Republic of Iraq
Ministry of Higher Education and Scientific Research
University of Babylon
College of Education for Pure Sciences
Department of Mathematics

On Estimator of Mutation Rates

A Research

Submitted to the Faculty of the Department of Mathematics, College of Education for Pure Science, University of Babylon as a Partial Fulfillment of the Requirement for the Degree of High Diploma Education / Mathematics

by

Amel Alwan Sleibi Kazem

Supervised by

Dr. Ali Hussein Mahmood Al-Obaidi

2021 A.D. 1443 A.H

بسم الله الرحمن الرحيم

((فَتَعَالَى اللَّهُ الْمَلِكُ الْحَقُّ ۗ وَلَا تَعْجَلْ بِالْقُرْآنِ مِن قَبْلِ أَن يُقْضَىٰ إِلَيْكَ وَحْيُهُ ۖ وَقُل رَّبِّ زِدْنِي عِلْمًا))

صدق الله العظيم

(سورة طه ١١٤)

Supervisor Certification

I certify that this research which is entitled **''On Estimator of Mutation Rates**" was prepared by the student " **Amel Alwan Slebi Kazem**" under my supervision at the University of Babylon, College of Education for Pure Sciences as a partial requirement for the degree of High Diploma Education /Mathematics.

Signature:

Name: **Dr. Ali Hussein Mahmood Al-Obaidi**

Scientific grade: **Leet**.

Date: **/ /2021**

According to the available recommendation, I forward this research for debate by the examining committee.

Signature:

Name: **Dr. Azal Jafar Musa**

Scientific grade: **Asst. Prof.**

Date: **/ /2021**

Certification of Scientific Expert

I certify that I have read this research entitled "On Estimator of Mutation Rates" and found it is qualified for debate.

Signature:

Name:

Title:

Date: / / 2021

Certification of Linguistic Expert

I certify that I have read this research entitled **''On Estimator of Mutation Rates**" and corrected its grammatical mistakes; therefore, it has qualified for debate.

Signature:

Name:

Title:

Date: **/ / 2021**

Examining Committee Certification

We certify that we have read the research entitled **''On Estimator of Mutation Rates**", as a committee examined the student "**Amel Alwan Slebi Kazem**" in its contents and that in our opinion, it meets the standard of research for the degree of High Diploma Education in Mathematics.

Chairman

Signature:

Name:

Title:

Date: **/ / 2021**

Member

Signature:

Name:

Title:

Date**: / / 2021**

Member

Signature:

Name:

Title:

Date: **/ / 2021**

Supervisor

Signature:

Name:

Title:

Date**: / / 2021**

Approved by the Dear of the College

Signature:

Name: **Dr. Bahaa Hussien Salih Rabee**

Scientific grade: **Professor**

Address: **Dean of the College of Education for Pure Sciences**

Date: **/ / 2021**

Table of Contents

Dedication

I dedicate my research to

my family and many friends.

Special gratitude to

my loving parents, whose words of encouragement

and push for tenacity ring in my ears.

My sisters and brothers have never left

my side and are incredibly special

Acknowledgements

I would like to acknowledge everyone who played a role in my academic accomplishments. First of all, my parents supported me with love and understanding. Without you, I could never have reached this current level of success.

Secondly, my advisor and committee members, each of whom has provided patient advice and guidance throughout the research process. Thank you all for your unwavering support.

Abstract

In this research, the best estimator of mutation rate μ under specific conditions are discussed. The division of an organism into at most two progeny are considered. The probability law that governs this production is subjected to the binomial distribution with parameters 2 and $v = 1 - \mu$. The calculation of μ as an explicit formula is given in two different ways. Moreover, the formula of μ is good to use in any generation because it is fixed. The sample mean is suggested as the estimator of μ. Some essential properties of this estimator are presented, like unbiasedness and consistency.

Introduction

The finding of mutation rates is serious to consider molecular apparatuses related to the fundamental biological processes of DNA replication and repair. These procedures influence the existence and flexibility of simple creatures, affecting genomic constancy that can imply cancer and genetic disorders in humans [1,4,5].

