11 April 2018

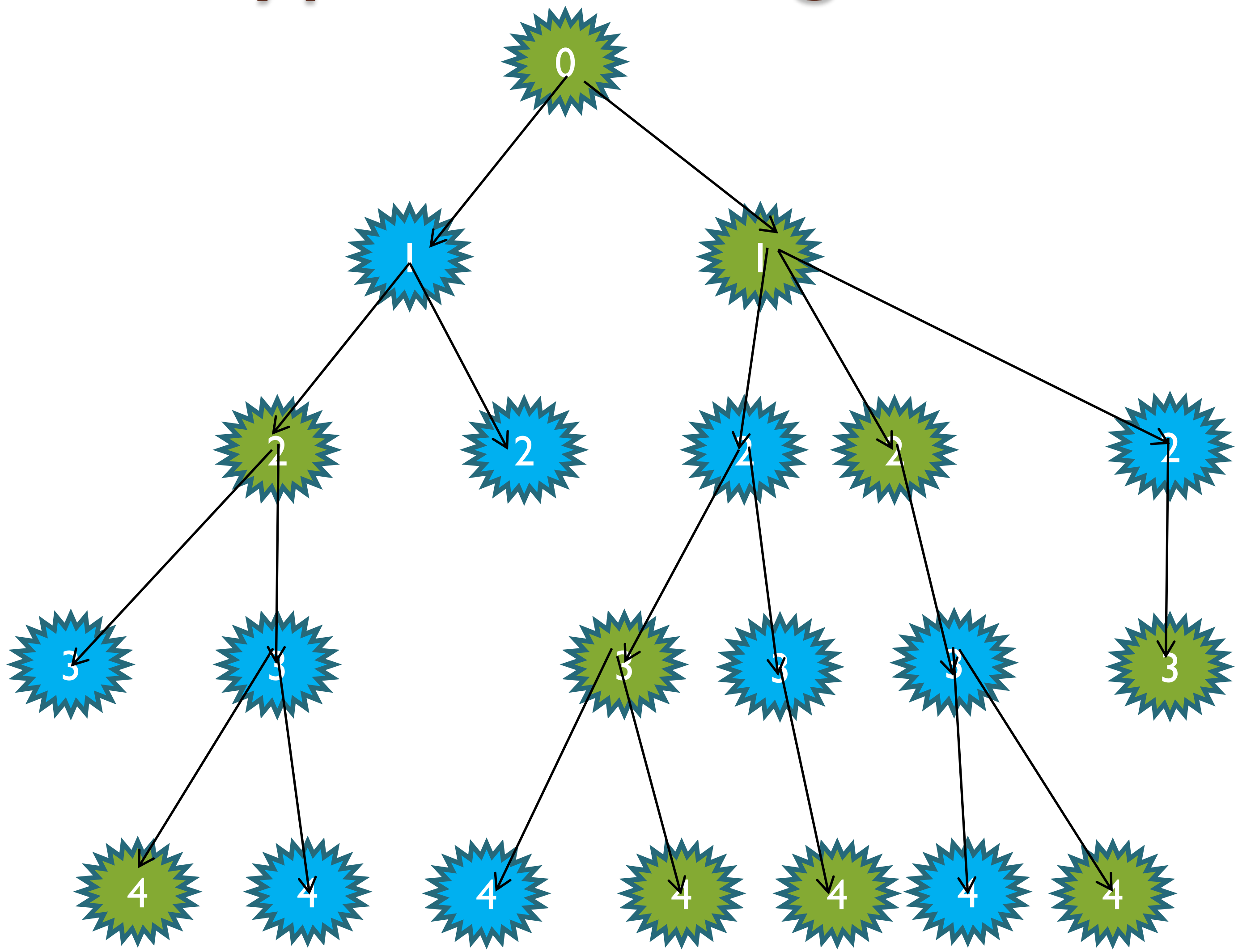
Criticality and Forefather distribution in a variant of Galton Watson Branching Process

