]$ defined in (5.3.9), satisfies*

$$a_{r,n}[\Upsilon^{(2r)}] = \begin{cases} O(n^{E_{max}}) & \text{if } t \geqslant 1 \\ O(n^{r+\alpha}) & \text{if } t = 0, \end{cases} \tag{5.3.21}$$

*where $\alpha$ is the number of diamond blocks of type (ii) in the diamond partition $\overline{\Upsilon}^{(r)}$ determined by the tilde partition $\Upsilon^{(2r)}$.*

**Proof:** First, let $\overline{\Upsilon}^{(r)} = \{\overline{\upsilon}_1, \cdots, \overline{\upsilon}_\beta\}$ denote the diamond partition determined by the tilde partition $\Upsilon^{(2r)}$. Without loss of generality it is assumed that $\overline{\upsilon}_1, \cdots, \overline{\upsilon}_\alpha$ are type (ii) diamond blocks and $\overline{\upsilon}_{\alpha+1}, \cdots, \overline{\upsilon}_\beta$ are the singleton diamond blocks, with $t_0 = \beta - \alpha$. Note that for each diamond singleton block, $\{\overline{L}_u\}$ say, there are $O(n^3)$ ways of choosing the indices $i_u$, $j_u$ and $k_u$. Therefore the number of ways of choosing the indices for the $t_0$ diamond singleton blocks is

$$\{O(n^3)\}^{t_0} = O(n^{3t_0}). \tag{5.3.22}$$

Since the blocks $\overline{\upsilon}_\gamma$ are disconnected from each other, it follows that, as $n \to \infty$,

$$\begin{aligned} a_{r,n}[\Upsilon^{(2r)}] &= O(n^{3t_0})O(\prod_{\gamma=1}^{\alpha} \mathfrak{f}_\gamma) \\ &= O(n^{3t_0})O(n^{\alpha+\sum_{\gamma=1}^{\alpha} b_\gamma+\sum_{\gamma=1}^{\alpha} I(t_\gamma>1)}) \\ &= O(n^E) \end{aligned}$$

where $\mathfrak{f}_\gamma$ is defined in Lemma 5.1 and

$$E = 3t_0 + \alpha + \sum_{\gamma=1}^{\alpha} b_\gamma + \sum_{\gamma=1}^{\alpha} I(t_\gamma > 1).$$

Now, $r = t_0 + \sum_{\gamma=1}^{\alpha} b_\gamma$ from the definition of these quantities. Moreover, $\alpha \leqslant (r - t_0)/2$ since each type (ii) block contains at least two elements of $\{\overline{L}_1, ..., \overline{L}_r\}$. Therefore,

$$\begin{aligned} E &\leqslant 3t_0 + \left(\frac{r-t_0}{2}\right) + (r - t_0) + \sum_{\gamma=1}^{\alpha} I(t_\gamma > 1) \\ &= \frac{3r}{2} + \frac{3t_0}{2} + \sum_{\gamma=1}^{\alpha} I(t_\gamma > 1). \end{aligned}$$

Now consider a block $\overline{v}_\gamma$ of type (ii) such that $t_\gamma \geqslant 2$. If, for a given $\gamma$, we move all (if $t_\gamma$ is positive and even) or all but one (if $t_\gamma \geqslant 3$ is odd) of the singletons in block $\gamma$ into blocks of type (i), then we decrease the sum of indications by 1, but we increase $E$ by at least $\frac{3}{2}$ because $t_0$ increases by at least 1. Therefore, to maximise $E$ we should move all the tilde singletons into blocks of type (i) if $t$ is even, and if $t$ is odd we should move all but one of the tilde singletons into blocks of type (i). The maximum value of $E$ given $t = 2t_0 + \sum_{\gamma=1}^{\alpha} t_\gamma$ is therefore as stated in (5.3.20).
On the other hand, if $t = 0$ then $t_0 = 0$, $t_\gamma = 0$, $\gamma = 1, \cdots, \alpha$, $r = \sum_{\gamma=1}^{\alpha} b_\gamma$ and then $E = \alpha + r$, so the second part of (5.3.21) follows immediately. □

### 5.3.5 Counting Lemma for Proposition 5.2

Lemma 5.2 plays a key role in the proof of Proposition 5.1. Lemma 5.3 and Lemma 5.4, stated below, play exactly analogous roles in the proofs of Proposition 5.2 and Proposition 5.3, respectively, as we shall now discuss.
Lemma 5.3 is relevant to Proposition 5.2 in which powers $C_3^r$ of $C_3$ defined in (5.2.5) are considered. In this case we defined the tilde relation on the set $\{L_1, \cdots, L_{3r}\}$, where for $u = 1, \cdots, r$,

$$L_{3(u-1)+v} = \begin{cases} \{i_u, j_u\} & \text{if } v = 1 \\ \{j_u, k_u\} & \text{if } v = 2 \\ \{k_u, i_u\} & \text{if } v = 3 \end{cases} \tag{5.3.23}$$

The equivalence relation is defined as before: for $L_u$ and $L_v$ in $\{L_1, \cdots, L_{3r}\}$: We define $L_u \sim L_v$ if we have the set equality $L_u = L_v$; and, since tilde defines an equivalence relation, a given set of equalities involving the $L_u$ will lead to a partition of $\{L_1, \cdots, L_{3r}\}$. A typical such partition will be denoted by $\Upsilon^{(3r)}$; note the analogy with $\Upsilon^{(2r)}$ above. Moreover, in exactly the same way as before, $\Upsilon^{(3r)}$ determines a diamond partition $\overline{\Upsilon}^{(r)} = \{\overline{v}_1, \cdots, \overline{v}_\beta\}$ on the set $\{\overline{L}_1, \cdots, \overline{L}_r\}$, where now $\overline{L}_u = \{i_u, j_u, k_u\}$, $u = 1, \cdots, r$, with no need for the semi-colon. Let $a_{r,n}[\Upsilon^{(3r)}]$ denote the number of choices of $\{i_u, j_u, k_u : u = 1, \cdots, r\} \subseteq \mathcal{N}^{3r}$, see (5.3.3), which satisfy the equality constrains implied by $\Upsilon^{(3r)}$ and satisfy $i_u \neq j_u \neq k_u \neq i_u$ for $u = 1, \cdots, r$. Then the following analogue of Lemma 5.2 holds.

The set $\{\overline{L}_1, \cdots, \overline{L}_r\}$ is defined in exactly the same way, and the diamond relation is now defined by

$L_u \Diamond L_v$ if at least one of $\{i_u, j_u\}$, $\{j_u, k_u\}$ or $\{k_u, i_u\}$ is equal to at least one of $\{i_v, j_v\}$, $\{j_v, k_v\}$ or $\{k_v, i_v\}$.

**Lemma 5.3**

*Let $t$ denote the number of tilde singletons in the partition $\Upsilon^{(3r)}$ of the set $\{L_1, \cdots, L_{3r}\}$ defined by (5.3.23). Then*

$$a_{r,n}[\Upsilon^{(3r)}] = \begin{cases} O(n^{E_{max}}) & \textit{if } t \geqslant 1 \\ O(n^{r+\alpha}) & \textit{if } t = 0, \end{cases}$$

*where, as before, $E_{max}$ is given by (5.3.20) and $\alpha$ is the number of diamond blocks of type (ii) in the diamond partition $\overline{\Upsilon}^{(r)}$ determined by tilde partition $\Upsilon^{(3r)}$.*

**Proof:** The statement and proof of the analogue of Lemma 5.1 required here is identical to that of Lemma 5.1. The proof of Lemma 5.3 is the same as that of Lemma 5.2, even though the definition of the underlying set $\{L_1, \cdots, L_{3r}\}$ is slightly different. □

### 5.3.6 Counting Lemma for Proposition 5.3

In the final case, Proposition 5.3, moments of the form $C_2^r C_3^s$ are considered. Here, the underlying set is defined by

$$\{L_1, \cdots, L_{2r}, L_{2r+1}, \cdots, L_{2r+3s}\}$$

where, for $u = 1, \cdots, r$,

$$L_{2(u-1)+v} = \begin{cases} \{i_u, j_u\} & \text{if } v = 1 \\ \{i_u, k_u\} & \text{if } v = 2, \end{cases} \tag{5.3.24}$$

and for $u = r+1, \cdots, r+s$,

$$L_{2r+3(u-r-1)+v} = \begin{cases} \{i_u, j_u\} & \text{if } v = 1 \\ \{j_u, k_u\} & \text{if } v = 2, \\ \{k_u, i_u\} & \text{if } v = 3, \end{cases} \tag{5.3.25}$$

The tilde relation is defined in the same way as before, i.e. $L_u \sim L_v$ if the set equality $L_u = L_v$ holds. Partitions of the set $\{L_1, \cdots, L_{2r+3s}\}$ are written $\Upsilon^{(2r,3s)}$. The diamond relation is defined in the same way as before but now on the set $\{\overline{L}_1, \cdots, \overline{L}_{r+s}\}$. Let $a_{r,s,n}[\Upsilon^{(2r,3s)}]$ denote the number of choices of $\{i_u, j_u, k_u : u = 1, \cdots, r+s\} \subseteq \mathcal{N}^{r+s}$, see (5.3.3), with $i_u \neq j_u \neq k_u \neq i_u$ such that the equalities implies by the partition $\Upsilon^{(2r,3s)}$ are respected. The analogue of Lemma 5.2 in this case is as follows.

Finally we give the counting lemma needed for Proposition 5.3.

**Lemma 5.4**

*Let $t$ denote the number of tilde singletons in the partition $\Upsilon^{(2r,3s)}$ of the set $\{L_1, \cdots, L_{2r+3s}\}$ defined by (5.3.24) and (5.3.25). Then*

$$a_{r,s,n}[\Upsilon^{(2r,3s)}] = \begin{cases} O(n^{E_{\max}}) & \text{if } t \geqslant 1 \\ O(n^{r+s+\alpha}) & \text{if } t = 0, \end{cases}$$

*where $\alpha$ is the number of diamond blocks of type (ii) in the diamond partition*

$\Upsilon^{(r,s)}$ *determined by tilde partition* $\Upsilon^{(2r,3s)}$, *and*

$$E_{\max} = \frac{3}{2}r + \frac{3}{2}s + \frac{3}{2}\left\lfloor \frac{t}{2} \right\rfloor$$

**Proof:** The details of the proof are very similar to those of Lemma 5.2 and Lemma 5.3, and are omitted.

## 5.4 Proof of Propositions 5.1-5.3

**Proof of Proposition 5.1:** Let $S(\Upsilon^{(2r)})$ denote the number of tilde singletons in the partition $\Upsilon^{(2r)}$. For the moment, fix $\Upsilon^{(2r)}$ and suppose that $S(\Upsilon^{(2r)}) = t \geqslant 1$. The contribution of $\Upsilon^{(2r)}$ to the expectation $E[C_2^r|m]$ is given by

$$\left\{ \frac{2^{-1}}{n(n-1)(n-2)} \right\}^{r/2} \sum_{A \in \mathcal{A}_{r,n}[\Upsilon^{(2r)}]} E[T(y,A)|m]; \tag{5.4.1}$$

see (5.3.5), (5.3.6) and (5.3.10).

From the general conditional moment results in Theorem 4.1,

$$E[T(y,A)|m] = O(n^{-2\lfloor (t+1)/2 \rfloor}) \tag{5.4.2}$$

Moreover, from Lemma 5.2 with $t \geqslant 1$, the cardinality of the set $\mathcal{A}_{r,n}[\Upsilon^{(2r)}]$, namely $a_{r,n}[\Upsilon^{(2r)}]$, is of order given by

$$a_{r,n}[\Upsilon^{(2r)}] = O\left\{ n^{3r/2+3\lfloor t/2 \rfloor/2} \right\} \tag{5.4.3}$$

Therefore, using the fact that $|\sum_{j=1}^{n} c_j| \leqslant \sum_{j=1}^{n} |c_j|$, (5.4.1) is bounded as follows:

$$
\begin{aligned}
& \left| \left\{ \frac{2^{-1}}{n(n-1)(n-2)} \right\}^{r/2} \sum_{A \in \mathcal{A}[\Upsilon^{(2r)}]} E[T(y,A)|m] \right| \\
\leqslant & \left\{ \frac{2^{-1}}{n(n-1)(n-2)} \right\}^{r/2} \sum_{A \in \mathcal{A}[\Upsilon^{(2r)}]} \Big| E[T(y,A)|m] \Big| \\
= & \ O(n^{-3r/2}) a_{r,n}[\Upsilon^{(2r)}] O(n^{-2\lfloor (t+1)/2 \rfloor}), \\
= & \ O(n^{-3r/2}) O(n^{3r/2+3\lfloor t/2 \rfloor/2}) O(n^{-2\lfloor (t+1)/2 \rfloor}), \qquad (5.4.4) \\
= & \ O(n^{-3r/2+3r/2+3\lfloor t/2 \rfloor/2-2\lfloor (t+1)/2 \rfloor}), \\
= & \ O(n^{3\lfloor t/2 \rfloor/2-2\lfloor (t+1)/2 \rfloor}), \\
= & \begin{cases} O(n^{-t/4}) & \text{if } t \geqslant 1 \text{ is even}, \\ O(n^{-(t+7)/4}) & \text{if } t \geqslant 1 \text{ is odd}, \end{cases} \\
= & \ O(n^{-t/4}).
\end{aligned}
$$

In (5.4.4), the first term comes from the factor outside the sum in the definition of $C_2^r$, the second term comes from Lemma 5.2 and the third term comes from Theorem 4.1.

Since, for given $r$, the number of partitions of the set with $2r$ elements is finite, we may conclude that the total contribution of all partitions $\Upsilon^{(2r)}$ with $\mathcal{S}(\Upsilon^{(2r)}) = t \geqslant 1$ to the sum (5.4.1) is also $O(n^{-t/4})$ for fixed $r$, and therefore this contribution is negligible as $n \to \infty$.

Now consider those partitions $\Upsilon^{(2r)}$ with $\mathcal{S}(\Upsilon^{(2r)}) = t = 0$. Using Lemma 5.2 again, along with the fact that $\alpha \leqslant \lfloor r/2 \rfloor$, where $\alpha$ is the number of diamond blocks of type (ii),

$$
\begin{aligned}
\left\{ \frac{2^{-1}}{n(n-1)(n-2)} \right\}^{r/2} \textstyle\sum_{A \in \mathcal{A}[\Upsilon^{(2r)}]} \Big| E[T(y,A)|m] \Big| &= O(n^{-3r/2}) O(n^{r+\alpha}) O(1), \\
&= O(n^{\alpha - r/2}), \\
&= o(1). \qquad (5.4.5)
\end{aligned}
$$

unless $\alpha = r/2$, in which case (5.4.5) is $O(1)$. Since, as noted above, $\alpha \leqslant \lfloor r/2 \rfloor$, $\alpha = r/2$ is only possible when $r$ is even. Therefor when $r$ is odd, the sum (5.3.10) convergence to 0 as $n \to \infty$.

In the remainder of the proof, we only need to consider those partitions $\Upsilon^{(2r)}$ with $r$ even, $t = 0$ and $\alpha = r/2$.
The number of ways of dividing $r = 2s$ objects into $s$ blocks of size 2 is given by

$$(r-1)!! = (r-1)(r-3)\cdots 3.1. \tag{5.4.6}$$

This result follows easily from an induction argument.
We also need to include a factor $2^s$ due to the fact that there are two ways of forming diamond block of size 2 (see Figure 5.5) with no singletons from $\overline{L}_u = \{i_u; j_u, k_u\}$ and $\overline{L}_v = \{i_v; j_v, k_v\}$ :

$$\text{pair } \{i_u, j_u\} \text{ with } \{i_v, j_v\} \text{ and pair } \{i_u, k_u\} \text{ with } \{i_v, k_v\};$$

$$\text{or pair } \{i_u, j_u\} \text{ with } \{i_v, k_v\} \text{ and pair } \{i_u, k_u\} \text{ with } \{i_v, j_v\}.$$

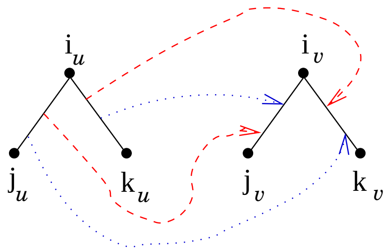


**Figure 5.5:** There are two ways of forming diamond block of size 2 with no singletons.

when calculating $E[C_2^r|m]$, where $C_2^r$ is written as in (5.3.1), we have shown that with $r$ even, we may restrict attention to the situation where all diamond blocks are of size 2 and no tilde singletons are present. In this case, using the size 2 block structure, with tilde singletons absent, we may write a typical term in the expectation of the multiple sum on the RHS of (5.3.1) as

$$\begin{aligned} E\left[\prod_{u=1}^{s}(y_{i_u j_u}-p)^2(y_{i_u k_u}-p)^2\middle| m\right] &\sim \prod_{u=1}^{s} E\left[(y_{i_u j_u}-p)^2(y_{i_u k_u}-p)^2\middle| m\right] \\ &\sim \left\{p^2(1-p)^2\right\}^s \\ &= p^r(1-p)^r, \end{aligned} \tag{5.4.7}$$

where $r = 2s$. Therefore, using (5.4.6), we may conclude that, still assuming$r = 2s$ is even,

$$\begin{aligned} E[C_2^r|m] &\sim 2^{-s} n^{-3r/2} n^{3r/2} 2^s p^r (1-p)^r (r-1)!!, \\ &= (r-1)!! p^r (1-p)^r, \\ &\xrightarrow{p} (r-1)!! p_0^r (1-p_0)^r \end{aligned} \tag{5.4.8}$$

as required, since $p \xrightarrow{p} p_0$ as $n \to \infty$ by assumption. Therefore, Proposition 5.1 holds and the proof is complete. □

**Proof of Proposition 5.2:** This has exactly the same structure as the proof of Proposition 5.1. However, there are some minor differences in the details, which we now list.

1. The set on which the tilde equivalence relation is defined is expanded from $\{L_1, \cdots, L_{2r}\}$ to $\{L_1, \cdots, L_{3r}\}$; see (5.3.23)

2. As before, we define the diamond equivalence relation on $\overline{L}_1, \cdots, \overline{L}_r$, but now we may define $\overline{L}_u = \{i_u, j_u, k_u\}$, i.e. we do not need to distinguish $i_u$ from $j_u$ or $k_u$, because now all three pairs $\{i_u, j_u\}$, $\{j_u, k_u\}$ and $\{k_u, i_u\}$ are all present.