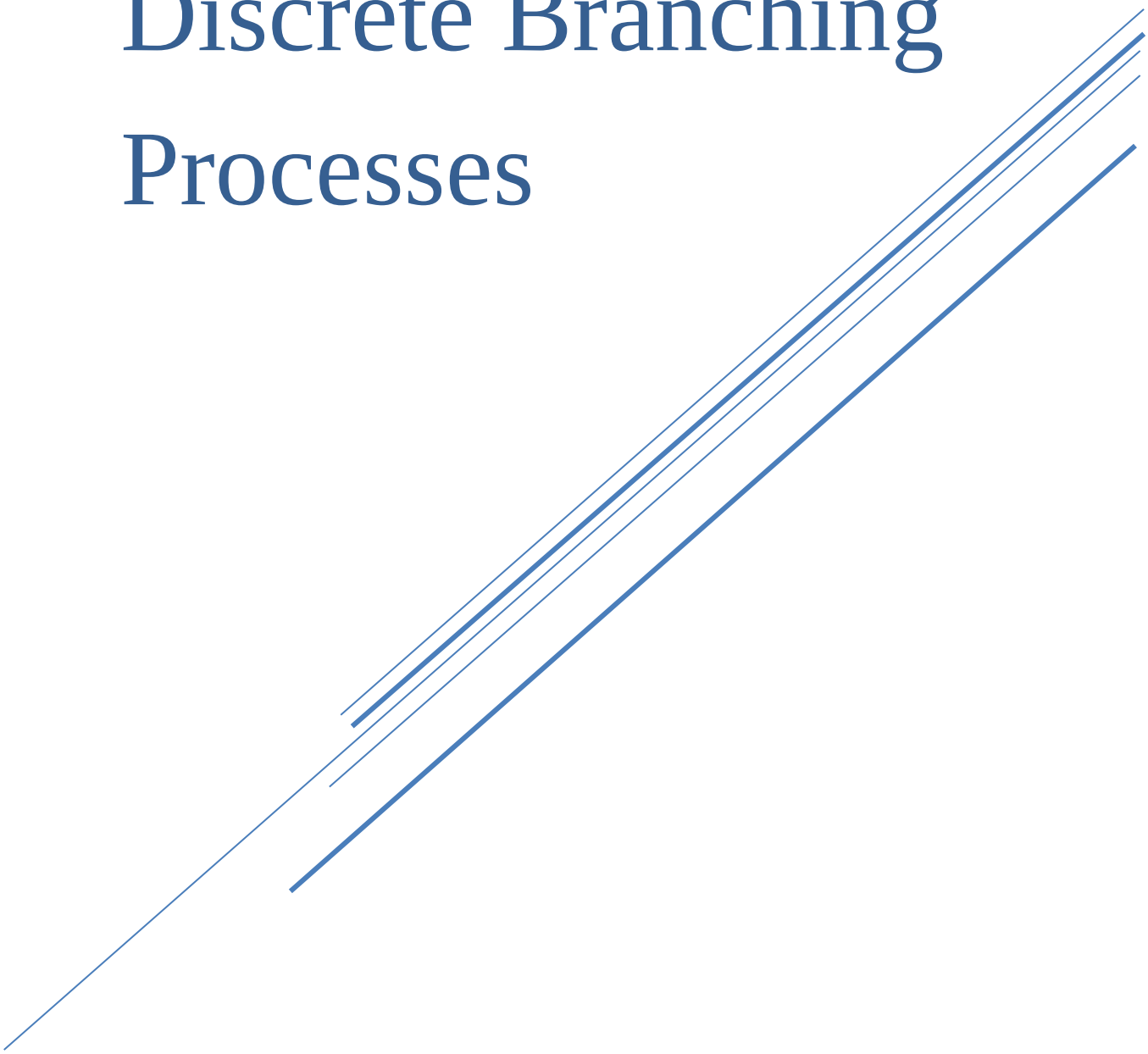
Recent ideas of the mutation rate for calculating bacterial mutation rates are deeply inspired by the attempt in 1943 when Luria and Delbrück realized the fluctuation test protocol and employed it to determine Escherichia coli mutation rates in the world's first fluctuation experiments [4]. However, mathematical mistakes in that significant paper produced incidents that harshly altered the mutation research. The primary incident is the lack of a firm definition of the mutation rate in genetics textbooks [6]. For example, the definition "The mutation rate is expressed as the number of mutations occurring in some units of time" in [7] seems acceptable, but it ignores the stochastic nature of the mutation.

To avoid this problem, some scholars adapted definitions under different considerations. However, this work studies and discusses the mutation rate and its estimator corresponding to the thought in [5]. Consequently, the mutation rate is subject to the binomial distribution.

The research includes three chapters. The first chapter has important ideas and background that are shown. For instance, the quantity of progenies and its distribution is described. Furthermore, some beliefs about these organisms are examined. Moreover, the size of the population is directed analogous to these assumptions. In addition, this size is exposed to time-homogeneous Markov chains. Finally, the expected number of progeny is got at a particular generation. The second chapter offers the mathematical system of the mutation rates by employing the branching process, especially the Galton-Watson process: Likewise, the mutation rate is determined as a ratio of the expected number of new mutants to the expected number of non-mutants in the nth generation. The crucial result is that this mutation rate is constant in every creation. As well, another approach to find this ratio is debated. Besides, the sample means estimator is investigated as an unbiased and consistent estimator for mutation rate.

Chapter One

Discrete Branching Processes



Chapter One

Discrete Branching Processes

In this chapter, the essential concepts and background are presented. For example, the number of progenies and its distribution is defined. Moreover, some assumptions about these creatures are discussed. Furthermore, the size of the population is indicated, corresponding to these hypotheses. Additionally, this size is subjected to time-homogeneous Markov chains. At the end of this chapter, the expected number of offspring is found at a particular generation. The chapter is a summary from the books in [2-4]

1.1. Preliminaries [2-4]

Suppose an organism at the end of its lifetime produces a random number Y of offspring with the probability distribution

$$p_k = P\{Y = K\}, K = 0,1, \ldots \qquad (1.1)$$

and with the corresponding pgf (probability generating function)

$$P(z) = E[Z^Y] = \sum_{k=0}^{\infty} p_k z^k, |z| \leq 1 \qquad (1.2)$$

We assume that all offspring act independently of each other and at the end of their lifetime (assuming that it is the same for all) have progeny by distribution in (1.1). Suppose that a population stems from a single organism. Thus, in the first generation, the population size is $X_1 = Y_1 \sim Y$ (we will write $X_1 \in [Y]$). Let X_n be the population size in the nth generation. Then,

$$X_0 = 1 \text{ almost surely (in short, a. s.)}, \qquad X_1 = Y_1 = Y_{X0}$$

and for the forthcoming generations,

$$X_{n+1} = \sum_{s=1}^{X_n} Y_s = Y_1 + \cdots . + Y_{X_n} \tag{1.3}$$

Where $\sum_{s=1}^{0} Y_s := 0$ and $Y_s \in [Y]$, i.e., Y_s are independent and identically distributed (in a word, iid) random variablers (briefly, r.v's) which belong to the same

equivalence class of r.v.'s induced by a generic Y with a probability distribution (See Figure 1.1.)