Arnab Kumar Laha and **Sumit Kumar Yadav**

Structure of the Presentation

- Basic Ideas of Branching Processes
- Problem Description
- Theoretical Results
- Simulation Results
- Scope of Further Work
- Questions and Suggestions

Multi-type Branching Process



Basic Ideas of Simple Branching Processes

- Particles or objects can generate objects similar to themselves
- A particle of n^{th} generation produces offsprings which belong to $n+1^{th}$ generation
- Essentially a Markov chain defined on non-negative integers
- A probability law is assigned for reproduction of offsprings
- All individuals follow this law independently of each other
- Conditional on the current generation, the generating function of the next generation can be determined

Problem Description

- Model the process in a way that mimics the real human population
- People can have discrete ages – 0, a,2a,3a,…ka
- We begin with one individual of age 0 in the 0^{th} generation
- Probability that an individual of age "i"a will survive in next generation is given by $p_{i,i+1}$
- After age "k"a, individuals and their possible offsprings are not taken into consideration for the model
- An individual of age "i"a gives birth to finite number of individuals of age 0 with mean λ_i
- Thus, if we model the process using multitype branching process, we get the following mean matrix

Mean Matrix for the Process

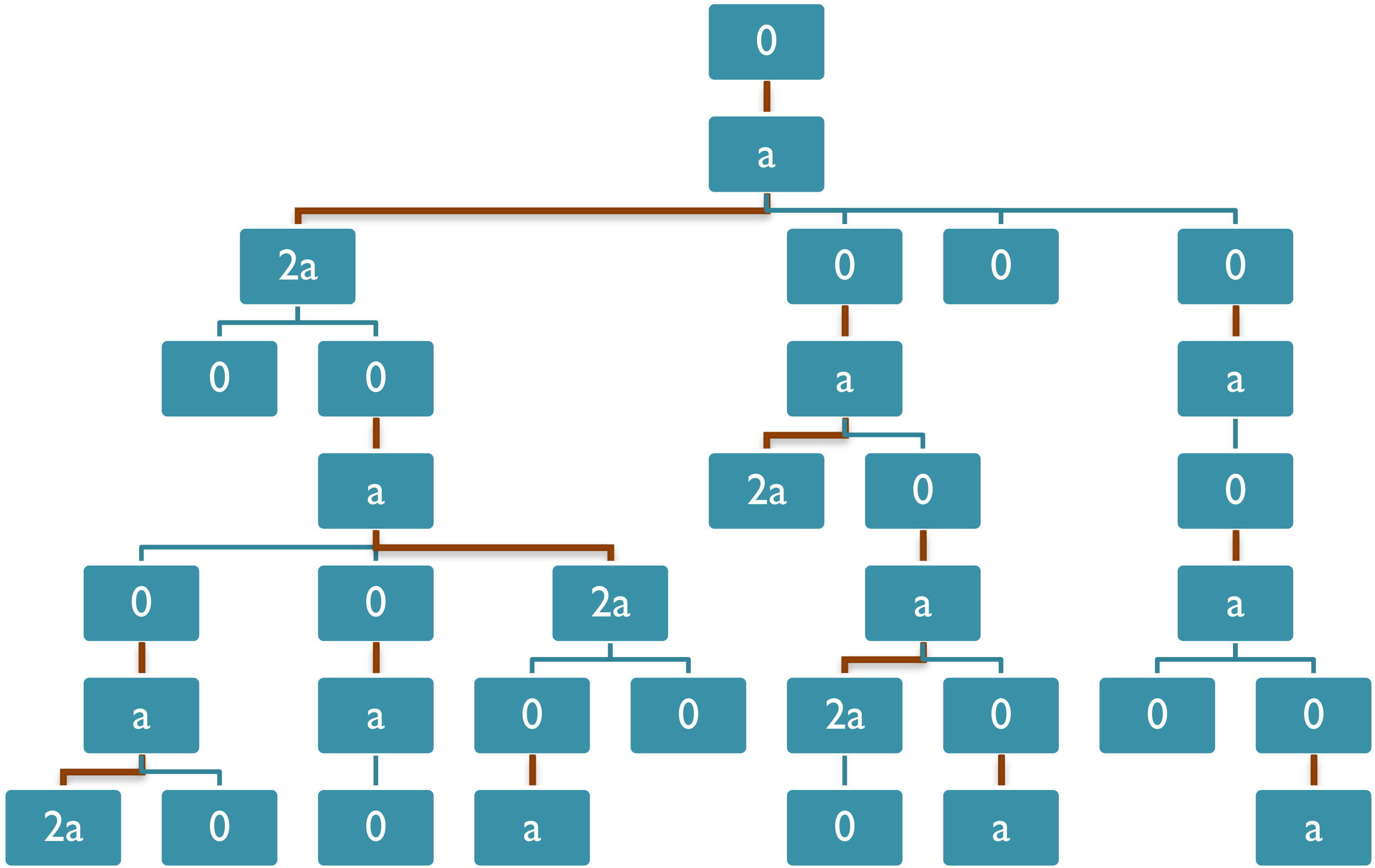
$$\begin{bmatrix} 0 & p_{01} & 0 & \cdots & \cdots & 0 & 0 \\ \lambda_1 & 0 & p_{12} & \cdots & \cdots & 0 & 0 \\ \lambda_2 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \lambda_{k-1} & 0 & 0 & \cdots & \cdots & 0 & p_{k-1,k} \\ \lambda_k & 0 & 0 & \cdots & \cdots & 0 & 0 \end{bmatrix}$$

