A Probability Model of a Family Using Branching Process in Relation to Poisson, Logarithmic and Negative Binomial Distribution

Bassa, Shiwaye Yakura *
Suleiman Mohammad Ismail
Department of Statistics and Operations Research,
Modibbo Adama University of Technology, Yola, Adamawa State, Nigeria.
**E-mail: fcbassa@gmail.com*

ABSTRACT

The focus of this work is on the application of a family branching process in Relation to Poisson, Logarithmic and Negative Binomial Distribution. The data used was obtained by personal interview and analyzed using probability generating function in order to obtain the fitted probability model of the family. The model obtained is found to be the probability generating function of negative binomial distribution. This probability model reveals that the number of descendants in every generation increases with the increase of the number of generation. This method employed in this work is suitable for the model of other systems with similar dynamics.

Keywords: *branching process, Poisson process, family, probability generating function.*

INTRODUCTION

Bienayme-Galton Watson Branching process with the nomenclature of population dynamics, is a discrete-time stochastic process that describes the evolution of a population in which each individual independently of the others gives rise to a random number of offspring (in accordance with a common reproduction law), and then dies (Miguel Gonzalez and Ines Ma del Puerto, 2010) or is not considered in the following counts. We shall give its formal definition and establish some interesting properties. Let $\{X_{nj} : n = 0, 1, ...; j = 0, 1, ...\}$ be non-negative integer valued independent and identically distributed (i.i.d.) random variables with probability distribution $\{P_k\}_{k \geq 0}$ i.e $P(X_{01} = k) = P_k, k \geq 0$. The BGWP is a stochastic process, $\{Z_n\}_{n \geq 0}$, defined recursively as follows:

$$Z_0 = N \in N,\ Z_{n+1} = \sum_{j=1}^{Z_n} X_{nj},\ n \geq 1 \quad (1)$$

Where $\sum_{j=1}^{0} X_{nj}$ is defined to be zero. Thus, X_{nj} represents the number of offspring produced by the *jth* individual in the *nth* generation, and Z_n represents the number of individuals in the *nth* generation. We refer to $\{P_k\}_{k \geq 0}$ as the offspring distribution or law, with P_k being interpreted as the probability that an individual has k offspring, in the simplest case to determine the fitted probability model of the family as a population that does not vary from individual to individual.

In species with sexual reproduction, the population sizes depend on the formation of couples. In many populations, mating is an important factor that cannot be neglected. Bisexual branching processes take this into account explicitly. In general, these processes start with N couples. Each couple has random numbers of female and male offspring which form new couples in accordance with a deterministic or stochastic function, and so on (Haccou, Jagers and Vatutin, 2005).

The aim of this study is to obtain the fitted probability model of occurrence of birth in a family. Therefore the objectives of this study are as follows:

i. To determine the average number of offspring in every generation
ii. To obtain a model that will be used in predicting number of offspring at higher generations
iii. To obtain the probability of extinction of the derived model

METHOD

The method adapted in this paper is primary method of data collection by personal interview and tool used for data analysis was probability generating function.

Notations and Terminologies

P_j is the individual probability of producing j offspring in the family

i. X_n is the generation size
ii. $G(S)$ = Probability generating function.
iii. j number of offspring that produce another offspring

Procedure for Obtaining the Fitted Probability Model Using Probability Generating Function

Assume that an individual has a known probability of producing a number of descendants at a given time and produce no other descendant. In turn these descendants each produce further descendant at the next subsequent time with the same probability. This process creates a successive generation (Bashir, 2015).

At each step there is probability P_j that any individual create j descendants, who is assumed to be the same for every individual at every generation. Let X_n be a discrete random variable representing the population size of the *nth* generation taking values in the non-negative integers 0, 1, 2, 3, 4, …), then the probability generating function of the descendant numbers X_n is define as:

$$G(S) = E(S^j) = \sum_{j=0}^{\infty} P_j S^j$$

$$\text{With } G_1(S) = pgf(X_0) \qquad (2)$$

and $G_2(S) = pgf(X_2)$, which is the sum of X_1 random variables (the descendants of X_0), which in turn we denoted by independent random variables say, $Y_1, Y_2, \ldots, Y_{x1}$.
So that, $X_2 = Y_1 + Y_2 + \ldots + Y_{x1}$.
Let $P_j = P(Y_k = j)$; $j = 0, 1, 2 \ldots$
$P_r = P(X_1 = r)$ and $P_n = P(X_2 = n)$

Using partition law

$$P_n = \sum_{r=0}^{\infty} P\left(\frac{X_2 = n}{X_1 = r}\right) P(X_1 = r)$$

$$P_n = \sum_{r=0}^{\infty} P_r P\left(\frac{X_2 = n}{X_1 = r}\right)$$

Multiply both side by S^n and sum over n. i.e

$$\sum_{n=0}^{\infty} P_n S^n = \sum_{r=0}^{\infty} P_r \sum_{n=0}^{\infty} P\left(\frac{X_2 = n}{X_1 = r}\right) S^n$$

$$G_2(S) = \sum_{r=0}^{\infty} P_r \sum_{n=0}^{\infty} P\left(\frac{X_2 = n}{X_1 = r}\right) S^{x2} \qquad (3)$$

From equation (3) above,

$$\sum_{n=0}^{\infty} P\left(\frac{X_2 = n}{X_1 = r}\right) S^{x2} = E\left(S^{X2}\right)$$