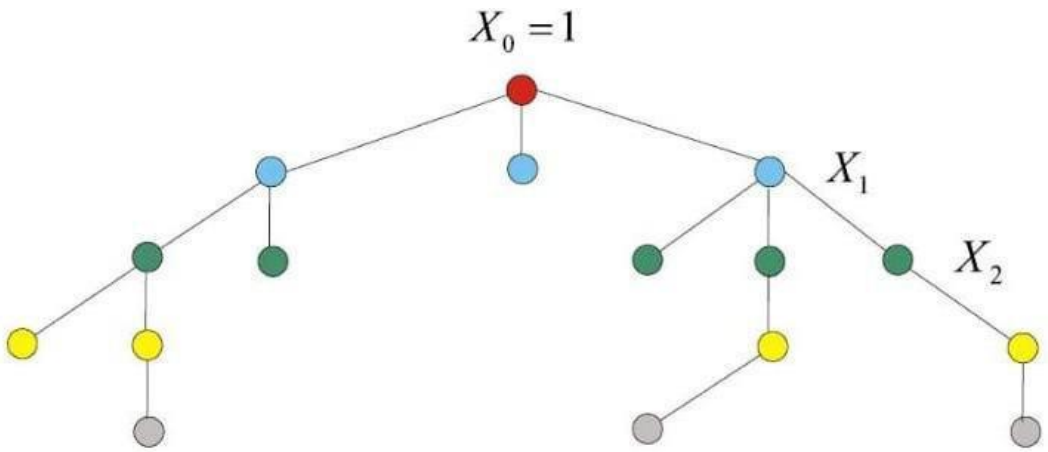


Figure 1-1 Bramching Process

Equation (1.3) can be understood as follows. If $X_n = j$ Organisms and $j > 0$, then each one of them individually will produce several offspring: the first of j organisms will have $Y_1^{(n)} \sim Y_1 \sim Y$, the second one will produce $Y_2^{(n)} \sim Y_2 \sim Y$, and finally the j th will produce $Y_j^{(n)} \sim Y_j \sim Y$ offspring. As we see it, Y_s should generally depend on n , but we dropped this dependence, thereby reducing equation (1.3) to its more straightforward form. Notice that we count only the number of offspring , if the patriarchs vanish. For example, if an organism divides in Y progeny, the following generation depends only on Y and not $Y + 1$ offspring.

Now, since the number of offspring at the $(n + 1)$st generation merely depends on the number of offspring at the nth generation, the process X_n is Markov with transition probabilities.

$$P\{X_{n+1=}j|X_{0,\ \ldots\ ,}X_{n-1,}X_n = i\} = P\{X_{n+1} = j|X_n = i\} = P\{Y_1 + \ldots + Y_i = j\} \qquad (1.4)$$

It also is time-homogeneous because, due to the above assumption (no dependence on n), the formation of progeny by an individual is stochastically equivalent to a Yand it is good at any generation.

1.2. The Expected Number of Progenies[2-4]

Now, we get one more result. Suppose we need to find $Ez^{Y_1+\cdots+Y_i}$. Since $Y_i \in [Y]$, $i = 1, 2, ..$ and they are independent

$$Ez^{Y_1+\cdots+Y_i} = [p(z)]^i \quad (1.5)$$

Let the probability generating function of the size of the population be given as follow

$$g_n(z) := Ez^{X_n} = \sum_{k=0}^{\infty} P\{X_n = k\}Z^k, n = 0,1,\ldots (1.6)$$

If we assume that the population stems from a single individual, i.e. $X_0 = 1$, a.s., then

$$g_{0(z)} = Ez^1 = z, \quad (1.7)$$

while

$$g_{1(z)=} Ez^{X_1} = Ez^{Y_1} = p(z). \quad (1.8)$$

Now we will find the pgf of X_{n+1} expressed in terms of the probability generating function (pgf) of X_n as

$$g_{n+1(z)=EZ^{X_{n+1}}}$$

Using the double-expectation formula and by equation (1.5), we have

$$g_{n+1}(z) = E\left[E[z^{X_{n+1}}|X_n]\right] = E\left[E[z^{Y_1+\cdots+Y_{X_n}}|X_n]\right] = E[[p(z)]^{X_n}] = E[[p(z)]^{X_n}] = gn(P(z))$$

In summary, we proved that

$$g_{n+1}(z) = g_n\big(p(z)\big). \qquad (1.9)$$

Now let us assume that the initial population size X_0 equals α almost surely (in short, a.s.) (an integer constant). Then

$$g_0(z) = Ez^a = z^a \qquad (1.10)$$

and

$$g_1(z) = Ez^{X_1} = Ez^{X_1+\cdots+Y_a} = (p(z))^a \qquad (1.11)$$

It is easy to verify that equation (1.9) still holds in this more general case as well. While equation (1.9) is somewhat implicit regarding the distribution of X_n (leaving us with recursive equations), we can still utilize it for the calculation of the mean and variance of X_n.

$$g'_{n+1}(z) = g'_{n+1}\big(p(z)\big)p'(z) \qquad (1.12)$$

and thus,

$$v_{n+1} = g'_{n+1}(1-) = \lim_{n\to 0} \acute{g}_{n+1}(x) = v_n v_1 = v_{n-1} v_1 v_1$$

By induction, we get that

$$v_{n+1} = v_1^{n+1} \qquad (1.13)$$

Where

$$v_n = EX_n \tag{1.14}$$

and

$$v_1 = EX_1 = g_1(1) = ap'(1) = aEY_1 = aEY \tag{1.15}$$

Therefore, if the expected number of progeny EY from one organism exists, then the expected number of progeny at the nth generation is

$$v_n = v_1^n = (aEY)^n \tag{1.16}$$

Where $a = X_0$ is the initial number of the organisms.

