กฎอย่างเข้มของเลขจำนวนมากบนกราฟเชิงสุ่มและกลุ่มของลูกผสมของเซลล์ร่างกายเชิงสุ่ม

นางสาวน้ำค้าง แสงศรีจันทร์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาคณิตศาสตร์ ภาควิชาคณิตศาสตร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2550

STRONG LAW OF LARGE NUMBERS ON RANDOM GRAPHS AND RANDOM SOMATIC CELL HYBRID PANELS

Miss Namkharng Sangsrijan

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Mathematics
Department of Mathematics
Faculty of Science
Chulalongkorn University
Academic Year 2007

Thesis Title	STRONG LAW OF LARGE NUMBERS ON RANDOM GRAPHS AND RANDOM SOMATIC CELL HYBRID PANELS
By	Miss Namkharng Sangsrijan
Field of Study	Mathematics
Thesis Advisor	Professor Kritsana Neammanee, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

.... Dean of the Faculty of Science

(Professor Supot Hannongbua, Ph.D.)

THESIS COMMITTEE

.... Chairman

(Kittipat Wong, Ph.D.)

.... Thesis Principal Advisor

(Professor Kritsana Neammanee, Ph.D.)

.... Member

(Chariya Uiyyasathian, Ph.D.)

.... External Member

(Assistant Professor Kanint Teerapabolarn, Ph.D.)

น้ำค้าง แสงศรีจันทร์ : กฎอย่างเข้มของเลขจำนวนมากบนกราฟเชิงสุ่มและกลุ่มของลูกผสมของเซลล์ร่างกายเชิงสุ่ม. (STRONG LAW OF LARGE NUMBERS ON RANDOM GRAPHS AND RANDOM SOMATIC CELL HYBRID PANELS) อ.ที่ปรึกษา : ศ. ดร. กฤษณะ เนียมมณี, 34 หน้า.

ในวิทยานิพนธ์ฉบับนี้เราได้ทำการตรวจสอบกฎอย่างเข้มของเลขจำนวนมากของจำนวนจุดยอดซึ่งมีดีกรีที่กำหนด จำนวนของต้นไม้เอกเทศซึ่งมีจำนวนจุดตามที่กำหนดและจำนวนชุดเอกเทศของกราฟเชื่อมโยงที่กำหนดในกราฟเชิงสุ่ม นอกจากนี้เราได้ทำการตรวจสอบกฎอย่างเข้มของเลขจำนวนมากของจำนวนคู่ของโครโมโซมในกลุ่มของลูกผสมของเซลล์ร่างกายเชิงสุ่มที่มีค่าระยะทางแบบแฮมมิ่งน้อยกว่าค่าที่กำหนดด้วย

ภาควิชา ...**คณิตศาสตร์**...	ลายมือชื่อนิสิต..............................
สาขาวิชา ...**คณิตศาสตร์**...	ลายมือชื่ออาจารย์ที่ปรึกษา..............................
ปีการศึกษา ......2550......	ลายมือชื่ออาจารย์ที่ปรึกษาร่วม..............................

##4972337323 : MAJOR MATHEMATICS
KEY WORDS : STRONG LAW OF LARGE NUMBERS / RANDOM GRAPH/ VERTEX WITH DEGREE d / ISOLATED TREES / ISOLATED COPIES OF CONNECTED GRAPH / SOMATIC CELL HYBRID / HAMMING DISTANCE

NAMKHARNG SANGSRIJAN : STRONG LAW OF LARGE NUMBERS ON RANDOM GRAPHS AND RANDOM SOMATIC CELL HYBRID PANELS. THESIS ADVISOR : PROF. KRITSANA NEAMMANEE, Ph.D., 34 pp.

We investigate the strong law of large numbers of the number of vertices with a fixed degree, the number of isolated trees with a fixed order, and the number of isolated copies of a fixed connected graph in a random graph, and also investigate the strong law of large numbers of the number of pairs of chromosomes in a random panel of n distinct hybrid clones for which the Hamming distance is less than a fixed Hamming distance.

Department	...Mathematics...	Student's Signature	N. Sangsrijan
Field of Study	...Mathematics...	Advisor's Signature	Kritsana Neammanee
Academic Year	.........2007..........		

ACKNOWLEDGEMENTS

I am deeply indebted to Professor Dr. Kritsana Neammanee, my thesis advisor, for his willingness to sacrifice his time to suggest and advise me in preparing and writing this thesis. I would like to thank Dr. Kittipat Wong, Dr. Chariya Uiyyasathian, Assistant Professor Dr. Kanint Teerapabolarn, my thesis committee, for their suggestions on this thesis. I am also sincerely grateful to all of my teachers and all the lecturers during my study.

In particular, I warmly thank to my dear friends for giving me good experiences at Chulalongkorn university.

Finally, I would like to express my deep gratitude to my beloved family for their love and encouragement throughout my graduate study.

CONTENTS

CHAPTER I
INTRODUCTION

Law of large numbers is a fundamental concept in statistics and probability. It includes various theorems that make statements about the convergence of the average of a random sample to the mean of the whole population. Usually two major categories are distinguished: *Weak Laws* versus *Strong Laws*. The Weak Laws deal with limits of probabilities involving average of random sample while the Strong Laws deal with probabilities involving limits of average of random sample.

In this thesis, we consider only the strong law of large numbers (SLLN) which is stated as follows.

Let (X_n) be a sequence of random variables with finite expectations in a probability space and $S_n = \sum_{i=1}^{n} X_i$. We say that (X_n) satisfies the strong law of large numbers (or the sequence S_n obeys the strong law of large numbers) if

$$\frac{1}{n}\left[S_n - E\left(S_n\right)\right] \xrightarrow{a.s.} 0$$

where a.s. stands for convergence almost surely. When the random variables are identically distributed, with the expectation μ, the law becomes:

$$\frac{1}{n} S_n \xrightarrow{a.s.} \mu. \tag{1.1}$$

The strong law of large numbers (1.1) was originally proved by Borel in the case of X_i's being independent Bernoulli random variables while the general form of (1.1) was proved by Kolmogorov.

Let $\mathbb{G}(n,p)$ be a graph on n labeled vertices $\{1, 2, ..., n\}$ where each possible edge, $\{i, j\}$, is present randomly and independently with a probability p, of $0 < p < 1$. Our main results below are obtained from the investigation of the strong law of large numbers of the number of vertices with a fixed degree, the number of isolated trees with a fixed order, and the number of isolated copies of a fixed connected graph in $\mathbb{G}(n,p)$.

Theorem 1.1. *The number of vertices with degree d in $\mathbb{G}(n,p)$ obeys the strong law of large numbers in the following cases:*

1. *p is a constant.*
2. *$p = \dfrac{1}{n^{\delta}}$, where $\delta > 1$ and $d \geq 1$.*

Corollary 1.1. *If p is a constant then the number of isolated vertices in $\mathbb{G}(n,p)$ obeys the strong law of large numbers.*

Theorem 1.2. *Let k be a positive integer and $k \geq 2$. The number of isolated trees with order k in $\mathbb{G}(n,p)$ obeys the strong law of large numbers in the following cases:*

1. *p is a constant.*
2. *$p = \dfrac{1}{n^{\delta}}$, where $\delta > \dfrac{1}{k-1}$.*

Theorem 1.3. *Let H be a fixed connected graph consisting of $k \geq 2$ vertices and $\ell \geq 1$ edges. The number of isolated copies of H in $\mathbb{G}(n,p)$ obeys the strong law of large numbers in the following cases:*

1. *p is a constant.*
2. *$p = \dfrac{1}{n^{\delta}}$, where $\delta > \dfrac{1}{\ell}$.*

The following result is obtained from our study of model of somatic cell hybrid panels in [14],

Theorem 1.4. *Let d be a fixed Hamming distance. Then the number of pairs of chromosomes in a random panel of n distinct hybrid clones for which the Hamming distance is less than d, $W_{n,d}$, obeys the strong law of large numbers when the retention probability p_0 is a constant , of $0 < p_0 < 1$.*

This thesis is organized as follows. Preliminaries are in Chapter 2. The strong law of large numbers of the number of vertices with a fixed degree, the number of isolated trees with a fixed order, and the number of isolated copies of a fixed connected graph in a random graph are investigated in Chapter 3. In Chapter 4, we prove that the number of pairs of chromosomes in a random panel of n distinct hybrid clones for which the Hamming distance is less than a fixed Hamming distance d obeys the strong law of large numbers.

CHAPTER II
PRELIMINARIES

In this chapter, we review some basic knowledges in probability theory which will be used in our study. The proof is omitted but can be found in many probability theory text books.

The definitions and theorems below are very useful for our study.

Definition 2.1 *A measure space* $(\Omega, \mathcal{F}, P)$ *is said to be a* **probability space** *if* $P(\Omega) = 1$.

If $(\Omega, \mathcal{F}, P)$ is a probability space then the measure P is called a **probability measure** and the set Ω will be referred as a **sample space** and its elements are called **points** or **elementary events**. The elements of $\mathcal{F}$ are called **events** and for any $A \in \mathcal{F}$, the value $P(A)$ is called the **probability of** A.

Definition 2.2 *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space. A function* $X : \Omega \to \mathbb{R}$ *is called a* **random variable** *if for every Borel set* B *in* $\mathbb{R}$, $X^{-1}(B) \in \mathcal{F}$.

We shall use the notation $P(X \in B)$ in place of $P(\{\omega \in \Omega | X(\omega) \in B\})$. In case of $B = \{a\}$, $(-\infty, a]$ or $[a, b]$, $P(X \in B)$ is denoted by $P(X = a)$, $P(X \leq a)$ or $P(a \leq X \leq b)$, respectively.

