

Branching Processes and Their Applications

Group 9C - Lucy Ellis, Billie Harris, Oliver Johnson, Jennifer Laws,
Matthew Speers & Thomas Whitfield

Supervisor: Professor Dmitry Korshunov

February 14, 2020

Abstract

Studied within this report is the idea of family lineage and how surnames can become extinct over time. To begin, the report discusses the history of branching processes, relating to the likelihood of extinction of a surname within a region, in Section 1. Subsequently, the definition of branching processes and their properties are discussed in Sections 2.1 and 2.2 respectively, before exploring the Galton-Watson process. This is done through the discussion of the formation of generating functions in Section 3.1, the total number of individuals in Section 3.2, and the ultimate extinction in Section 3.3. Other applications of branching processes are then discussed including their applications to nuclear chain reactions in Section 4.1 and genes and mutations in Section 4.2.

1 Introduction

In 1873, Francis Galton proposed a problem regarding the likelihood of extinction of a surname within a region. The problem was originally proposed in a similar manner to the following:

A large population of men, where each man in the first generation has a different surname, has reproduction laws where $a_k\%$ of men of a generation have k male children. Find what proportion of these surnames are extinct after a set number of generations; and how many instances there are of the same surname being held by a set number of people in a generation.

Whilst the original problem specified a maximum of 5 male children, Reverend H. W. Watson and Galton extended it to an unknown large number, which is so large it is inconsequential to the distribution. Since Watson and Galton's initial discussion in their 1874 paper [1] this problem has been further discussed and expanded by reducing assumptions to produce a more accurate model whilst still following some of the initial ideas.

In this report we will first define branching processes, in particular the Galton-Watson process, in the context of the original problem and discuss its properties. We will then explore the Galton-Watson process as detailed in the original paper [1] and how it relates to the ultimate extinction theorem. Finally we will discuss further applications of branching processes outside the extinction of surnames. These applications include nuclear chain reactions and genetic mutations.

2 Galton-Watson Process

2.1 Definition

Branching processes are mathematical representations depicting how populations grow over multiple generations [2]. In each generation, n , the members give birth to a certain number of offspring (subject to laws of chance), making up the $(n + 1)^{\text{th}}$ generation.

In order for us to study these branching processes, a number of observations (outlined by Fewster in her lecture notes [3] from the University of Auckland) must be made.

1. Firstly, it is vital to assume that members reproduce individually of one another in order to uphold the probabilities of reproduction - and therefore extinction.

2. Another assumption is that the size of a generation is independent of the size of the previous generations.
3. In addition, we assume that the number of offspring of different individuals are independent, identically distributed random variables.

In practical terms this means that the size of a generation doesn't depend on the size of the generations that came before it. In addition to this, the number of offspring a particular member has, is not affected by the number of members currently present.

In reality however, these may not always be realistic assumptions. For example, an individual with fewer siblings may be more likely to have fewer children themselves as this is the type of family environment they are used to. Or in-fact, the number of cancerous cells in one's body may be restricted by the number of cancerous cells currently surrounding it. Nonetheless, they are invaluable when trying to model real-world applications.

One particular branching process is the Galton-Watson process, which can be expressed using a fundamental formula, given by Zitkovic [4] from the University of Texas at Austin.

Definition 2.1. Under the above assumptions, a discrete time process, Z_t , is called a (Galton-Watson) branching process if $Z_0 = 1$ and the population of the n^{th} generation, Z_n for $n \geq 1$, is given by the formula:

$$Z_n = \sum_{j=1}^{Z_{n-1}} Z_{(n-1)j},$$

where $Z_{(n-1)j}, n \geq 1, j \geq 1$ are independent copies of an integer random variable Z .

This naturally allows for the possibility of extinction (a concept we will explore further in later sections) because if $Z_n = 0$ for a particular generation, then $Z_q = 0$ for all $q > n$.

It is as a result of certain properties relating to the Galton-Watson process that the above definition can be determined.

2.2 Properties

Branching processes satisfy a number of properties particularly relating to Z_n , the size of the population at the n^{th} generation. These properties include the below and

are inspired by lecture notes by Zitkovic [4] from the University of Texas at Austin and Fewster [3] from the University of Auckland.