Chapter Two

The Estimation of Mutation Rates

Chapter Two

The Estimation of Mutation Rates

This chapter presents the mathematical model of the mutation rates by using the branching process, especially the Galton-Watson process. Moreover, the mutation rate is calculated as a ratio of the expected number of new mutants to the expected number of non-mutants in the nth generation. The important conclusion is that this mutation rate is fixed in every generation. Additionally, another way to find the latter ratio is discussed. Furthermore, the sample means estimator is studied as an unbiased and consistent estimator for mutation rate. This chapter is summary of [2,4,5]

2.1. Preliminaries [2,4,5]

Mathematical modelling with branching processes leads to one of the most efficient methods of statistical determination of mutation rates. We assume (which applies to our case) that a population of organisms such as bacteria multiply by division, forming a deterministic branching process so that each organism produces exactly two progeny in a unit of time. In typical situations, one may observe a cultre of Escherichia coli bacteria that divides approximately every 20-40 minutes depending on the growth media. If a culture originates from a single non-mutant

organism, i.e. $X_0 = 1$. Then in the nth generation, the total number of bacteria will be exactly $c_n = 2^n$.

Now, let us assume that, in every generation stemming from a patriarch bacterium, each progeny mutates with probability $\mu \in (0,1)$ and that this probability (commonly referred to as the mutation rate) is constant and dependent only on a particular organism and on a chemical agent if used in the underlying experiment. In other words, if in the nth generation, there are X_n non-mutants of the total of 2^n organisms, then in the next generation, each of the $2X_n$ recently divided bacteria can mutate with common probability μ and generate a new mutant.

Once a bacterium becomes a mutant, all of its offspring are also mutants that we call preexisting mutants. Now the quantity of a new mutant in the nth generation is a random number denoted by NM_n. Note that the unknown probability μ is being always called the mutation rate in the biological literature. Even though is also the expected value of an underlying Bernoulli random vanable with this probability, this still does not warrant the usage of rate which has been assigned in physics and mathematics to quantify related occurrences per unit time. Hence the term rate is improper except for its historical appeal. As we recall from chapter 1, in the general setting of population growth modelled by a branching process, it is assumed that an organism at the end of its lifetime number Y of offspring with a probability distribution that gives in equation (1.1) and with the associated pgf in equation (1.2)

In our case, each bacterium produces exactly two offspring. However, the number of non-mutants in each generation and it forms a random branching process, known as the Galton-Watson process. More specifically, each bacterium upon its replication produces a random number Y of non-mutants ranging from 0 to 2, obviously, binomially distributed with parameters 2 and $v = 1 - \mu$, where v is the probability that a newly replicated bacterium does not mutate. Hence, while the process of proliferation is ly deterministic, the quantity of mutants (including new mutants and preexisting mutants) and non-mutants are random, and each one of them forms a Galton-Watson process. See Figure 2.1 below.