Definition 2.3 *A random variable* X *is said to be a* **discrete random variable** *if its image is countable.*

Definition 2.4 *Let* X *be a discrete random variable. A function* $f : \mathbb{R} \to [0, 1]$ *defined by*

$$f(x) = P(X = x)$$

is called the **probability function** *of* X.

Definition 2.5 *Let* E *be an event from a probability space* $(\Omega, \mathcal{F}, P)$. *A function* $X : \Omega \to \mathbb{R}$ *defined by*

$$X(\omega) = \begin{cases} 1, & \text{if } \omega \in E \text{ ;} \\ 0, & \text{if } \omega \in E, \end{cases}$$

is called an **indicator random variable**.

Definition 2.6 *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and* $\mathcal{F}_\alpha \subseteq \mathcal{F}$ *for all* $\alpha \in \Lambda$. *Then* $\{\mathcal{F}_\alpha : \alpha \in \Lambda\}$ *is said to be* **independent** *if and only if for each nonempty finite subset* $j = \{j_1, j_2, ..., j_k\}$ *of* Λ,

$$P\left(\bigcap_{m=1}^{k} A_m\right) = \prod_{m=1}^{k} P(A_m)$$

for all $A_m \in \mathcal{F}_{j_m}$ *and* $m = 1, 2, ..., k$.

Definition 2.7 *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space. A collection of events* $\{E_\alpha \in \mathcal{F} : \alpha \in \Lambda\}$ *is* **independent** *if and only if* $\{\sigma(E_\alpha) : \alpha \in \Lambda\}$ *is independent.*

Theorem 2.1 *A family of events* $\{E_\alpha : \alpha \in \Lambda\}$ *is* **independent** *if and only if*

$$P(\bigcap_{\alpha \in \Lambda} E_\alpha) = \prod_{\alpha \in \Lambda} P(E_\alpha).$$

Definition 2.8 *Let* $\{X_\alpha : \alpha \in \Lambda\}$ *be a family of random variables on a probability space* $(\Omega, \mathcal{F}, P)$. *Then* $\{X_\alpha : \alpha \in \Lambda\}$ *is said to be* **independent** *if and only if* $\{\sigma(X_\alpha) : \alpha \in \Lambda\}$ *is independent, where* $\sigma(X_\alpha) = \{X_\alpha^{-1}(B) : B \text{ is a Borel set on } \mathbb{R}\}$.

Definition 2.9 *Let* X *be a discrete random variable with its probability function* f. *If* $\sum_{x \in ImX} |x| f(x) < \infty$ *then the* **expected value** *of* X, *denoted by* $E(X)$ *is defined by*

$$E(X) =: \sum_{x \in \operatorname{Im} X} x f(x).$$

and the **variance** *of* X, *in notation* $\mathrm{Var}(X)$ *is defined by*

$$\mathrm{Var}(X) = E\left[(X - E(X))^2\right].$$

Theorem 2.2 *Let* X *be a random variable such that* $E(X^2) < \infty$. *Then*

$$\mathrm{Var}(X) = E(X^2) - E^2(X).$$

Theorem 2.3 (Chebyšhev inequality) *Let* X *be a random variable with finite expectation and variance. Then for each* $\epsilon > 0$

$$P(\{\omega \in \Omega \ : \ |X(\omega) - E(X)| \geq \epsilon\}) \leq \frac{\mathrm{Var}(X)}{\epsilon^2}.$$

Theorem 2.4 *Let* X, X_1, ...,X_n *be random variables on a same probablility space and* a, b *be any real numbers. If* $E(X) < \infty$ *and* $E(X_i) < \infty$ *for* $i = 1, 2, ..., n$ *then we have the followings;*

1. $E(aX + b) = aE(X) + b$
2. $E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n)$.

Definition 2.10 *Let* X *and* Y *be random variables on the same probability space. If* $E(X^2) < \infty$ *and* $E(Y^2) < \infty$, *then the* **covariance** *of* X *and* Y *is defined by*

$$\mathrm{Cov}(X, Y) = E\left[(X - E(X))(Y - E(Y))\right].$$

Theorem 2.5 *Let* X, Y, X_1, ...X_n *be random variables on the same probablity space. If* $E(X^2)$, $E(Y^2)$, $E(X_i^2) < \infty$ *for* $i = 1, 2, ..., n$ *then we have the followings;*

1. $\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)$
2. $\mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{j=1}^{n}\sum_{k=1}^{n} \mathrm{Cov}(X_j, X_k)$.

Definition 2.11 *Let* $X,\ X_1,\ X_2,\ \ldots$ *be random variables on a probability space* $(\Omega, \mathcal{F}, P)$*. Then we say that* (X_n) **converges to** X **almost surely**, *in notation* $X_n \xrightarrow{a.s.} X$ *if*

$$P(\{\omega \in \Omega \ : \ X_n(\omega) \to X(\omega) \text{ as } n \to \infty\}) = 1.$$

CHAPTER III
STRONG LAW OF LARGE NUMBERS ON RANDOM GRAPHS

3.1 Introduction and Main Results

A random graph is a graph generated by some random procedure. In other words, a random graph is a collection of vertices with edges connecting pairs of them at random.

Random graphs are widely used in probabilistic method, where one tries to prove the existence of graphs with certain properties. In application, Random graphs have been used as models of networks in diverse areas of science, engineering and sociology, for examples models of food webs ([22]), networks of telephone calls ([1]), networks of friendships within a variety of communities ([2],[11]), etc.

The study of random graphs has long history. The notion of random graphs was first introduced in 1947 paper of Erdös ([4]). A decade later, the theory of random graphs had been developed by 1959-1968 papers of Erdös and Rényi ([5]-[10]).

The theory of random graphs lines intersection between graph theory and probability theory, and studies the properties of typical random graphs. Different random graph models produce different probability distribution on graphs.

The simple model introduced by Erdös is very natural and can be described as choosing a graph at random, with equal probabilities, from the set of all $2^{\binom{n}{2}}$ graphs whose vertex set is $\{1, 2, ..., n\}$. Nowadays, among several models of random graphs, there are two basic ones, the binomial model and the uniform model, both models were introduced by Erdös (1947).

In this work, we study on the binomial model that can be described as the following definition.

Definition 3.1.1. *A* **binomial random graph** *(or simply* **random graph***) on* n **vertices**, *denoted by* $\mathbb{G}(n,p)$, *is a graph on* n *labeled vertices* $\{1,2,...,n\}$ *where each possible edge,* $\{i,j\}$, *is present randomly and independently with a probability* p, *of* $0 \leq p \leq 1$.

The probability space of this model is $(\Omega, \mathcal{F}, P)$ *where* Ω *is the set of all* $2^{\binom{n}{2}}$ *graphs whose vertex set is* $\{1,2,...,n\}$, *and* $\mathcal{F}$ *is the family of all subsets of* Ω *and for each* $G \in \Omega$,

$$P(G) = p^{e_G}\,(1-p)^{\binom{n}{2}-e_G}$$

where e_G *is the number of edges in a graph* G.

Example 3.1.1. *Let* $n = 4$, $p = \dfrac{1}{4}$. According to Definition 3.1.1 we have,
$\Omega = \{G : G \text{ is a graph with vertex set } \{1,2,3,4\}\} =: \{G_1, G_2, ..., G_k, ..., G_{2^6}\}$

G_1 G_2 G_k G_{2^6}

$$\begin{aligned} P(G_k) &= p^{e_{G_k}}\,(1-p)^{\binom{n}{2}-e_{G_k}} \\ &= (\frac{1}{4})^4 \cdot (1-\frac{1}{4})^{6-4} \\ &= (\frac{1}{4})^4 \cdot (\frac{3}{4})^2 \\ &= 0.0021973. \end{aligned}$$

In this chapter, we investigate the strong law of large numbers of some sequence of random variables, defined on the sample space of random graphs. The necessary definitions in the graph theory which are related to our study are as follows.

Next, we denote the set of all two-element subsets of a set A by $A^{(2)}$.

Definition 3.1.2. *A* **graph** *consists of two things: a nonempty set V and a (possibly empty) subset E of $V^{(2)}$. Typically written $G = (V, E)$, the elements of V are the* **vertices** *(or* **nodes***) of G, and the elements of E are its* **edges***. When more than one graph is under consideration, it may be useful to write $V(G)$ and $E(G)$ for its vertex and edge sets, respectively.*

If $e = \{u, v\} \in E(G)$, then vertices u and v are said to be **adjacent** *(to each other) and* **incident** *to e.*

Definition 3.1.3. *Let G and H be graphs. Then H is a* **subgraph** *of G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$.*

For any subset W of $V(G)$, the **subgraph of** G **induced** *(or* **spanned***) by W is $G[W] = (W, E(G) \cap W^{(2)})$. An induced subgraph $G[W]$ is said to be* **isolated** *if any vertex in W is not adjacent to a vertex in $V(G) \backslash W$.*

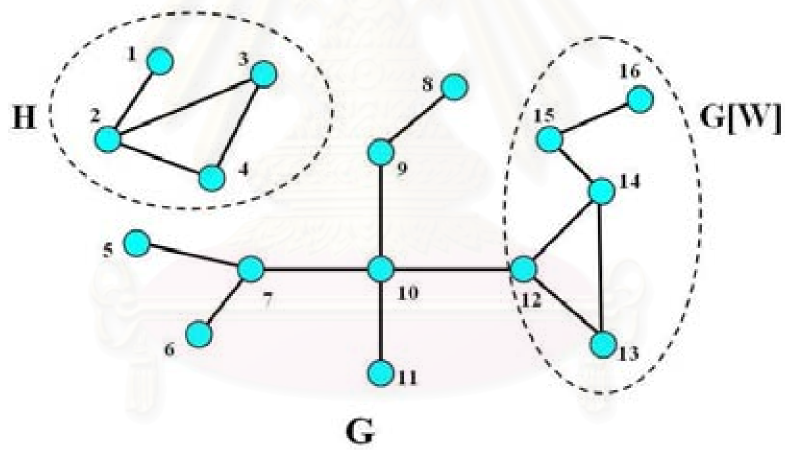
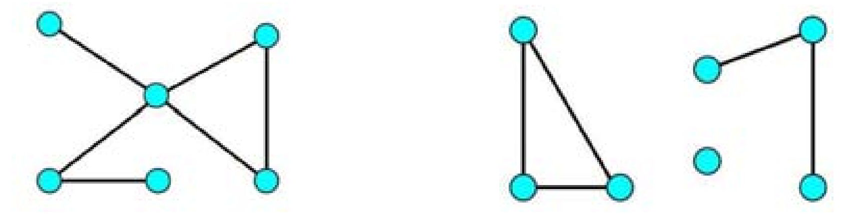


Figure 3.1 An example of a graph G containing induced subgraph $G[W]$ where $W = \{12, 13, 14, 15, 16\}$ and the isolated subgraph H.