1. When a new generation is formed, the members of the previous generation die out leaving only the members of this new generation.
2. Relating to the population sizes, Z_n , of the generations $n = 1$ onwards:
 - The generation $n = 1$ will die out if $Z_1 = 0$, and generations $n \geq 2$ will not exist, resulting in $Z_n = 0$ for all $n \geq 2$.
 - If $Z_1 \neq 0$, then all members of Z_1 will produce a random number of members who, collectively, will make up the generation $n = 2$, Z_2 . For example, the first member of Z_1 produces Z_{11} members, the second member of Z_1 produces Z_{12} members, etc. In general, the j^{th} member of Z_1 produces Z_{1j} members, where the largest value j is Z_1 . Following from the assumption in Section 2.1 that the number of descendants of different members of a given generation are independent and identically distributed, we know that the number of members of the generation $n = 2$, Z_2 , is:

$$Z_2 = \sum_{j=1}^{Z_1} Z_{1j}. \tag{2.1}$$

All successive generations after Z_2 are constructed in a similar way. As previously stated in Section 2.1, if a given generation, $n - 1$, dies out, i.e. $Z_{n-1} = 0$, then $Z_q = 0$ for all $q > n - 1$. It is this property, therefore, that leads to the formula for Z_n seen above in the definition.

Having an understanding of branching processes and their properties allows us to further explore the Galton-Watson Process through generating functions. It specifically helps in determining the total number of individuals in a generation as well as the probability of ultimate extinction.

3 Exploring the Galton-Watson Process

When studying family lineage using the Galton-Watson Process there are a number of additional assumptions that we must make. The main one of these, stated by T. E. Harris [2], is that we are only interested in male descendants and assume that only males are needed for reproduction. It should be mentioned that the Bisexual Galton-Watson Process is deemed by some to be a more realistic model that only allows

couples to reproduce. However we shall not review this here and shall instead focus our efforts into exploring the original Galton-Watson Process detailed in 1873.

3.1 Forming Generating Functions

In order to study the extinction probability of a particular surname, we must look at the random sequence Z_0, Z_1, Z_2, \dots of generation sizes. This will require the use of generating functions. A definition of generating functions, inspired by Wilf's work in generatingfunctionology [5], is given below.

Definition 3.1. A generating function of an infinite sequence of real numbers, $G(s)$, is a formal power series represented by

$$G(s) = \sum_{i=0}^{\infty} a_i s^i, \quad |s| \leq 1.$$

The probability generating function of Z_n is denoted by $G_n(s) = \mathbb{E}(s^{Z_n})$. We are interested in extending this to the $(m+n)$ th generation, where m and n represent random generations with the m th generation preceding the n th generation. We can consider the generating function of the $(m+n)$ th generation as follows. The following theorem and its proof uses ideas from G. R. Grimmett and D. R. Stirzaker [6].

Theorem 3.1. *The generating function of a branching process satisfies the equality $G_{m+n}(s) = G_m(G_n(s))$, thus $G_n(s)$ is the n -fold iterate of G .*

Proof. For every member of the $(m+n)$ th generation, there exists a unique member of the m th generation. Therefore,

$$Z_{m+n} = X_1 + X_2 + \dots + X_{Z_m}$$

where X_j = the number of members from the $(m+n)$ th generation originating from the j th member of the m th generation. By the assumptions above, this sum of random variables is independent and identically distributed with an equal distribution to Z_n (the number of n th generation particles originating from the very first particle in Z_0). We can then obtain that $G_{m+n}(s) = G_m(G_{X_1}(s))$ where $G_{X_1}(s) = G_n(s)$. If we iterate this, we get

$$G_n(s) = G_1(G_{n-1}(s)) = G_1(G_1(G_{n-2}(s))) = G_1(G_1(\dots(G_1(s))\dots)). \quad (3.1)$$

Note that $G_1(s)$ is equivalent to $G(s)$ as before, therefore the theorem is proven. \square

This theorem demonstrates that theoretically, we could learn a lot about Z_n and its distribution using generating functions.

