EM 502: Optimization

Project : Branching process

E/14/339

Branching process is a stochastic model used to describe the a population growth. In general it is assumed that the individuals reproduce and die independently according to some predefined probability distribution. This model is used to find quantities such as the number of off-springs produced, the total number of individuals in a particular generation, and the total number of individuals up to a certain generation. Furthermore, it is specifically used to answer an interesting question regarding the life time of a branching process and the probability that it may die off. The model was introduced by F. Galton, in late 1870s to study the disappearance of family names.

Branching processes can be divided in many ways

- Discrete time / Continuous time
- Single type / Multi-type
- Identical reproduction / Varying reproduction
- Effects of environment

For simplicity let us consider the simplest form of branching process : Galton–Watson process. This is a discrete time, single type process, with identically distributed reproduction probability and infinite state space.

The Galton-Watson branching process

The Galton-Watson(GW) branching process can be identified using the following characteristics.

- **Discrete time** process, where each time steps n represents a generation.
- Each individual has a life time of **one unit** at the end of which it may or may not reproduce.
- Each individual produces a random number of offspring in the next generation, **independently** of other individuals.

- The probability mass function for the number of offspring is known as the **offspring distribution**. The offspring distribution is denoted by the random variable ξ and is given by,

$$p_k = P[\xi_i = k] , \quad k \geq 0 \tag{1}$$

  where k is a non-negative integer.

In general, it is assumed that $p_0 < 1$ and $p_1 < 1$ to eliminate the trivial case. Figure 1 depicts a GW branching process with one individual at generation 0. As mentioned before, there are certain questions of interest in respect to branching process. Let us consider them next.

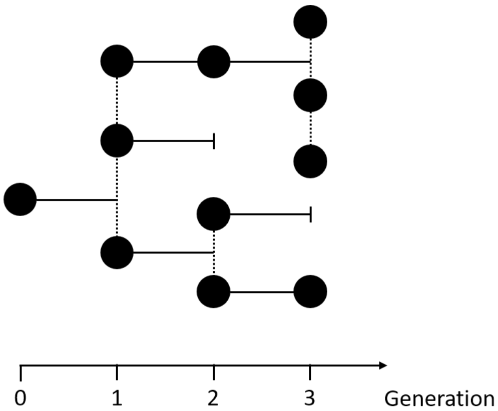


Figure 1: Galton-Watson(GW) branching process.

1. Population size

The random variable X_n, denotes the number of individuals at generation n. This is called the **population size** at generation n and it given by,

$$X_n = \sum_{i=0}^{X_n - 1} \xi_i , \quad n \geq 1 \tag{2}$$

where ξ_i is the number of offsprings of individual i from the previous generation. Note that $\xi_1, \xi_2, \dots, \xi_{X_n - 1}$ area the i.i.d copies of ξ from Eq. (1).

Now let us consider a branching process in which the tree starts with one individual. i.e. $X_0 = 1$. Consider the following probability.

$$P[X_{n+1} = k] = \sum P(X_{n+1} = k | X_n = i, X_{n-1} = j, \dots, X_0 = 1)$$

However, it was initially assumed that each individual reproduces independently and the reproduction is not affected by any external factors (including prior generations). Thus intuitively, it is evident that the population size of the next generation would only depend on the population size of the current generation (and not any prior generations). i.e.:

$$P[X_{n+1} = k] = \sum_{i=1}^{\infty} P(X_{n+1} = k | X_n = i)$$

Therefore, the stochastic process defined satisfies the Markov property. Furthermore, note that since the reproduction rate is independent of the generation, the chain is also time-homogeneous. Thus, the process X_n is a **discrete time Markov chain** on a **state space** of $\mathcal{S} = \{0, 1, 2, \ldots\}$ with **countably infinite non-negative integers**. For example, Figure 1 shows a branching process with $X_0 = 1$, $X_1 = 3$, $X_2 = 3$, and $X_3 = 4$. Let us look at the properties of these states.

- If $p_0 = 0$, then the branching process will continuously expand and never die.
- If $p_0 \geq 0$, then there is always a non-zero probability that the process may die at any generation. i.e.

$$P(X_{n+1} = 0 \,|\, X_n = k) \geq 0, \quad \forall\, n, k$$

- If X_n reaches zero, then it stays there. i.e

$$P(X_{n+1} = k \,|\, X_n = 0) = 0, \quad \forall\, n, k$$

- Therefore, from these properties, we can deduce that 0 is an **absorbing state**, and all other states are **transient states**.