Long run behaviour of the process

- Usually, it is helpful to classify the process as sub-critical, critical or supercritical (Probability of extinction < 1)
- In multitype branching process, if eigenvalue of mean matrix >1, it implies process is super critical
- We provide a proof of the intuitive condition that if one individual produces more than one individual on an average during it's entire life, it implies the process is super critical

- 1.) $\lambda_1 * p_{01} + \lambda_2 * p_{01} * p_{12} + \ldots + \lambda_k * p_{01} * p_{12} * \cdots * p_{k-1,k} > 1$ is equivalent to eigenvalue of mean martix greater than 1, hence supercritical
- 2.) $\lambda_1 * p_{01} + \lambda_2 * p_{01} * p_{12} + \ldots + \lambda_k * p_{01} * p_{12} * \cdots * p_{k-1,k} < (=) 1$ is equivalent to eigenvalue of mean martix less than (equal to) , hence subcritical (critical)

Forefathers



Question

- In a generation, how many individuals have a given number of forefathers??

Random variable $X_{i,j}^{n}$ = no. of individuals of age i in the n^{th} generation having j forefathers

- properties of the random variables $X_{i,j}^{n}$
- E($X_{i,j}^{n}$), Var($X_{i,j}^{n}$), P($X_{i,j}^{n}$=r) and so on
- Limiting behavior of these variables

Recurrence Relations for Expected Value

- Let us denote the expected number of individuals in the n^{th} generation of age "i"a having j forefathers be denoted by $\boldsymbol{E^n_{i,j}}$ = $E(\boldsymbol{X^n_{i,j}})$
- $E^{n+1}_{0,j+1} = \boldsymbol{\lambda_1} * E^n_{1,j} + \boldsymbol{\lambda_2} * E^n_{2,j} + \ldots\ldots + \boldsymbol{\lambda_k} * E^n_{k,j}$
- $E^{n+1}_{1,j} = p_{01} * E^n_{0,j}$
- $E^{n+1}_{2,j} = p_{12} * E^n_{1,j}$
- $E^{n+1}_{3,j} = p_{23} * E^n_{2,j}$

  ⋮

  ⋮
- $E^{n+1}_{k,j} = p_{k-1,k} * E^n_{k-1,j}$
- **The equations can be combined to obtain the following expression –**
- $E^{n+1}_{0,j+1} = \boldsymbol{\lambda_1} * p_{01} * E^{n-1}_{0,j} + \boldsymbol{\lambda_2} * p_{01} * p_{12} E^{n-2}_{0,j} + \ldots\ldots + \boldsymbol{\lambda_k} * p_{01} * p_{12} * \cdots . * p_{k-1,k} * E^{n-k}_{0,j}$