Definition 3.1.4. *A graph G is* **connected** *if for any given pair of vertices a and b there is a finite sequence of distinct vertices and edes of the form $v_{i_0}, e_{i_1}, v_{i_1}, ..., e_{i_n}, v_{i_n}$ where $v_{i_0} = a$, $v_{i_n} = b$, and $e_{i_1} = \{v_{i_0}, v_{i_1}\}$, $e_{i_2} = \{v_{i_1}, v_{i_2}\}, ..., e_{i_n} = \{v_{i_{n-1}}, v_{i_n}\}$, and otherwise, G is* **disconnected**.



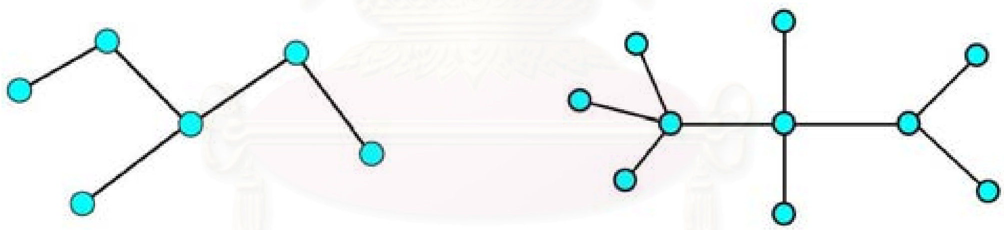
Connected graph　　　　Disconnected graph

Figure 3.2

Definition 3.1.5. *The* **degree** *of a vertex v in a graph G, denoted by $deg(v)$, is the number of edges incident to v. If $deg(v) = 0$, then v is said to be an isolated vertex.*

Definition 3.1.6. *A* **cycle** *is a connected graph in which every vertices has degree 2.*

Definition 3.1.7. *A* **tree with order** k *is a connected graph with k vertices containing no cycles.*



Tree with order 6　　　　Tree with order 10

Figure 3.3

Definition 3.1.8. *A graph G_1 is* **isomorphic** *to a graph G_2 (or G_1 is a* **copy** *of G_2) if there is a one-to-one function f from $V(G_1)$ onto $V(G_2)$ such that $\{u, v\} \in E(G_1)$ if and only if $\{f(u), f(v)\} \in E(G_2)$. If such a function exists, it is called an* **isomorphism** *from G_1 to G_2.*

An isomorphism from G_1 to itself is called an **automorphism** *of G_1.*

Definition 3.1.9. *Let G_0 be a fixed graph. A subgraph H in a graph G is an* **isolated copy of** G_0 *if H is isolated subgraph of G and isomorphic to G_0.*

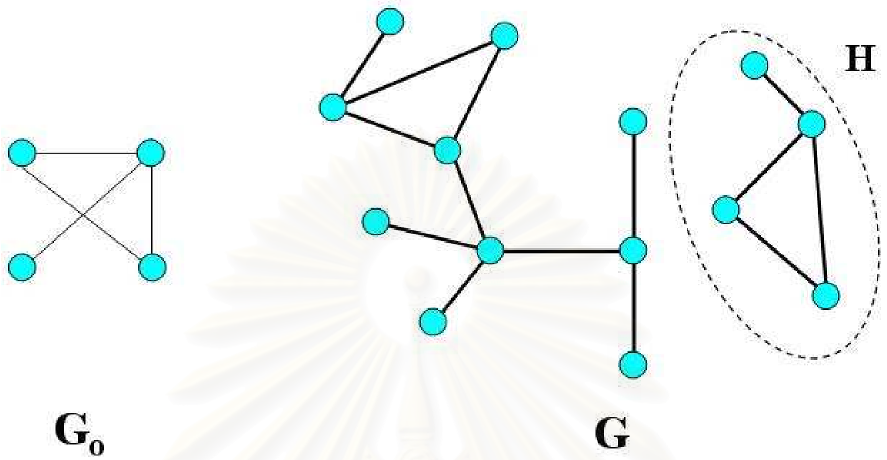


Figure 3.4 An example of a fixed graph G_0 and a graph G containing subgraph H which is an isolated copy of G_0.

Our objective is to investigate the strong law of large numbers of the number of vertices with a fixed degree, the number of isolated trees with a fixed order, and the number of isolated copies of a fixed connected graph in $\mathbb{G}(n, p)$ where $0 < p < 1$. The followings are our main results.

Theorem 3.1.1. *The number of vertices with degree d in $\mathbb{G}(n, p)$ obeys the strong law of large numbers in the following cases:*

1. *p is a constant.*
2. *$p = \dfrac{1}{n^\delta}$, where $\delta > 1$ and $d \geq 1$.*

Corollary 3.1.1. *If p is a constant then the number of isolated vertices in $\mathbb{G}(n, p)$ obeys the strong law of large numbers.*

Theorem 3.1.2. *Let k be a positive integer and $k \geq 2$. The number of isolated trees with order k in $\mathbb{G}(n,p)$ obeys the strong law of large numbers in the following cases:*

1. *p is a constant.*
2. *$p = \dfrac{1}{n^{\delta}}$, where $\delta > \dfrac{1}{k-1}$.*

Theorem 3.1.3. *Let H be a fixed connected graph consisting of $k \geq 2$ vertices and $\ell \geq 1$ edges. The number of isolated copies of H in random graph $\mathbb{G}(n,p)$ obeys the strong law of large numbers in the following cases:*

1. *p is a constant.*
2. *$p = \dfrac{1}{n^{\delta}}$, where $\delta > \dfrac{1}{\ell}$.*

3.2 Proof of main results

Throughout our study, we use the Proposition 3.2.2 as our tool for proving our main results. So we begin this section with the following facts that give us this tool.

Lemma 3.2.1. *For each $\epsilon \in \mathbb{R}^+$, let $B(\epsilon)$ be the set depend on ϵ. If $B(\epsilon)$ is increasing with ϵ then $\displaystyle\bigcap_{\epsilon>0} B(\epsilon) = \bigcap_{n\in\mathbb{N}} B\left(\frac{1}{n}\right)$.*

Proof. We assume that $B(\epsilon)$ is increasing with ϵ.
Then $\displaystyle\bigcap_{\epsilon>0} B(\epsilon) = \bigcap_{\epsilon\in(0,1]} B(\epsilon)$. It's clear that $\displaystyle\bigcap_{\epsilon\in(0,1]} B(\epsilon) \subseteq \bigcap_{n\in\mathbb{N}} B\left(\frac{1}{n}\right)$.
By Archimedean property, for each $\epsilon \in (0,1]$ there exists $n \in \mathbb{N}$ such that $\dfrac{1}{n} < \epsilon$.
Thus for each $\epsilon \in (0,1]$ there exists $n_\epsilon \in \mathbb{N}$ such that $B(\epsilon) \supseteq B\left(\dfrac{1}{n_\epsilon}\right)$.
Hence $\displaystyle\bigcap_{\epsilon\in(0,1]} B(\epsilon) \supseteq \bigcap_{n_\epsilon} B\left(\frac{1}{n_\epsilon}\right) \supseteq \bigcap_{n\in\mathbb{N}} B\left(\frac{1}{n}\right)$.
Therefore $\displaystyle\bigcap_{\epsilon>0} B(\epsilon) = \bigcap_{\epsilon\in(0,1]} B(\epsilon) = \bigcap_{n\in\mathbb{N}} B\left(\frac{1}{n}\right)$. □

Proposition 3.2.1. (The first Borel-Cantelli Lemma) *Let* $A_1, A_2, ...$ *be sequence of events from a probability space* $(\Omega, \Im, P)$, *if* $\sum_{n=1}^{\infty} P(A_n) < \infty$, *then*

$P\left(\bigcap_{n\in\mathbb{N}} \bigcup_{m=n}^{\infty} A_m\right) = 0.$

Proof. (see [12], pp. 320) □

Proposition 3.2.2. *Let* (X_n) *be a sequence of finite variance random variables on a probability space* $(\Omega, \mathcal{F}, P)$ *and* (a_n) *be a sequence of positive real numbers. If* $\sum_{n=1}^{\infty} \frac{\operatorname{Var} X_n}{a_n^2} < \infty$ *then* $\frac{1}{a_n}[X_n - E(X_n)] \xrightarrow{a.s.} 0$.