3.2 Total Number of Individuals

To eventually be able to evaluate the total number of individuals with ease, we must firstly introduce the moments of Z_n as follows. This allows us to calculate the expectation and variance of the number of people in a given generation.

Lemma 3.2. *Let $\mu = \mathbb{E}(Z_1)$ and $\sigma^2 = \text{Var}(Z_1)$. Then,*

$$\mathbb{E}(Z_n) = \mu^n \tag{3.2}$$

$$\text{Var}(Z_n) = \begin{cases} n\sigma^2 & \mu = 1; \\ \sigma^2(\mu^n - 1)\mu^{n-1}(\mu - 1)^{-1} & \mu \neq 1. \end{cases} \tag{3.3}$$

Proof. We will prove this lemma using ideas from E. J. McCoy [7]. For the expectation, recall that $\mu = \mathbb{E}(Z_1)$ by substituting $n = 1$ into (3.2). Then,

$$\mu = G'_n(1).$$

Recall that $G_n(s) = G_1(G_{n-1}(s))$ from (3.1). Differentiating the last equality gives:

$$\begin{aligned} G'_n(s) &= G'_{n-1}[G(s)]G'(s) \\ G'_n(1) &= G'_{n-1}[G(1)]G'(1) \\ &= G'_{n-1}(1)G'(1). \end{aligned}$$

Therefore $\mu = \mu_{n-1}\mu = \mu_{n-2}\mu^2 = \dots = \mu^n$, and (3.2) is proven.

For the variance, recall that $\sigma^2 = \text{Var}(Z_1)$ by substituting $n = 1$ into (3.3). Furthermore, let $\sigma_n^2 = \text{Var}(Z_n)$. Differentiating the generating function twice gives:

$$G''_n(s) = G''_{n-1}[G(s)]G'(s)^2 + G'_{n-1}[G(s)]G''(s). \tag{3.4}$$

Recall that $G(1) = 1$ and $G'(1) = \mu$. Then $G''(1) = \sigma^2 - \mu + \mu^2$.

In addition, we have $\sigma_n^2 = G''_n(1) + \mu_n - \mu_n^2$. Then,

$$G''_n(1) = \sigma_n^2 - \mu^n + \mu^{2n}$$

and

$$G''_{n-1}(1) = \sigma_{n-1}^2 - \mu^{n-1} + \mu^{2n-2}.$$

From (3.4),

$$\begin{aligned} G_n''(1) &= G_{n-1}''(1)G'(1)^2 + G_{n-1}'(1)G''(1) \\ \sigma_n^2 - \mu^n + \mu^{2n} &= (\sigma_{n-1}^2 - \mu^{n-1} + \mu^{2n-2})\mu^2 + \mu^{n-1}(\sigma^2 - \mu + \mu^2). \end{aligned}$$

This implies that $\sigma_n^2 = \mu^2\sigma_{n-1}^2 + \mu^{n-1}\sigma^2$, which leads to:

$$\sigma_n^2 = \mu^{n-1}\sigma^2(1 + \mu + \mu^2 + \dots + \mu^{n-1}).$$

Therefore (3.3) is proven. □

The above lemma details how to obtain the expected number of people in a given generation. Subsequently, we shall calculate the number of people to have ever carried a particular surname. Let T_n be the total number of individuals up to and including the n^{th} generation, that means, $T_n = Z_0 + Z_1 + Z_2 + \dots + Z_n$. Then, by the linearity of expectation and inspired by lecture notes [7] from Imperial College London,

$$\begin{aligned} \mathbb{E}(T_n) &= \mathbb{E}(Z_0) + \mathbb{E}(Z_1) + \mathbb{E}(Z_2) + \dots + \mathbb{E}(Z_n) \\ &= 1 + \mu + \mu^2 + \dots + \mu^n \\ &= \sum_{i=0}^n \mu^i \end{aligned}$$

by Lemma (3.2). Hence,

$$\mathbb{E}(T_n) = \begin{cases} n + 1 & \mu = 1; \\ \frac{\mu^{n+1} - 1}{\mu - 1} & \mu \neq 1. \end{cases} \quad (3.5)$$

3.3 Ultimate Extinction

In order to study the evolution of family surnames in a given area, we must first define the probability of ultimate extinction that governs this.