3. The role that Lemma 5.2 plays in the proof of Proposition 5.1 is played by Lemma 5.3 in the proof of Proposition 5.2.

4. As in the proof of Proposition 5.1, it is established that when $r$ is odd, $E[C_3^r|m] \to 0$ as $n \to \infty$ and when $r = 2s$ is even, the only non-negligible contributions are those corresponding to all diamond blocks being of size 2 with no tilde singletons present. In this case, a typical term in the expectation of the multiple sum on the RHS of (5.3.2) is given by

$$\begin{aligned} &E\left[\prod_{u=1}^{s} (y_{i_u j_u} - p)^2 (y_{j_u k_u} - p)^2 (y_{k_u i_u} - p)^2 \middle| m\right] \\ &\sim \prod_{u=1}^{s} E\left[(y_{i_u j_u} - p)^2 (y_{i_u k_u} - p)^2 (y_{k_u i_u} - p)^2 \middle| m\right] \\ &\sim \left\{p^3 (1-p)^3\right\}^s \\ &= p^{3r/2} (1-p)^{3r/2}, \end{aligned} \tag{5.4.9}$$

5. Using (5.3.2), (5.4.6) and (5.4.9), and still assuming $r = 2s$ is even, the analogue of (5.4.6) is

$$\begin{aligned} E[C_3^r|m] &\sim 6^{-s} n^{-3r/2} n^{3r/2} 6^s p^{3r/2} (1-p)^{3r/2} (r-1)!!, \\ &= (r-1)!! p^{3r/2} (1-p)^{3r/2}, \\ &\xrightarrow{p} (r-1)!! p_0^{3r/2} (1-p_0)^{3r/2}, \end{aligned} \tag{5.4.10}$$

as required, since $p \xrightarrow{p} p_0$ as $n \to \infty$ by assumption. □

**Proof of Proposition 5.3:**

The structure of this proof is exactly the same as that of Proposition 5.1 and Proposition 5.2, but there are a few minor differences in the details which are now explained.

1. The tilde equivalence relation is defined on the set $\{L_1, \cdots, L_{2r}, L_{2r+1}, \cdots, L_{2r+3s}\}$; see (5.3.24) and (5.3.25).

2. The diamond equivalence relation is defined on $\{\overline{L}_1, \cdots, \overline{L}_r, \overline{L}_{r+1}, \cdots, \overline{L}_{r+s}\}$ where

$$\overline{L}_u = \begin{cases} \{i_u; j_u, k_u\} & u = 1, \cdots, r, \\ \{i_u, j_u, k_u\} & u = r+1, \cdots, r+s. \end{cases}$$

   The role of the semi-colon in $\{i_u; j_u, k_u\}$ has been explained above in Subsection 5.3.2.

3. The role that Lemma 5.2 plays in the proof of Proposition 5.1 is played by Lemma 5.4 in the proof of Proposition 5.3.

4. Lemma 5.4 and Theorem 4.1 together imply that $E[C_2^r C_3^s|m] \to 0$ unless both $r$ and $s$ are even. When $r$ and $s$ are even, the only non-negligible contributions are from those diamond partitions in which all diamond blocks are of size 2 and no tile singletons are present.

5. When $r$ and $s$ are both even and $t = 0$, the only ways to arrange $\{\overline{L}_1, \cdots, \overline{L}_r, \overline{L}_{r+1}, \cdots, \overline{L}_{r+s}\}$ into blocks of size 2 are as follows: each block is either of the form $\{\overline{L}_u, \overline{L}_v\}$ where either $u, v \in \{1, \cdots, r\}$ or $u, v \in \{r+1, \cdots, r+s\}$. In this case $\alpha$, the number of diamond blocks, is given

by $\alpha = (r+s)/2$. Then

$$\begin{aligned} E[C_2^r C_3^s|m] &\sim 2^{-r/2} n^{-3r/2} 6^{-s/2} n^{-3s/2} 2^{r/2} 6^{s/2} n^{r+s+(r+s)/2} \\ &\quad p^r(1-p)^r p^{3s/2}(1-p)^{3s/2}(r-1)!!(s-1)!! \\ &\sim (r-1)!!(s-1)!! p^r(1-p)^r p^{3s/2}(1-p)^{3s/2}, \\ &\xrightarrow{p} (r-1)!!(s-1)!! p_0^r(1-p_0)^r p_0^{3s/2}(1-p_0)^{3s/2} \end{aligned}$$

as required, since $p \xrightarrow{p} p_0$ as $n \to \infty$ by assumption. □

## 5.5 Numerical Results

The purpose of numerical study to be presented in this section is to examine how accurate the conditional Gaussian approximation based on Theorem 5.1 is for particular choices of $n$ and $m = Np$.

We implemented the R Core Team (2014) program version 3.1.2 for simulating random graphs $RG(n, m)$ which assign equal probabilities to graphs with $n$ nodes and exactly $m$ edges. We chose the values $n = 100, 200, 500, 1000$ and $p = m/N = 0.1, 0.3, 0.5, 0.7$ and $0.9$, where $N = n(n-1)/2$. In each case we ran $M = 1000$ Monte Carlo repititions, and random vectors

$$\overline{C}^{(1)} = \begin{pmatrix} \overline{C}_2^{(1)} \\ \overline{C}_3^{(1)} \end{pmatrix}, \cdots, \overline{C}^{(M)} = \begin{pmatrix} \overline{C}_2^{(M)} \\ \overline{C}_3^{(M)} \end{pmatrix}$$

were simulated, where $\overline{C}_2^{(i)}$ and $\overline{C}_3^{(i)}$ are the statistics defined in (4.6.1) and (4.6.2) respectively.

We also calculated the Mahalanobis statistics

$$Z^{(i)} = \left(\underline{\overline{C}}^{(i)} - \underline{\mu}_{n,p}\right)^T \underline{V}_{n,p}^{-1} \left(\underline{\overline{C}}_{(i)} - \underline{\mu}_{n,p}\right)$$

where, from (4.7.1) and (4.7.2),

$$\underline{\mu}_{n,p} = \begin{pmatrix} E[\overline{C}_2^{(i)}] \\ E[\overline{C}_3^{(i)}] \end{pmatrix} = \begin{pmatrix} -p(1-p)/(N-1)) \\ 2p(1-p)(1-2p)/\{(N-1)(N-2)\} \end{pmatrix}$$

and from (4.7.3)-(4.7.5),

$$\underline{V}_{n,p} = \begin{pmatrix} \mathrm{Var}(\overline{C}_2^{(i)}) & \mathrm{Cov}(\overline{C}_2^{(i)}, \overline{C}_3^{(i)}) \\ \mathrm{Cov}(\overline{C}_2^{(i)}, \overline{C}_3^{(i)}) & \mathrm{Var}(\overline{C}_3^{(i)}) \end{pmatrix} = \begin{pmatrix} \frac{2p^2(1-p)^2}{n(n-1)(n-2)} & 0 \\ 0 & \frac{6p^3(1-p)^3}{n(n-1)(n-2)} \end{pmatrix}.$$

Note that under Theorem 5.1, the $Z^{(i)}$ should have approximately a $\chi^2_2$ distribution.

Based on the 1000 Mont Carlo runs for each combination of $n$ and $m = Np$ considered, we produced the following output.

**(A)** Normal QQ plots for $\overline{C}_2$ and $\overline{C}_3$ and $\chi^2_2$ QQ plots for $Z$.

**(B)** Approximate confidence intervals for the three quartiles of $Z$.

**(A): QQ plots for $\overline{C}_2, \overline{C}_3$ and $Z$.**

In Figures 5.6-5.8 QQ plots are presented for $p = 0.1, 0.5$ and 0.9 and for $n = 100, 500, 1000$. The QQ plots for other choices of $n$ and $p$ were not presented because they were broadly similar. In QQ plots for $\overline{C}_2$ and $\overline{C}_3$ the relationship between the empirical and theoretical quantiles is approximately linear when the theoretical quantiles of $N(0,1)$ lie between approximately -2 and 2. However, there is some modest departure from linearity in the extreme tails, as might be expected.

In the $\chi^2_2$ quantile plots for $Z$, in the bottom row of Figures 5.6-5.8, some care is needed in interpreting the results because the scale of the horizontal axis changes. However, it is seem that the region over which the plot is linear is increasing as $n$ increases in each case.

**(B): Approximate confidence intervals for the three quartiles of $Z$.**

To compare the empirical quartiles of $Z$ with the theoretical quartiles of $\chi^2_2$, we constructed confidence intervals for the theoretical quartiles based on $Z^{(1)}, \cdots, Z^{(m)}$. Let $Z_{(1)} \leqslant \cdots \leqslant Z_{(m)}$ denote the order statistic based on $Z^{(1)}, \cdots, Z^{(m)}$. Then we constructed the approximate confidence intervals according to

$$Z_{(mr)} \pm 1.96\sqrt{\mathrm{Var}(Z_{(mr)})} \quad r \in (0,1), \tag{5.5.1}$$

where $\mathrm{Var}(Z_{(mr)})$ is approximated by

$$\mathrm{Var}(Z_{(mr)}) \approx \frac{r(1-r)}{mf(Q_j)}, \tag{5.5.2}$$

where $f$ is the probability density function of $\chi_2^2$. In the above, $r = 0.25, 0.5$ and $0.75$ and $Q_j$ is the $j^{th}$ quartile of $\chi_2^2$ (i.e. $j = r/0.25$ where $r = 0.25, 0.5, 0.75$). The approximations (5.5.1) and (5.5.2) are based on central limit theorems for quantiles; see e.g. van der Vaart (2000).