Proof. We assume that $\sum_{n=1}^{\infty} \frac{\operatorname{Var} X_n}{a_n^2} < \infty$.

Let $\epsilon > 0$ be arbitrary and $A_n(\epsilon) = \left\{\omega \in \Omega : \left|\frac{X_n - E(X_n)}{a_n}\right| \geq \epsilon\right\}$.

By Chebyšhev's inequality, we get

$$\sum_{n=1}^{\infty} P(A_n(\epsilon)) = \sum_{n=1}^{\infty} P\left(\left|\frac{X_n - E(X_n)}{a_n}\right| \geq \epsilon\right) \leq \frac{1}{\epsilon^2} \sum_{n=1}^{\infty} \frac{\operatorname{Var}(X_n)}{a_n^2} < \infty.$$

Hence by the first Borel-Cantelli Lemma, $P\left(\bigcap_{n\in\mathbb{N}} \bigcup_{m=n}^{\infty} A_m(\epsilon)\right) = 0$.

That is $P\left(\bigcup_{n\in\mathbb{N}} \bigcap_{m=n}^{\infty} A_m^c(\epsilon)\right) = 1$ for any $\epsilon > 0$.

Since $\bigcup_{n\in\mathbb{N}} \bigcap_{m=n}^{\infty} A_m^c(\epsilon)$ is increasing with ϵ, it follows from Lemma 3.2.1 that

$\bigcap_{\epsilon>0} \bigcup_{n\in\mathbb{N}} \bigcap_{m=n}^{\infty} A_m^c(\epsilon) = \bigcap_{k\in\mathbb{N}} \bigcup_{n\in\mathbb{N}} \bigcap_{m=n}^{\infty} A_m^c\left(\frac{1}{k}\right).$

Note that $\bigcup_{n\in\mathbb{N}} \bigcap_{m=n}^{\infty} A_m^c\left(\frac{1}{k}\right)$ is decreasing with k.

We obtain that

$$
\begin{aligned}
&P\left(\left\{\omega \in \Omega \ : \ \left[\frac{X_n - E(X_n)}{a_n}\right](\omega) \to 0 \ \text{ as } n \to \infty\right\}\right) \\
&= P\left(\left\{\omega \in \Omega : \ \forall \epsilon > 0, \ \exists n \in \mathbb{N}, \ \forall m \geq n, \ \left|\frac{X_m(\omega) - E(X_m)}{a_m}\right| < \epsilon\right\}\right) \\
&= P\left(\bigcap_{\epsilon>0} \bigcup_{n\in\mathbb{N}} \bigcap_{m=n}^{\infty} \left\{\omega \in \Omega \ : \ \left|\frac{X_m(\omega) - E(X_m)}{a_m}\right| < \epsilon\right\}\right) \\
&= P\left(\bigcap_{\epsilon>0} \bigcup_{n\in\mathbb{N}} \bigcap_{m=n}^{\infty} A_m^c(\epsilon)\right) \\
&= P\left(\bigcap_{k\in\mathbb{N}} \bigcup_{n\in\mathbb{N}} \bigcap_{m=n}^{\infty} A_m^c\left(\frac{1}{k}\right)\right) \\
&= \lim_{k\to\infty} P\left(\bigcup_{n\in\mathbb{N}} \bigcap_{m=n}^{\infty} A_m^c\left(\frac{1}{k}\right)\right) \\
&= 1.
\end{aligned}
$$

That implies $\frac{1}{a_n}[X_n - E(X_n)] \xrightarrow{a.s.} 0$. □

Now, we are ready to prove our main theorems. From now on, we let $q = 1 - p$ and $0 < p < 1$.

3.2.1 Proof of Theorem 3.1.1

Proof. Let a non negative integer d be fixed and $S_{n,d}$ be the number of vertices with degree d in $\mathbb{G}(n, p)$. For each $n \in \mathbb{N}$ and $i \in \{1, 2, ..., n\}$, we define the indicator random variable

$$
X_{in} = \begin{cases} 1, & \text{if the vertex } i \text{ in } \mathbb{G}(n,p) \text{ has degree } d; \\ 0, & \text{otherwise.} \end{cases}
$$

It's clearly that

$$
S_{n,d} = \begin{cases} \sum_{i=1}^{n} X_{in}, & \text{if } \ n > d; \\ 0, & \text{if } \ n \leq d. \end{cases} \tag{3.1}
$$

Barbour, Karonski and Rucinski ([3]) show that

$$E(S_{n,d}) = E(\sum_{i=1}^{n} X_{in}) = nE(X_{in}) = n\binom{n-1}{d} p^d q^{n-1-d}, \tag{3.2}$$

and

$$\begin{aligned}\mathrm{Var}(S_{n,d}) =& \frac{n}{n-1}\binom{n-1}{d}^2 (d-(n-1)\,p)^2 p^{2d-1} q^{2(n-d)-3} \\ &+ E(S_{n,d}) - \frac{1}{n}E^2(S_{n,d}).\end{aligned} \tag{3.3}$$

For $d \geq 1$, it follows from (3.1) and (3.2) -(3.3) that $\mathrm{Var}(S_{n,d}) = 0$ for every $n \leq d$ and for each $n > d$,

$$\begin{aligned}\mathrm{Var}(S_{n,d}) &= \frac{n}{n-1}\binom{n-1}{d}^2 [d-(n-1)\,p]^2 p^{2d-1} q^{2(n-d)-3} \\ &\quad + E(S_{n,d}) - \frac{1}{n}E^2(S_{n,d}) \\ &= \frac{n}{n-1}\binom{n-1}{d}^2 [d-(n-1)\,p]^2 p^{2d-1} q^{2(n-d)-3} \\ &\quad + n\binom{n-1}{d} p^d q^{n-1-d} - n\binom{n-1}{d}^2 p^{2d} q^{2(n-d)-2} \\ &= \frac{n}{n-1}\binom{n-1}{d}^2 \left[d^2 - 2dp\,(n-1) + p^2\,(n-1)^2\right] p^{2d-1} q^{2(n-d)-3} \\ &\quad + n\binom{n-1}{d} p^d q^{n-1-d} - n\binom{n-1}{d}^2 p^{2d} q^{2(n-d)-2} \\ &\leq (n^2-n)\binom{n-2}{d-1}^2 p^{2d-1} q^{2(n-d)-3} + (n^2-n)\binom{n-1}{d}^2 p^{2d+1} q^{2(n-d)-3} \\ &\quad + n\binom{n-1}{d} p^d q^{n-1-d} \\ &\leq n^2\binom{n-2}{d-1}^2 p^{2d-1} q^{2(n-d)-3} + n^2\binom{n-1}{d}^2 p^{2d+1} q^{2(n-d)-3} \\ &\quad + n\binom{n-1}{d} p^d q^{n-1-d} \\ &\leq n^2\left(\frac{n^{2(d-1)}}{[(d-1)!]^2}\right) p^{2d-1} q^{2(n-d)-3} + n^2\left(\frac{n^{2d}}{(d!)^2}\right) p^{2d+1} q^{2(n-d)-3} \\ &\quad + n\left(\frac{n^d}{d!}\right) p^d q^{n-1-d} \\ &= \frac{1}{[(d-1)!]^2} n^{2d} p^{2d-1} q^{2(n-d)-3} + \frac{1}{(d!)^2} n^{2d+2} p^{2d+1} q^{2(n-d)-3} \\ &\quad + \frac{1}{d!} n^{d+1} p^d q^{n-1-d} \\ &=: \ a_n + \ b_n + \ c_n\end{aligned} \tag{3.4}$$

where $a_n =: \dfrac{1}{[(d-1)!]^2} n^{2d} p^{2d-1} q^{2(n-d)-3}$, $b_n =: \dfrac{1}{(d!)^2} n^{2d+2} p^{2d+1} q^{2(n-d)-3}$

and $\quad c_n =: \dfrac{1}{d!} n^{d+1} p^d q^{n-1-d}$.

Case 1: Let p be a constant . To prove that $S_{n,d}$ obeys the strong law of large numbers, it suffices to show that $[S_{n,d} - E(S_{n,d})] \xrightarrow{a.s.} 0$.
We will show that $\sum_{n=1}^{\infty} \mathrm{Var}(S_{n,d}) < \infty$ by considering the case $d = 0$ and $d \geq 1$.

For $d = 0$, it follows from (3.2) and (3.3) that

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathrm{Var}(S_{n,0}) &= \sum_{n=1}^{\infty} \left[\frac{n}{n-1} (n-1)^2 pq^{2n-3} + E(S_{n,0}) - n^{-1} E^2(S_{n,0}) \right] \\
&= \sum_{n=1}^{\infty} \left[n^2 pq^{2n-3} - npq^{2n-3} + nq^{n-1} - nq^{2n-2} \right] \\
&= \sum_{n=1}^{\infty} \left[nq^{n-1} + n^2 pq^{2n-3} - n(p+q) q^{2n-3} \right] \\
&= \sum_{n=1}^{\infty} \left[nq^{n-1} + n^2 pq^{2n-3} - nq^{2n-3} \right] \\
&= \frac{1}{q} \sum_{n=1}^{\infty} nq^n + \frac{p}{q^3} \sum_{n=1}^{\infty} n^2 q^{2n} - \frac{1}{q^3} \sum_{n=1}^{\infty} nq^{2n}.
\end{aligned}$$

Each series on the right side converges by the ratio test, which implies $\sum_{n=1}^{\infty} \mathrm{Var}(S_{n,0}) < \infty$.