Definition 3.2. The probability of ultimate extinction is given by:

$$\mathbb{P}(\text{ultimate extinction}) = \mathbb{P}\left(\lim_{n \rightarrow \infty} Z_n = 0\right) = \lim_{n \rightarrow \infty} \mathbb{P}(Z_n = 0).$$

Also as $G_n(0) = \mathbb{P}(Z_n = 0)$, this implies that

$$\mathbb{P}(\text{ultimate extinction}) = \lim_{n \rightarrow \infty} G_n(0).$$

This allows us to conclude the following theorem, as stated by Feller [8].

Theorem 3.3 (Ultimate Extinction). *If $\mu < 1$, the probability of ultimate extinction is 1 as extinction is guaranteed. If $\mu = 1$, extinction is also guaranteed (unless $Z_n = 1$ at every generation, n). If $\mu > 1$, the probability of ultimate extinction is between 0 and 1.*

For the cases when $\mu < 1$ or $\mu > 1$, we can clearly show why this would be the case by using the Equation (3.5) and letting n tend to infinity:

$$\lim_{n \rightarrow \infty} \mathbb{E}(T_n) = \begin{cases} \frac{1}{1 - \mu} & \mu < 1; \\ \infty & \mu > 1. \end{cases}$$

When $\mu < 1$, the total number of family members converges as $n \rightarrow \infty$. Therefore, we are certain that in this case the family surname will eventually die out.

For the case where $\mu > 1$, there is a non zero probability that the family name will not become extinct. It is expected that in general every generation will contain a non-empty set of members leading to an infinite number of males carrying the surname, as $n \rightarrow \infty$.

For all values of μ , we can use the fact that $\mathbb{E}(Z_n) = \mu^n$ to show the size of the n^{th} generation as n tends to infinity,

$$\mu_n = \mu^n \rightarrow \begin{cases} 0 & \mu < 1; \\ 1 & \mu = 1; \\ \infty & \mu > 1. \end{cases}$$

At first glance, it would appear that the number of people with a specific surname would either eventually be 0, if $\mu < 1$, or increase unboundedly, if $\mu > 1$. However, when $\mu = 1$ it is more complicated. In general, when $\mu = 1$ the surname will die out because at some point $Z_n < 1$ ie. Z_n will be 0 for one of the generations. But clearly when $Z_n = 1$ for every generation this is not the case as there will always be 1 male to carry on the family name, and therefore the surname will not become extinct.

Having looked at the ideas behind the above theorem, we can now formally prove this using graphical methods with a proof inspired by Chapter 7 of Fewster's lecture notes [9].

Proof. In order to prove Theorem 3.3, we shall study graphs we have drawn depicting curves of $G(s)$ for the different values of μ .

In each instance, the graph of $G(s)$ satisfies the following underlying conditions:

1. $G(s)$ is increasing and strictly convex, as long as Z_n can be ≥ 2 .
2. $G(0) = \mathbb{P}(Z_n = 0) \geq 0$.
3. $G(1) = 1$.
4. $G'(1) = \mu$, so the slope of $G(s)$ at $s = 1$ gives the value μ .
5. The extinction probability, γ , is the smallest value ≥ 0 for which $G(s) = s$.

By evaluating each possible value of μ individually, we can determine the extinction probability in instance.

We first look at the case where $\mu < 1$. As shown in our graph, the curve, $G(s)$, is above the line $y = s$ for all values $s < 1$. Therefore, in the interval $s \in [0, 1)$, there is no other place where the curve could cut the line $y = s$ before hitting the y axis. This shows that here, the ultimate extinction probability, γ , is 1, as this is the smallest value of $s > 0$ where $G(s) = s$.

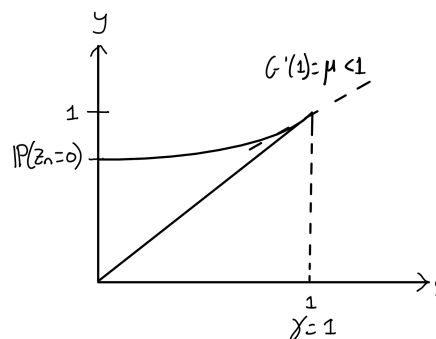


Figure 1: Graph of $y = G(s)$ when $\mu < 1$.