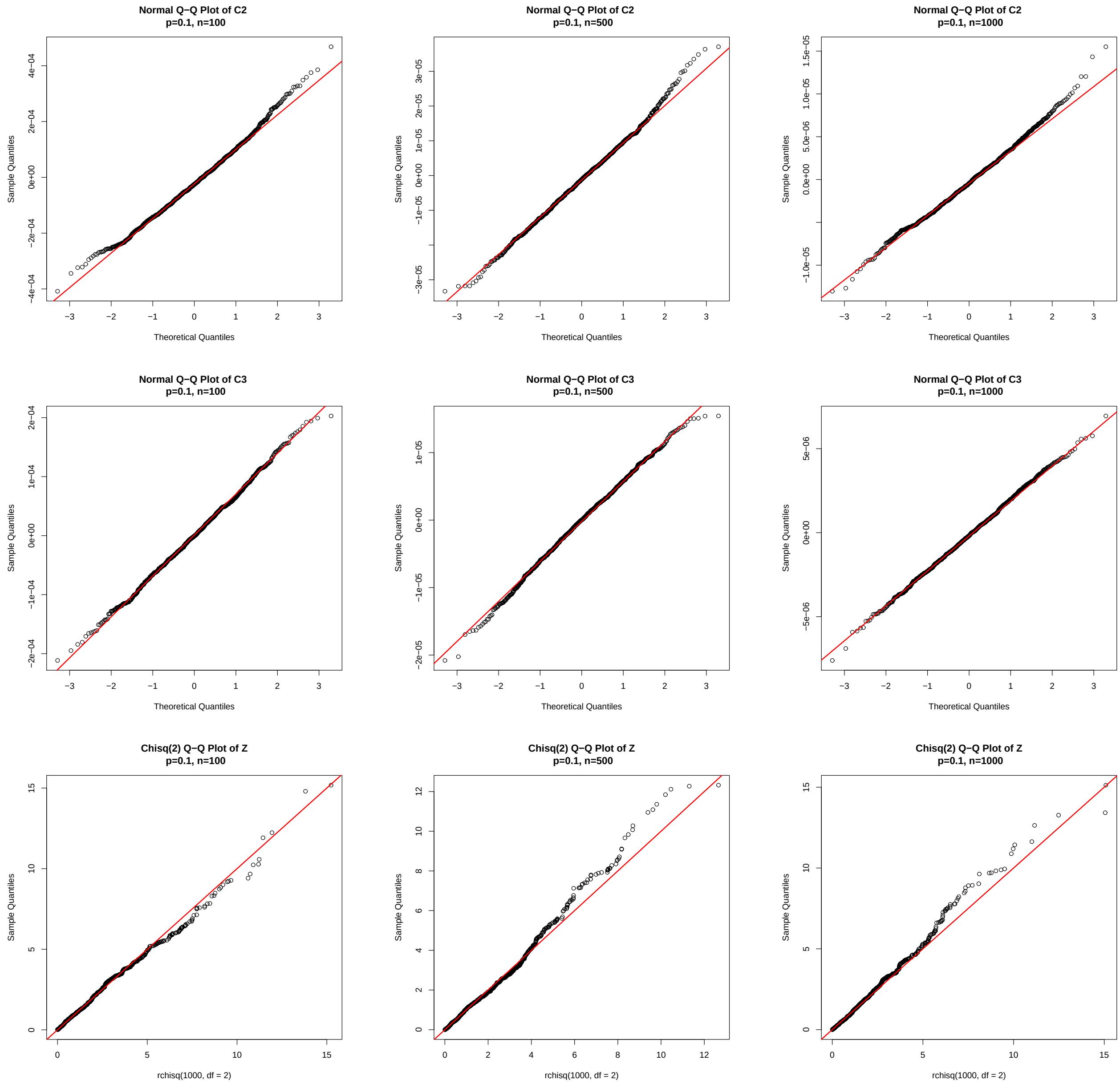


**Figure 5.6:** Q-Q Plots of $\overline{C}_2, \overline{C}_3$, and $Z$ when $p = 0.1$ and $n = 100, 500, 1000$.

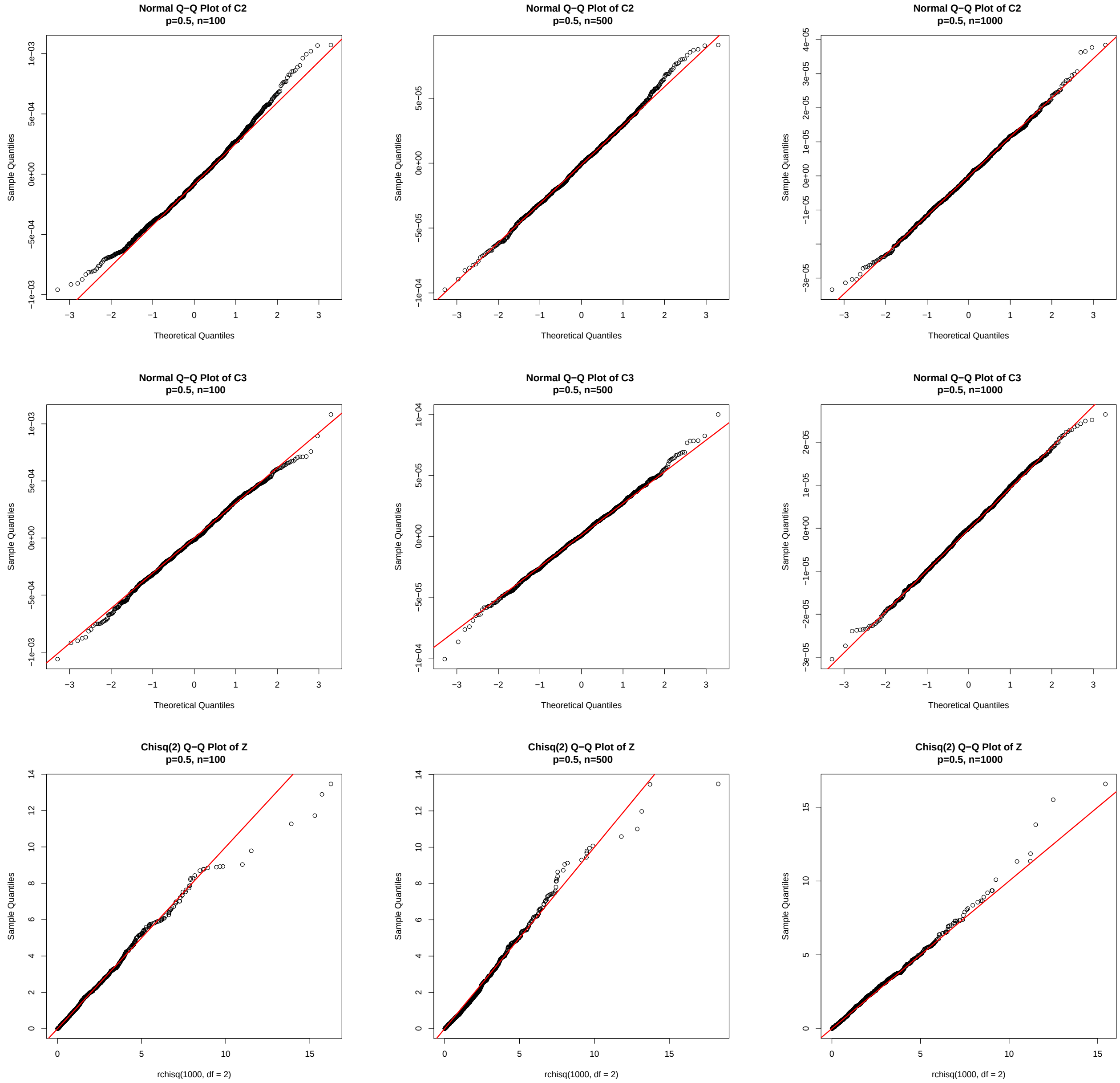


**Figure 5.7:** Q-Q Plots of $\overline{C}_2, \overline{C}_3$, and $Z$ when $p = 0.5$ and $n = 100, 500, 1000$.

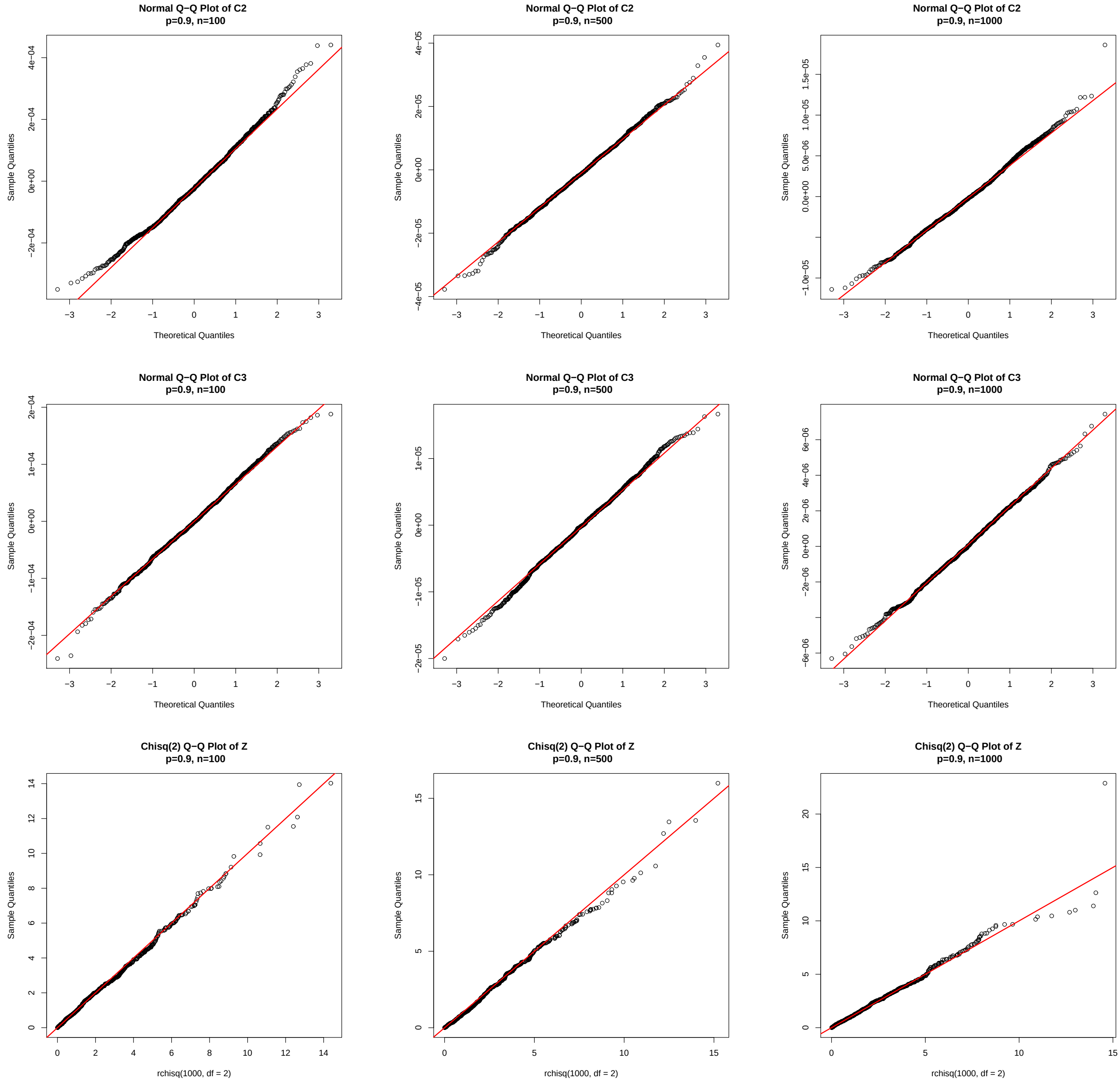


**Figure 5.8:** Q-Q Plots of $\overline{C}_2, \overline{C}_3$, and $Z$ when $p = 0.9$ and $n = 100, 500, 1000$.

In Figure 5.9, we illustrate the confidence intervals for $Q_1, Q_2, Q_3$ of $\chi^2_2$, but only displayed when $p = 0.1, 0.5, 0.9$. We note that the most intervals contain the theoretical quartiles of $\chi^2_2$, which are:

$$\begin{aligned} Q_1 &= 0.58 \\ Q_2 &= 1.39 \\ Q_3 &= 2.77. \end{aligned}$$

The width of the confidence intervals of $Q_1$ is less than the width of confidence intervals of $Q_2$, and the width of the confidence intervals of $Q_2$ is less than the width of confidence intervals of $Q_3$.

In conclusions, the data of $\overline{C}_2$ and $\overline{C}_3$ in Q-Q plots appear to be normally distributed. Whereas, the data of $Z$ in Q-Q plots appear to be $\chi^2_{(2)}$ distributed.

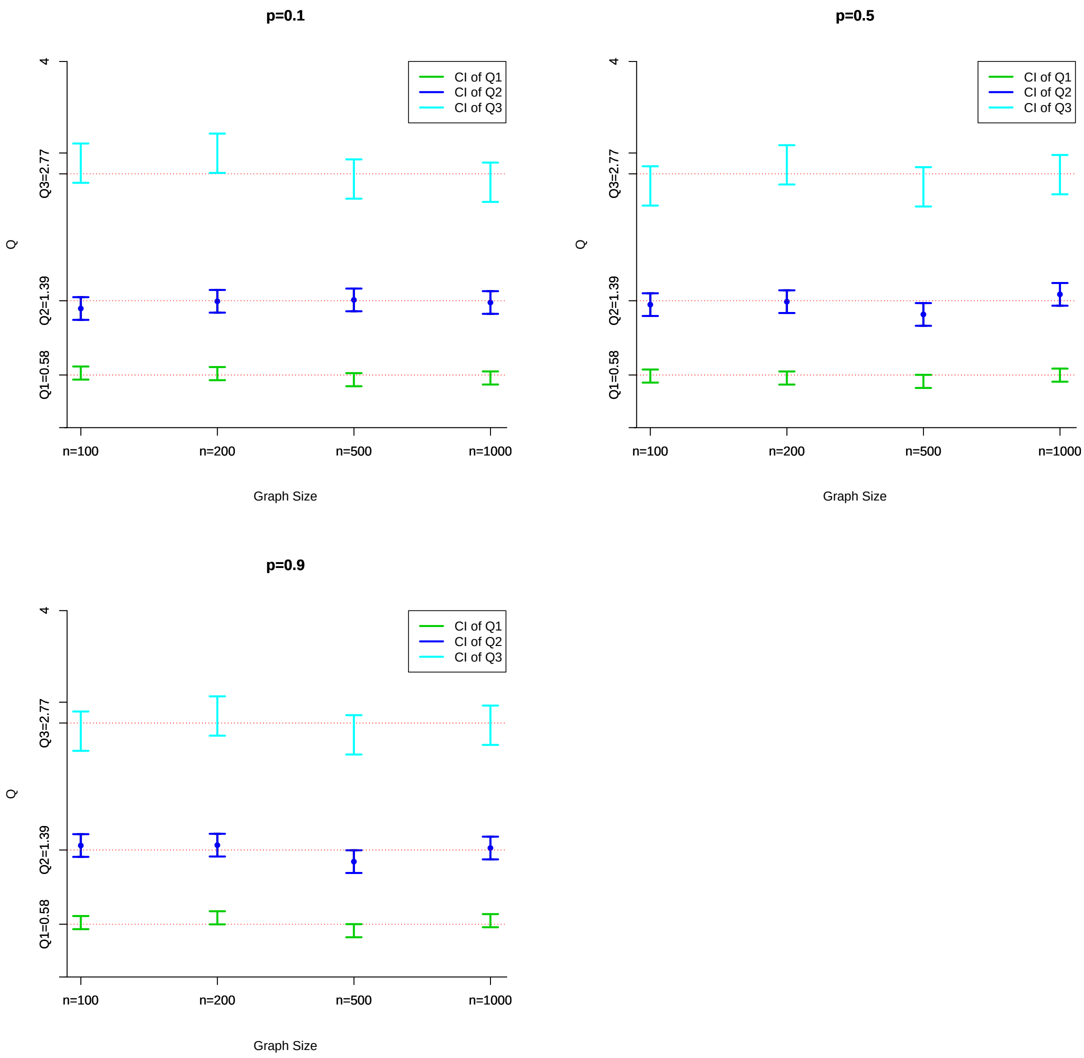


**Figure 5.9:** Error Bar Plots for Confidence Intervals of $Q_1, Q_2, Q_3$ of $Z$,when $n = 100, 200, 500, 1000$, for $p = 0.1, 0.5, 0.9$.

## 5.6 Block Graph Models and Numerical results

In this section, we consider a random graph model with vertices split into classes. Such models are known as block models. A general introduction to block models was given by Faust and Wasserman (1992).

Here we consider graphs with $h$ different types of vertices. We are interested in developing conditional tests based on Theorem 5.1 for statistics such as the density of 2-stars and the density of triangles given the number of edges. Let $A_\alpha$ denote the set of vertices of type $\alpha$, $\alpha = 1, 2, ..., h$. So, $A_\alpha \bigcap A_\beta = \phi$ if $\alpha \neq \beta$, and $\bigcup_{\alpha=1}^{h} A_\alpha = V_n$, the full set of vertices. Let $n_\alpha$ and $m_\alpha$ denote the number of vertices and edges, respectively, in block $\alpha$ and define $p_\alpha = m_\alpha/N_\alpha$ where $N_\alpha = n_\alpha(n_\alpha - 1)/2$.

We will look at the density of 2-stars and triangles among $A_\alpha$ to $A_\alpha$ connections, and also look at the density of 2-stars among $A_\alpha$ to $A_\beta$ connection ($\alpha \neq \beta$). In the first case, we shall condition on the number of $A_\alpha$ to $A_\alpha$ edges, and in the second case we condition on the number of $A_\alpha$ to $A_\beta$ edges.

Applying Lemma 4.5 and Lemma 4.6 to block $\alpha$, it is seen that

$$\begin{aligned}
\overline{C}_2^{\alpha\alpha} &= \frac{2}{n_\alpha(n_\alpha-1)(n_\alpha-2)} \sum_{\substack{i=1 \\ i\in A_\alpha}}^{n} \sum_{\substack{j<k \\ j,k\in A_\alpha\setminus\{i\}}} (y_{ij}-p_\alpha)(y_{ik}-p_\alpha), \\
&= \frac{2}{n_\alpha(n_\alpha-1)(n_\alpha-2)} \sum_{\substack{i=1 \\ i\in A_\alpha}}^{n} \sum_{\substack{j<k \\ j,k\in A_\alpha\setminus\{i\}}} (y_{ij}y_{ik}-p_\alpha^2), \qquad (5.6.1) \\
&= \overline{T}_2^{\alpha\alpha}
\end{aligned}$$

and

$$\begin{aligned}
\overline{C}_3^{\alpha\alpha} &= \frac{6}{n_\alpha(n_\alpha-1)(n_\alpha-2)} \sum_{\substack{1\leqslant i<j<k\leqslant n \\ i,j,k\in A_\alpha}} (y_{ij}-p_\alpha)(y_{jk}-p_\alpha)(y_{ki}-p_\alpha) \\
&= \frac{6}{n_\alpha(n_\alpha-1)(n_\alpha-2)} \sum_{\substack{1\leqslant i<j<k\leqslant n \\ i,j,k\in A_\alpha}} (y_{ij}y_{jk}y_{ki}-p_\alpha^3) - 3p_\alpha \\
&\quad \left( \frac{6}{n_\alpha(n_\alpha-1)(n_\alpha-2)} \sum_{i=1}^{n} \sum_{\substack{j<k \\ j,k\in A_\alpha\setminus\{i\}}} (y_{ij}y_{jk}-p_\alpha^2) \right) \\
&= \overline{T}_3^{\alpha\alpha} - 3p_\alpha \overline{T}_2^{\alpha\alpha}
\end{aligned}$$

where $\overline{T}_2^{\alpha\alpha}$ is defined in (5.6.1) and $\overline{T}_3^{\alpha\alpha}$ is defined by

$$\overline{T}_3^{\alpha\alpha} = \frac{6}{n_\alpha(n_\alpha-1)(n_\alpha-2)} \sum_{\substack{1\leqslant i<j<k\leqslant n \\ i,j,k\in A_\alpha}} (y_{ij}y_{jk}y_{ki}-p_\alpha^3).$$

Let $m_{\alpha\beta}$ denote the number of edges connecting vertices in $A_\alpha$ with vertices in $A_\beta$, and define $p_{\alpha\beta} = m_{\alpha\beta}/n_\alpha n_\beta$. Then, using similar calculations to these used to prove Lemma 4.5,

$$\begin{aligned}
\overline{C}_2^{\alpha\beta} &= \frac{2}{n_\alpha n_\beta(n_\beta - 1)} \sum_{\substack{i=1 \\ i\in A_\alpha}}^{n} \sum_{\substack{j<k \\ j,k\in A_\beta}} (y_{ij} - p_{\alpha\beta})(y_{ik} - p_{\alpha\beta}) \\
&= \frac{2}{n_\alpha n_\beta(n_\beta - 1)} \sum_{\substack{i=1 \\ i\in A_\alpha}}^{n} \sum_{\substack{j<k \\ j,k\in A_\beta}} (y_{ij}y_{ik} + p_{\alpha\beta}^2) - \frac{2p_{\alpha\beta}}{n_\alpha n_\beta(n_\beta - 1)} \sum_{\substack{i=1 \\ i\in A_\alpha}}^{n} \sum_{\substack{j<k \\ j,k\in A_\beta}} (y_{ij} + y_{ik}) \\
&= \left\{ \frac{2}{n_\alpha n_\beta(n_\beta - 1)} \sum_{\substack{i=1 \\ i\in A_\alpha}}^{n} \sum_{\substack{j<k \\ j,k\in A_\beta}} (y_{ij}y_{ik} + p_{\alpha\beta}^2) \right\} - 2p_{\alpha\beta}^2, \\
&= \frac{2}{n_\alpha n_\beta(n_\beta - 1)} \sum_{\substack{i=1 \\ i\in A_\alpha}}^{n} \sum_{\substack{j<k \\ j,k\in A_\beta}} (y_{ij}y_{ik} - p_{\alpha\beta}^2), \\
&= \overline{T}_2^{\alpha\beta}.
\end{aligned}$$

The quantities $\overline{C}_2^{\beta\alpha}$ and $\overline{T}_2^{\beta\alpha}$ are defined similarly. Note that usually $\overline{T}_2^{\alpha\beta} \neq \overline{T}_2^{\beta\alpha}$ and $\overline{C}_2^{\alpha\beta} \neq \overline{C}_2^{\beta\alpha}$.

### 5.6.1 Numerical Results

We aim in this section to use the conditional central limit theorem, Theorem 5.1, to examine goodness-of-fit of a random block model applied to real data. Recall the statistics $\overline{C}_2$ and $\overline{C}_3$, first defined in (4.6.1) and (4.6.2), respectively. Here we work with standardised versions, $Z_2$ and $Z_3$, defined by

$$Z_j = (\overline{C}_j - E(\overline{C}_j))/Var(\overline{C}_j), \quad j = 2, 3, \tag{5.6.2}$$

where the conditional means and variances of $\overline{C}_2$ and $\overline{C}_3$ are given in Proposition 4.1.

The approach we use is to obtain blocks, or subsets, of vertices is *community detection algorithms*. The problem of detecting communities in networks has received a lot of recent interest, for example, Arias-Castro and Verzelen (2013) and Birmele et al. (2012). There are different types of algorithms to find the

community structure, for instance, *FastGreedy*, *Walktrap*, *edge.betweenness* and *spinglass* algorithms; see e.g. Lancichinetti and Fortunato (2009). We use the *igraph* package to implement the algorithms. Community structure refers to the occurrence of subsets of vertices in a graph that are more densely connected internally than with the rest of the graph. This inhomogeneity of connections suggests that the network has certain natural divisions within it. We give very partial descriptions, so the reader might want to search the review by Fortunato (2010) to find more information concerning community detection.
First of all we apply the approach to a simulated Erdös-Rényi-Gilbert random graph in order to illustrate and clarify what the procedure is. The main output is in Table 5.2. Then we apply the procedure to a real dataset, the main outputs here being Table 5.3, Figure 5.12 and Figure 5.13.

#### 5.6.1.1 Block Graph Models in a Simulated Random Graph

We use the R program, in particular the *igraph* package, to generate an Erdös-Rényi-Gilbert random graph. Set $n = 20$, $m = 24$, i.e. $RG(20, 24)$, then $N = n(n-1)/2 = 190$, is the maximum number of edges, and $p = m/N = 0.1263158$; see Figure 5.10. The standardized statistics $Z_2$ and $Z_3$ are defined in (5.6.2), and we obtained the following results for the whole random graph.

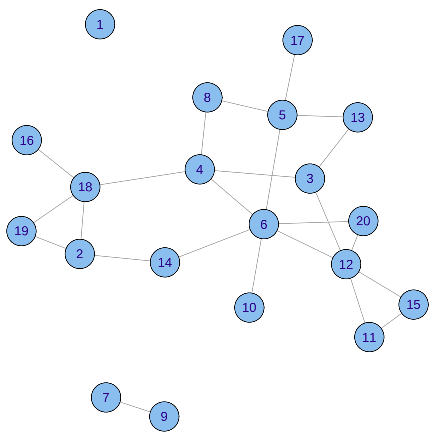


**Figure 5.10:** A simulated Erdos-Renyi-Gilbert random graph with $n = 20$ vertices and $m = 24$ edges.

| n | m | p | 2-stars | triangles | $Z_2$ | $Z_3$ | $Z$ |
|---|---|---|---|---|---|---|---|
| 20 | 24 | 0.1263158 | 56 | 3 | 0.5312366 | 0.4170583 | 0.45615 |

We notice the random graph in Figure 5.10 has 56 2-stars and 3 triangles, while the probability of an edge being present is $p = 0.13$. Then we implemented the Fastgreedy algorithm, see Lancichinetti and Fortunato (2009), to partition the graph into subgraphs with dense connections within the subgroups and sparser connections between them. Thus we get 6 types of vertices as shown in Figure 5.11 with difference sizes as Table 5.1.

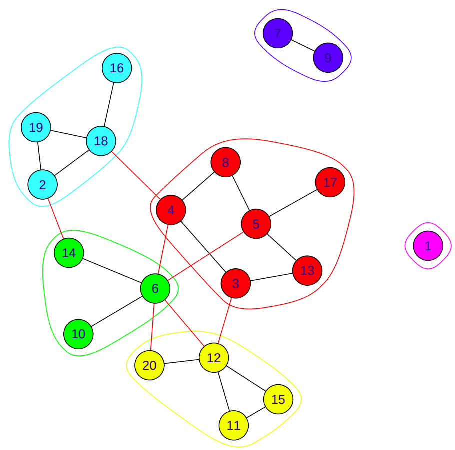


**Figure 5.11:** Blocks obtained using the Fastgreedy algorithm in Erdos-Renyi-Gilbert random graph with $n = 20$ vertices and $m = 24$ edges.

**Table 5.1:** The sizes of blocks in Erdös-Rényi-Gilbert random graph with $n = 20$ vertices and $m = 24$ edges.

| Block | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $A_5$ | $A_6$ |
|---|---|---|---|---|---|---|
| Size | 6 | 4 | 3 | 4 | 2 | 1 |

Then we calculated $Z_2$ and $Z_3$ within the subgraphs. Thus we obtained the following results in Table 5.2.