Let us define the transition probability between two states as follows.

$$p_{ij} = P(X_{n+1} = i \,|\, X_n = j), \quad i, j \in \mathcal{S}$$

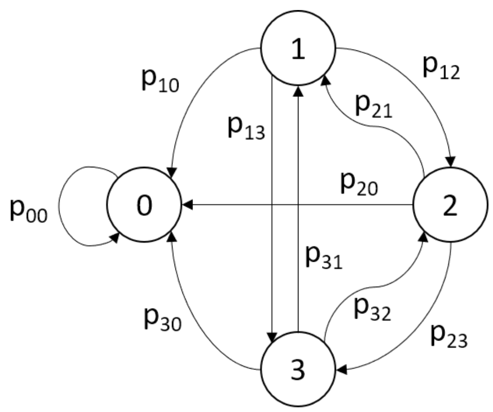


Figure 2: State transition.

Figure 2 shows a graphical representation of the states, and the transition probabilities for branching process with a subset of the state space containing $\{0, 1, 2, 3\}$. Note that as shown in the diagram, 0 is both **recurrent** and **absorbing**, while all other states are **transient**. Thus, branching process in **reducible**.

2. Mean and Variance

Let μ and σ^2 be the mean and the variance of the offspring distribution: ξ. Since $X_0 = 1$, the distribution of X_1 and ξ correspond to each other, and the mean and variance can be given as,

$$E[X_1] = E[\xi] = \mu$$
$$Var(X_1) = Var(\xi) = \sigma$$

where

$$\mu = \sum_{k=0}^{\infty} k\, P(\xi = k) = \sum_{k=0}^{\infty} k\, p_k \tag{3}$$

Now, from Eq. (2) we observe that the population size X_{n+1} can be written as a summation of X_n number of i.i.d offspring distributions : ξ. Thus, from theorem 1, the mean and variance of the population size can be given as,

$$E[X_n + 1] = E[X_n]\, \mu$$
$$Var(X_n + 1) = E[X_n]\, \sigma^2 + Var(X_n)\, \mu^2$$

However, $E[X_0] = 1$ and $Var(X_0) = 0$. Therefore, by considering these initial conditions, the mean and variance can be derived recursively as follows (Refer theorem 2 for proof).

$$E[X_n] = \mu^n \tag{4}$$

$$Var(X_n) = \begin{cases} \frac{\sigma^2\, \mu^{n-1}(1-\mu^n)}{1-\mu} & \text{if } \mu \neq 1 \\ n\sigma^2 & \text{if } \mu = 1 \end{cases} \tag{5}$$

3. Probability generating function of branching process.

Consider the transition probability for this Markov chain (branching process).

$$p_{ij} = P(X_{n+1} = i | X_n = k) = P(\sum_{j=1}^{k} \xi_j = i)$$

This has a rather complicated form. Therefore,let us consider the probability generating function (p.g.f) of the branching process to analyze the properties of the transition probability.

The **p.g.f** can be given as,

$$\Phi_Z(s) = E[s^Z] = \sum_{z=0}^{\infty} P(Z = z)\, s^z \tag{6}$$

Let Φ be the p.g.f of the offspring distribution : ξ.

$$\Phi(s) = E\left[s^{\xi}\right] = \sum_{k=0}^{\infty} P(\xi = k)\, s^k = \sum_{k=0}^{\infty} p_k\, s^k \tag{7}$$

Note that, if $X_0 = 1$, then the p.g.f of X_1 corresponds to ξ. Thus, by comparing terms, the p.g.f of X_1 and X_n can be given as,

$$\Phi_{X_1}(s) = \Phi(s) = \sum_{k=0}^{\infty} P(X_1 = k)\, s^k$$
$$\Phi_{X_n}(s) = \sum_{k=0}^{\infty} P(X_n = k)\, s^k \tag{8}$$