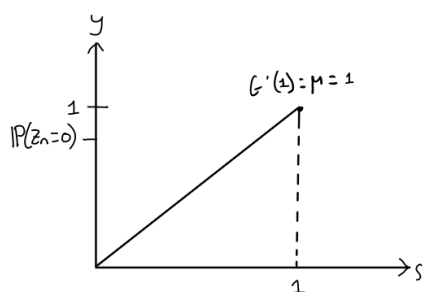


Figure 2: Graph of $y = G(s)$ when $\mu = 1$.

When $\mu = 1$, the situation is equivalent to the case when $\mu < 1$, as the curve $G(s)$ would again lie above the line $y = s$. Therefore, $\gamma = 1$ here too.

The exception to this is when $Z_n = 1$ for all values of $n \in \mathbb{N}_0$. Then $G(s) = s$ for all values of $s \in [0, 1]$ and so the smallest value of s where $G(s) = s$ is 0. This shows that the probability of ultimate extinction, γ , is 0.

Conversely, when $\mu > 1$, the gradient of $G(s)$ is steeper than the line $y = s$ at the point $s = 1$. Therefore, the curve sits below the line from this point, meaning that in order to intersect the y axis at $\mathbb{P}(Z_n = 0)$, the curve must once again cut the line at another point. As a result, there are two roots to the equation $G(s) = s$, namely 1 and s_1 where $0 < s_1 < 1$. Hence, here the extinction probability, γ , is a positive value less than 1.

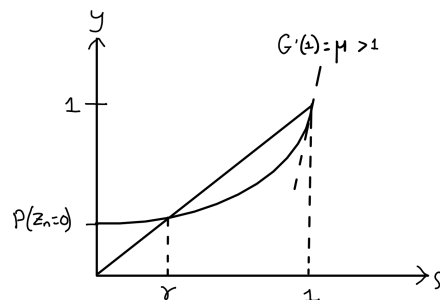


Figure 3: Graph of $y = G(s)$ when $\mu > 1$.

Conclusively, the graphs show that when μ decreases, $\mathbb{P}(Z_n = 0)$ increases. \square

This section has explored many theorems relating to branching processes in the context of passing on family surnames to successive generations. However, as stated earlier, this is not the only application of branching processes - an idea which will be studied in further detail in the following section.

4 Further Applications

There are numerous other applications of branching processes in addition to Galton's investigation into the extinction of family names. This section gives a brief overview of some of these.

The processes described in Sections 4.1 and 4.2 each operate under this simple model proposed by Feller [8]:

- (i) the first generation consists of 1 particle,
- (ii) the probability of a particle producing exactly k particles, is fixed and denoted p_k ($k = 0, 1, 2, \dots$),
- (iii) the $(n + 1)^{\text{th}}$ generation is formed from the offspring of the n^{th} ,
- (iv) the particles act independently of each other.

4.1 Nuclear Chain Reactions

The above model can be used to describe the behaviour of neutrons in nuclear fission reactions, such as that in an atomic bomb. In this case, the particles being considered are the neutrons. Their offspring are the neutrons released from the splitting of larger nuclei into smaller nuclei by collision with a neutron. This problem was initially discussed by Feller [8].

Let Z_{nj} denote the number of offspring for the j^{th} particle of the n^{th} generation. Assume that all collisions produce the same number of neutrons, say m . Also assume that each neutron collides with a larger nuclei with probability p . Then the probability $\mathbb{P}(Z_{nj} = m) = p_m$ that a particle (neutron) produces m offspring is equal to p . Also, the probability $\mathbb{P}(Z_{nj} = 0) = p_0$ that each particle produces zero offspring is given by $1 - p$. Clearly, the only possible numbers of offspring are 0 and m , giving a probability density function of

$$\mathbb{P}(Z_{nj} = i) = p_i = \begin{cases} p & i = m \\ 1 - p & i = 0 \\ 0 & \text{otherwise} \end{cases}$$

with $\mathbb{E}(Z_{nj}) = \mu = mp + 0 \cdot (1 - p) = mp$. Also, $\mathbb{E}(Z_n) = \mu^n = (mp)^n = m^n p^n$, where Z_n is the number of particles in the n^{th} generation.