**Table 5.2:** The results of block graph models of Erdös-Rényi-Gilbert random graph with $n = 20$ vertices and $m = 24$ edges.

| | $n$ | $m$ | $p$ | $Z_2$ | $Z_3$ | $Z$ |
|---|---|---|---|---|---|---|
| $A_1$ | 6 | 6 | 0.4 | -0.845294 | -0.0200631 | 0.7149244 |
| $A_2$ | 4 | 4 | 0.6666667 | 0.2598076 | 4.914392 | 24.21875 |
| $A_3$ | 3 | 2 | 0.6666667 | 0 | 10.25305 | 105.125 |
| $A_4$ | 4 | 4 | 0.6666667 | 0.2598076 | 4.914392 | 24.21875 |
| $A_5$ | 2 | 1 | 1 | NA | NA | NA |
| $A_6$ | 1 | 0 | NA | NA | NA | NA |

From the results in Table 5.2, we found the values of $Z_2$ within the subgraphs is reasonable and consistence with Theorem 5.1, the conditional central limit theorem. However, most the values of $Z_3$ within the subgraphs are large and are not consistence with Theorem 5.1. We believe this is a small sample effect. We note that, when the number of 2-stars or triangle equal zero, $Z_2$ or $Z_3$ , respectively, yield NA, meaning ' Not Available'.

#### 5.6.1.2 Block Graph Models applies to a Real Data Set

We use the *immuno* dataset from built-in data set in *blockmodeling* package. We first of all calculated $Z_2$ and $Z_3$ using the full dataset and obtained the following results.

| dataset | n | m | p | 2-stars | triangles | $Z_2$ | $Z_3$ |
|---|---|---|---|---|---|---|---|
| *immuno* | 1316 | 6300 | 0.0073 | 58656 | 9485 | -6.597 | 781.6 |

Both $Z_2$ and $Z_3$ are outside the numerical range, but note that $Z_3$ is much further outside than $Z_2$.

To investigate the goodness-of-fit test in block models, we analyse community structures of this dataset. A network is said to have community structure if it can be divided into subsets of vertices with dense connections within the subsets

and sparser connections between the subsets.
We used algorithms for community detection to partition the set of vertices with dense connections internally and spares connections between communities. In particular, the *Fastgreedy* algorithm was implemented and produced the following block sizes.

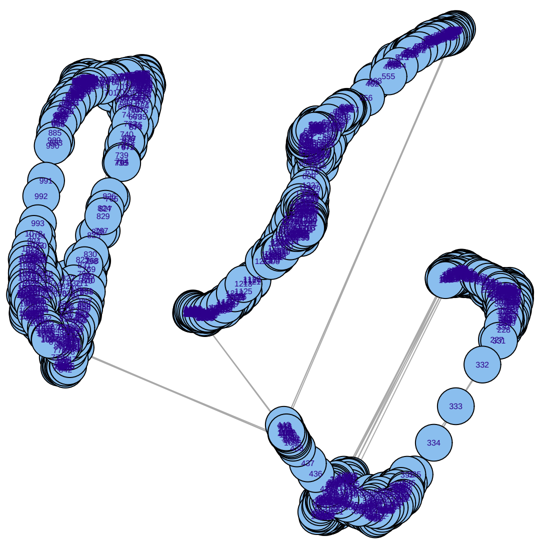

**Figure 5.12:** Graph of immuno dataset.

| Block | $A_1$ | $A_2$ | $A_3$ | $A_4$ | $A_5$ | $A_6$ |
|---|---|---|---|---|---|---|
| Size | 136 | 106 | 355 | 100 | 299 | 320 |

We also implement *Walktrap* algorithm for forming blocks. This gave different block sizes but broadly similar results in terms of goodness-of-fit.

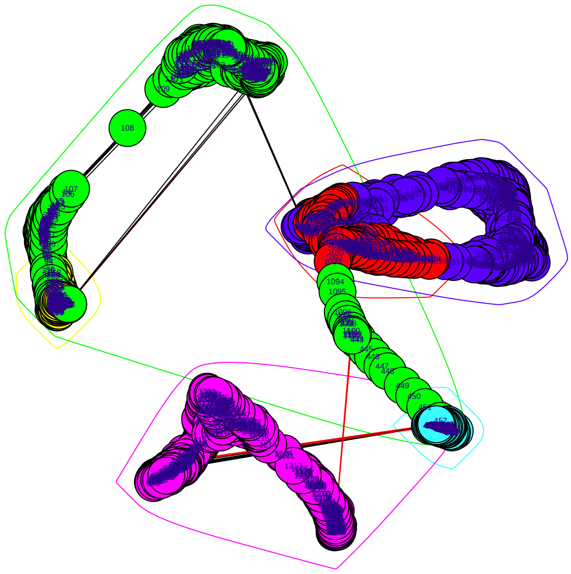

**Figure 5.13:** Applying Fastgreedy algorithm in immuno dataset.

**Table 5.3:** The results of block graph models of immuno dataset to obtain 6 blocks.

| | n | m | $p$ | $Z_2$ | $Z_3$ | $Z$ |
|---|---|---|---|---|---|---|
| $A_1$ | 136 | 597 | 0.06503268 | -1.739917 | 79.61699 | 6341.892 |
| $A_2$ | 160 | 503 | 0.09038634 | -2.010289 | 61.87462 | 3832.51 |
| $A_3$ | 355 | 1644 | 0.02616376 | -1.588977 | 210.0418 | 44120.08 |
| $A_4$ | 100 | 450 | 0.09090909 | -0.851095 | 58.47941 | 3420.565 |
| $A_5$ | 299 | 1454 | 0.03263675 | -1.797357 | 179.9231 | 32375.54 |
| $A_6$ | 320 | 1489 | 0.0291732 | -3.072335 | 187.5388 | 35180.25 |

If the Erdös-Rényi-Gilbert model, referred to below as the null model, holds within subgraphs then $Z_2$ and $Z_3$ in each row will each be standard normal and independent. It is noteworthy that in all rows in Table 5.3, $Z_2$ is either within range of a standard normal or just outside, indicating that the number of 2-stars is not much different to what we would expect under the null model, given the number of edges present in each subgraph. In contrast,the statistic $Z_3$ is way out of the range of a standard normal, indicating there are many more triangles than one would expect under the null model. Similar findings - that $Z_2$ is approximately within rang and negative, while $Z_3$ is way out of the range and positive - were obtained when looking at other dataset. This suggest that the triangle statistic will often be more sensitive to departures from the null model than the 2-star statistic.

## 5.7 Summary

In this chapter we proved Theorem 5.1, a conditional central limit theorem for the number of 2-stars and the number of triangles given the number of edges. The result was proved under the Erdös-Rényi-Gilbert random graph model assuming that the number of vertices, $n$, goes to infinity. In Section 5.6 we applied a goodness-of-fit statistic to blocks in a fitted block model based on real-world network data.

The main purpose of the following chapter is to explore three new composite likelihood methods for Exponential Random Graph Model (ERGM) defined in (6.1.1) and compare their performances.

Chapter 6

# Composite Likelihood for Exponential Random Graph Models

## 6.1 Introduction

Exponential random graph models (*ERGMs*) have already been reviewed in Chapter 2; see Section 2.2 and Section 2.6. In this chapter we focus on a widely-studied 3-parameter *ERGM* which was mentioned in (2.2.10) above and is stated here again for convenience:

$$P_\theta\{Y = y\} = \exp\{\theta_1 u_1(y) + \theta_2 u_2(y) + \theta_3 u_3(y) - \psi(\theta)\} \tag{6.1.1}$$

where $y = (y_{ij})_{1\leqslant i<j\leqslant n}$ is the adjacency matrix of the random graph of $n$ vertices, the parameter vector to be estimated is $\theta = (\theta_1, \theta_2, \theta_3)^T$, $u_1(y) = \sum_{1\leqslant i<j\leqslant n} y_{ij}$ is the number of edges, $u_2(y) = \sum_{i=1}^{n} \sum_{i\neq j<k\neq i} y_{ij}y_{ik}$ is the number of 2-stars and $u_3(y) = \sum_{1\leqslant i<j<k\leqslant n} y_{ij}y_{jk}y_{ki}$ is the number of triangles.
Exact maximum likelihood estimation of $\theta$ is intractable unless $n$, the number of vertices, is quite small, because evaluation of $\psi(\theta)$ requires summation of $2^N$ terms where $N = n(n-1)/2$. A widely-used method of parameter estimation for model (6.1.1) is composite likelihood; see Section 2.5 of Chapter 2. However, so far as we are aware, the only type of composite likelihood that has been used for estimating the vector parameter, $\theta$, in (6.1.1) is the one where the component likelihoods are the conditional distribution of each edge given knowledge of all

the other edges. More specifically, the composite likelihood $L_c^{[1]}(\theta, y))$ given by

$$L_c^{[1]}(\theta; y) = \prod_{1 \leqslant i < j \leqslant n} L_{ij}(\theta), \tag{6.1.2}$$

where $L_{ij}(\theta) = P(y_{ij}|rest, \theta)$, where 'rest' means all $y_{\alpha\beta}$ with $\alpha < \beta$ and $(\alpha, \beta) \neq (i, j)$. This composite likelihood has been used by a number of authors; see for example Snijders and Van Duijn (2002).

The purpose of this chapter is to explore three new composite likelihoods for model (6.1.1) and compare their performance with that of (6.1.2). The first of these new composite likelihoods is:

$$L_c^{[2]}(\theta; y) = \prod_{i=1}^{n} \prod_{i \neq j < k \neq i} L_{i|jk}(\theta), \tag{6.1.3}$$

where $L_{i|jk}(\theta) = P(y_{ij}, y_{ik}|rest, \theta)$, and 'rest' now means all $y_{\alpha\beta}$, $1 \leqslant \alpha < \beta \leqslant n)$, except for $y_{ij}$ and $y_{ik}$. The second new composite likelihood is given by

$$L_c^{[3]}(\theta; y) = \prod_{1 \leqslant i < j < k \leqslant n} L_{ijk}(\theta), \tag{6.1.4}$$

where $L_{ijk}(\theta) = P(y_{ij}, y_{jk}, y_{ki}|rest, \theta)$, and 'rest' now means all $y_{\alpha\beta}$, $1 \leqslant \alpha < \beta \leqslant n)$, except for $y_{ij}$, $y_{jk}$ and $y_{ki}$. Finally, The third new composite likelihood is given by

$$L_c^{[4]}(\theta; y) = \prod_{1 \leqslant i < j < k < l \leqslant n} L_{ijkl}(\theta), \tag{6.1.5}$$

where $L_{ijkl}(\theta) = P(y_{ij}, y_{ik}, y_{il}, y_{jk}, y_{jl}, y_{kl}|rest, \theta)$, and 'rest' now means all $y_{\alpha\beta}$, $1 \leqslant \alpha < \beta \leqslant n)$, except for $y_{ij}$, $y_{jk}$,$y_{il}$, $y_{jk}$, $y_{jl}$, $y_{kl}$.

The motivation for considering the new composite likelihoods (6.1.3)-(6.1.5) is that, including the joint distribution on two or more $y_{ij}$'s conditional on the rest we may hope to retain more information concerning the dependence structure of the $y_{ij}$. As will be seen, from computational point of view, the composite likelihoods (6.1.3)-(6.1.5) are tractable. Unfortunately, the standard asymptotic theory for composite likelihood in Chapter 2 can not be applied because the required independence assumptions do not hold, and the asymptotic theory of (6.1.3)-(6.1.5) as $n \to \infty$ is unknown. However, it is possible to study the nu-

merical performance of the new estimators in simulation studies and this we do later in the chapter.
The outline of this chapter is as follows. In Section 6.2 we determine the composite likelihoods in (6.1.2)-(6.1.5) more explicitly and provide algorithms for calculation them. In Section 6.3 we present simulation results concerning the performance of the estimators. Finally, we briefly explore connections with the work of Chatterjee and Diaconis (2013).

## 6.2 Composite Likelihood for *ERGMs*

In this section we derive explicit expressions for the composite likelihoods (6.1.2)-(6.1.5). We also present computational algorithms for calculating them.

### 6.2.1 Calculation of Composite Likelihood (6.1.2)

Define

$$P(y_{ij}|rest,\theta) = P_\theta\{Y_{ij} = y_{ij}|Y_{hk} = y_{hk} \text{ for all } \{h,k\} \neq \{i,j\}\}.$$

Then

$$\begin{aligned} P(y_{ij}|rest,\theta) &\propto \exp\{\theta_1 y_{ij} + \theta_2 y_{ij}(\sum_{k\neq i,j}(y_{ik}+y_{jk}) + \theta_3 y_{ij}\sum_{k\neq i,j} y_{ik}y_{jk}\} \\ &\propto \exp\{\theta_1 y_{ij} + \theta_2 y_{ij}S_{ij} + \theta_3 y_{ij}T_{ij}\} \\ &\propto \exp\{y_{ij}\widetilde{\theta}_{ij}\} \end{aligned}$$

where

$$\widetilde{\theta}_{ij} = \theta_1 + \theta_2 S_{ij} + \theta_3 T_{ij} \tag{6.2.1}$$

where

$$S_{ij} = \sum_{k\neq i,j}(y_{ik}+y_{jk}), \tag{6.2.2}$$

and

$$T_{ij} = \sum_{k \neq i,j} y_{ik} y_{jk} \qquad (6.2.3)$$

Therefore,

$$P(y_{ij}|rest, \theta) = \frac{\exp\{y_{ij}\widetilde{\theta}_{ij}\}}{1 + \exp\{\widetilde{\theta}_{ij}\}}.$$

**Lemma 6.1**

*The composite log-likelihood function, $l_1(\theta)$, for the composite likelihood (6.1.2)*

$$l_1(\theta) = \sum_{1 \leq i < j \leq n} \left( y_{ij}\widetilde{\theta}_{ij} - \log\{1 + \exp(\widetilde{\theta}_{ij})\} \right),$$

*where $\widetilde{\theta}_{ij}$ is defined in (6.2.1).*

An algorithm for calculating $l_1(\theta)$ is now given in Algorithm 1.

---

**Algorithm 1** Calculation of the composite likelihood $l_1(\theta)$ in (6.1.2).

---

**Step 0** Input the adjacency matrix $y = (y_{ij})_{1 \leq i,j \leq n}$ of an undirected graph,

**Step 1** Calculate $y_{i+} = \sum_{k=1}^{n} y_{ik}$, $i = 1, \ldots, n$, and $T(i,j) = \sum_{k=1}^{n} y_{ik} y_{jk}$, $1 \leq i < j \leq n$,

**Step 2** Set $l_1 = 0$,

**Step 3** For $1 \leq i < j \leq n$, calculate
$D = \theta_1 + \theta_2(y_{i+} - y_{ij} + y_{j+} - y_{ij}) + \theta_3 T(i,j)$,
$l_1 = l_1 + (y_{ij} D - log(c))$
where $c = 1 + e^D$,

**Step 4** Return $l_1$.

---

Note that, Algorithm 1 avoids the storage of the $S_{ij}$ and $T_{ij}$, which results in greater computationally efficiency.

### 6.2.2 Calculation of Composite Likelihood (6.1.3)

Continuing with the model (6.1.1), we now explore composite likelihood based on pairs $y_{ij}$ and $y_{\alpha\beta}$ given the rest. These probabilities are of the form

$$\begin{aligned} & P(y_{ij}, y_{\alpha\beta}|rest, \theta) \\ &= P_\theta\{Y_{ij} = y_{ij}, Y_{\alpha\beta} = y_{\alpha\beta}|Y_{hk} = y_{hk} \text{ for all } \{h,k\} \neq \{i,j\} \text{ and } \{h,k\} \neq \{\alpha,\beta\}\}, \end{aligned}$$

and we have two cases:

**Case 1**: $\alpha$ and $\beta$ are both different to $i$ and $j$.

**Case 2**: One of $\alpha$ and $\beta$ is equal to one of $i$ or $j$.

The composite likelihood (6.1.3) in Case 1 is

$$\begin{aligned} P(y_{ij}, y_{\alpha\beta}|rest, \theta) &\propto \exp\{\theta_1(y_{ij} + y_{\alpha\beta}) + \theta_2(y_{ij} \sum_{k\neq i,j}(y_{ik} + y_{jk}) + y_{\alpha\beta} \sum_{k\neq\alpha,\beta}(y_{\alpha k} + y_{\beta k})) + \\ &\quad \theta_3(y_{ij} \sum_{k\neq i,j} y_{ik}y_{jk} + y_{\alpha\beta} \sum_{k\neq\alpha,\beta} y_{\alpha k}y_{\beta k})\} \\ &\propto \exp\{\theta_1(y_{ij} + y_{\alpha\beta}) + \theta_2(y_{ij}S_{ij} + y_{\alpha\beta}S_{\alpha\beta}) + \theta_3(y_{ij}T_{ij} + y_{\alpha\beta}T_{\alpha\beta})\} \\ &\propto \exp\{y_{ij}\widetilde{\theta}_{ij} + y_{\alpha\beta}\widetilde{\theta}_{\alpha\beta}\} \end{aligned}$$

where

$$\begin{aligned} \widetilde{\theta}_{ij} &= \theta_1 + \theta_2 S_{ij} + \theta_3 T_{ij}, \\ \widetilde{\theta}_{\alpha\beta} &= \theta_1 + \theta_2 S_{\alpha\beta} + \theta_3 T_{\alpha\beta}, \end{aligned}$$

where $S_{ij}$ and $S_{\alpha\beta}$ are defined by (6.2.2) and $T_{ij}$ and $T_{\alpha\beta}$ are defined by (6.2.3). We notice that, in Case 1, the two random variable $y_{ij}$ and $y_{\alpha\beta}$ are treated as independent. Therefore, it is unlikely that Case 1 will lead to an improvement over $l_1(\theta)$ in Lemma 6.1. We therefore explore Case 2.