Expected number of forefathers

- We derive an exact expression from this recurrence relation for $\boldsymbol{E^{n}_{0,j}}$ which equals

- $\sum_{i_1,i_2,\ldots i_{k-2}\in C} \frac{j!}{(3j-n+\sum_{l=1}^{k-2} l.i_l)!(\prod_{l=1}^{k-2} i_l!)} {\lambda'}_1^{(3j-n+\sum_{l=1}^{k-2} l.i_l)} {\lambda'}_2^{(n-2j-\sum_{l=1}^{k-2}(l+1).i_l)} {\lambda'}_3^{i_1} {\lambda'}_4^{i_2} \ldots {\lambda'}_k^{i_{k-2}}$

- C = $\{ i_1, i_2, \ldots i_{k-2} : i_1 \geq 0, i_2 \geq 0, \ldots i_{k-2} \geq 0, 3j - n + \sum_{l=1}^{k-2} l.\, i_l \geq 0,\ n - 2j - \sum_{l=1}^{k-2} (l+1).\, i_l \geq 0 \}$

- ${\lambda'}_l = \lambda_l * p_{01} * p_{12} * \cdots * p_{l-1,l}$

A smaller version of the problem

- Discretizing the population to start with a simple model (k=2, a =20)
- We assume that one generation is equal to 20 years
- There are three kinds of people in the population

  Age 0, Age 20 and Age 40
- We assume that Age 0 don't give birth, they turn to Age 20 after one generation with probability p
- Age 20 gives birth to Age 0 with mean λ, they turn Age 40 after one generation with probability q
- Age 40 gives birth to Age 0 with mean μ, they are then taken out of consideration for the model

Problem Description

- Thus, the problem can be formulated as a 3 type branching process with types being

  a.) Age 0; b.) Age 20; c.) Age 40

- Mean Matrix ($M = (m_{ij})$)

$$\begin{bmatrix} 0 & p & 0 \\ \lambda & 0 & q \\ \mu & 0 & 0 \end{bmatrix}$$

Expectation Expression

- We just take type "Age 0" for illustration
- All other types can be easily derived from this
- Recurrence Relation :
- $E_{0,j+1}^{n+1}$ =λ∗p ∗ $E_{0,j}^{n-1}$ + μ*p*q* $E_{0,j}^{n-2}$

- **<u>Note :</u>** To start the process, we assume that the $(0)^{th}$ generation contains one individual of Age 0

Expectation Expression

- $E^{n}_{0,r} = {}^{r}C_{n-2r+1}\ (\lambda p)^{3r-n-1}(\mu pq)^{n-2r+1}$

- Note: we have not assumed any distribution, the only assumption made is that mean of the offspring distribution exists

- r varies from [[n/3] +1, $\lceil n/2 \rceil$]