Under this model, when $p > 1/m$ the probability of the reaction continuing indefinitely is non-zero (since $\mathbb{E}(Z_{nj}) > 1$). In reality, for large numbers of particles the probability of collision does not remain constant and the collisions are no longer independent of each other, so the model can no longer be applied. Instead, Feller attributes this mathematical idea of an indefinite reaction to the physical idea of an explosion.

4.2 Genes and Mutations

Also discussed by Feller [8] is the application of branching processes in modelling the survival of genes between generations of organisms. Here the Z_0 particle represents a new mutant gene in a single organism which has some chance p_i (for $i = 1, 2, \dots$) to appear in i direct descendants. These probabilities are fixed and apply to all organisms that carry the gene (assumption (ii)).

This idea is best illustrated by an example. Consider a species of animal. Assume that each member of the species always produces 200 direct descendants and that there is a $1/200$ chance of gene being represented in each descendant.

Clearly $Z_{nj} \sim \text{Bin}(200, 1/200)$, and so

$$\mathbb{P}(Z_{nj} = i) = p_i = \binom{200}{i} \left(\frac{1}{200}\right)^i \left(1 - \frac{1}{200}\right)^{200-i}$$

with $\mathbb{E}(Z_{nj}) = \mu = 200 \cdot 1/200 = 1$ and $\mathbb{E}(Z_n) = \mu^n = 1^n = 1$.

Alternatively, since $n = 200$ is sufficiently large, the distribution of p_i could be approximated by a Poisson(1) distribution. This is justified by Figures 4 and 5 (each 10000 trials). These were generated in R using the code included in the appendix.

So in this case the the probability that the gene will die out after a sufficiently large number of generations is 1, since $\mathbb{E}(Z_{nj}) = 1$. This is interpreted by Feller as the gene not carrying a biological advantage. For this to be the case, Z_n must be distributed by Poisson(λ) for $\lambda > 1$.

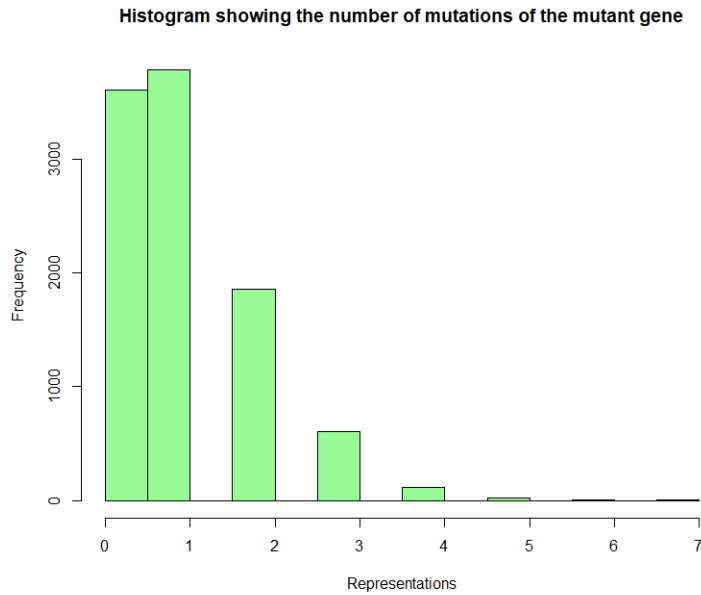


Figure 4: Bin(200,1/200)