For $d \geq 1$, It follows from (3.4) that

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathrm{Var}(S_{n,d}) &= \sum_{n=d+1}^{\infty} \mathrm{Var}(S_{n,d}) \\
&\leq \sum_{n=d+1}^{\infty} a_n + \sum_{n=d+1}^{\infty} b_n + \sum_{n=d+1}^{\infty} c_n
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{n=d+1}^{\infty} \frac{1}{[(d-1)!]^2} n^{2d} p^{2d-1} q^{2(n-d)-3} \\
&\quad + \sum_{n=d+1}^{\infty} \frac{1}{(d!)^2} n^{2d+2} p^{2d+1} q^{2(n-d)-3} \\
&\quad + \sum_{n=d+1}^{\infty} \frac{1}{d!} n^{d+1} p^{d} q^{n-1-d} \\
&\leq C_1 \sum_{n=1}^{\infty} n^{2d} q^{2n} + C_2 \sum_{n=1}^{\infty} n^{2d+2} q^{2n} + C_3 \sum_{n=1}^{\infty} n^{d+1} q^{n}
\end{aligned}
$$

where C_1, C_2 and C_3 are certain positive constants. Then $\sum_{n=1}^{\infty} \text{Var}(S_{n,d}) < \infty$ by applying the ratio test with each series on the right side.

According to the proposition 3.2.2, we get $[S_{n,d} - E(S_{n,d})] \xrightarrow{a.s.} 0$ when $d \geq 0$ and p be a constant.

Case 2: We let $p = \frac{1}{n^{\delta}}$, $\delta > 0$ and $d \geq 1$.
We will show that $\frac{1}{n}[S_{n,d} - E(S_{n,d})] \xrightarrow{a.s.} 0$. It follows from (3.4) that

$$
\begin{aligned}
\sum_{n=1}^{\infty} \frac{\text{Var}(S_{n,d})}{n^2} &= \sum_{n=d+1}^{\infty} \frac{\text{Var}(S_{n,d})}{n^2} \\
&\leq \sum_{n=d+1}^{\infty} \frac{a_n}{n^2} + \sum_{n=d+1}^{\infty} \frac{b_n}{n^2} + \sum_{n=d+1}^{\infty} \frac{c_n}{n^2} \\
&= \frac{a_{d+1} + b_{d+1}}{(d+1)^2} + \frac{1}{[(d-1)!]^2} \sum_{n=d+2}^{\infty} n^{2d-2} p^{2d-1} q^{2(n-d)-3} \\
&\quad + \frac{1}{(d!)^2} \sum_{n=d+2}^{\infty} n^{2d} p^{2d+1} q^{2(n-d)-3} + \frac{1}{d!} \sum_{n=d+1}^{\infty} n^{d-1} p^{d} q^{n-1-d}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{a_{d+1}+b_{d+1}}{(d+1)^2} + \frac{1}{[(d-1)!]^2}\sum_{n=d+2}^{\infty}\frac{1}{n^{\delta(2d-1)-2d+2}}(1-\frac{1}{n^{\delta}})^{2(n-d)-3} \\
&\quad + \frac{1}{(d!)^2}\sum_{n=d+2}^{\infty}\frac{1}{n^{\delta(2d+1)-2d}}(1-\frac{1}{n^{\delta}})^{2(n-d)-3} \\
&\quad + \frac{1}{d!}\sum_{n=d+1}^{\infty}\frac{1}{n^{\delta d-d+1}}(1-\frac{1}{n^{\delta}})^{n-1-d} \\
&\le \frac{a_{d+1}+b_{d+1}}{(d+1)^2} + \frac{1}{[(d-1)!]^2}\sum_{n=d+2}^{\infty}\frac{1}{n^{\delta(2d-1)-2d+2}} \\
&\quad + \frac{1}{(d!)^2}\sum_{n=d+2}^{\infty}\frac{1}{n^{\delta(2d+1)-2d}} + \frac{1}{d!}\sum_{n=d+1}^{\infty}\frac{1}{n^{\delta d-d+1}}.
\end{aligned}
$$

Each series on the right side converges when $\delta > 1$. By the proposition 3.2.2, $\frac{1}{n}\left[S_{n,d} - E(S_{n,d})\right] \xrightarrow{a.s.} 0$ where $\delta > 1$. □

3.2.2 Proof of Corollary 3.1.1

Proof. It follows directly from Theorem 1.1 in the case of $d = 0$. □

3.2.3 Proof of Theorem 3.1.2

Proof. Let a positive integer $k \ge 2$ be fixed and $S_{n,k}$ be the number of isolated trees of order k in $\mathbb{G}(n,p)$. For each $n \ge k$, we define

$$D_{n,k} =: \left\{\vec{i} = (i_1, i_2, ..., i_k) : \ 1 \le i_1 < i_2 < ... < i_k \le n\right\}$$

be the set of all possible combinations of k vertices and for each $\vec{i} \in D_{n,k}$, we define

$$
X_{\vec{i}n} = \begin{cases} 1, & \text{if the induced subgraph which is spanned by all vertices} \\ & \quad \text{of } \vec{i}\text{, is an isolated trees in } \mathbb{G}(n,p); \\ 0, & \text{otherwise.} \end{cases}
$$

Stien ([18]) shows that for $n \geq k$

$$E(X_{\vec{i}n}) = P\left(X_{\vec{i}n} = 1\right) = k^{k-2}p^{k-1}q^{k(n-k)+\binom{k}{2}-(k-1)}. \tag{3.5}$$

It's obvious that

$$S_{n,k} = \begin{cases} \sum_{\vec{i} \in D_{n,k}} X_{\vec{i}n}, & \text{if} \quad n \geq k \quad ; \\ 0, & \text{if} \ \ n < k. \end{cases}$$

Thus

$$\mathrm{Var}(S_{n,k}) = 0 \ \text{ for every } \ n < k. \tag{3.6}$$

Barbour, Karonski and Rucinski ([3]) show that for each $n \geq k$,

$$\mathrm{Cov}(X_{\vec{i}n}, X_{\vec{j}n}) = \begin{cases} E(X_{\vec{i}n}) - E^2(X_{\vec{i}n}), & \text{if } \vec{i} = \vec{j} \ ; \\ E(X_{\vec{i}n})E(X_{\vec{j}n})(q^{-k^2} - 1), & \text{if } \vec{i} \text{ and } \vec{j} \text{ have disjoint vertices ;} \\ -E(X_{\vec{i}n})E(X_{\vec{j}n}), & \text{if } \vec{i} \neq \vec{j} \text{ and } \vec{i},\ \vec{j} \text{ have at least one} \\ & \ \text{vertex in common.} \end{cases}$$

For each $\vec{i} \in D_{n,k}$, we let

$$L_{\vec{i}} =: \{\vec{j} \in D_{n,k} : \ \vec{i} \text{ and} \vec{j} \text{ have disjoint vertices}\},$$

and

$$L'_{\vec{i}} =: \{\vec{j} \in D_{n,k} : \vec{i} \neq \vec{j} \text{ and } \vec{i}, \vec{j} \text{ have at least one vertex in common}\}.$$

Hence by (3.5),

$$\mathrm{Cov}(X_{\vec{i}n}, X_{\vec{j}n}) = \begin{cases} k^{k-2}p^{k-1}q^{k(n-k)+\binom{k}{2}-k+1} & \\ -k^{2(k-2)}p^{2k-2}q^{2k(n-k)+2\binom{k}{2}-2k+2}, & \text{if } \vec{j} = \ \vec{i} \ ; \\ & \\ k^{2(k-2)}p^{2k-2}q^{2k(n-k)+2\binom{k}{2}-2k+2}\left(q^{-k^2} - 1\right), & \text{if } \vec{i} \in L_{\vec{i}} \ ; \\ & \\ -k^{2(k-2)}p^{2k-2}q^{2k(n-k)+2\binom{k}{2}-2k+2}, & \text{if } \vec{j} \in L'_{\vec{i}} \ . \end{cases}$$

Then we obtain that for each $n \geq k$,

$$
\begin{aligned}
\operatorname{Var}(S_{n,k}) &= \sum_{\vec{i}\in D_{n,k}} \sum_{\vec{j}\in D_{n,k}} \operatorname{Cov}(X_{\vec{i}n}, X_{\vec{j}n}) \\
&= \sum_{\vec{i}\in D_{n,k}} \sum_{\vec{j}=\vec{i}} \operatorname{Cov}(X_{\vec{i}n}, X_{\vec{j}n}) + \sum_{\vec{i}\in D_{n,k}} \sum_{\vec{j}\in L_{\vec{i}}} \operatorname{Cov}(X_{\vec{i}n}, X_{\vec{j}n}) \\
&\quad + \sum_{\vec{i}\in D_{n,k}} \sum_{\vec{j}\in L'_{\vec{i}}} \operatorname{Cov}(X_{\vec{i}n}, X_{\vec{j}n}) \\
&= \binom{n}{k} [k^{k-2} p^{k-1} q^{k(n-k)+\binom{k}{2}-k+1} - k^{2(k-2)} p^{2k-2} q^{2k(n-k)+2\binom{k}{2}-2k+2}] \\
&\quad + \binom{n}{k}\binom{n-k}{k} [\, k^{2(k-2)} p^{2k-2} q^{2k(n-k)+2\binom{k}{2}-2k+2} (q^{-k^2}-1) \,] \\
&\quad - \binom{n}{k} \sum_{r=1}^{k-1} \binom{k}{r}\binom{n-k}{k-r} k^{2(k-2)} p^{2k-2} q^{2k(n-k)+2\binom{k}{2}-2k+2} \\
&\leq \binom{n}{k} k^{k-2} p^{k-1} q^{k(n-k)+\binom{k}{2}-k+1} \\
&\quad + \binom{n}{k}\binom{n-k}{k} k^{2(k-2)} p^{2k-2} q^{2kn-2k^2-3k+2} \\
&=: r_n + t_n \qquad (3.7)
\end{aligned}
$$

where $r_n =: \binom{n}{k} k^{k-2} p^{k-1} q^{k(n-k)+\binom{k}{2}-k+1}$,

and $\quad t_n =: \binom{n}{k}\binom{n-k}{k} k^{2(k-2)} p^{2k-2} q^{2kn-2k^2-3k+2}$.