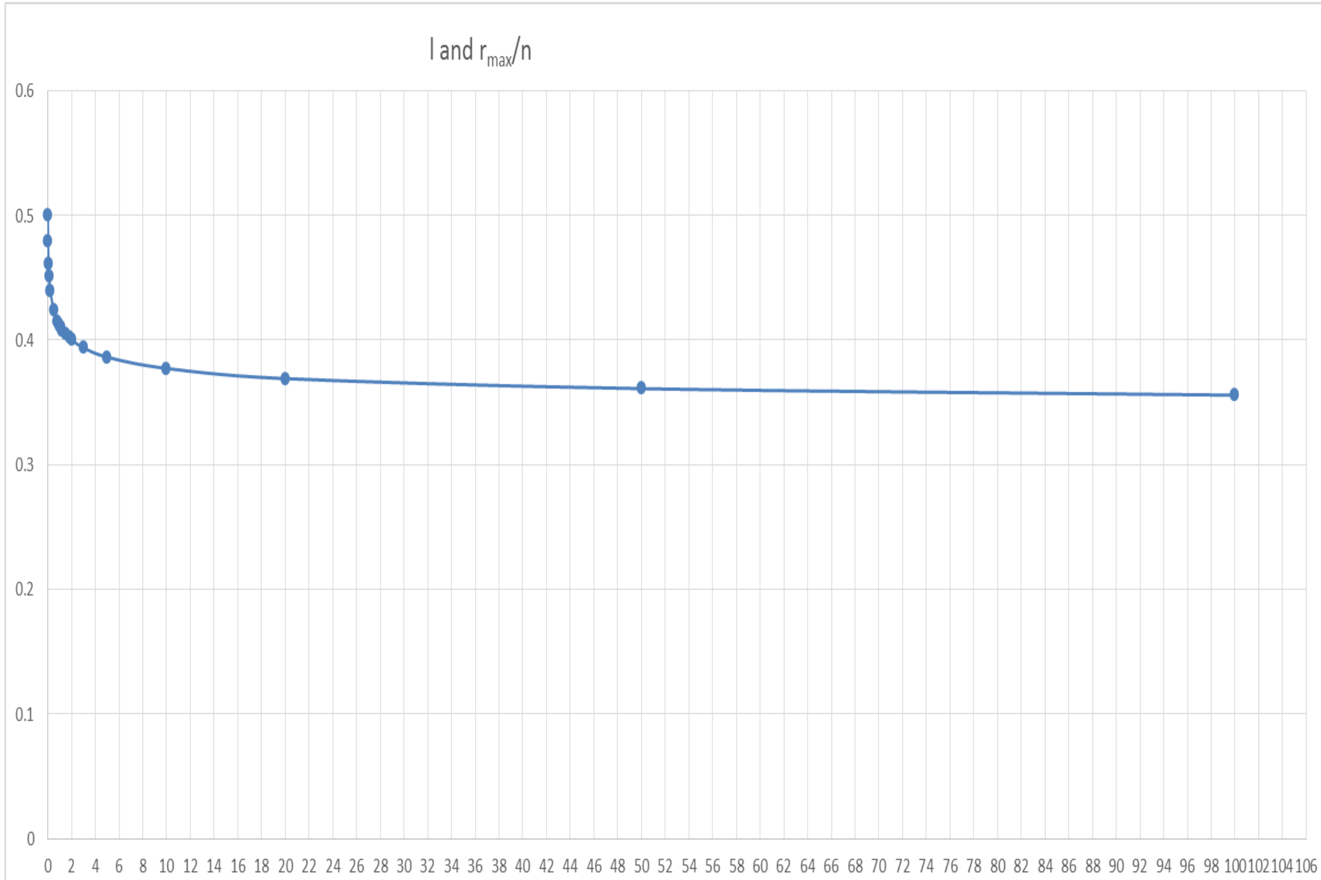
- Another quantity of interest could be to find where does the value of $E^{n}_{0,r}$ attain it's maximum

Maximum number of forefathers

- $\frac{E_{0,r+1}^{n}}{E_{0,r}^{n}}$ tells us that for large n, the expression is unimodal with respect to r
- Let r_{max} be the index for which the expected value of number of forefathers is maximum
- We are interested in r_{max} / n.
- Define a quantity I = $\frac{\mu^2 q^2}{\lambda^3 p}$

Plot of l vs r_{max}/n ; l $\{ l = \frac{\mu^2 q^2}{\lambda^3 p} \}$

0	0.5
0.01	0.479
0.05	0.461
0.1	0.451
0.2	0.439
0.5	0.4235
0.8	0.415
0.9	0.413
1	0.4115
1.1	0.4099
1.2	0.408
1.5	0.4047
1.8	0.4017
2	0.4
3	0.39368
5	0.38624
10	0.37718
20	0.369
50	0.361
100	0.3557
∞	0.33333333