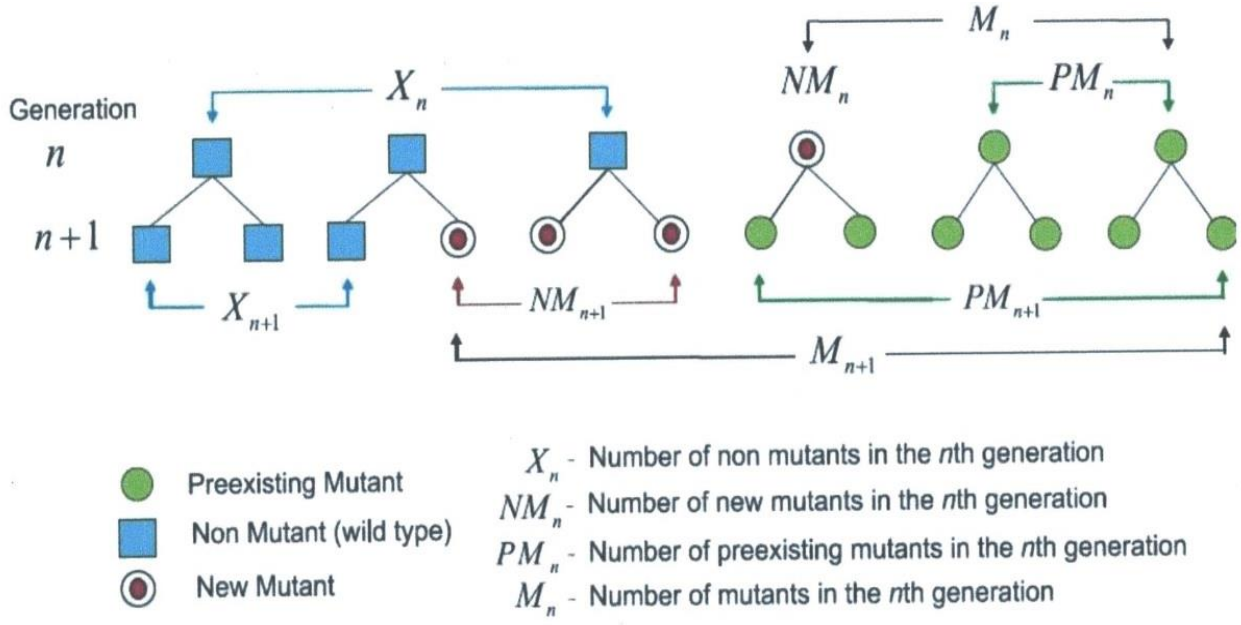


Figure 2-1 Mutation Model

2.2. Mutation Rates [4,5]

Let X_n be the number of non-mutants, M_n be the number of all mutants, and NM_n be the number of new mutants in the nth generation. Then X_n will double to $2X_n$ upon the $(n+1)$st division, of which nowNM_{n+1} are newly mutated.

The mutation frequency in any generation, say $n+1$, can be determined based on the quantity of new mutants (NM_{n+1}) and the number of newly divided cells $2X_n$. (The preexisting mutants PM_{n+1} are naturally excluded from counting). Thus, $NM_{n+1}/2X_n$ is the proportion of the new mutants from the total of $2X_n$ divided cells. We call this ratio (which depends on n) the random mutation frequency. The number $NM_{n+1}/2X_n$ is obviously not μ (that we are looking for), as it random, because X_n and NM_{n+1} vary, and because they change from experiment to experiment and from generation.

It stands for a reason to try the ratio of their (weighted) averages, EX_n and ENM_{n+1} (mathematical expectations) instead, in the hope to attain μ, which we initially conjectured as being a constant. This way, we eliminate the randomness of X_n and NM_{n+1}, which are functions along with their distributions. But what about n?

In following proposition below, the mutation rate μ for a fixed organism in every division is indeed in the form of the ratio of the two expectations

$$\mu = \frac{E[NM_{n+1}]}{2E[X_n]}$$

and that this ratio does not depend on n.

Proposition 2.1. [5]

Suppose a population of bacteria starts with a single non-mutant organism and then periodically by dividing in two. Assume that the forthcoming generations of bacteria mutate each with a constant rate μ independent of each other and constant in any generation. Then, for each $n = 1,2,...$ if it holds that

$$\mu = \frac{E[NM_{n+1}]}{2E[X_n]} \tag{2.1}$$

where NM_{n+1} is the number of new mutants in the $(n+1)$st generation

Proof.

Recall that by our assumption, a population begins with a single organism. Thus, in the first generation, the population size is $X_1 = Y_1 \in [Y]$ (where $[Y]$ The equivalence class of r.v.'s induced by Y and Y is the number of non-mutants. Produced by a single non-mutant). If X_n is the number of non-mutants in the nth generation, then,

$$X_{n+1} = \sum_{i=1}^{X_n} Y_{i,} \tag{2.2}$$

where $Y_i \in [Y] = [B(2, v)]$, i.e. it is a binomial r.v. with a maximum of two non-mutant bacteria produced by a single non-mutant upon replication.

If $g_{n(z)}$ is the pgf of X_n and$P(z)$ is the pgf of Y (as per (1.2), we have from equation (1.9) that

$$Ez^{X_{n+1}} = g_{n+1}(z) = g_n\big(P(z)\big) = g_n((vz + \mu)^2). \tag{2.3}$$

Indeed, since Y is binomial with parameters 2 and v and so are $Y_i's$, we have

$$P(z) = (vz + \mu)^2 \tag{2.6}$$

and $EY = 2v$ turning equation (2.4) to (2.3) and furthermore,

$$g'_n(1-) = EX_n = (2v)^n \tag{2.5}$$

Notice that $g'_n(1-)$ denotes the limit of $g'_n(z)$ when z converges to 1, along any part, from the unit circle in the complex plane. Alternatively, by equation (1.16), with $a = X_0 = 1$, we have

$$v_n = EX_n = (EY)^n = v_1^n = (2v)^n. \tag{2.6}$$

which is the expected number of progeny in the nth generation.

The number of all mutants M_n in the nth generation also forms another Galton-Watson process complementing $X_n$to $C_n = 2^n$.

Non-mutants can disappear in some generation (with a very small but positive probability), say in generation nth, with $X_n = 0$. Then, $X_{n+1} = 0$, and all forthcoming generations will contain zero non-mutant cells. Hence, the whole population will consist of entire mutants, and the forthcoming evolution of mutants will run a deterministic process. We will elaborate on this possibility later on.

Now, let NM_ndenote the total number of new mutants in the nth generation (see Figure 2.1 again), that is

$$NM_{n+1} = 2X_n - X_{n+1}. \tag{2.7}$$

We get now, since $EX_n = (2v)^n$, then

$$ENM_{n+1} = 2EX_n - EX_{n+1} = 2(1-v)(2v)^n.$$

Therefore, the ratio

$$\frac{E[NM_{n+1}]}{2E[X_n]} = 1 - v = \mu \qquad \blacksquare$$

Remark 2.2.