Now, we suppose that p is a constant. It suffices to prove that $[S_{n,k} - E(S_{n,k})] \xrightarrow{a.s.} 0$. From (3.6) and (3.7), we get

$$
\begin{aligned}
\sum_{n=1}^{\infty} \operatorname{Var}(S_{n,k}) &= \sum_{n=k}^{\infty} \operatorname{Var}(S_{n,k}) \\
&\leq \sum_{n=k}^{\infty} r_n + \sum_{n=k}^{\infty} t_n \\
&= \sum_{n=k}^{\infty} \binom{n}{k} k^{k-2} p^{k-1} q^{k(n-k)+\binom{k}{2}-k+1} \\
&\quad + \sum_{n=2k}^{\infty} \binom{n}{k}\binom{n-k}{k} k^{2(k-2)} p^{2k-2} q^{2kn-2k^2-3k+2}
\end{aligned}
$$

$$
\begin{aligned}
&\le \sum_{n=k}^{\infty} \frac{n^k}{k!} k^{k-2} p^{k-1} q^{k(n-k)+\binom{k}{2}-k+1} \\
&\quad + \sum_{n=2k}^{\infty} \frac{n^{2k}}{(k!)^2} k^{2(k-2)} p^{2k-2} q^{2kn-2k^2-3k+2} \\
&\le c_1 \sum_{n=k}^{\infty} n^k q^{kn} + c_2 \sum_{n=2k}^{\infty} n^{2k} q^{2kn}
\end{aligned}
$$

where $c_1$and c_2 are positive constants. Thus $\sum_{n=1}^{\infty} \mathrm{Var}(S_{n,k}) < \infty$ by applying the ratio test with each series on the right side.
According to the proposition 3.2.2, $[S_{n,k} - E(S_{n,k})] \xrightarrow{a.s.} 0$ where p be a constant.

Next, suppose that $p = \frac{1}{n^\delta}$, $\delta > 0$. It follows from (3.6) and (3.7) that

$$
\begin{aligned}
\sum_{n=1}^{\infty} \frac{\mathrm{Var}(S_{n,k})}{\binom{n}{k}^2} &= \sum_{n=k}^{\infty} \frac{\mathrm{Var}(S_{n,k})}{\binom{n}{k}^2} \\
&\le \sum_{n=k}^{\infty} \frac{r_n}{\binom{n}{k}^2} + \sum_{n=k}^{\infty} \frac{t_n}{\binom{n}{k}^2} \\
&= \sum_{n=k}^{\infty} \frac{\binom{n}{k}}{\binom{n}{k}^2} k^{k-2} p^{k-1} q^{k(n-k)+\binom{k}{2}-k+1} + \sum_{n=k}^{2k-1} \frac{t_n}{\binom{n}{k}^2} \\
&\quad + \sum_{n=2k}^{\infty} \frac{\binom{n}{k}\binom{n-k}{k}}{\binom{n}{k}^2} k^{2(k-2)} p^{2k-2} q^{2kn-2k^2-3k+2} \\
&\le \sum_{n=k}^{\infty} \frac{k^{k-2}}{n^{\delta(k-1)}} \left(1 - \frac{1}{n^\delta}\right)^{k(n-k)+\binom{k}{2}-k+1} \\
&\quad + \sum_{n=2k}^{\infty} \frac{k^{2(k-2)}}{n^{\delta(2k-2)}} \left(1 - \frac{1}{n^\delta}\right)^{2kn-2k^2-3k+2} \\
&\le k^{k-2} \sum_{n=k}^{\infty} \frac{1}{n^{\delta(k-1)}} + k^{2(k-2)} \sum_{n=2k}^{\infty} \frac{1}{n^{\delta(2k-2)}} \\
&< \infty
\end{aligned}
$$

whenever $\delta > \frac{1}{k-1}$. According to the proposition 3.2.2, we get
$\frac{1}{\binom{n}{k}} [S_{n,k} - E(S_{n,k})] \xrightarrow{a.s.} 0$ where $\delta > \frac{1}{k-1}$. □

3.2.4 Proof of Theorem 3.1.3

Proof. Let $S_{n,H}$ be the number of isolated copies of H in $\mathbb{G}(n,p)$.

For each $n \geq k$, let $D_{n,k}$, $L_{\vec{i}}$ and $L'_{\vec{i}}$ be defined as in the proof of Theorem 3.1.2. Then for each $\vec{i} \in D_{n,k}$, we define

$$Y_{\vec{i}n} = \begin{cases} 1, & \text{if the induced subgraph which is spanned by all vertices} \\ & \text{ of } \vec{i}, \text{ is an isolated copy of } H \text{ in } \mathbb{G}(n,p); \\ 0, & \text{otherwise.} \end{cases}$$

Note that there are $\dfrac{k!}{aut(H)}$ possible copies of H which spanned by all vertices of $\vec{i}$, where $aut(H)$ stands for the number of automorphisms of H ([13], pp 141). Then we get

$$E(Y_{\vec{i}n}) = P\left(Y_{\vec{i}n} = 1\right) = \frac{k!}{aut(H)} p^{\ell} q^{k(n-k)+\binom{k}{2}-\ell}$$

and $E(Y_{\vec{i}n}) = E(Y_{\vec{j}n})$ for any $\vec{i},\ \vec{j} \in D_{n,k}$.

It's clearly that

$$S_{n,H} = \begin{cases} \displaystyle\sum_{\vec{i} \in D_{n,k}} Y_{\vec{i}n}, & \text{if } \ n \geq k; \\ 0, & \text{if } \ n < k. \end{cases}$$

Thus

$$\text{Var}(S_{n,H}) = 0 \ \text{ for every } \ n < k. \tag{3.8}$$

Now, we will determine $\mathrm{Var}(S_{n,H})$ for $n \geq k$. Consider, if $\vec{i} = \vec{j}$ then $E(Y_{\vec{i}n}Y_{\vec{j}n}) = E(Y_{\vec{i}n}^2) = E(Y_{\vec{i}n})$ and hence,

$$\begin{aligned}
\mathrm{Cov}(Y_{\vec{i}n}, Y_{\vec{j}n}) &= E(Y_{\vec{i}n}Y_{\vec{j}n}) - E(Y_{\vec{i}n})E(Y_{\vec{j}n}) \\
&= \frac{k!}{aut(H)} p^{\ell} q^{k(n-k)+\binom{k}{2}-\ell} - (\frac{k!}{aut(H)})^2 p^{2\ell} q^{2k(n-k)+2\binom{k}{2}-2\ell} \\
&= \frac{k!}{aut(H)} p^{\ell} q^{kn-k^2+\binom{k}{2}-\ell} - (\frac{k!}{aut(H)})^2 p^{2\ell} q^{2kn-k^2-k-2\ell}. \qquad (3.9)
\end{aligned}$$

In case of $\vec{j} \in L'_{\vec{i}}$ we have $E(Y_{\vec{i}n}Y_{\vec{j}n}) = 0$ which implies that

$$\begin{aligned}
\mathrm{Cov}(Y_{\vec{i}n}, Y_{\vec{j}n}) &= E(Y_{\vec{i}n}Y_{\vec{j}n}) - E(Y_{\vec{i}n})E(Y_{\vec{j}n}) \\
&= -(\frac{k!}{aut(H)})^2 p^{2\ell} q^{2k(n-k)+2\binom{k}{2}-2\ell} \\
&= -(\frac{k!}{aut(H)})^2 p^{2\ell} q^{2kn-k^2-k-2\ell}. \qquad (3.10)
\end{aligned}$$

In case of $\vec{j} \in L_{\vec{i}}$ we get

$$E(Y_{\vec{i}n}Y_{\vec{j}n}) = P\left(Y_{\vec{i}n} = 1, Y_{\vec{j}n} = 1\right) = (\frac{k!}{aut(H)})^2 p^{2\ell} q^{2k(n-2k)+\binom{2k}{2}-2\ell}$$

and hence,

$$\begin{aligned}
\mathrm{Cov}(Y_{\vec{i}n}, Y_{\vec{j}n}) &= E(Y_{\vec{i}n}Y_{\vec{j}n}) - E(Y_{\vec{i}n})E(Y_{\vec{j}n}) \\
&= (\frac{k!}{aut(H)})^2 p^{2\ell} q^{2k(n-2k)+\binom{2k}{2}-2\ell} - (\frac{k!}{aut(H)})^2 p^{2\ell} q^{2k(n-k)+2\binom{k}{2}-2\ell} \\
&= (\frac{k!}{aut(H)})^2 p^{2\ell} q^{2kn-2k^2-k-2\ell} - (\frac{k!}{aut(H)})^2 p^{2\ell} q^{2kn-k^2-k-2\ell}. \qquad (3.11)
\end{aligned}$$