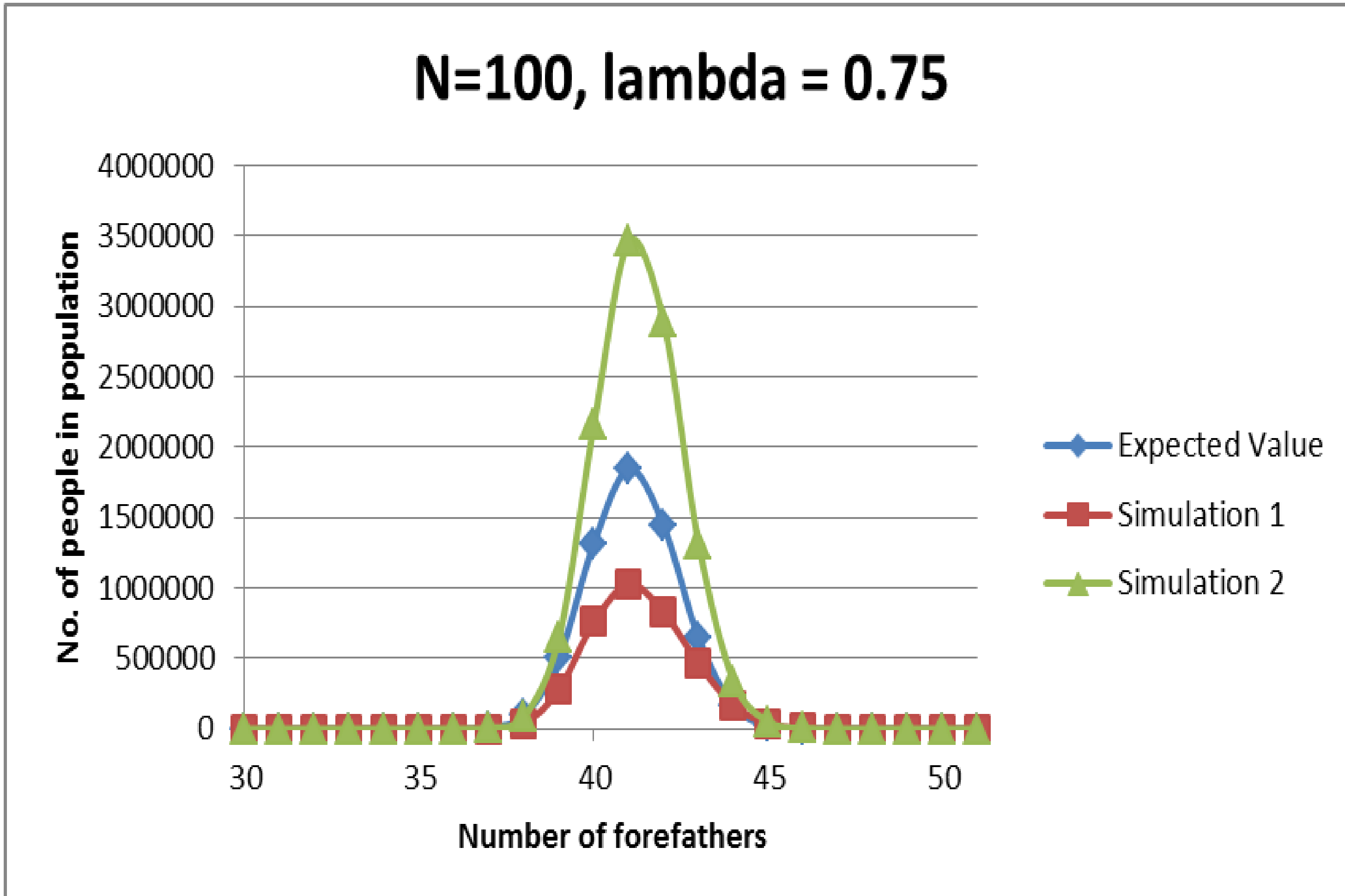


Simulation Results

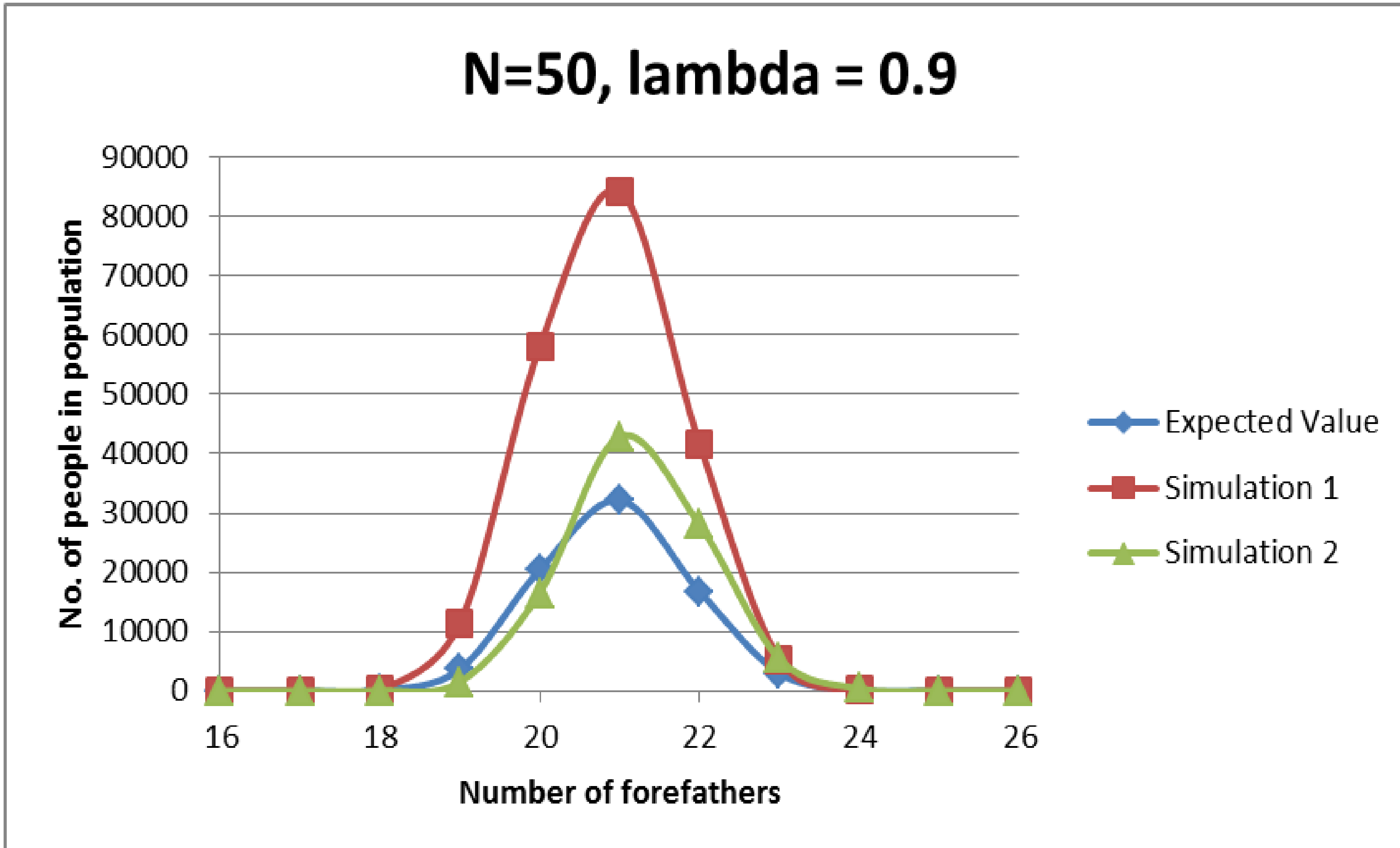
- The process was simulated using R Software
- The number of generations till which simulation was carried out and value of λ was varied
- For the results shown, we keep

$$\lambda = \mu, p = q = 1$$

Simulation Results



Simulation Results



Future Work

- Variance to be calculated by assuming some distribution (Poisson, Binomial, etc.)
- Inference problem can be studied
- The population can be discretized even further, so as to give better approximation to human population
- The model can be used as approximation to continuous time branching processes with a different probability of giving birth and dying at different times
- Modelling some marketing phenomenon using this model and checking how good the model fits

Thank you :)