We will assume that $0 \leq s \leq 1$, and observe that for such s this power series converges. Next let us derive a recursive equation for the p.g.f in Eq. (8) as,

$$\begin{aligned}
\Phi_{X_n}(s) &= E[s^{X_n}] \\
&= \sum_{k=0}^{\infty} E[s^{X_n} \mid X_{n-1} = k]\, P(X_{n-1} = k) \\
&= \sum_{k=0}^{\infty} E[s^{\xi_1+\xi_2+\ldots+\xi_k}]\, P(X_{n-1} = k) \\
&= \sum_{k=0}^{\infty} E(s^{\xi_1})E(s^{\xi_2})\ldots E(s^{\xi_k})\, P(X_{n-1} = k) && (\xi_i : \text{i.i.d}) \\
&= \sum_{k=0}^{\infty} \Phi_{\xi}(s)^k\, P(X_{n-1} = k) && (\text{from Eq. (7)}) \\
&= \Phi_{X_n-1}(\Phi(s)) && (\text{from Eq. (8)})
\end{aligned}$$

By analyzing this further, we can see that this is a recursive function which can be written as,

$$\begin{aligned}
\Phi_{X_2}(s) &= \Phi_{X_1}(\Phi(s)) = \Phi(\Phi(s)) \\
\Phi_{X_3}(s) &= \Phi(\Phi(\Phi(s))) \\
&\vdots \\
\Phi_{X_n}(s) &= \underbrace{\Phi(\Phi(\ldots\Phi}_{n \text{ times}}(s)))
\end{aligned}$$

which can also be written as,

$$\Phi_{X_n}(s) = \Phi(\Phi_{X_{n-1}}(s)) \tag{9}$$

Now, let us investigate the properties of the p.g.f. Φ in Eq. (7).

$$\Phi(0) = p_0 \geq 0$$
$$\Phi(1) = \sum_{k=0}^{\infty} p_k = 1$$

Furthermore,

$$\Phi'(s) = \sum_{k=0}^{\infty} k\, p_k\, s^{k-1} \geq 0 \tag{10}$$

$$\Phi'(1) = \sum_{k=0}^{\infty} k\, p_k = \mu \tag{11}$$

and,

$$\Phi''(s) = \sum_{k=0}^{\infty} k(k-1)\, p_k s^{k-2} \geq 0 \tag{12}$$

Therefore, since the second derivative in Eq. (12) is non-negative, $\Phi(s)$ is convex.

4. Extinction probability

The next important analysis in a branching process is to find whether the population would become extinct (the process would die out) at a particular generation. i.e: $X_n = 0$. Let δ_n be the probability that the population is extinct by generation n where,

$$\delta_n = P(X_n = 0)$$

Then the probability that the branching process dies out is given by,

$$\gamma = \lim_{n\to\infty} P(X_n = 0) = \lim_{n\to\infty} \delta_n$$

4..1 Mean and variance analysis

By analyzing the mean and variance, an important observation from Eq. (4) is that if $\mu < 1$, then,

$$\lim_{n\to\infty} E[X_n] = \lim_{n\to\infty} \mu^n = 0$$

thus,

$$\gamma = \lim_{n\to\infty} \delta_n = 1$$

Intuitively this can be viewed that on average if an individual produces less than one offspring, then the population will definitely be extinct.

On the other hand if $\mu > 1$, the infinite limit of the expected value goes to infinity. Therefore, extinction is not guaranteed since the offspring distribution may vary depending on the variance σ^2.

4..2 p.g.f analysis

Section 3. lists certain important properties of the p.g.f of the offspring distribution. In addition to these, another crucial observation of the p.g.f of the population size X_n is,

$$\delta_n = \Phi_{X_n}(0)$$

Thus, the extinction probability at generation n can be calculated iteratively, by starting at $n = 0$ and computing the n^{th} iteration of Φ. Furthermore, it is also evident that δ is a non-decreasing sequence (since $X_n = 0$ implies that $X_{n-1} = 0$).