From (3.9)-(3.11) we obtain that for $n \geq k$,

$$\begin{aligned}
\mathrm{Var}(S_{n,H}) &= \sum_{\vec{i}\in D_{n,k}} \sum_{\vec{j}\in D_{n,k}} \mathrm{Cov}(Y_{\vec{i}n}, Y_{\vec{j}n}) \\
&= \sum_{\vec{i}\in D_{n,k}} \sum_{\vec{j}=\vec{i}} \mathrm{Cov}(Y_{\vec{i}n}, Y_{\vec{j}n}) + \sum_{\vec{i}\in D_{n,k}} \sum_{\vec{j}\in L_{\vec{i}}} \mathrm{Cov}(Y_{\vec{i}n}, Y_{\vec{j}n}) \\
&\quad + \sum_{\vec{i}\in D_{n,k}} \sum_{\vec{j}\in L'_{\vec{i}}} \mathrm{Cov}(Y_{\vec{i}n}, Y_{\vec{j}n}) \\
&= \binom{n}{k} \Big[\frac{k!}{aut(H)} p^{\ell} q^{kn-k^2+\binom{k}{2}-\ell} - \Big(\frac{k!}{aut(H)}\Big)^2 p^{2\ell} q^{2kn-k^2-k-2\ell} \Big] \\
&\quad + \binom{n}{k}\binom{n-k}{k} \Big[\Big(\frac{k!}{aut(H)}\Big)^2 p^{2\ell} q^{2kn-2k^2-k-2\ell} - \Big(\frac{k!}{aut(H)}\Big)^2 p^{2\ell} q^{2kn-k^2-k-2\ell} \Big] \\
&\quad - \binom{n}{k} \sum_{r=1}^{k-1} \binom{k}{r}\binom{n-k}{k-r} \Big(\frac{k!}{aut(H)}\Big)^2 p^{2\ell} q^{2kn-k^2-k-2\ell} \\
&\leq \binom{n}{k} \frac{k!}{aut(H)} p^{\ell} q^{kn-k^2+\binom{k}{2}-\ell} + \binom{n}{k}\binom{n-k}{k} \Big(\frac{k!}{aut(H)}\Big)^2 p^{2\ell} q^{2kn-2k^2-k-2\ell} \\
&=: w_n + z_n \qquad (3.12)
\end{aligned}$$

where $w_n =: \binom{n}{k} \frac{k!}{aut(H)} p^{\ell} q^{kn-k^2+\binom{k}{2}-\ell}$,

and $\quad z_n =: \binom{n}{k}\binom{n-k}{k} \Big(\frac{k!}{aut(H)}\Big)^2 p^{2\ell} q^{2kn-2k^2-k-2\ell}$.

Now, suppose that p is a constant. It suffices to show that $[S_{n,H} - E(S_{n,H})] \xrightarrow{a.s.} 0$. From (3.8) and (3.12) we get

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathrm{Var}(S_{n,H}) &= \sum_{n=k}^{\infty} \mathrm{Var}(S_{n,H}) \\
&\leq \sum_{n=k}^{\infty} \binom{n}{k} \frac{k!}{aut(H)} p^{\ell} q^{kn-k^2+\binom{k}{2}-\ell} \\
&\quad + \sum_{n=2k}^{\infty} \binom{n}{k}\binom{n-k}{k} \Big(\frac{k!}{aut(H)}\Big)^2 p^{2\ell} q^{2kn-2k^2-k-2\ell} \\
&\leq C_1 \sum_{n=k}^{\infty} n^k q^{kn} + C_2 \sum_{n=2k}^{\infty} n^{2k} q^{2kn}
\end{aligned}$$

where C_1, C_2 are certain positive constants. Then $\sum_{n=1}^{\infty} \mathrm{Var}(S_{n,H}) < \infty$ by using the ratio test. According to the proposition 3.2.2, $[S_{n,H} - E(S_{n,H})] \xrightarrow{a.s.} 0$, where p is a constant.

Next, suppose that $p = \dfrac{1}{n^{\delta}}$, $\delta > 0$. Then from (3.8) and (3.12), we obtain

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{\mathrm{Var}(S_{n,H})}{\binom{n}{k}^2} &= \sum_{n=k}^{\infty} \frac{\mathrm{Var}(S_{n,H})}{\binom{n}{k}^2} \\
&\leq \sum_{n=k}^{\infty} \frac{w_n}{\binom{n}{k}^2} + \sum_{n=k}^{\infty} \frac{z_n}{\binom{n}{k}^2} \\
&= \frac{k!}{aut(H)} \sum_{n=k}^{\infty} \frac{\binom{n}{k}}{\binom{n}{k}^2} p^{\ell} q^{kn-k^2+\binom{k}{2}-\ell} + \sum_{n=k}^{2k-1} \frac{z_n}{\binom{n}{k}^2} \\
&\quad + (\frac{k!}{aut(H)})^2 \sum_{n=2k}^{\infty} \frac{\binom{n}{k}\binom{n-k}{k}}{\binom{n}{k}^2} p^{2\ell} q^{2kn-2k^2-k-2\ell} \\
&\leq \frac{k!}{aut(H)} \sum_{n=k}^{\infty} \frac{1}{n^{\delta\ell}} \left(1 - \frac{1}{n^{\delta}}\right)^{kn-k^2+\binom{k}{2}-\ell} \\
&\quad + (\frac{k!}{aut(H)})^2 \sum_{n=2k}^{\infty} \frac{1}{n^{2\delta\ell}} \left(1 - \frac{1}{n^{\delta}}\right)^{2kn-2k^2-k-2\ell} \\
&\leq \frac{k!}{aut(H)} \sum_{n=k}^{\infty} \frac{1}{n^{\delta\ell}} + (\frac{k!}{aut(H)})^2 \sum_{n=2k}^{\infty} \frac{1}{n^{2\delta\ell}} \\
&< \infty
\end{aligned}$$

where $\delta > \dfrac{1}{\ell}$. According to the proposition 3.2.2, $S_{n,H}$ obeys the strong law of large numbers whenever $\delta > \dfrac{1}{\ell}$. $\square$

CHAPTER IV
STRONG LAW OF LARGE NUMBERS ON RANDOM SOMATIC CELL HYBRID PANELS

4.1 Introduction and main result

Somatic cell hybrids are usually used to assign particular human genes to specific human chromosomes ([16, 17, 20, 21, 23]). The potential of human gene localization by rodent-human somatic cell hybrids has been confirmed since the pioneering work of Weiss and Green ([21]). Rodent-human somatic cell hybrids are formed by fusing normal diploid human somatic cells with permanently transformed rodent cells. The resulting hybrid somatic cells retrain all of the rodent chromosomes while losing random subsets of the human chromosomes. A few generation after their formation, a collection of different hybrid clones are analyzed for the expression of the human gene and for the presence of each of the 24 distinct human chromosomes. The chromosomes bearing the interested gene are consistently present in the hybrid clones expressing the gene and consistently absent in that clones not expressing it. From this pattern one can assign the gene to the particular chromosome. Since the Y chromosome bear few gene of interest, we will focus on somatic hybrid clones derived from human female cells that give total of 23 different chromosome types: 22 autosomes and the X chromosome.

Throughout this chapter we study on the mathematical models for the design of hybrid clone panels in the paper of Lange ([14]). There are three assumptions that should be satisfied when somatic cell hybrid panels are randomly created. First, each human chromosome is lost or retrained independently during the formation of a stable clone. Second, there is a common retention probability p_0 where $0 < p_0 < 1$, applying to all chromosome pairs. Third, different clones behave independently in their patterns.

0	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0	1	1	0	1	1	1	1
1	0	1	0	1	1	0	0	1	0	0	0	0	1	0	0	1	0	1	0	1	1	1
0	1	1	1	1	0	1	0	0	0	0	0	1	0	0	1	1	0	1	1	0	1	1
1	1	1	0	0	1	1	0	0	1	0	1	0	0	0	1	1	1	0	0	1	0	1
0	0	0	1	1	1	1	0	0	0	1	1	1	1	1	0	1	0	0	0	1	1	0
0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0
0	0	1	0	1	0	1	1	0	1	1	1	0	0	0	0	1	1	1	1	1	0	0
0	0	0	1	0	1	1	1	0	0	0	1	0	1	1	1	1	0	1	0	1	0	1
1	0	0	0	1	1	0	0	0	1	0	1	1	0	1	0	1	0	1	1	0	0	1

Figure 4.1: An example of karyotype matrix of a somatic cell hybrid panel

Let n denote the number of distinct hybrid clones in a panel. We construct a karyotype matrix of this panel when each clone in the panel is assayed for the presence of each of 23 chromosomes. It consists of n rows and 23 columns and its entry in the row i and column j is 1 if the clone i contains chromosome j ; otherwise it is 0. We also construct an additional test column of 0's and 1's when each clone is assayed for the presence of a given human gene. Barring assay errors or failure of critical assumptions ([14, 15]), the test column will uniquely match one of the columns of the matrix. In this case the gene is assigned to the corresponding chromosome. If two columns of the karyotype matrix of a panel are identical, then the gene assignment becomes ambiguous for any gene lying on one of the two corresponding chromosomes. Figure 1 depicts the karyotype matrix of a hybrid panel with $n = 9$. This panel has an unusual property that every pair of columns differs at least three entries. This level of redundancy is useful. If a single assay error is made in creating a test column for a human gene, then the program of gene assignment must be successful. In practice, the level of redundancy is random. Minimum Hamming distance is a natural measure of the redundancy of a panel.

Let c_s^n denote the column s of the karyotype matrix of a random panel with n

distinct hybrid clones. The Hamming distance $\rho\,(c_s^n, c_t^n)$ between the chromosomes s and t is just the number of entries in which c_s^n and c_t^n differ.