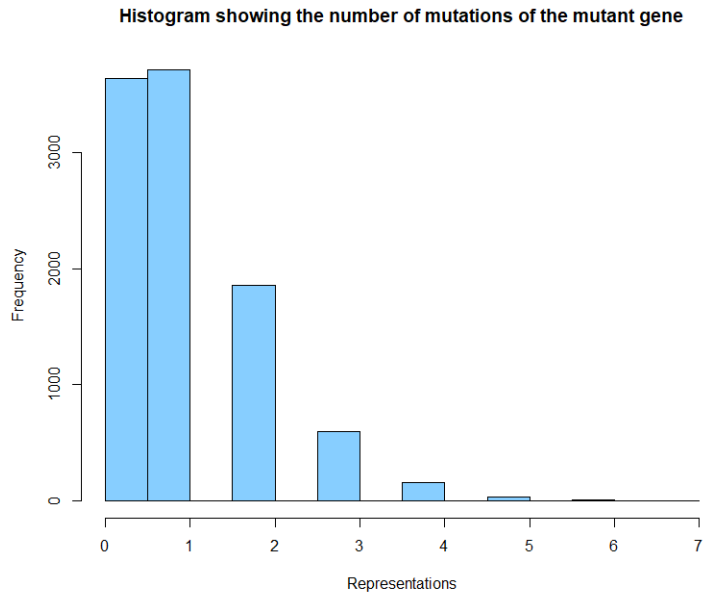


Figure 5: Poisson(1)

5 Conclusion

Throughout this report we have seen how branching processes and their given properties allow us to explore the growth of various populations over multiple generations. Through the use of generating functions, we have shown how to calculate the number of members in a particular generation, allowing for the exploration of when a particular population is expected to become extinct (if ever). In general, we conclude that when the expected size of each generation is less than or equal to 1, it is certain that this population will eventually die out. And therefore generations must be of notable size (have an expectation greater than 1) in order to have a chance of surviving.

When determining the extinction probability of a population, this report has mainly focused on the Galton-Watson process - a model used to determine the expected lifetime of a particular family name, carried by reproducing males. But Section 4 outlines a mere handful of other applications that branching processes can be used to model. Other notable applications not included in this report are the spread of bacteria and the expected time people spend waiting in lines.

However, many assumptions have been made in order for us to study these - with

a number of such assumptions being unrealistic compared to how these populations act in real-life situations. We therefore acknowledge that it would be an interesting exercise to repeat our investigation without these assumptions and compare these new results with those explored here.

References

- [1] H. W. Watson and F. Galton. On the Probability of the Extinction of Families. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 4, 1875. Pages: 138–144. 2
- [2] T. E. Harris. *The Theory of Branching Processes*, volume 119 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag Berlin Heidelberg, 1963. 2, 4
- [3] R. Fewster. STATS 325 Stochastic Processes, 2014. Chapter 6: Branching Processes: The Theory of Reproduction. 2, 4
- [4] G. Zitkovic. M362K Intro to Stochastic Processes, Fall 2014. Lecture 7: Branching Processes. 3, 4
- [5] H. Wilf. *generatingfunctionology*. Academic Press, second edition, 1994. 5
- [6] G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes*. Oxford University Press, 1992. 5
- [7] E. J. McCoy. M3S4/M4S4 Applied Probability, 2008. Chapter 4: Branching processes. 6, 7
- [8] W. Feller. *An Introduction into Probability Theory and its Applications, Vol 1*. John Wiley & Sons, 1950. 8, 10, 11
- [9] R. Fewster. STATS 325 Stochastic Processes, 2014. Chapter 7: Extinction in Branching Processes. 8

Appendix A R Code for Figure 4

```
1 ## binomial sample
2 bin_sample <- rbinom(size=200,n=10000,p=1/200)
3 hist(bin_sample,col="palegreen",xlab="representations",main="
  Histogram of number of mutations of mutant gene")
```

This is the R Code used to generate Figure 4. A random sample of size 10000 is taken from a binomial distribution with parameters $n = 200$ and $p = 1/200$. A histogram is then generated from this sample.

Appendix B R Code for Figure 5

```
1 ## poisson sample
2 poi_sample <- rpois(10000,1)
3 hist(poi_sample,col="skyblue1",xlab="representations",main="
  Histogram of number of mutations of mutant gene")
```

This is the R Code used to generate Figure 5. Here a random sample of 10000 is taken from a Poisson distribution with parameter $\lambda = 1$. This sample is then also used to generate a histogram.