However, the same result can be obtained more often, which yields some valuable discussions. First, notice that unlike $\{X_n\}$ and $\{M_n\}$, the process $\{NM_n\}$ is not branching. Furthermore, It is not even Markov as it can be readily seen. We will use a different approach to calculate a recursive equation for its pgf's. We start with the introduction of

$$\xi_i^{[n]} = \begin{cases} 1, \text{If the } i\text{th If the ith non – mutant cell in the nth generation does not mutate} \\ 0, 0\ otherwise \end{cases} \tag{2.8}$$

So that $Y_i^{[n]} = \xi_{1i}^{[n]} + \xi_{2i}^{[n]}$ and $\xi_{ij}^{[n]} \sim \xi_i^{[n]}$. (Of course, we have dropped the superscript n in $Y_t's$.) Due to the above assumption, $\left\{\xi_i^{[n]}\right\}$ is a double sequence of iid Bernoulli r.v.'s. Each with parameter $v = 1 - \mu$. Then,

$$X_{n+1} \sim \sum_{i=1}^{2X_n} \xi_i^{[n]}, n = 0,1, \dots, X_0 = 1 \tag{2.9}$$

and

$$NM_{n+1} \sim 2X_n - \sum_{i=1}^{2X_n} \xi_i^{[n]} \tag{2.10}$$

Here, without loss of generality, we abbreviate $\xi_i = \xi_i^{[n]}$. Now using the double expectation equation and notation $h_n(z) = Ez^{NM_n}$ we have

$$h_{n+1}(z) = Ez^{NM_{+1n}} = E\left[E\left[z^{2X_n - \sum_{i=}^{2X_n} \xi i} | X_n\right]\right]$$

$$= E\left[z^{2X_n}(Ez^{-\xi i})^{2X_n}\right] = E\left[(zEz^{-\xi i})^{2X_n}\right]$$

$$= E\left[(z(v\frac{1}{z} + \mu))^{2X_n}\right] = g_n((v + \mu z)^2, n = 0,1, \ldots \tag{2.11}$$

Notice that unlike equation (2.8), inside g_n we have the pgf of a $B(2, \mu)$ r.v... Differentiating (2.11), we have

$$E[NM_{n+1}] = h'_{n+1}(-1) = g'_n((v + \mu)^2)2\mu$$

and using (2.6)

$$E[NM_{n+1}] = (2v)^n 2\mu = 2\mu EX_n, n = 0,1, \ldots$$

The latter yields the same equation for the mutation rate

$$\mu = E[NM_{n+1}] / 2EX_n, n = 0,1,..$$

2.3. Estimation of Mutation Rates [4,5]

Equation (2.1) is intuitively clear for the case if the processes $\{X_n\}$ were deterministic. If the number of non-mutants at the nth division is X_n, it would double

to $2X_n$ at the $(n+1)$st division, of which now NM_{n+1} get newly mutated. Then $NM_{n+1}/2X_n$is the proportion of the new mutants from the total of $2X_n$, divided cells. and from the deterministic point of view, it is not a surprise to see the mutation rate equal $NM_{n+1}/2X_n$, and being a constant. However, with $\{X_n\}$ and $\{NM_n\}$ being stochastic, the fact that the same ratio holds true for the expectations of the associated r.v.'s is less obvious. Furthermore, in equation (2.1), it follows that the ratio indeed does not depend on n.

The good news is that the ratio $E[NM_{n+1}]/2EX_n$, no longer depends on n and is valid for every experiment a given organism. However, we cannot find the named expectation experimentally. In other words, we cannot statistically "observe" the above ratio, thus making equation (2.1) impractical.

If we, however, replace $E[NM_{n+1}/2EX_n]$ with $NM_{n+1}/2EX_n$ (the random mutation frequency) we return to a similar random ratio, which can be regarded as an estimator of μ, in notation $\hat{\mu_n}$i.e.

$$\hat{\mu_n} = \frac{NM_{n+1}}{2X_n} \tag{2.12}$$

which can be converted to an estimate of μ

$$\hat{\mu_n} = \frac{nm_{n+1}}{2x_n} \tag{2.13}$$

Replacing statistic $\hat{\mu_n}$, with an observed value of this statistic. Of course, after we showed that $\mu = E\ [\ NM_{n+1}]/2EX_n$. the estimator $\hat{\mu_n}$ of μ gains in

credibility. But how credible is the estimator $\hat{\mu_n}$ in light of the traditional merits? One of the desired qualities of $\hat{\mu_n}$ is that it is "unbiased," i.e., if $\grave{E}\mu_n$ Even more significant was that $\hat{\mu_n}$ be "consistent," i.e., if for a large $n, \hat{\mu_n} \approx \mu$. In practice, most desirable is when $\hat{\mu_n} \approx \mu$ for not very large n. All the wishful properties of $\hat{\mu_n} \approx \mu$ are established, beginning with Theorem 2.3.

Theorem 2.3. [5]

The estimator $\hat{\mu_n}$ is unbiased and consistent on a non-extinction set.

Proof.

1) Unbiasedness

From (2.12), we have

$$NM_{n+1} \sim 2X_n - \sum_{i=1}^{2X_n} \xi_i$$

i.e., NM_{n+1} depends only on X_n. Thus given $X_n, NM_{n+1}$is a binomial r.v. with parameters $(2X_n, \mu)$. Consequently, the conditional expectation is

$$E[NM_{n+1}|X_n] = 2X_n\mu \tag{2.14}$$

Therefore,

$$E\left[\hat{\mu_n}\right] = E\left[E\left[\hat{\mu_n}|X_n\right]\right] = E\left[E\left[\frac{NM_{n+1}}{2X_n}\right]\right]$$

$$= \frac{1}{2}E\left[\frac{1}{X_n}E[NM_{n+1}|X_n]\right] = \frac{1}{2}E\left[\frac{1}{X_n}2X_n\mu\right] = \mu. \tag{2.15}$$

2) Consistency

Here is how we establish this very significant property of $\hat{\mu_n}$. Let $(\Omega, F(\Omega), (F_n; n = 0,1, \ldots), P)$ be a filtered probability space and let $\{W_n\}$be a discrete-time parameter process on this space adapted to the filtration $\{F_n\}$. This process is called a **martingale** if for each pair of nonnegative integers n and r :

$$E[W_{n+r}|F_n] = W_n \; a.s. \tag{2.16}$$

The following martingale convergence theorem was once formulated (for martingales) and established by Joseph Doob.