Let $\Gamma = \{\alpha = s, t \ : \ 1 \leq s \neq t \leq 23\}$. The minimum Hamming distance of a panel is defined as $\min\limits_{\{s,t\}\in\Gamma} \rho\,(c_s^n, c_t^n)$.

Let d be a fixed Hamming distance and $W_{n,d}$ denote the number of pairs of chromosomes in the panel for which the Hamming distance is less than d.

Clearly, $W_{n,d}$ must be 0 when the minimum Hamming distance of the panel equals or exceeds d.

In this Chapter, we investigate the strong law of large numbers of $W_{n,d}$. The following is our main result.

Theorem 4.1.1. *$W_{n,d}$ obeys the strong law of large numbers when the retention probability p_0 is a constant , $0 < p_0 < 1$.*

4.2 Proof of Main Result

Proof. It suffices to prove that $[W_{n,d} - E(W_{n,d})] \xrightarrow{a.s.} 0$. Let p_0 be a constant and $0 < p_0 < 1$. For each $n \in \mathbb{N}$ and $\alpha = \{s, t\} \in \Gamma$, we define

$$X_{\alpha n} = \begin{cases} 1, & \text{if } \rho\,(c_s^n, c_t^n) < d\ ; \\ 0, & \text{otherwise.} \end{cases}$$

Then $X_{\alpha n}$'s are dependent identically distributed. Teerapabolarn and Neammanee ([19]) show that for $n \geq d$,

$$p =: E(X_{\alpha n}) = P(X_{\alpha n} = 1) = \sum_{i=0}^{d-1} \binom{n}{i} q^i (1-q)^{n-i}. \tag{4.1}$$

where $q = 2p_0(1 - p_0)$ is the probability that c_s^n and c_t^n differ in each entry.

It is obvious that

$$W_{n,d} = \begin{cases} \sum_{\alpha\in\Gamma} X_{\alpha n}, & \text{if } \ n \geq d \ ; \\ \binom{23}{2}, & \text{if } \ n < d. \end{cases}$$

i.e., $\mathrm{Var}(W_{n,d}) = 0$ if $n < d$ and for each $n \geq d$,

$$\begin{aligned} \mathrm{Var}(W_{n,d}) &= \mathrm{Var}(\sum_{\alpha\in\Gamma} X_{\alpha n}) \\ &= \sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma} \mathrm{Cov}(X_{\alpha n}, X_{\beta n}) \\ &= \sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma} [E(X_{\alpha n}X_{\beta n}) - E(X_{\alpha n})E(X_{\beta n})] \qquad (4.2) \end{aligned}$$

By Hölder's inequality and (4.1), we obtain that for any $\alpha, \ \beta \in \Gamma$,

$$\begin{aligned} E(X_{\alpha n}X_{\beta n}) &\leq \sqrt{E(X_{\alpha n}^2)}\sqrt{E(X_{\beta n}^2)} \\ &= \sqrt{E(X_{\alpha n})E(X_{\beta n})} \\ &= \sqrt{p^2} \\ &= p. \qquad (4.3) \end{aligned}$$

It follows from (4.1), (4.2) and (4.3) that for $n \geq d$,

$$\begin{aligned} \mathrm{Var}(W_{n,d}) &\leq \sum_{\alpha\in\Gamma}\sum_{\beta\in\Gamma} [p - E(X_{\alpha n})E(X_{\beta n})] \\ &= |\Gamma|^2 \left[p - p^2\right] \\ &\leq |\Gamma|^2 p \\ &= |\Gamma|^2 \sum_{i=0}^{d-1} \binom{n}{i} q^i (1-q)^{n-i} \qquad . \end{aligned}$$

Thus we have

$$\begin{aligned}
\sum_{n=1}^{\infty} \operatorname{Var} W_{n,d} &= \sum_{n=d}^{\infty} \operatorname{Var} W_{n,d} \\
&\leq |\Gamma|^2 \sum_{n=d}^{\infty} \sum_{i=0}^{d-1} \tbinom{n}{i} q^i (1-q)^{n-i} \\
&\leq |\Gamma|^2 \sum_{n=d}^{\infty} \sum_{i=0}^{d-1} \frac{n^i}{i!} q^i (1-q)^{n-i} \\
&= |\Gamma|^2 \sum_{i=0}^{d-1} \sum_{n=d}^{\infty} \frac{n^i}{i!} q^i (1-q)^{n-i} \\
&= |\Gamma|^2 \sum_{i=0}^{d-1} \frac{1}{i!} \left(\frac{q}{1-q}\right)^i \sum_{n=d}^{\infty} n^i (1-q)^n \\
&\leq |\Gamma|^2 \sum_{i=0}^{d-1} \frac{1}{i!} \left(\frac{q}{1-q}\right)^i \sum_{n=1}^{\infty} n^i (1-q)^n.
\end{aligned}$$

Since $0 < p_0 < 1$ and $q = 2p_0(1-p_0)$, $|1-q| < 1$. Hence for each $i = 0, 1, 2, ..., d-1$, the series $\sum_{n=1}^{\infty} n^i (1-q)^n$ converges by the ratio test. That implies $\sum\limits_{n=1}^{\infty} \operatorname{Var} W_{n,d} < \infty$.

According to the proposition 3.2.2, we obtain the required result. □

REFERENCES

[1] Aiello, W., Chung, F. and Lu, L.: A random graph model for massive graphs, *Proc. of the* 32^{th} *Annual ACM Symposium on Theory of Computing*, 171-180 (2000).

[2] Amaral, A. and et al.: Studying social relations croosculturally. *Ethnology.* **2**, 155-179 (1998).

[3] Barbour, A.D., Karoński, M. and Ruciński A.: A central limit theorem for decomposable random variables with application to random graphs. *J. Comb. Th. Ser. B.* **47**, 125-145 (1989).

[4] Erdös, P.: Some remarks on the theory of graphs. *Bull. Amer. Math. Soc.* **53**, 292-294 (1947).

[5] Erdös, P. and Rényi, A.: On random graphs. I. *Publ. Math. Debrecen.* **6**, 290-297 (1959).

[6] Erdös, P. and Rényi, A.: On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.* **5**, 17-61 (1960).

[7] Erdös, P. and Rényi, A.: On the strenght of connectedness of a random graph. *Acta Math. Acad. Sci. Hungar.* **12**, 261-267 (1961).

[8] Erdös, P. and Rényi, A.: On random matrices. *Publ. Math. Inst. Hungar. Acad. Sci.* **8**, 455-461 (1964).

[9] Erdös, P. and Rényi, A.: On the existence of a factor of degree one of a connected random graph. *Acta Math. Acad. Sci. Hungar.* **17**, 359-368 (1966).

[10] Erdös, P. and Rényi, A.: On random matrices II. *Studia Sci. Math. Hungar.* **3**, 459-464 (1968).

[11] Fararo, T. J. and Sunshine, M.: *A study of a biased friendship network.* Syracuse University Press, New York (1964).

[12] Grimmett, G. R. and Stirzaker, D. R.: *Probability and Random Processes.* Oxford University Press Inc., New York (2001).

[13] Janson, S., Luzak, T. and Rucinski, A.: *Random graphs.* John Wiley & Sons Inc., New York (2000).

[14] Lange, K.: Applications of coding theory to the design of somatic cell hybrid panel. *Mathematical Biosciences.* **91**, 201-219 (1988).

[15] Lange, K.: *Applied Probability.* Springer, New York (2003).

[16] Mohandas, T. and et al.: Assignment of the human gene for galactose-1-phosephate uridryltransferase to chromosome 9:Studies with chinese hamster-human somatic cell hybrids. *Proc. Nat. Acad. Sci. U.S.A.* **74**, 5628-5631 (1977).

[17] Reddy, U. R. and et al.: Localization of the human Ror1 gene(NTRKR1) to chromosome 1p31-p32 by fluorescence *in situ* hybridrization and somatic cell hybrid analysis. *Genomics.* **41**, 283-285 (1997).

[18] Stien, C. M.: *Approximate computation of expectations.* IMS, Hayward, California (1986).

[19] Teerapabolarn, K. and Neammanee, K.: A non-uniform bound on Poisson approximation in somatic cell hybrid model. *Mathematical Biosciences.* **195**, 56-64 (2005).

[20] Tischfield, R.J., Ricciuti, F. and Ruddle, F. H.: Chromosome assignments of genes in man using mouse-human somatic cell hybrids: Mitochondrial superoxide dimutase (Indophenol oxidase-B, Tetrameric) to chromosome 6. *Humangenetik.* **20**, 203-209 (1973).

[21] Wiess, M. and Green, H.: Human-mouse hybrid cell lines containing partial complements of human chromosomes and functioning human genes. *Proc. Nat. Acad. Sci. U.S.A.* **58**, 1104-1111 (1967).

[22] Williams, R.J. and Martinez, N.: Simple rules yield complex food webs. *Nature.* **404**, 180-183 (2000).

[23] Vora, S. and Francke, U.: Assignment of the human gene for liver-type-6-phosphofructokinase Isozyyme(PFKL) to chromosome 21 by using somatic cell hybrids and monoclonal anti-L antibody. *Proc. Nat. Acad. Sci. U.S.A.* **78**, 3738-3742 (1981).

VITA

Name	: Miss Namkharng Sangsrijan
Date of Birth	: May 5, 1980
Place of Birth	: Phayao, Thailand
Previous Studies	: B.S.(Mathematics), Chiang Mai University, 2003.
Work Experience	: Chun Wittayakhom School.