Consider Case 2 with $A = \{\{i,\alpha\},\{i,\beta\}\}$, $y_A = \{y_{u,v} : \{u,v\} \in A\}$ and $y_{A^c} = \{y_{u,v} : \{u,v\} \in A^c\}$, where $A^c$ is the compliment of $A$ in the edge set $E$. In this case the conditional distribution of $y_{A^c}$ is

$$\begin{aligned}
P_\theta\{Y_A = y_A | Y_{A^c} = y_{A^c}\} \quad &\propto \quad \exp\{\theta_1(y_{i\alpha} + y_{i\beta}) + \theta_2(y_{i\alpha}y_{i\beta} + y_{i\alpha}(\sum_{k\neq i,\alpha,\beta} y_{ik} + \sum_{k\neq i,\alpha} y_{\alpha k}) \\
&\quad + y_{i\beta}(\sum_{k\neq i,\alpha,\beta} y_{ik} + \sum_{k\neq i,\beta} y_{\beta k})) + \theta_3(y_{i\alpha}y_{i\beta}y_{\alpha\beta} \\
&\quad + y_{i\alpha}\sum_{k\neq i,\alpha,\beta} y_{ik}y_{\alpha k} + y_{i\beta}\sum_{k\neq i,\alpha,\beta} y_{ik}y_{\beta k})\} \\
&\propto \quad \exp\{\theta_1(y_{i\alpha} + y_{i\beta}) + \theta_2(y_{i\alpha}y_{i\beta} + y_{i\alpha}S_{i\alpha,\beta} + y_{i\beta}S_{i\beta,\alpha}) \\
&\quad + \theta_3(y_{i\alpha}y_{i\beta}y_{\alpha,\beta} + y_{i\alpha}T_{i\alpha,\beta} + y_{i\beta}T_{i\beta,\alpha})\} \\
&\propto \quad \exp\{y_{i\alpha}\widetilde{\theta}_{1,i\alpha,\beta} + y_{i\beta}\widetilde{\theta}_{1,i\beta,\alpha} + y_{i\alpha}y_{i\beta}\widetilde{\theta}_{2,i\alpha\beta}\}
\end{aligned}$$

where

$$\widetilde{\theta}_{1,i\alpha,\beta} = \theta_1 + \theta_2 S_{i\alpha,\beta} + \theta_3 T_{i\alpha,\beta}, \tag{6.2.4a}$$

$$\widetilde{\theta}_{1,i\beta,\alpha} = \theta_1 + \theta_2 S_{i\beta,\alpha} + \theta_3 T_{i\beta,\alpha}. \tag{6.2.4b}$$

$$\widetilde{\theta}_{2,i\alpha\beta} = \theta_2 + \theta_3 y_{\alpha\beta}, \tag{6.2.4c}$$

and

$$S_{i\alpha,\beta} = \sum_{k\neq i,\alpha,\beta} y_{ik} + \sum_{k\neq i,\alpha} y_{\alpha k} = y_{i+} + y_{\alpha+} - 2y_{i\alpha} - y_{i\beta} \tag{6.2.5}$$

and

$$T_{i\alpha,\beta} = \sum_{k\neq i,\alpha,\beta} y_{ik}y_{\alpha k} = T_{i\alpha} - y_{i\beta}y_{\alpha\beta}, \tag{6.2.6}$$

with corresponding definition for $S_{i\beta,\alpha}$ and $T_{i\beta,\alpha}$. Therefore

$$P_\theta\{Y_A = y_A | Y_{A^c} = y_{A^c}\} = \frac{\exp\{y_{i\alpha}\widetilde{\theta}_{1,i\alpha} + y_{i\beta}\widetilde{\theta}_{1,i\beta} + y_{i\alpha}y_{i\beta}\widetilde{\theta}_{2,i\alpha\beta}\}}{\sum_{y_{i\alpha}=0}^{1}\sum_{y_{i\beta}=0}^{1} \exp\{y_{i\alpha}\widetilde{\theta}_{1,i\alpha} + y_{i\beta}\widetilde{\theta}_{1,i\beta} + y_{i\alpha}y_{i\beta}\widetilde{\theta}_{2,i\alpha,\beta}\}}. \tag{6.2.7}$$

From (6.2.7), we construct a composite log-likelihood function as follows.

**Lemma 6.2**

*The composite log-likelihood function, $l_2(\theta)$, for the composite likelihood (6.1.3) is given by*

$$l_2(\theta) = \sum_{i=1}^{n} \sum_{\alpha<\beta, i\neq\alpha,\beta} y_{i\alpha}\widetilde{\theta}_{1,i\alpha,\beta} + y_{i\beta}\widetilde{\theta}_{1,i\beta,\alpha} + y_{i\alpha}y_{i\beta}\widetilde{\theta}_{2,i\alpha\beta} - \log(c_{i\alpha,i\beta})$$

*where $c_{i\alpha,i\beta}$ is the normalization constant*

$$c_{i\alpha,i\beta} = \sum_{y_{i\alpha}=0}^{1} \sum_{y_{i\beta}=0}^{1} \exp\{y_{i\alpha}\widetilde{\theta}_{1,i\alpha,\beta} + y_{i\beta}\widetilde{\theta}_{1,i\beta,\alpha} + y_{i\alpha}y_{i\beta}\widetilde{\theta}_{2,i\alpha,\beta}\}$$

*where $\widetilde{\theta}_{1,i\alpha,\beta}$, $\widetilde{\theta}_{1,i\beta,\alpha}$, and $\widetilde{\theta}_{2,i\alpha,\beta}$ are defined in* (6.2.4).

An algorithm for calculating $l_2(\theta)$ is now given in Algorithm 2.

**Algorithm 2** Calculation of the composite likelihood $l_2(\theta)$ in (6.1.3).

**Step 0** Input the adjacency matrix $y = (y_{ij})_{1\le i,j\le n}$, of an undirected graph,

**Step 1** Calculate $y_{i+} = \sum_{k=1}^{n} y_{ik}$, $i = 1,\dots,n$, and $T(i,j) = \sum_{k=1}^{n} y_{ik}y_{jk}$, $1 \le i < j \le n$.

**Step 2** Set $l_2 = 0$.

**Step 3** For $i = 1,\dots,n$, $\alpha < \beta$, $i \ne \alpha$, $i \ne \beta$, calculate
$D_1 = \theta_1 + \theta_2[(y_{i+} - y_{i\alpha} - y_{i\beta}) + (y_{\alpha+} - y_{i\alpha})] + \theta_3(T(i,\alpha) - y_{i\beta}y_{\alpha\beta})$
$D_2 = \theta_1 + \theta_2[(y_{i+} - y_{i\alpha} - y_{i\beta}) + (y_{\beta+} - y_{i\beta})] + \theta_3(T(i,\beta) - y_{i\alpha}y_{\beta\alpha})$
$D_3 = \theta_2 + \theta_3 y_{\alpha,\beta}$
$l_2 = l_2 + D_1 y_{i\alpha} + D_2 y_{i\beta} + D_3 y_{i\alpha} y_{i\gamma} - \log(c)$
where $c = 1 + e^{D_1} + e^{D_2} + e^{D_1+D_2+D_3}$

**Step 4** Return $l_2$.

### 6.2.3 Calculation of Composite Likelihood (6.1.4)

Let $A = \{\{i,\alpha\},\{i,\beta\},\{\alpha,\beta\}\}$, $y_A = \{y_{u,v} : \{u,v\} \in A\}$, $y_{A^c} = \{y_{u,v} : \{u,v\} \in A^c\}$, and define

$$P(y_A|y_{A^c,\theta}) = P_\theta\{Y_A = y_A|Y_{A^c} = y_{A^c}\}.$$

Than

$$\begin{aligned} P(y_A|y_{A^c,\theta}) &\propto \exp\{\theta_1(y_{i\alpha}+y_{i\beta}+y_{\alpha\beta})+\theta_2((y_{i\alpha}y_{i\beta}+y_{i\alpha}y_{\alpha\beta}+y_{i\beta}y_{\alpha\beta}) \\ &\quad +y_{i\alpha}\sum_{k\neq i,\alpha,\beta}(y_{ik}+y_{\alpha k})+y_{i\beta}\sum_{k\neq i,\alpha,\beta}(y_{ik}+y_{\beta k})+y_{\alpha\beta}\sum_{k\neq i,\alpha,\beta}(y_{\alpha k}+y_{\beta k})) \\ &\quad +\theta_3((y_{i\alpha}y_{i\beta}y_{\alpha\beta})+y_{i\alpha}\sum_{k\neq i,\alpha,\beta}y_{ik}y_{\alpha k}+y_{i\beta}\sum_{k\neq i,\alpha,\beta}y_{ik}y_{\beta k}+y_{\alpha\beta}\sum_{k\neq i,\alpha,\beta}y_{\alpha k}y_{\beta k})\} \\ &\propto \exp\{\theta_1(y_{i\alpha}+y_{i\beta}+y_{\alpha\beta})+\theta_2((y_{i\alpha}y_{i\beta}+y_{i\alpha}y_{\alpha\beta}+y_{i\beta}y_{\alpha\beta}) \\ &\quad +y_{i\alpha}\grave{S}_{i\alpha,\beta}+y_{i\beta}\grave{S}_{i\beta,\alpha}+y_{\alpha\beta}\grave{S}_{\alpha\beta,i})+\theta_3((y_{i\alpha}y_{i\beta}y_{\alpha\beta}) \\ &\quad +y_{i\alpha}T_{i\alpha,\beta}+y_{i\beta}T_{i\beta,\alpha}+y_{\alpha\beta}T_{\alpha\beta,i})\} \\ &\propto \exp\{y_{i\alpha}\widetilde{\theta}_{1,i\alpha}+y_{i\beta}\widetilde{\theta}_{1,i\beta}+y_{\alpha\beta}\widetilde{\theta}_{1,\alpha\beta}+\theta_2(y_{i\alpha}y_{i\beta}+y_{i\alpha}y_{\alpha\beta}+y_{i\beta}y_{\alpha\beta}) \\ &\quad +\theta_3(y_{i\alpha}y_{i\beta}y_{\alpha\beta})\} \end{aligned}$$

where

$$\widetilde{\theta}_{1,i\alpha,\beta} = \theta_1+\theta_2\grave{S}_{i\alpha,\beta}+\theta_3T_{i\alpha,\beta}, \tag{6.2.8a}$$

$$\widetilde{\theta}_{1,i\beta,\beta} = \theta_1+\theta_2\grave{S}_{i\beta,\alpha}+\theta_3T_{i\beta,\alpha}, \tag{6.2.8b}$$

$$\widetilde{\theta}_{1,\alpha\beta,i} = \theta_1+\theta_2\grave{S}_{\alpha\beta,i}+\theta_3T_{\alpha\beta,i}, \tag{6.2.8c}$$

where

$$\grave{S}_{i\alpha,\beta} = \sum_{k\neq i,\alpha,\beta}(y_{ik}+y_{\alpha k}), \tag{6.2.9}$$

with corresponding definition for $\grave{S}_{i\beta,\alpha}$ and $\grave{S}_{\alpha\beta,i}$, and $T_{i\alpha,\beta}$ is defined in (6.2.6) with corresponding definition for $T_{i\beta,\alpha}$ and $T_{\alpha\beta,i}$. Therefore

$$P_\theta\{Y_A=y_A|Y_{A^c}=y_{A^c}\} = c_{i\alpha\beta}^{-1}\exp\{y_{i\alpha}\widetilde{\theta}_{1,i\alpha}+y_{i\beta}\widetilde{\theta}_{1,i\beta}+y_{\alpha\beta}\widetilde{\theta}_{1,\alpha\beta}+\theta_2(y_{i\alpha}y_{i\beta}+y_{i\alpha}y_{\alpha\beta}+y_{i\beta}y_{\alpha\beta})+\theta_3y_{i\alpha}y_{i\beta}y_{\alpha\beta}\} \tag{6.2.10}$$

where $c_{i\alpha\beta}$ is the normalization constant given by

$$c_{i\alpha\beta} = \sum_{yi\alpha=0}^{1} \sum_{yi\beta=0}^{1} \sum_{y\alpha\beta=0}^{1} \exp\{y_{i\alpha}\widetilde{\theta}_{1,i\alpha} + y_{i\beta}\widetilde{\theta}_{1,i\beta} + y_{\alpha\beta}\widetilde{\theta}_{1,\alpha\beta} + \theta_2(y_{i\alpha}y_{i\beta} + y_{i\alpha}y_{\alpha\beta} + y_{i\beta}y_{\alpha\beta}) + \theta_3(y_{i\alpha}y_{i\beta}y_{\alpha\beta})\} \quad (6.2.11)$$

From 6.2.10, we construct a composite log-likelihood function as follows.

**Lemma 6.3**
*The composite log-likelihood function, $l_3(\theta)$, for the Calculation of composite likelihood (6.1.4) as following*

$$l_3(\theta) = \sum_{i<\alpha<\beta} y_{i\alpha}\widetilde{\theta}_{1,i\alpha} + y_{i\beta}\widetilde{\theta}_{1,i\beta} + y_{\alpha\beta}\widetilde{\theta}_{1,\alpha\beta} + \theta_2(y_{i\alpha}y_{i\beta} + y_{i\alpha}y_{\alpha\beta} + y_{i\beta}y_{\alpha\beta}) + \theta_3(y_{i\alpha}y_{i\beta}y_{\alpha\beta}) - \log(c_{i\alpha\beta})$$

*where $c_{i\alpha\beta}$ is defined in (6.2.11), and $\widetilde{\theta}_{1,i\alpha}$, $\widetilde{\theta}_{1,i\beta}$, and $\widetilde{\theta}_{2,i\alpha,\beta}$ are defined in* (6.2.8).

A convenient algorithm for calculating $l_3(\theta)$ is given in Algorithm 3.

**Algorithm 3** Calculation of the composite likelihood $l_3(\theta)$ in (6.1.4).

**Step 0** Input the adjacency matrix $y = (y_{ij})_{1\leq i,j\leq n}$, of an undirected graph,

**Step 1** Calculate $y_{i+} = \sum_{k=1}^{n} y_{ik}$, $i = 1,\ldots,n$, and $T(i,j) = \sum_{k=1}^{n} y_{ik}y_{jk}$, $1 \leq i < j \leq n$,

**Step 2** Set $l_3 = 0$,

**Step 3** For $1 \leq i < \alpha < \beta \leq n$, calculate

$$\begin{aligned} D_1 &= \theta_1 + \theta_2[(y_{i+} - y_{i\alpha} - y_{i\beta}) + (y_{\alpha+} - y_{\alpha i} - y_{\alpha\beta})] + \theta_3\{T(i,\alpha) - y_{i\beta}y_{\alpha\beta}\} \\ D_2 &= \theta_1 + \theta_2[(y_{i+} - y_{i\alpha} - y_{i\beta}) + (y_{\beta+} - y_{\beta i} - y_{\beta\alpha})] + \theta_3\{T(i,\beta) - y_{i\alpha}y_{\beta\alpha}\} \\ D_3 &= \theta_1 + \theta_2[(y_{\alpha+} - y_{\alpha i} - y_{\alpha\beta}) + (y_{\beta+} - y_{\beta i} - y_{\beta\alpha})] + \theta_3\{T(\alpha,\beta) - y_{\alpha i}y_{\beta i}\} \end{aligned}$$

**Step 4** Calculate

$$\begin{aligned} c = \sum_{yi\alpha=0}^{1}\sum_{yi\beta=0}^{1}\sum_{y\alpha\beta=0}^{1} & \exp\{y_{i\alpha}\widetilde{\theta}_{1,i\alpha,\beta} + y_{i\beta}\widetilde{\theta}_{1,i\beta,\alpha} + y_{\alpha\beta}\widetilde{\theta}_{1,\alpha\beta,i} \\ & + \theta_2(y_{i\alpha}y_{i\beta} + y_{i\alpha}y_{\alpha\beta} + y_{i\beta}y_{\alpha\beta}) + \theta_3 y_{i\alpha}y_{i\beta}y_{\alpha\beta} \end{aligned}$$

and

$$\begin{aligned} l_3 = l_3 + D_1 y_{i\alpha} + D_2 y_{i\beta} + D_3 y_{\alpha\beta} + \theta_2(y_{i\alpha}y_{i\beta} + y_{i\alpha}y_{\alpha\beta} + y_{i\beta}y_{\alpha\beta}) \\ + \theta_3 y_{i\alpha}y_{i\beta}y_{\alpha\beta} - \log(c) \end{aligned}$$

**Step 5** Return $l_3$.

### 6.2.4 Calculation of Composite Likelihood (6.1.5)

Let $A = \{\{i,j\},\{i,\alpha\},\{i,\beta\},\{j,\alpha\},\{j,\beta\},\{\alpha,\beta\}\}$, $y_A = \{y_{i,j} : \{i,j\} \in A\}$ and $y_{A^c} = \{y_{i,j} : \{i,j\} \in A^c\}$, and define