Next, substituting in Eq. (9) with $s = 0$,

$$\Phi_{X_n}(0) = \Phi(\Phi_{X_{n-1}}(0))$$
$$\delta_n = \Phi(\delta_{n-1})$$

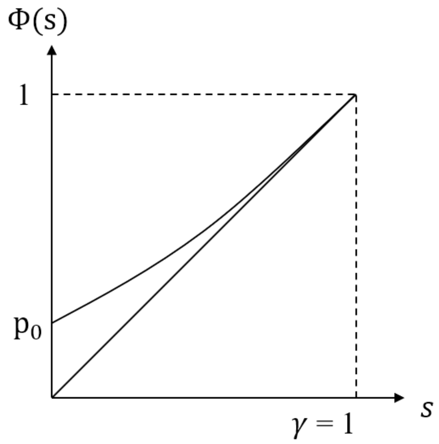
and then considering the infinite convergence,

$$\lim_{n\to 0} \delta_n = \lim_{n\to 0} \Phi(\delta_{n-1})$$
$$\gamma = \Phi(\gamma)$$

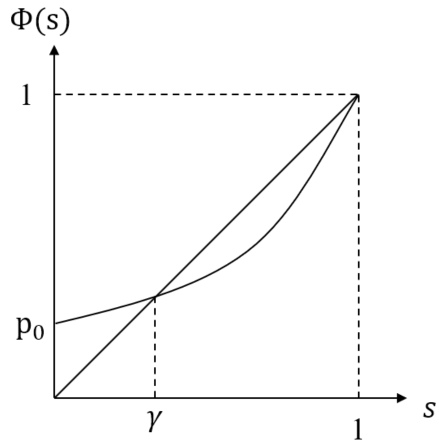
Thus, we arrive at the following lemma.

The extinction probability γ is the minimal non-negative solution of the fixed point equation $s = \Phi(s)$

In practice $s = \Phi(s)$ cannot be solved explicitly. The extinction probability γ is obtained by solving $\delta_n = \Phi_{X_n}(0)$ iteratively based on Eq. (9). However, for this analysis let us consider the following 2 cases.



(a) $\mu \leq 1$



(b) $\mu > 1$

Figure 3: Extinction probability.

- $\mu \leq 1$: From Eq. (11), in this case, $\Phi(1) \leq 1$. This occurs only if Φ is above the diagonal as shown in Figure 3a. i.e : $\Phi(s) \geq s$, $\forall s \in [0, 1]$. In this case δ_n converges to 1, so $\gamma = 1$.
- $\mu > 1$: From Eq. (11), $\Phi(1) > 1$. In this case, Φ is not always above the diagonal as shown in Figure 3b. There exists exactly one point $s < 1$ which solves $s = \Phi(s)$. As δ_n converges to this solution, we observe that $\gamma < 1$ and is given by the intersection point : $s = \Phi(s)$.

Therefore, we arrive at the following conclusion.

1. If $\mu \leq 1$, extinction is definite.
2. If $\mu > 1$, extinction in uncertain. If extinction doesn't occur, the process grows infinitely.

5. Stationary distribution

From the properties of the transition probability, we can construct the following transition matrix.

$$P = \begin{bmatrix} 1 & 0 & 0 & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ p_{20} & p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Note that the first row corresponds to $p_{0i} = P(X_{n+1} = i|X_n = 0)$ which has zero for $i > 0$ since 0 is an absorbing state. Furthermore, we initially considered $0 < p_0 < 1$, to make the problem non-trivial. Thus, the elements p_{i0} in the matrix P must be non-zero.

By definition. the stationary distribution (denoted by π) of the markov chain (branching process) is a probability distribution that remains unchanged. i.e.

$$\pi = \pi P$$

where π is a row vector whose entries are the probabilities summing to 1. Let $\pi = [\pi_0, \pi_1, \pi_2, \ldots]$ such that $\sum_{i=0}^{\infty} \pi_i = 1$. Since this is a matrix multiplication operation, let us consider the first element π_0.

$$\pi_0 = \sum_{i=0}^{\infty} \pi_i p_{i0}$$

$$\pi_0 = \pi_0 + \sum_{i=1}^{\infty} \pi_i p_{i0}$$

$$0 = \sum_{i=1}^{\infty} \pi_i p_{i0}$$

Note that $0 < p_{i0} < 1$ (non-trivial) and $0 \leq \pi_i \leq 1$ (π is a probability vector). Therefore, the only solution for this problem is $\pi_i = 0, \forall\, i > 0$ and thus, $\pi_0 = 1$. i.e.

$$\pi = [1, 0, 0, \ldots]$$

However, this is the trivial case where the process is always dead and this contradicts with a crucial assumption of the GW branching process ($X_0 = 1$). Thus, the only solution is trivial.

There exists no (non-trivial) stationary distribution for a GW branching process.