Theorem 2.4. [5]

Let $\{W_n\}$ be a martingale relative to a filtration $\{F_n\}$ such that $\{W_n \geq 0\}$ and $sup\{EW_n: n = 0,1, \ldots\} < \infty$. Then, there is an integrable r.v. W_∞, i.e. $W_\infty \in L^1$ i.e. $W_\infty \in L^1(\Omega, F(\Omega)\, P; R_+)$ such that $\lim_{n\to\infty} W_n = W_\infty$ a.s.

Now we return to the Galton-Watson process $\{X_n\}$ and let $\{F_n\}$be the natural filtration (i.e. the filtration generated by $\{X_n\}$). Is it a Markov chain? For any $n = 1,2, \ldots$, we have by the elementary Markov property.

$$E[X_{n+r}|F_n] = E[X_{n+r}|X_n] \text{ a.s.} \tag{2.17}$$

Then, from the last equation (2.17) and (2.2)

$$E[X_{n+1}|F_n] = \left[\sum_{i=1}^{X_n} Y_i|X_n\right] = \sum_{i=1}^{X_n} EY_i = 2vX_n \text{ a. s.} \tag{2.18}$$

Furthermore, for $r \geq 1$, we note that

$$E[X_{n+r}|F_n] = E[E[X_{n+r}|F_{n+r-1}]|F_n] = 2v[EX_{n+r-1}|F_n] \text{ a.s.}$$

Continuing this procedure backwards, by induction, we get

$$E[X_{n+r}|F_n] = (2v)^r X_n a.s. \tag{2.19}$$

Now, define $W_n = X_n/(2v)^n$. Obviously, $\{W_n\}$ is adapted to $\{F_n\}$ and from (2.19).

$$E[X_{n+r}|F_n] = \frac{1}{(2v)^{n+r}} E[X_{n+r}|F_n] = \frac{X_n}{(2v)^n} = W_n \; a.s. \tag{2.20}$$

Implying that $\{W_n\}$ is a nonnegative martingale. Furthermore,$EW_n = 1$. Therefore, the conditions of Doob's martingale convergence Theorem 2.3 are met, and hence there is a nonnegative r.v. W_∞ such that $\lim_{n\to\infty} W_n = W_\infty$ a.s. Because of $NM_{n+1} = 2X_n - X_{n+1}$, (2.20) implies that

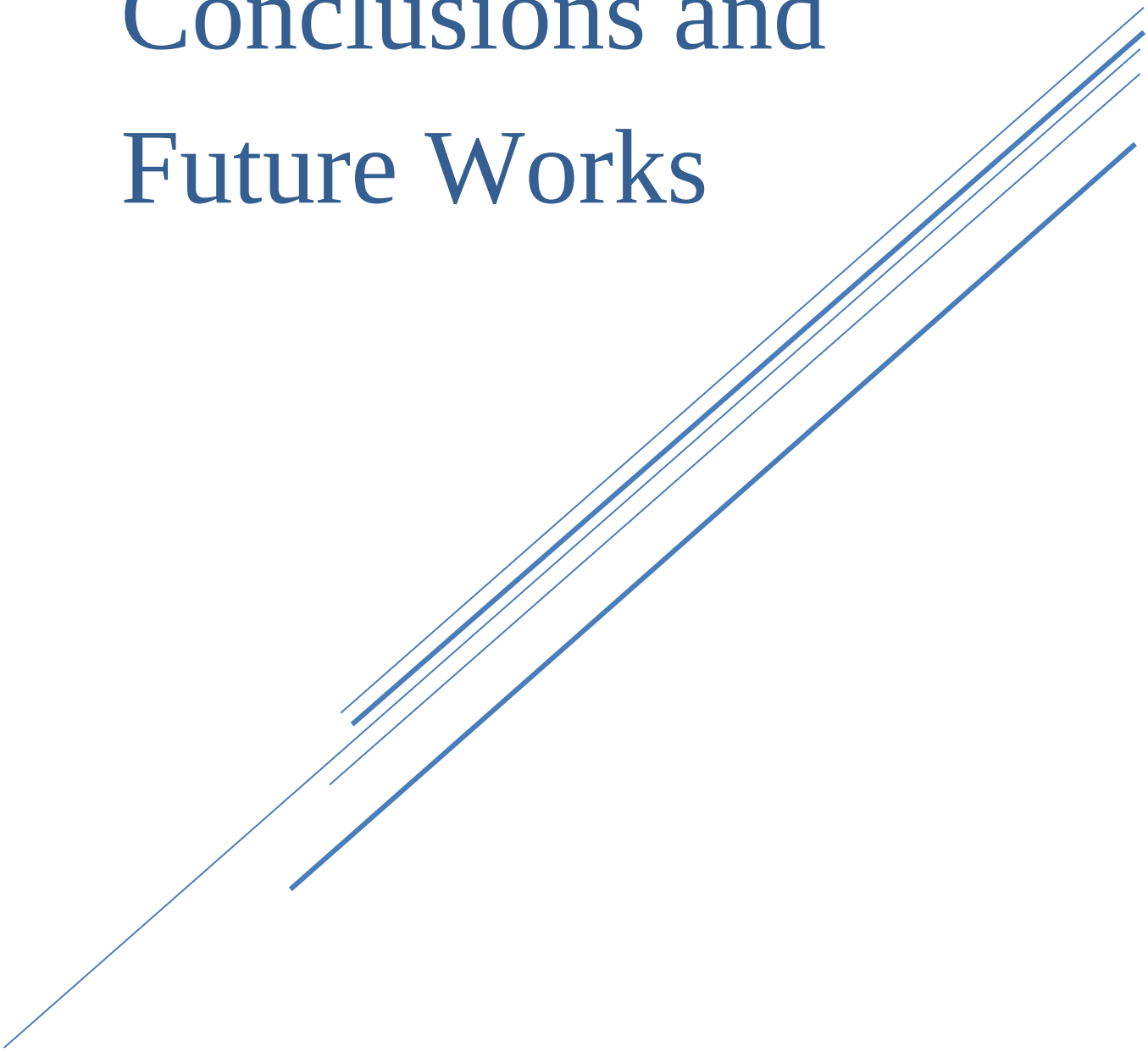
$$\hat{\mu_n} = \frac{NM_{n+1}}{2X_n} = \frac{2X_n - X_{n+1}}{2X_n} = 1 - \frac{X_{n+1}}{2X_n}$$