$$P(y_A|y_{A^c}) = P_\theta\{Y_A = y_A|Y_{A^c} = y_{A^c}\}.$$

Then

$$\begin{aligned}
P(y_A|y_{A^c}) \quad \propto \quad & exp\{\theta_1(y_{ij}+y_{i\alpha}+y_{i\beta}+y_{i\alpha}+y_{j\beta}+y_{\alpha\beta}) \\
& +\theta_2((y_{ij}y_{j\alpha}+y_{ij}y_{j\beta}+y_{ij}y_{i\beta}+y_{ij}y_{i\alpha}+y_{j\alpha}y_{\alpha\beta} \\
& +y_{j\alpha}y_{\alpha i}+y_{j\alpha}y_{j\beta}+y_{\alpha\beta}y_{\alpha i}+y_{\alpha\beta}y_{\beta i}+y_{\alpha\beta}y_{\beta j}+y_{\beta i}y_{\beta j}+y_{\beta i}y_{i\alpha}) \\
& +y_{ij}\sum_{k\neq i,j,\alpha,\beta}(y_{ik}+y_{jk})+y_{i\alpha}\sum_{k\neq i,j,\alpha,\beta}(y_{ik}+y_{\alpha k}) \\
& +y_{i\beta}\sum_{k\neq i,j,\alpha,\beta}(y_{ik}+y_{\beta k})+y_{j\alpha}\sum_{k\neq i,j,\alpha,\beta}(y_{jk}+y_{\alpha k}) \\
& +y_{j\beta}\sum_{k\neq i,j,\alpha,\beta}(y_{jk}+y_{\beta k})+y_{\alpha\beta}\sum_{k\neq i,\alpha,\beta}(y_{\alpha k}+y_{\beta k})) \\
& +\theta_3((y_{ij}y_{i\alpha}y_{j\alpha}+y_{ij}y_{i\beta}y_{j\beta}+y_{i\alpha}y_{i\beta}y_{\alpha\beta}+y_{j\alpha}y_{j\beta}y_{\alpha\beta}) \\
& +y_{ij}\sum_{k\neq i,j,\alpha\beta}y_{ik}y_{jk}+y_{i\alpha}\sum_{k\neq i,j,\alpha,\beta}y_{ik}y_{\alpha k}+y_{i\beta}\sum_{k\neq i,j,\alpha,\beta}y_{ik}y_{\beta k} \\
& +y_{j\alpha}\sum_{k\neq i,j,\alpha,\beta}y_{jk}y_{\alpha k}+y_{j\beta}\sum_{k\neq i,j,\alpha,\beta}y_{jk}y_{\beta k}+y_{\alpha\beta}\sum_{k\neq i,j,\alpha,\beta}y_{\alpha k}y_{\beta k})\} \\
\propto \quad & exp\{\theta_1(y_{ij}+y_{i\alpha}+y_{i\beta}+y_{j\alpha}+y_{j\beta}+y_{\alpha\beta}) \\
& +\theta_2((y_{ij}y_{j\alpha}+y_{ij}y_{j\beta}+y_{ij}y_{i\beta}+y_{ij}y_{i\alpha}+y_{j\alpha}y_{\alpha\beta} \\
& +y_{j\alpha}y_{\alpha i}+y_{j\alpha}y_{j\beta}+y_{\alpha\beta}y_{\alpha i}+y_{\alpha\beta}y_{\beta i}+y_{\alpha\beta}y_{\beta j}+y_{\beta i}y_{\beta j}+y_{\beta i}y_{i\alpha}) \\
& +y_{ij}S_{ij,\alpha\beta}y_{i\alpha}S_{i\alpha,j\beta}+y_{i\beta}S_{i\beta,j\alpha}+y_{j\alpha}S_{j\alpha,i\beta}+y_{j\beta}S_{j\beta,i\alpha}+y_{\alpha\beta}S_{\alpha\beta,i}) \\
& +\theta_3((y_{ij}y_{i\alpha}y_{j\alpha}+y_{ij}y_{i\beta}y_{j\beta}+y_{i\alpha}y_{i\beta}y_{\alpha\beta}+y_{j\alpha}y_{j\beta}y_{\alpha\beta}) \\
& +y_{ij}T_{ij,\alpha\beta}+y_{i\alpha}T_{i\alpha,\beta}+y_{i\beta}T_{i\beta,\alpha}+y_{j\alpha}T_{j\alpha,i\beta}+y_{j\beta}T_{j\beta,i\alpha}+y_{\alpha\beta}T_{\alpha\beta},i)\} \\
\propto \quad & exp\{y_{ij}\widetilde{\theta}_{1,ij}+y_{i\alpha}\widetilde{\theta}_{1,i\alpha}+y_{i\beta}\widetilde{\theta}_{1,i\beta}+y_{j\alpha}\widetilde{\theta}_{1,j\alpha}+y_{j\beta}\widetilde{\theta}_{1,j\beta}+y_{\alpha\beta}\widetilde{\theta}_{1,\alpha\beta} \\
& +\theta_2(y_{ij}y_{j\alpha}+y_{ij}y_{j\beta}+y_{ij}y_{i\beta}+y_{ij}y_{i\alpha}+y_{j\alpha}y_{\alpha\beta} \\
& +y_{j\alpha}y_{\alpha i}+y_{j\alpha}y_{j\beta}+y_{\alpha\beta}y_{\alpha i}+y_{\alpha\beta}y_{\beta i}+y_{\alpha\beta}y_{\beta j}+y_{\beta i}y_{\beta j}+y_{\beta i}y_{i\alpha}) \\
& +\theta_3(y_{ij}y_{i\alpha}y_{j\alpha}+y_{ij}y_{i\beta}y_{j\beta}+y_{i\alpha}y_{i\beta}y_{\alpha\beta}+y_{j\alpha}y_{j\beta}y_{\alpha\beta})\}
\end{aligned}$$

where

$$\begin{aligned}
S_{ij,\alpha\beta} &= \sum_{k\neq i,j,\alpha,\beta}(y_{ik}+y_{jk}), \\
T_{ij,\alpha\beta} &= \sum_{k\neq i,j,\alpha,\beta}y_{ik}y_{jk}, \qquad (6.2.12) \\
\widetilde{\theta}_{1,ij} &= \theta_1+\theta_2S_{ij,\alpha\beta}+\theta_3T_{ij,\alpha\beta},
\end{aligned}$$

with corresponding definition for $S_{i\alpha,j\beta}$, $S_{i\beta,j\alpha}$, $S_{j\alpha,i\beta}$, $S_{j\beta,i\alpha}$ and $S_{\alpha\beta,ij}$, and with corresponding definition for $T_{i\alpha,j\beta}$, $T_{i\beta,j\alpha}$, $T_{j\alpha,i\beta}$, $T_{j\beta,i\alpha}$ and $T_{\alpha\beta,ij}$, and with corresponding definition for $\widetilde{\theta}_{1,i\alpha}$, $\widetilde{\theta}_{1,i\beta}$, $\widetilde{\theta}_{1,j\alpha}$, $\widetilde{\theta}_{1,j\beta}$, $\widetilde{\theta}_{1,\alpha\beta}$.

Therefore

$$P_\theta\{Y_A = y_A|Y_{A^c} = y_{A^c}\} = c^{-1}exp\{y_{ij}\widetilde{\theta}_{1,ij} + y_{i\alpha}\widetilde{\theta}_{1,i\alpha} + y_{i\beta}\widetilde{\theta}_{1,i\beta} + y_{j\alpha}\widetilde{\theta}_{1,j\alpha} + y_{j\beta}\widetilde{\theta}_{1,j\beta} + y_{\alpha\beta}\widetilde{\theta}_{1,\alpha\beta} + \theta_2 A + \theta_3 B\}.$$

where

$$\begin{aligned} A =& y_{ij}y_{j\alpha} + y_{ij}y_{j\beta} + y_{ij}y_{i\beta} + y_{ij}y_{i\alpha} + y_{j\alpha}y_{\alpha\beta} + y_{j\alpha}y_{\alpha i} + \\ & y_{j\alpha}y_{j\beta} + y_{\alpha\beta}y_{\alpha i} + y_{\alpha\beta}y_{\beta i} + y_{\alpha\beta}y_{\beta j} + y_{\beta i}y_{\beta j} + y_{\beta i}y_{i\alpha}, \\ B =& y_{ij}y_{i\alpha}y_{j\alpha} + y_{ij}y_{i\beta}y_{j\beta} + y_{i\alpha}y_{i\beta}y_{\alpha\beta} + y_{j\alpha}y_{j\beta}y_{\alpha\beta}, \\ c =& \sum_{y_{ij}=0}^{1}\sum_{y_{i\alpha}=0}^{1}\sum_{y_{i\beta}=0}^{1}\sum_{y_{j\alpha}=0}^{1}\sum_{y_{j\beta}=0}^{1}\sum_{y_{\alpha\beta}=0}^{1} \exp\{y_{ij}\widetilde{\theta}_{1,ij} + y_{i\alpha}\widetilde{\theta}_{1,i\alpha} \\ & + y_{i\beta}\widetilde{\theta}_{1,i\beta} + y_{j\alpha}\widetilde{\theta}_{1,j\alpha} + y_{j\beta}\widetilde{\theta}_{1,j\beta} + y_{\alpha\beta}\widetilde{\theta}_{1,\alpha\beta} + \theta_2 A + \theta_3 B\} \end{aligned} \tag{6.2.13}$$

**Lemma 6.4**

*The composite log-likelihood function, $l_c(\theta)$, for the composite likelihood (6.1.5) as following*

$$l_c(\theta) = \sum_{i<j<\alpha<\beta} L_A(\theta)$$

*where $L_A(\theta)$ is the log of the composite likelihood (6.1.5)*

$$L_A(\theta) = y_{ij}\widetilde{\theta}_{1,ij} + y_{i\alpha}\widetilde{\theta}_{1,i\alpha} + y_{i\beta}\widetilde{\theta}_{1,i\beta} + y_{j\alpha}\widetilde{\theta}_{1,j\alpha} + y_{j\beta}\widetilde{\theta}_{1,j\beta} + y_{\alpha\beta}\widetilde{\theta}_{1,\alpha\beta} + \theta_2 A + \theta_3 B - \log(c)$$

*where $A$, $B$ and $c$ are given in (6.2.13), and $\widetilde{\theta}_{1,ij}, \widetilde{\theta}_{1,i\alpha}, \widetilde{\theta}_{1,j\alpha}, \widetilde{\theta}_{1,j\beta}, \widetilde{\theta}_{1,\alpha\beta}$ are defined in* (6.2.12)

A convenient algorithm for calculating $l_4(\theta)$ is given in Algorithm 4.

**Algorithm 4** Calculation of the composite likelihood $l_4(\theta)$ in (6.1.5).

**Step 0** Input the adjacency matrix $y = (y_{ij})_{1\le i,j\le n}$, of an undirected graph,

**Step 1** Calculate $y_{i+} = \sum_{k=1}^{n} y_{ik}$, $i = 1, \dots, n$, and $T(i,j) = \sum_{k=1}^{n} y_{ik}y_{jk}$, $1 \le i < j \le n$,

**Step 2** Set $l_4 = 0$,

**Step 3** For $1 \le i < j < \alpha < \beta \le n$, calculate

$$\begin{aligned}
D_{ij} =& \theta_1 + \theta_2[(y_{i+} - y_{ij}y_{i\alpha} - y_{i\beta}) + (y_{j+} - y_{ji} - y_{j\alpha} - y_{j\beta})] \\
&+ \theta_3\{T(i,j) - y_{i\alpha}y_{j\alpha} - y_{i\beta}y_{j\beta}\} \\
D_{i\alpha} =& \theta_1 + \theta_2[(y_{i+} - y_{ij}y_{i\alpha} - y_{i\beta}) + (y_{\alpha+} - y_{\alpha i} - y_{\alpha j} - y_{\alpha\beta})] \\
&+ \theta_3\{T(i,\alpha) - y_{ij}y_{\alpha j} - y_{i\beta}y_{\alpha\beta}\} \\
D_{i\beta} =& \theta_1 + \theta_2[(y_{i+} - y_{ij}y_{i\alpha} - y_{i\beta}) + (y_{\beta+} - y_{\beta i} - y_{\beta j} - y_{\beta\alpha})] \\
&+ \theta_3\{T(i,\beta) - y_{ij}y_{\beta j} - y_{i\alpha}y_{\beta\alpha}\} \\
D_{j\alpha} =& \theta_1 + \theta_2[(y_{j+} - y_{ji}y_{j\alpha} - y_{j\beta}) + (y_{\alpha+} - y_{\alpha i} - y_{\alpha j} - y_{\alpha\beta})] \\
&+ \theta_3\{T(i,\alpha) - y_{ij}y_{\alpha j} - y_{i\beta}y_{\alpha\beta}\} \\
D_{j\beta} =& \theta_1 + \theta_2[(y_{j+} - y_{ji}y_{j\alpha} - y_{j\beta}) + (y_{\beta+} - y_{\beta i} - y_{\beta j} - y_{\beta\alpha})] \\
&+ \theta_3\{T(j,\beta) - y_{ji}y_{\beta i} - y_{j\alpha}y_{\beta\alpha}\} \\
D_{\alpha\beta} =& \theta_1 + \theta_2[(y_{\alpha+} - y_{\alpha i}y_{\alpha j} - y_{\alpha\beta}) + (y_{\beta+} - y_{\beta i} - y_{\beta j} - y_{\beta\alpha})] \\
&+ \theta_3\{T(\alpha,\beta) - y_{\alpha i}y_{\beta i} - y_{\alpha j}y_{\beta j}\}.
\end{aligned}$$

Calculate $A$ and $B$ defined in (6.2.13).

**Step 4** Calculate $c$ defined in (6.2.13), and

$$l_4 = l_4 + y_{ij}D_{ij} + y_{i\alpha}D_{i\alpha} + y_{i\beta}D_{i\beta} + y_{j\alpha}D_{j\alpha} + y_{j\beta}D_{j\beta} + y_{\alpha\beta}D_{\alpha\beta} \\ + \theta_2 A + \theta_3 B - \log(c)$$

**Step 5** Return $l_4$.

## 6.3 Simulation Studies of the Composite Likelihood

### 6.3.1 Introduction

In this section we explore by simulation the four composite likelihood estimators derived in Section 6.2 and computed using Algorithm 1 - Algorithm 4. As before,

$y = (y_{ij})_{1\le i,j\le n}$ denotes the adjacency matrix for a random graph with $n$ vertices, and $y_{ij} = y_{ji}$ and $y_{ii} = 0$ as we only consider simple undirected graphs here.
We focus on the fitting of two models: the 3-parameter model

$$P_\theta\{Y = y\} = \exp\{\theta_1 u_1(y) + \theta_2 u_2(y) + \theta_3 u_3(y) - \psi_3(\theta)\}, \tag{6.3.1}$$

and 2-parameter model with probabilities given by

$$P_\theta\{Y = y\} = \exp\{\theta_1 u_1(y) + \theta_3 u_3(y) - \psi(\theta)\}, \tag{6.3.2}$$

where, as before, $u_1(y)$, $u_2(y)$ and $u_3(y)$ denote, respectively, the number of edges, the number of 2-stars and the number of triangles. In subsection 6.3.2 we present simulation results for the four composite likelihood estimators of the 3-parameter model (6.3.1), and in subsection 6.3.3 we present simulation results for the corresponding estimators for the 2-parameter model (6.3.2).
In all cases we have simulated from the homogeneous Bernoulli random graph model with $\theta_2 = \theta_3 = 0$ in (6.3.1) and $\theta_3 = 0$ in (6.3.2). Ideally, we would also have explored simulations from models (6.3.1) and (6.3.2) with non-zero $\theta_2$ and $\theta_3$. However, this would have required the use of an iterative MCMC simulation method such as Metropolis-Hastings or Gibbs sampler. Not only would this have greatly increased the computing resources needed, but there would also have been uncertainty about whether the MCMC procedure had converged. In our current work we have opted to obtained reliable results for more limited set of cases ($\theta_2 = \theta_3 = 0$), rather than obtain possibly unreliable results in a broader set of cases. However, simulation studied with $\theta_2 \neq 0$ and/or $\theta_3 \neq 0$ is an interesting topic for further work.
Throughout this section, the number of Monte Carlo runs is $M = 100$ in each case.

### 6.3.2 Numerical Results for the 3-Parameter Model

In this subsection, the Root Mean Squared Error (RMSE) for an estimator $\widehat{\theta}^{(k)}$ based on Monte Carlo realisations

$$\widehat{\theta}^{(k)}[j] = (\widehat{\theta}_1^{(k)}[j], \widehat{\theta}_2^{(k)}[j], \widehat{\theta}_3^{(k)}[j])^T, \quad j = 1, \cdots, M,$$

is given by

$$RMSE(k) = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left\{(\widehat{\theta}_1^{(k)}[j]-\theta_1)^2+(\widehat{\theta}_2^{(k)}[j]-\theta_2)^2+(\widehat{\theta}_3^{(k)}[j]-\theta_3)^2\right\}}, \tag{6.3.3}$$

where in all examples $\theta = (\theta_1, \theta_2, \theta_3)^T$ is of the form $(\theta_1, 0, 0)^T$, and $k = 1, 2, 3, 4$ corresponds to estimators based on the composite likelihoods calculated in Algorithm 1- Algorithm 4, respectively.

Numerical results are shown in Table 6.1. When $n = 10$, RMSE(4) is the smallest in each case, followed by RMSE(3), with the difference being greatest when $\theta_1 = 2$. However, when $n \geqslant 20$, the RMSE results for the four estimators are very similar indeed. We also note that as $n$ increases, the RMSE for each estimator decreases steadily, suggesting that the estimators are consistent under the homogeneous Bernoulli model at least, a result we have not proved theoretically. A further conclusion we draw is that for these estimators performance as measured by RMSE becomes worse as $\theta_1$ increases.

**Table 6.1:** The Root Mean Squared Error (RMSE) of different composite likelihood estimators of the four methods, implemented using the algorithms mentioned in Section 6.2, assuming $\theta_2 = \theta_3 = 0$ and $\theta_1$ varying, with $n$ the number of nodes, and the number of Monte Carlo runs $M = 100$. Cases with $n = 50$ and $n = 100$ were not calculated for RMSE(4) due to the run time being too long.

| $\theta_1$ | | **n=10** | **n=20** | **n=30** | **n=50** | **n=100** |
|---|---|---|---|---|---|---|
| -2 | RMSE(1) | 16.79983 | 7.616962 | 1.650574 | 0.5476716 | 0.3746696 |
| | RMSE(2) | 15.3608 | 7.449942 | 1.788648 | 0.5474275 | 0.3745891 |
| | RMSE(3) | 15.02113 | 7.521577 | 1.735983 | 0.5470767 | 0.3744534 |
| | RMSE(4) | 14.98221 | 7.669413 | 1.951661 | 0.5465144 | ——— |
| -1 | RMSE(1) | 12.90397 | 1.63242 | 1.039710 | 0.6503136 | 0.4377969 |
| | RMSE(2) | 8.796941 | 1.616878 | 1.037407 | 0.6494742 | 0.4377857 |
| | RMSE(3) | 8.533732 | 1.606362 | 1.034880 | 0.6491427 | 0.4377313 |
| | RMSE(4) | 8.475723 | 1.589709 | 1.030610 | ——— | ——— |
| 0 | RMSE(1) | 4.983311 | 2.295181 | 2.094724 | 1.347965 | 0.7484166 |
| | RMSE(2) | 4.526295 | 2.271271 | 2.085905 | 1.3472 | 0.7483615 |
| | RMSE(3) | 4.309999 | 2.257489 | 2.078797 | 1.346123 | 0.7482923 |
| | RMSE(4) | 4.036978 | 2.22743 | 2.06983 | ——— | ——— |
| 1 | RMSE(1) | 57.68832 | 6.030356 | 3.659141 | 2.858981 | 1.835285 |
| | RMSE(2) | 58.27078 | 5.988627 | 3.645089 | 2.857446 | 1.835245 |
| | RMSE(3) | 57.61023 | 5.963112 | 3.637414 | 2.855695 | 1.834834 |
| | RMSE(4) | 55.2161 | 5.913037 | 3.617684 | ——— | ——— |
| 2 | RMSE(1) | 275.3757 | 121.9528 | 14.1641 | 8.708826 | 5.96737 |
| | RMSE(2) | 313.8732 | 120.5589 | 14.15409 | 8.70749 | 5.973889 |
| | RMSE(3) | 160.8919 | 124.9955 | 14.14478 | 8.705931 | 5.961482 |
| | RMSE(4) | 120.8304 | 126.3075 | 13.92554 | ——— | ——— |

Monte Carlo estimators of the correlation matrices of the $\widehat{\theta}^{(k)}$, $k = 1, 2, 3, 4$, are given below in (6.3.4)-(6.3.7) respectively for the case $n = 30$ and $\theta_1 = 0$, values we chose because they correspond to the center of Table 6.1.
The correlation matrix of $\widehat{\theta}^{(k)}$ based on Monte Carlo realisations $\widehat{\theta}^{(k)}[1], \cdots, \widehat{\theta}^{(k)}[M]$

was calculated as

$$Corr^{(k)} = D^{(k)}\left\{\frac{1}{M}\sum_{j=1}^{M}\left(\widehat{\theta}^{(k)}[j](\widehat{\theta}^{(k)}[j])^T - \overline{\theta}^{(k)}(\overline{\theta}^{(k)})^T\right)\right\}D^{(k)}$$

where

$$\overline{\theta}^{(k)} = \frac{1}{M}\sum_{j=1}^{M}\widehat{\theta}^{(k)}[j], \quad D^{(k)} = \text{diag}\left\{\frac{1}{\sigma_1^{(k)}}, \frac{1}{\sigma_2^{(k)}}, \frac{1}{\sigma_3^{(k)}}\right\}$$

and, for $\alpha = 1, 2, 3$,

$$\sigma_\alpha^{(k)} = \frac{1}{M}\sum_{j=1}^{M}\left\{(\widehat{\theta}_\alpha^{(k)}[j])^2 - (\overline{\theta}_\alpha^{(k)})^2\right\},$$

where $\overline{\theta}^{(k)} = (\overline{\theta}_1^{(k)}, \overline{\theta}_2^{(k)}, \overline{\theta}_3^{(k)})^T$. The correlation matrices were found to be

$$Corr^{(1)} = \begin{pmatrix} 1.0000000 & -0.9275569 & 0.3470725 \\ -0.9275569 & 1.0000000 & -0.6686536 \\ 0.3470725 & -0.6686536 & 1.0000000 \end{pmatrix}, \tag{6.3.4}$$

$$Corr^{(2)} = \begin{pmatrix} 1.0000000 & -0.9275396 & 0.3486106 \\ -0.9275396 & 1.0000000 & -0.6698873 \\ 0.3486106 & -0.6698873 & 1.0000000 \end{pmatrix}, \tag{6.3.5}$$

$$Corr^{(3)} = \begin{pmatrix} 1.0000000 & -0.9273094 & 0.3502174 \\ -0.9273094 & 1.0000000 & -0.6716093 \\ 0.3502174 & -0.6716093 & 1.0000000 \end{pmatrix}, \tag{6.3.6}$$

and

$$Corr^{(4)} = \begin{pmatrix} 1.0000000 & -0.9273055 & 0.3517414 \\ -0.9273055 & 1.0000000 & -0.6728100 \\ 0.3517414 & -0.6728100 & 1.0000000 \end{pmatrix}. \tag{6.3.7}$$

There are two main points to note about (6.3.4)-(6.3.7). First, in all cases $\widehat{\theta}_1^{(k)}$ and $\widehat{\theta}_2^{(k)}$ are highly correlated, while in contrast $\widehat{\theta}_1^{(k)}$ and $\widehat{\theta}_2^{(k)}$ are both somewhat less correlated with $\widehat{\theta}_3^{(k)}$. Second, the four correlation matrices are remarkably similar. Scatterplots of the components of each of the four estimators based on Monte Carlo realisations are shown in Figure 6.1. These plots corroborate the findings in (6.3.4)-(6.3.7). Broadly similar results were obtained with other choices of $\theta_1$ and $n$, when $n \geqslant 20$.