6. Total number of individuals

Another important property of branching process is the total number of individuals up to a particular generation. The total number of individual up to generation n is denoted by the random variable T_n where,

$$T_n = \sum_{i=0}^{n} X_i$$

The mean of this random variable can be derived as follows.

$$\begin{aligned} E[T_n] &= E[X_0 + X_1 + X_2 + \ldots + X_n] \\ &= E[X_0] + E[X_1] + E[X_2] + \ldots + E[X_n] \\ &= 1 + \mu + \mu^2 + \ldots + \mu^n \\ &= \begin{cases} \frac{\mu^{n+1}-1}{\mu-1} & \text{if } \mu \neq 1 \\ n+1 & \text{if } \mu = 1 \end{cases} \end{aligned}$$

The infinite convergence can be given as,

$$\lim_{n\to\infty} E[T_n] = \begin{cases} \infty & \text{if } \mu \geq 1 \\ \frac{1}{1-\mu} & \text{if } \mu < 1 \end{cases}$$

7. Applications

Branching processes have many application in a wide variety of field including computer science, economics, biology, humanitarian studies, etc. In this report 2 of the major applications of branching process within the field of computer science is discussed briefly.

7..1 Tree algorithms

It is self explanatory that the branching process itself takes the form of a tree. As such, it is used in many algorithms based on different tree structures such as binary tree, tree searching, etc. A few such use cases are listed below.

1. Binary search tree: A binary tree is a tree structure in which each node has at most two children. A binary search tree (BST) is a data structure which has the form of a binary tree in which the elements are arranged in a certain order which makes searching efficient. The efficiency of searching would depend on the depth of the tree. Since the depth of the tree varies depending on the elements inserted and the order of insertion, the branching model is used to estimate the bounds on parameters such as the depth of the tree, the time for insertion, and the time for search.

2. Heuristic search: While the operations related to the binary trees are bounded by its depth, trees with infinite depths may not have such bounds. Algorithms such as depth first search (DFS) and breadth first search (BFS) which are used for search on trees (or connected graphs with recursive paths) use the branching process to estimate the bounds of the search efficiency. Furthermore, the bounds algorithms such as backtracking which make heuristic look-ahead decisions can also be estimated through the branching model.

In addition the branching process can be used to model algorithms such as quad tree search, branching random walk, etc. and find their corresponding search bounds [2].

7..2 Could computing

Cloud computing is a model which allows multiple remote users to access a pool of shared resources such as storage, network, application, devices, etc. There are 5 main characteristics of cloud model: on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. In order to ensure that these requirements are satisfied, the underlying network requires a robust and reliable model. One of the important components for this is the traffic prediction model which focuses on identifying the dynamic demand for specific resources which vary depend on time.

To design effective and efficient network solutions for cloud environment and to understand the bottlenecks and solve performance problems arising in communication networks providers require accurate models to describe network traffic. Unfortunately, it is extremely difficult to predict the exact performance characteristics and demands on the network at any particular time. The main problem is that it is hard to forecast the number of queries that are going to appear. However, this can be modeled using branching model [3]. The proposed model assumes that resources queries come in independently. However, it is fair to assume that the primary request would be accompanied by multiple following requests which request access to certain resources which would be related to the primary resource. As such it is possible to create a branching model in which the number of queries branch with time and thus, the network traffic can be estimated. This stochastic model helps the service providers to estimate the dynamic demand for cloud services and thus avoid any bottlenecks and improve the service quality.

References

[1] P. Haccou, P. Jagers, V.A. Vatutin (eds.). "Branching Processes : Variation, Growth and Extinction of Populations."Cambridge University Press, Cambridge, 2005.

[2] Devroye L. (1998) Branching Processes and Their Applications in the Analysis of Tree Structures and Tree Algorithms. In: Habib M., McDiarmid C., Ramirez-Alfonsin J., Reed B. (eds) Probabilistic Methods for Algorithmic Discrete Mathematics. Algorithms and Combinatorics, vol 16. Springer, Berlin, Heidelberg

[3] R. Victor, A. Varfolomeeva, and A. Koryakovskiy. "Branching processes theory application for cloud computing demand modeling based on traffic prediction." International Conference on Advanced Information Systems Engineering. Springer, Berlin, Heidelberg, 2012.