$$= 1 - \frac{X_{n+1}(2v)^{n+1}}{2X_n(2v)^{n+1}} = 1 - \frac{X_{n+1}}{(2v)^{n+1}} \cdot \frac{(2v)^n}{X_n} \; v \tag{2.21}$$

which converges to $1 - \frac{W_\infty}{W_\infty}(1-\mu) = \mu$ a. s. ■

Chapter Three

Conclusions and Future Works



Chapter Three

Conclusions and Future Works

3.1. Conclusions

This research included the study of mutation rate μ by adapting some essential assumptions about Escherichia coli. One crucial hypothesis is that each organism produces at most two progeny in a unit of time. Besides, this product is subject to a binomial distribution with parameters 2 and $v = 1 - \mu$. That is means each bacteria produces a random number of non-mutants varying from 0 to 2. Consequently, evaluating this ratio that argues under these conditions does not depend on the generation of creatures, and this ratio is static. Furthermore, the suggested estimator is a sample mean, and it is appropriate because it is an unbiased and consistent estimator

3.2. Future Works

The future work is that we desire to alter these conditions. For example, the production of organisms is subject to a binomial distribution with parameters n and v. That means each organism produces at most n progeny in a unit of time.

References

[1] Griffiths AJF, Miller JH, Suzuki DT, Lewontin RC, and Gelbart W (2000). An introduction to genetic analysis, 7th edn. W. H. Freeman and Company, New Work.

[2] Haccou, P., Haccou, P., Jagers, P., Vatutin, V. A., & Vatutin, V. (2005). Branching processes: variation, growth, and extinction of populations (No.5). Cambridge university press.

[3] Harris, T. E. (1963). The theory of branching processes (Vol. 6). Berlin: Springer.

[4]Kimmel, M., & Axelrod, D. E. (2015).Branching Processes in Biology. Springer, New York, NY.

[5] Luria, S. E., & Delbrück, M. (1943). Mutations of bacteria from virus sensitivity to virus resistance. Genetics, 28(6), 491.

[6] Niccum, B. A., Poteau, R., Hamman, G. E., Varada, J. C., Dshalalow, J. H., & Sinden, R. R. (2012). On an unbiased and consistent estimator for mutation rates. Journal of theoretical biology, 300, 360-367.

[7] Zheng, Q. (2017). Toward a unique definition of the mutation rate. Bulletin of mathematical biology, 79(4), 683-692.

الخلاصة

في هذا البحث, تم مناقشة افضل مقدر لنسبة الطفرات μ ضمن شروط معينة بحيث ان كل كائن حي مثل البكتريا ينقسم الى كائنين على الأكثر. اي ان الكائن اما يموت او لا ينقسم او ينقسم فيصبح لدينا كائنان. لذلك ان انقسام الكائنات هذا يخضع الى توزيع احتمالي وهذا التوزيع هو توزيع ذو الحدين بمعلمتين 2 و μ $v = 1 - \mu$. ناقشنا الصياغة الصريحة لنسبة الطفره حيث وجدناها ثابتة ولا تعتمد على الجيل الذي وجدت منه. كما درسنا مقترح حول افضل مقدر لهذا النسبة فكان متوسط العينة هو مقدر جيد من حيث عدم التحيز والاتساق.

جمهورية العراق

وزارة التعليم العالي والبحث العلمي

جامعة بابل

كلية التربية للعلوم الصرفة

قسم الرياضيات

حول مقدر معدل الطفرات

بحث

مقدمة الى مجلس كلية التربية للعلوم الصرفة في جامعة بابل

كجزء من متطلبات نيل درجة الدبلوم العالي في التربية / الرياضيات

مقدمة من قبل الطالبة

امل علوان صليبي كاظم

بإشراف الدكتور

علي حسين محمود العبيدي

1443 هـ **2021 م**