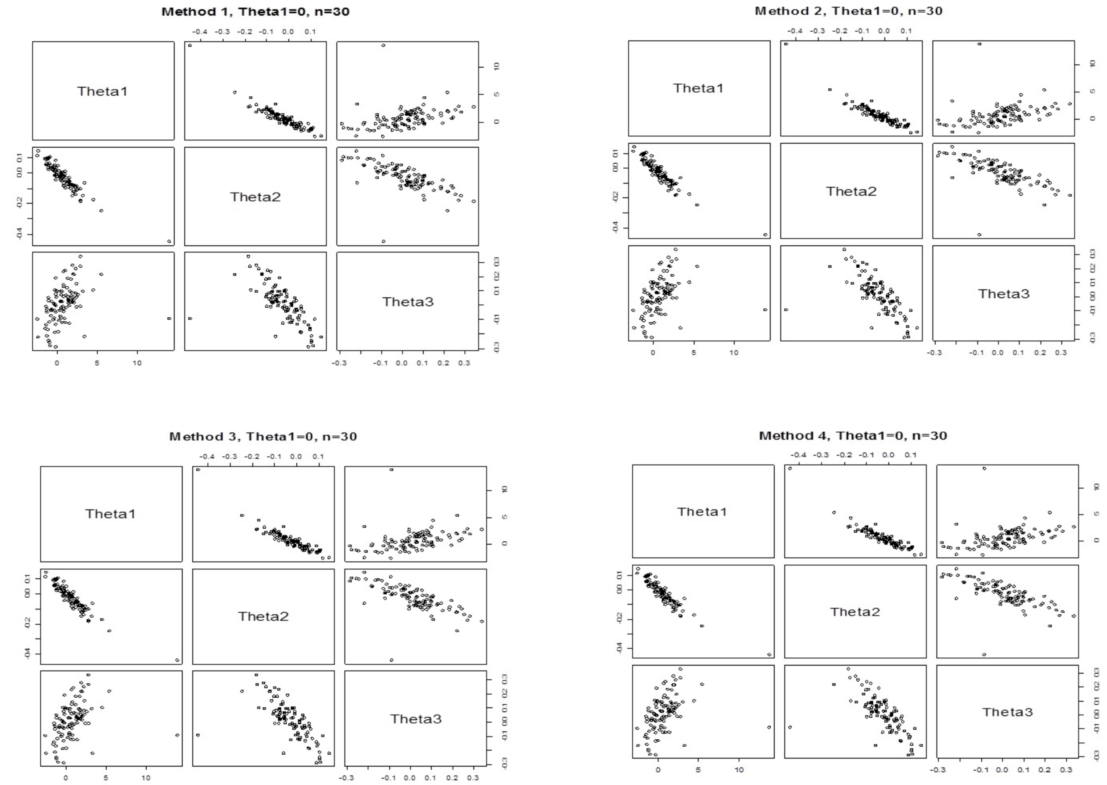


**Figure 6.1:** Scatter plots for components of $\widehat{\theta}_1$, $\widehat{\theta}_2$, $\widehat{\theta}_3$ and $\widehat{\theta}_4$ when $\theta_1 = 0$ and $n = 30$, based on $M = 100$ Monte Carlo runs.

In Figure 6.2, QQ plots are shown of the squared Mahalanobis distances against the $\chi^2_3$ quantiles. Let

$$V_{(k)} = \left\{ \frac{1}{M} \sum_{j=1}^{M} \widehat{\theta}^{(k)}[j](\widehat{\theta}^{(k)}[j])^T \right\} - \overline{\theta}^{(k)}(\overline{\theta}^{(k)})^T$$

denote the sample covariance matrix of $\widehat{\theta}^{(k)}$ based on Monte Carlo realisations $\widehat{\theta}^{(k)}[j]$, $j = 1, \cdots, M$. Then for each $\widehat{\theta}^{(k)}[j]$ define

$$\tau_j^{(k)} = (\widehat{\theta}^{(k)}[j] - \theta)^T (V_{(k)})^{-1} (\widehat{\theta}^{(k)}[j] - \theta),$$

where $\theta = (0, 0, 0)^T$ in the example considered. If the normal approximation

$\widehat{\theta}^{(k)} \approx N_3(\theta, V_{(k)})$ is good, then $\tau_j^{(k)}$, $j = 1, \cdots, M$, will be approximately $\chi_3^2$.

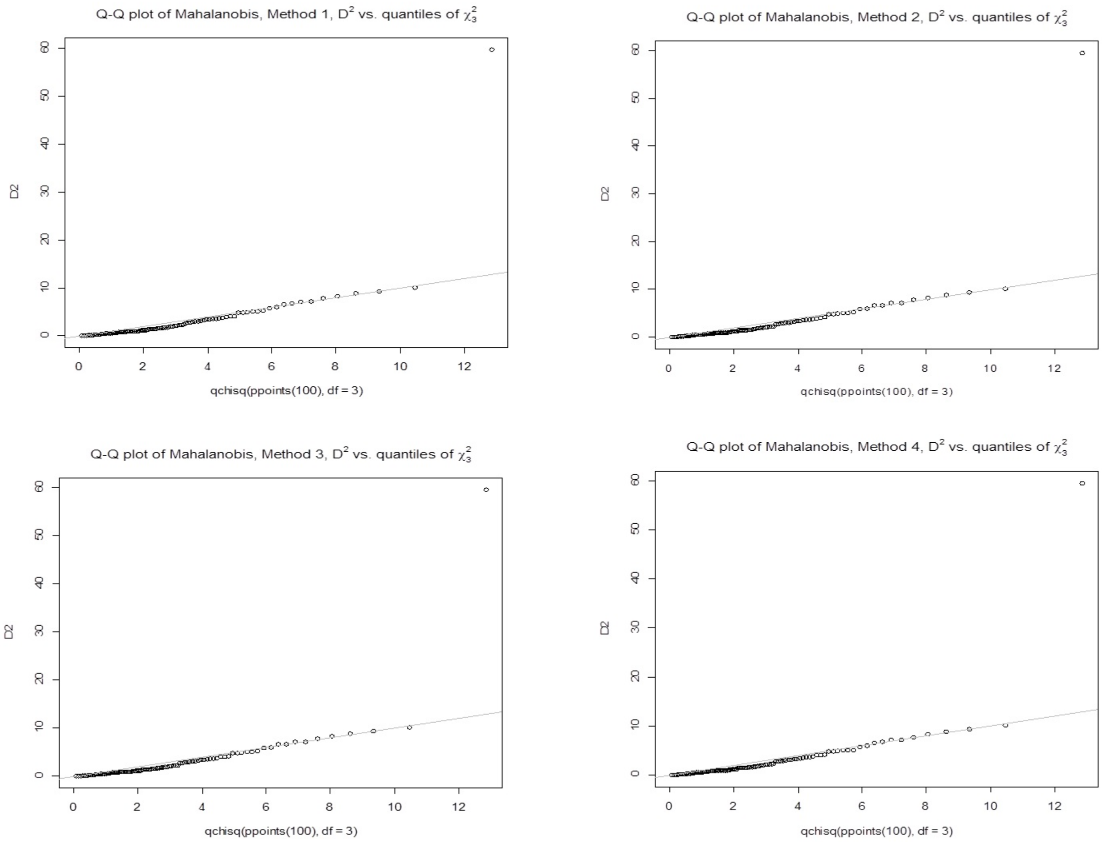


**Figure 6.2:** Mahalanobis to compare between CLEs for the full model, when $\theta_1 = 0$ and $n = 30$.

It is seen from Figure 6.2 that, apart from a single outlier which appears at the some location on all four plots, the $\chi_3^2$ approximation to the $\tau_j^{(k)}$ looks to be very good for each of the four estimators. Moreover, as found previously, the behavior of the four estimators is very similar.

Finally, we look at the sample correlation of $\theta_\alpha^{(1)}$, $\theta_\alpha^{(2)}$, $\theta_\alpha^{(3)}$ and $\theta_\alpha^{(4)}$ for $\alpha = 1, 2, 3$. The relevant correlation matrices were founded to be

$$Corr_1 = \begin{pmatrix} 1.0000000 & 0.9999919 & 0.9999844 & 0.9999646 \\ 0.9999919 & 1.0000000 & 0.9999958 & 0.9999858 \\ 0.9999844 & 0.9999958 & 1.0000000 & 0.9999910 \\ 0.9999646 & 0.9999858 & 0.9999910 & 1.0000000 \end{pmatrix}, \quad (6.3.8)$$

$$Corr_2 = \begin{pmatrix} 1.0000000 & 0.9999867 & 0.9999789 & 0.9999432 \\ 0.9999867 & 1.0000000 & 0.9999949 & 0.9999785 \\ 0.9999789 & 0.9999949 & 1.0000000 & 0.9999860 \\ 0.9999432 & 0.9999785 & 0.9999860 & 1.0000000 \end{pmatrix}, \quad (6.3.9)$$

$$Corr_3 = \begin{pmatrix} 1.0000000 & 0.9999816 & 0.9999801 & 0.9999324 \\ 0.9999816 & 1.0000000 & 0.9999986 & 0.9999805 \\ 0.9999801 & 0.9999986 & 1.0000000 & 0.9999826 \\ 0.9999324 & 0.9999805 & 0.9999826 & 1.0000000 \end{pmatrix}, \quad (6.3.10)$$

again in the case $\theta_1 = 0$ and $n = 30$. The correlations of each component across estimators are remarkably high, and certainly high than we would have expected. These finding are confirmed in the scatterplots in Figure 6.3.

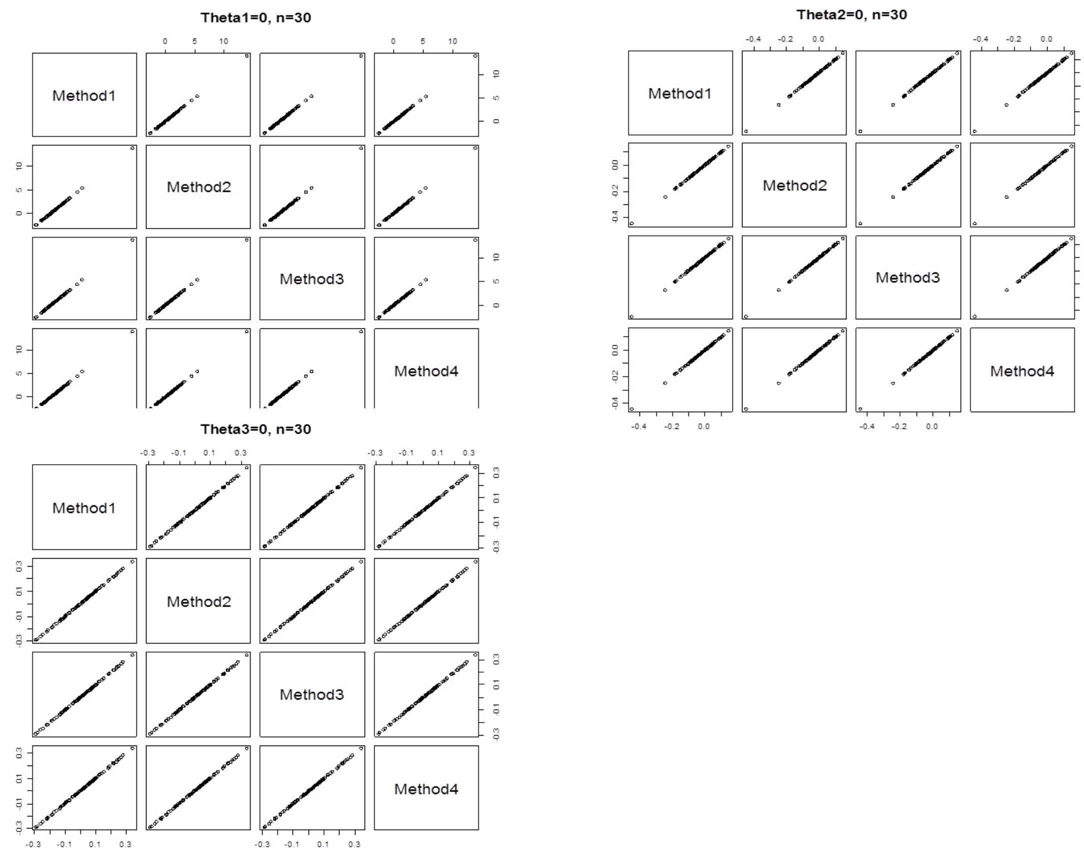


**Figure 6.3:** Scatterplots to compare between methods for the Full Model, when $\theta_1 = 0$ and $n = 30$.

### 6.3.3 Numerical Results for the 2-Parameter Model

In this case we just focused on two estimators of the parameters in model (6.3.2): these based on the composite likelihoods calculated in Algorithm 1 and Algorithm 3. The RMSE for an estimator $\widehat{\theta}^{(k)}$ based on Monte Carlo realisations $\widehat{\theta}^{(k)}[j]$, $j = 1, \cdots, M$, is given by

$$RMSE(k) = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left\{(\widehat{\theta}_1^{(k)}[j] - \theta_1)^2 + (\widehat{\theta}_3^{(k)}[j] - \theta_3)^2\right\}}, \tag{6.3.11}$$

where in all examples $\theta_3 = 0$ and here we limit attention to $k = 1$ and $k = 3$. The RMSEs are presented in Table 6.2 for this case. As can be seen from Table 6.2, there are some differences in RMSE(1) and RMSE(3) in the same cases when $n \leqslant 30$, and when they are different, RMSE(1) is smaller than RMSE(3). However, for each value of $\theta_1$, the two estimators are essentially identical when $n$ is sufficiently large. This last findings was calculated by further simulations which have not included here.

**Table 6.2:** The Root Mean Squared Error (RMSE) of two composite likelihood estimators in model (6.3.2), implemented using Algorithm 1 and Algorithm 3 mentioned in Section 6.2, assuming $\theta_3 = 0$ and $\theta_1$ varying, with $n$ the number of nodes, and the number of Monte Carlo runs $M = 100$.

| $\theta_1$ | | n=10 | n=20 | n=30 | n=50 | n=100 |
|---|---|---|---|---|---|---|
| -2 | RMSE1 | 12.87015 | 5.893478 | 2.801062 | 0.2449863 | 0.1195854 |
| | RMSE3 | 13.79034 | 5.943486 | 2.709793 | 0,2448274 | 0.1196195 |
| -1 | RMSE1 | 6.603481 | 0.4712673 | 0.3349784 | 0.2046826 | 0.1384643 |
| | RMSE3 | 6.360478 | 0.4651426 | 0.3337344 | 0.2043459 | 0.1384179 |
| 0 | RMSE1 | 1.287476 | 0.6367887 | 0.5146259 | 0.4166038 | 0.2379807 |
| | RMSE3 | 1.204182 | 0.628152 | 0.5112811 | 0.4158055 | 0.2377244 |
| 1 | RMSE1 | 4.507642 | 1.507897 | 1.273537 | 0.731839 | 0.5128981 |
| | RMSE3 | 4.168313 | 1.488465 | 1.255581 | 0.7301646 | 0.5118664 |
| 2 | RMSE1 | 51.15089 | 3.620943 | 3.360931 | 2.082709 | 1.182569 |
| | RMSE3 | 54.08247 | 3.568708 | 3.334704 | 2.082709 | 1.181996 |
| 4 | RMSE1 | 103.2746 | 153.56 | 139.8623 | 15.45466 | 9.316641 |
| | RMSE3 | 103.8711 | 185.2349 | 158.3944 | 15.41858 | 9.312107 |

## 6.4 Connection with Chatterjee and Diaconis (2013)

There has been some interesting and important recent theoretical work on model (6.3.2) by (Chatterjee and Diaconis (2013)). They consider the model on simple graphs with $n$ vertices given by

$$P_{\theta_1,\theta_3} = \exp\{2\theta_1 u_1(y) + \frac{6\theta_3}{n} u_3 - n^2\psi_n(\theta_1,\theta_3)\} \tag{6.4.1}$$

where, as before, $u_1(y)$ and $u_3(y)$ denote the number of edges and the number of triangles in the graph and $y$ is the adjacency matrix. They prove that, with high probability , when $n$ is large, $\theta_1 \in \mathbb{R}$ and $\theta_3 > 0$, a realisation $y$ from (6.4.1) is essentially the same as an Erdös-Rényi-Gilbert graph generated by including edges independently with probability that the maximizing value of the following function:

$$\psi_n(\theta_1, \theta_3) \simeq \sup_{0 \le u \le 1} \left( \theta_1 u + \theta_3 u^3 - \frac{1}{2} u log(u) - \frac{1}{2}(1-u) log(1-u) \right). \quad (6.4.2)$$

The optimum value is denoted $u^*(\theta_1, \theta_3)$. In other words, most realizations of this model look like realisations of the Erdös-Rényi-Gilbert simple model. Here, almost all graphs are essentially empty graphs or complete graphs. Chaterjee and Diaconis produced the first proofs of "degeneracy " observed in theses models.