Appendices

Theorem 1

Assume that $X, X_1, X_2, \ldots$ is an i.i.d sequence of random variables with a finite mean $E[X] = \mu$ and $Var(X) = \sigma^2$. Let N be a non-negative integer random variable, independent of all X_i , and let

$$S = \sum_{i=1}^{N} X_i$$

Then

$$E[S] = \mu E[N]$$
$$Var(S) = \sigma^2 E[N] + \mu^2 Var(N)$$

Proof

Let $S_n = X_1 + X_2 + \ldots + X_n$. Then,

$$E[S \mid N = n] = E[S_n] = n\,\mu$$
$$Var(S \mid N = n) = Var(S_n) = n\,\sigma^2$$

Therefore the mean of S is given by,

$$\begin{aligned} E[S] &= \sum_{n=0}^{N} E[S \mid N = n] P(N = n) \\ &= \sum_{n=0}^{N} n\mu\, P(N = n) \\ &= \mu \sum_{n=0}^{N} nP(N = n) \\ &= \mu\, E[N] \end{aligned}$$

The variance can be derived as,

$$Var(S) = E[S^2] - E[S]^2$$

where,

$$
\begin{aligned}
E[S^2] &= \sum_{n=0}^{N} E[S^2 \mid N = n] P(N = n) \\
&= \sum_{n=0}^{N} E(S_n^2) P(N = n) \\
&= \sum_{n=0}^{N} \left(Var(S_n^2) + E[S_n]^2\right) P(N = n) \\
&= \sum_{n=0}^{N} (n\,\sigma^2 + n^2\,\mu^2) P(N = n) \\
&= \sigma^2 E[N] + \mu^2 E[N^2]
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
Var(S) &= E[S^2] - (E[S])^2 \\
&= \sigma^2 E[N] + \mu^2 E[N^2] + \mu^2 E[N]^2 \\
&= \sigma^2 E[N] + \mu^2 Var(N)
\end{aligned}
$$

Theorem 2

The mean $E[X_n]$ and variance $Var(X_n)$ of population size can be given as,

$$
\begin{aligned}
E[X_n] &= \mu^n \\
Var(X_n) &= \begin{cases} \frac{\sigma^2\,\mu^{n-1}(1-\mu^n)}{1-\mu} & \text{if } \mu \neq 1 \\ n\sigma^2 & \text{if } \mu = 1 \end{cases}
\end{aligned}
$$

Proof

The mean and variance can be given by the recursive equation,

$$
\begin{aligned}
E[X_n + 1] &= E[X_n]\,\mu \\
Var(X_n + 1) &= E[X_n]\sigma^2 + Var(X_n)\,\mu^2
\end{aligned}
$$

In addition, for GW branching process, $E[X_0] = 0$ and $Var(X_0) = 0$. Thus the mean can be immediately derived as,

$$
\begin{aligned}
E[X_1] &= E[X_0]\mu = \mu \\
E[X_2] &= E[X_1]\mu = \mu^2 \\
&\vdots \\
E[X_n] &= E[X_{n-1}]\mu = \mu^n
\end{aligned}
$$

If $\mu = 1$, $E[X_n] = 1$ and recursively, the variance can be given as,

$$\begin{aligned} Var(X_1) &= E[X_0]\,\sigma^2 + Var(X_0)\,\mu^2 = \sigma^2 \\ Var(X_2) &= E[X_1]\,\sigma^2 + Var(X_1)\,\mu^2 = 2\,\sigma^2 \\ &\vdots \\ Var(X_n) &= E[X_{n-1}]\,\sigma^2 + Var(X_{n-1})\,\mu^2 = n\,\sigma^2 \end{aligned}$$

When $\mu \neq 1$, the solution has a general form of $Var(X_n) = A\mu^n + B\mu^{2n}$. Substituting in the variance equation,

$$A\mu^{n+1} = \sigma^2\mu^n + A\mu^{n+2}$$
$$A = \frac{\sigma^2}{\mu\,(1-\mu)}$$

From the initial point $V(X_0) = 0$, we obtain $A + B = 0$. Thus the variance can be derived as,

$$Var(X_n) = \frac{\sigma^2\,\mu^{n-1}(1-\mu^n)}{1-\mu}, \qquad \text{if } \mu \neq 1$$