After calculating $u^*$ from our simulation, we notice when that $\theta_3 < 0$ have different behavior, with most values of $u^*$ being greater than or equal 0.5 for $\theta_3 < 0$, and most values of $u^*$ being less than or equal 0.5 for $\theta_3 > 0$. These results fit in with the results in Chatterjee and Diaconis (2013), as can be seen in Figure 6.4 and Figure 6.5.

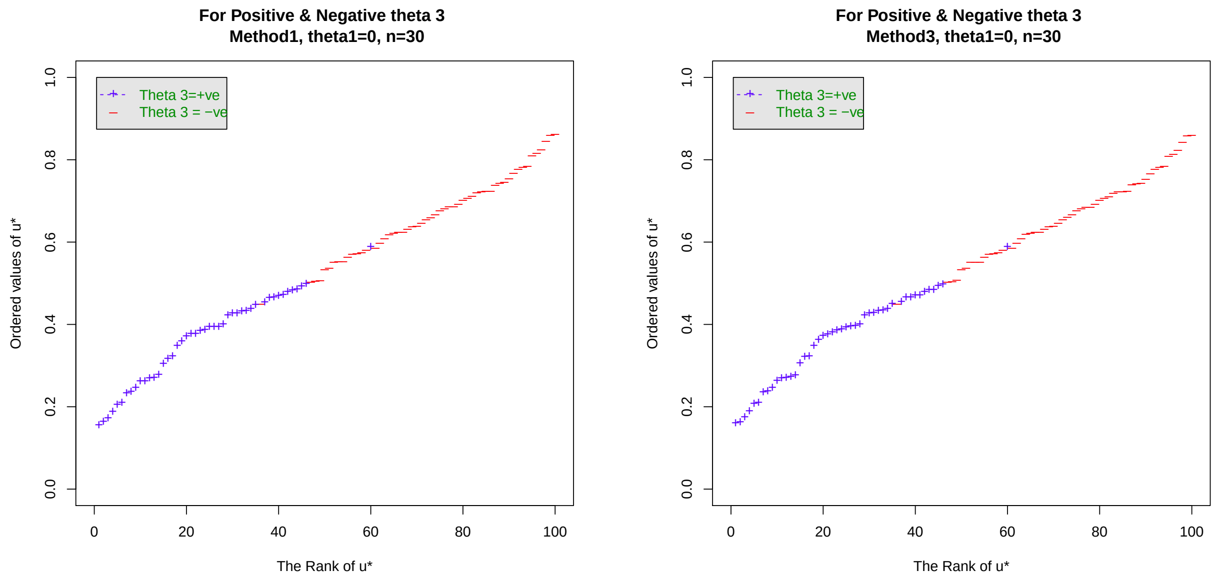


**Figure 6.4:** $u^*$ against the rank of $u^*$, $\theta_0 = (0,0)^T$ and $n = 30$.

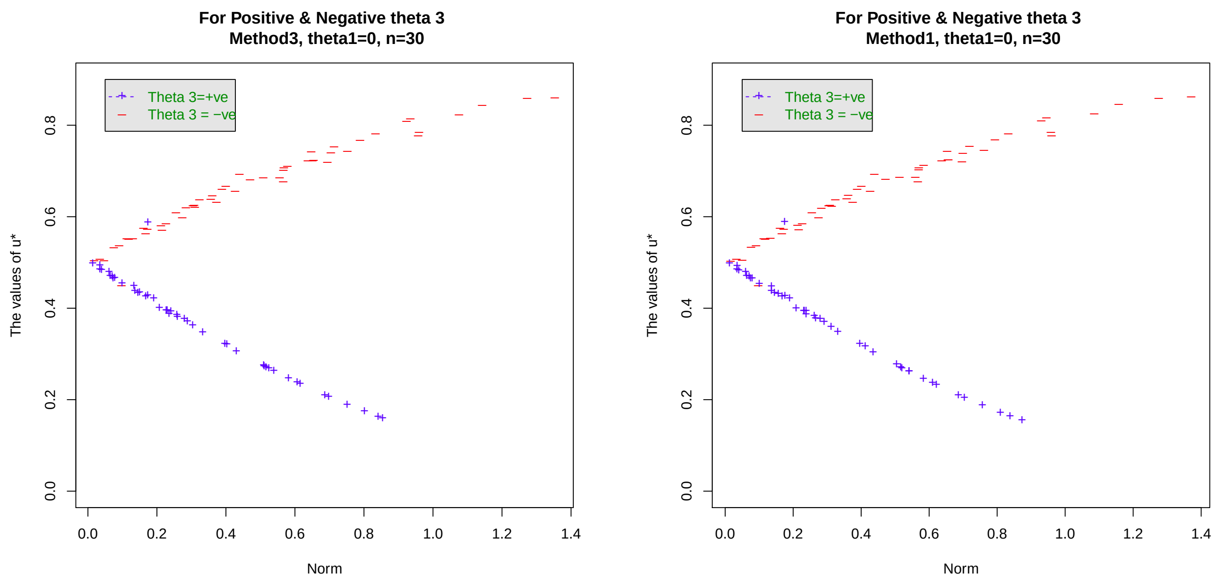


**Figure 6.5:** $u^*$ against the Norm of $u^*$, $\theta_0 = (0,0)^T$ and $n = 30$.

## 6.5 Summary

In this chapter we explore three new composite likelihoods, defined in (6.1.3)-(6.1.5), for the estimation of the parameters in the triad Exponential Random Graph Model (*ERGM*) given in (6.1.1). These new composite likelihoods are based on the conditional distributions of more complicated data structures than in the standard and widely-used composite likelihood in which the components consist the conditional distribution of an edge being present given knowledge of the rest of the edge data.

Our numerical results indicate that the new composite likelihoods perform well in the examples considered. However, our findings are inconclusive in that there is no evidence that the new composite likelihoods perform better than the standard one except possibly when the graph is small (e.g. with $n = 10$ vertices).

One limitation of our simulation study is that we only simulated from the homogeneous Bernoulli random graph model. This was due to the large amount of computer time that would be needed to simulate from a general triad *ERGM* using the Markov Chain Monte Carlo procedure.

Chapter 7

# Summary, Conclusion and Further Research

In this chapter we first summarize the main results and conclusions of the thesis. Then we discuss possible directions for further research.

## 7.1 Summary of the Thesis

The new material in the thesis is contained in Chapters 3-6. In Chapter 3, the main result is Theorem 3.1. This result gives a central limit theorem for three random graph statistics, the number of edges, $u_1$, the number of 2-stars, $u_2$, and the number of triangles, $u_3$. The results is proved under the Bernoulli random graph model in which the presence or absence of each potential edge is an independent Bernoulli random variable with fixed probability $p$ of an edge being present. Theorem 3.1 was proved using the projection method. That a joint central limit theorem holds for these statistics is not a surprise. The surprising aspect of this theorem is that the limiting covariance matrix has rank 1 as opposed to rank 3 and therefore the limiting trivariate normal distribution is degenerate. We have not been able to find mention of this result anywhere in the literature. From the point of view of the key statistical motivation for proving this central limit theorem, which is to construct goodness-of-fit tests, this degeneracy result is a negative one.

In order to see this degeneracy can be removed by conditioning we investigated whether it is possible to prove a central limit theorem for $u_2$ and $u_3$, the number of 2-stars and the number of triangles, respectively, conditional on $u_1$, the number of edges. It turns out that such a central limit theorem does holds and, moreover, the limiting conditional covariance matrix of $u_2$ and $u_3$, suitably standardised, has full rank 2, and therefore the limiting bivariate normal distribution is non-degenerate. However, we were could not see how to use the projection method in this case because conditioning on the number of edges induced dependency between the Bernoulli random variables. As an alternative we used the method of moments. The proof turned out to be very long and is covered in two chapters, Chapter 4 and Chapter 5.

The main result in Chapter 4, Theorem 4.1, gives the order of the expectation of a general product of central Bernoulli random variables subject to their (non-centred) sum, $\sum_{1 \leqslant i < j \leqslant n} y_{ij} = m$, being fixed. Theorem 4.1 plays a crucial role in the proof of the conditional central limit theorem and it may also be of independent interest.

Theorem 5.1, the most substantial result in the thesis, states the joint central limit theorem for $u_2$ and $u_3$ conditional on $u_1$, the number of edges. In addition to using Theorem 4.1, the proof of Theorem 5.1 depends on some fairly complicated counting lemmas. These counting lemmas are stated and proved in Chapter 5. In Section 5.6, goodness-of-fit tests based on Theorem 5.1 are applied to subgraphs of real network data via a fitted block model.

In Chapter 6, three new composite likelihood estimators were investigated for estimating the 3 parameters of the so-called triad Exponential Random Graph Model (*ERGM*). The three new composite likelihoods are based on the conditional likelihoods of more complex data structures than simply the conditional distribution of each edge given the rest of the edge data, which is what is done withe the standard composite likelihood estimator for the triad *ERGM*. The asymptotic theory of the three new estimators seems to be intractable but simulation results suggest that all of the new estimators perform well. However, the numerical results do not provide any evidence that the new estimators are better than the standard composite likelihood estimator apart from possibly the case $n = 10$. In fact, all four estimators have remarkably similar behavior in all cases considered when $n \geqslant 20$.

### 7.1.1 Discussion and Further Research

It is of interest to apply the conditional goodness-of-fit tests based on the Mahalanobis distance using the statistics $C_2$ and $C_3$ in the central limit result in Theorem 5.1. We made a start on this in Section 5.6 but it would be of interest to look at many other real network data examples, in conjunction with different ways of determining the blocks. While it is usually the case that the Erdös-Rényi-Gilbert random graph model will not be adequate for most real network data, we believe that for some block graph models the random graph hypothesis within blocks will sometimes be of interest.

Indeed, we could choose blocks to minimise a suitable goodness-of-fit statistic based on Theorem 5.1 using some kind of stochastic search method such as *simulated annealing*. Developing such a procedure would be of potential interest.

A limitation of our simulation study in Chapter 6 is that we only used data simulated from the homogeneous Bernoulli model. This was due to the fact that simulating from the general triad *ERGM* is expensive because a Markov Chain Monte Carlo (MCMC) method is required, and it would have been very expensive in computing time if we had simulated from the models using MCMC in the simulation study. In addition, there would have been uncertainty in the interpretation of the results due to the question of whether the MCMC simulations had converged. Nevertheless, our current results concerning the new estimators are inconclusive and it will be necessary to simulate from general models within the triad *ERGM* before the efficacy and usefulness of our new composite likelihood estimators can be fully evaluated.

One more direction for future research will now be mentioned. We have shown that conditioning on the number of edges makes the difference between a degenerate and non-degenerate central limit theorem for $u_2$ and $u_3$ in the Erdös-Rényi-Gilbert model. Chatterjee and Diaconis (2013) have shown that a 2-parameter submodel of the triad *ERGM* based on $u_1$ and $u_3$ exhibits certain pathologies. It would be interesting to know whether or not conditioning on $u_1$ removes these pathologies in the case of the 2-parameter submodel and the full 3-parameter triad *ERGM*. The results of this thesis gives some hope that conditioning will have a beneficial effect on statistical inference in the triad *ERGM*.

# Bibliography


Arias-Castro, E. and Verzelen, N. (2013), 'Community detection in random networks', *arXiv preprint arXiv:1302.7099* .

Ayres, F. (1965), *Modern Algebra*, McGraw-Hill: New York.

Bickel, P. J., Chen, A., Levina, E. et al. (2011), 'The method of moments and degree distributions for network models', *The Annals of Statistics* **39**, 2280–2301.

Billingsley, P. (2012), *Probability and Measure*, Wiley: New York.

Birmele, E. et al. (2012), 'Detecting local network motifs', *Electronic Journal of Statistics* **6**, 908–933.

Bloznelis, M. and Götze, F. (2001), 'Orthogonal decomposition of finite population statistics and its applications to distributional asymptotics', *Annals of Statistics* **29**, 899–917.

Bloznelis, M. and Götze, F. (2002), 'An edgeworth expansion for symmetric finite population statistics', *The Annals of Probability* **30**, 1238–1265.

Bollobás, B., Janson, S. and Riordan, O. (2007), 'The phase transition in inhomogeneous random graphs', *Random Structures & Algorithms* **31**(1), 3–122.

Bollobás, B. (2001), *Random Graphs*, Cambridge University Press: Cambridge.

Britton, T., Deijfen, M. and Martin-Löf, A. (2006), 'Generating simple random graphs with prescribed degree distribution', *Journal of Statistical Physics* **124**(6), 1377–1397.

Caimo, A. and Friel, N. (2013), 'Bayesian model selection for exponential random graph models', *Social Networks* **35**(1), 11–24.

Chatterjee, S. and Diaconis, P. (2013), 'Estimating and understanding exponential random graph models', *The Annals of Statistics* **41**, 2428–2461.

Chatterjee, S., Diaconis, P. and Sly, A. (2011), 'Random graphs with a given degree sequence', *The Annals of Applied Probability* **21**, 1400–1435.

Chung, K. L. (2001), *A Course in Probability Theory*, Academic press: New York.

Conway, J. H. and Guy, R. (1996), *The Book of Numbers*, Springer: New York.

Durrett, R. (2010), *Random Graph Dynamics*, Vol. 20 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press: Cambridge.

Erdős, P. and Rényi, A. (1959), 'On random graphs', *Publicationes Mathematicae* **6**, 290–297.

Erdős, P. and Rényi, A. (1960), 'On the evolution of random graphs', *Publ. Math. Inst. Hungar. Acad. Sci* **5**, 17–61.

Erdős, P. and Rényi, A. (1961), 'On the strength of connectedness of a random graph', *Acta Mathematica Hungarica* **12**, 261–267.

Faust, K. and Wasserman, S. (1992), 'Blockmodels: Interpretation and evaluation', *Social Networks* **14**, 5–61.

Fortunato, S. (2010), 'Community detection in graphs', *Physics Reports* **486**, 75–174.

Frank, O. and Strauss, D. (1986), 'Markov graphs', *Journal of the American Statistical Association* **81**, 832–842.

Geyer, C. J. and Thompson, E. A. (1992), 'Constrained Monte Carlo maximum likelihood for dependent data', *Journal of the Royal Statistical Society. Series B (Methodological)* **54**, 657–699.

Gilbert, E. N. (1959), 'Random graphs', *The Annals of Mathematical Statistics* **30**, 1141–1144.

Hoeffding, W. (1948), 'A class of statistics with asymptotically normal distribution', *The Annals of Mathematical Statistics* **19**, 293–325.

Janson, S. (1990), 'A functional limit theorem for random graphs with applications to subgraph count statistics', *Random structures & Algorithms* **1**(1), 15–37.

Janson, S. and Nowicki, K. (1991), 'The asymptotic distributions of generalized u-statistics with applications to random graphs', *Probability theory and related fields* **90**(3), 341–375.

Janson, S., Oleszkiewicz, K. and Ruciński, A. (2004), 'Upper tails for subgraph counts in random graphs', *Israel Journal of Mathematics* **142**(1), 61–92.

Kolaczyk, E. (2010), *Statistical Analysis of Network Data: Methods and Models*, Springer Series in Statistics, Springer: New York.

Kunegis, J. (2014), 'Polcovar: Software for computing the mean and variance of subgraph counts in random graphs', *arXiv preprint arXiv:1402.5835* .

Lancichinetti, A. and Fortunato, S. (2009), 'Community detection algorithms: a comparative analysis', *Physical Review E* **80**, 056117.

Mardia, K. V., Kent, J. and Bibby, J. (1980), *Multivariate Analysis*, Academic Press: New York.

Newman, M. (2010), *Networks: an Introduction*, Oxford University Press: Oxford.

Olhede, S. C. and Wolfe, P. J. (2012), 'Degree-based network models', *arXiv preprint arXiv:1211.6537* .

Olhede, S. C. and Wolfe, P. J. (2014), 'Network histograms and universality of blockmodel approximation', *Proceedings of the National Academy of Sciences* **111**, 14722–14727.

Picard, F., Daudin, J.-J., Koskas, M., Schbath, S. and Robin, S. (2008), 'Assessing the exceptionality of network motifs', *Journal of Computational Biology* **15**(1), 1–20.

R Core Team (2014), *R: A Language and Environment for Statistical Computing*, http://www.R-project.org R Foundation for Statistical Computing, Vienna, Austria.

Robins, G., Pattison, P., Kalish, Y. and Lusher, D. (2007), 'An introduction to exponential random graph models for social networks', *Social Networks* **29**, 173–191.

Robins, G., Snijders, T., Wang, P., Handcock, M. and Pattison, P. (2007), 'Recent developments in exponential random graph models for social networks', *Social networks* **29**, 192–215.

Ruciński, A. (1988), 'When are small subgraphs of a random graph normally distributed?', *Probability Theory and Related Fields* **78**(1), 1–10.

Snijders, T. A. (2002), 'Markov chain monte carlo estimation of exponential random graph models', *Journal of Social Structure* **3**, 1–40.

Snijders, T. A., Pattison, P. E., Robins, G. L. and Handcock, M. S. (2006), 'New specifications for exponential random graph models', *Sociological Methodology* **36**, 99–153.

Snijders, T. A. and Van Duijn, M. A. (2002), 'Conditional maximum likelihood estimation under various specifications of exponential random graph models', *Contributions to Social Network Analysis, Information Theory, and Other Topics in Statistics* pp. 117–134.

van der Vaart, A. W. (2000), *Asymptotic Statistics*, Cambridge University Press.

Varin, C., Reid, N. and Firth, D. (2011), 'An overview of composite likelihood methods', *Statistica Sinica* **21**, 5–42.