Where $P\left(\frac{X_2 = n}{X_1 = r}\right)$ is the conditional probability function for $\frac{X_2}{X_1}$

$$E\left(S^{X2}\right) = E\left(S^{Y_1 + Y_2 + \ldots + Y_{x1}}\right)$$

$$E\left(S^{X2}\right) = E\left(S^{Y_1} . S^{Y_2} \cdot S^{Y_3} \ldots . S^{Y_{x1}}\right)$$

$$= E\left(S^{Y_1}\right) E\left(S^{Y_2}\right) E\left(S^{Y_3}\right) \ldots . E\left(S^{Y_{x1}}\right)$$

Since, every individual in the process reproduced independently.

$$E\left(S^{X2}\right) = \{G(S).G(S).G(S) \ldots . G(S)\}$$

$$E\left(S^{X2}\right) = [G(S)]^r$$

Substitute back into equation (2)

$$G_2(S) = E[G(S)]^r = \sum_{r=0}^{\infty} P_r [G(S)]^r$$

$$G_2(S) = \sum_{r=0}^{\infty} P_r [G(S)]^r = \{G(G(S))\} \tag{4}$$

Equation (4) gives the probability generating function of second generation. The probability generating function of third generation using the same procedure is:

$$G_3(S) = \sum_{r=0}^{\infty} P_r [G(G(S))]^r .$$

In general, probability generating function of *nth* generation is given as:

$$G_n(S) = \sum_{r=0}^{\infty} P_r [G_{n-1}(G(S))]^r = G_{n-1}(G(S))$$

$$G_n(S) = [G(G \ldots (G(S)) \ldots)] \tag{5}$$

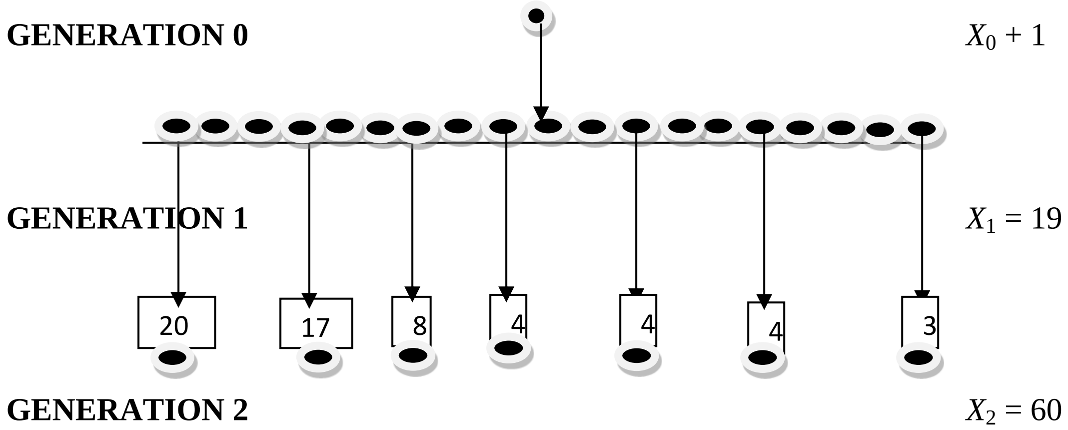

Figure 1: Diagrammatic Representation of Branching Process and Data Presentation

Let the probability generating function of the *nth* generation be

$$G_n(S) = E(S^j) = \sum_{j=0}^{\infty} P_j S^j$$

which is the probability generating function of producing *j* offspring in the next generation.

To find the fitted probability model for the nth generation we need to observe the behaviour of the branching system (assumption of branching process).

1 Giving birth in every woman occurs with time.
2 Every individual in the same generation have equal probability of producing number of offspring.
3 The occurrences are independent, that is given birth of a woman do not affect another woman.

These behaviours (assumptions) are exactly the same with the theoretical form of Poisson assumption which are as follows

1 An event occurs from time to time.
2 Events in a time interval have certain probability of occurrence.
3 The occurrences are independent.

Since, the behaviours of branching process are the same with Poisson assumption; we consider giving birth of individual as occurrences of event in Poisson process denoted by $N(t)$.

Let $N(t)$ be the number of events that have occurred in the interval $(0, t)$. Let event A denote the occurrence of exactly one event in the interval $(t, t + h)$. Similarly, let B and C respectively denote the occurrence of none and more than one event in the interval $(t, t + h)$ (Joyce, 2014).

Also let

$P(A = \textit{occurrence of one event}) = p(h)$

$P(B = \textit{no event}) = q(h)$

$P(C = \textit{more than one event}) = \in(h)$

Now, $N(t)$ form a Poisson process with the following four conditions

1 $N(0) = 0$

2 Events occurring in non-overlapping interval of time are mutually independence.

3 The probabilities $p(h)$, $q(h)$ and $\in(h)$ depend only on the length h of time interval and not on the time origin t.

4 For sufficiently small values of h, we can write for positive constant λ

$$p(h) = P[\textit{one event in the interval}(t, t, + h)] = \lambda h + 0(h)$$

$$q(h) = P[\textit{no event in the interval } (t, t, + h)] = 1 - \lambda h + 0(h)$$

$$\in(h) = P[\textit{more than one event in the interval } (t, t, + h)] = 0(h)$$

Where $lim_{h \to 0} \dfrac{0(h)}{h} = 0$

Let $P_n(t) = P[N(t) = n]$

From condition 1

$$P_0(0) = 1, \textit{and } P_n(0) = 0; n > 0$$

Now consider two successive non-overlapping interval $(0,t) and (t,t+\tau)$

To compute the probability that n events occur in the interval $(0,t+\tau)$ given by

$$P[n \, events \, in \, (0, t + \tau)] = P_n(t + \tau)$$

by total probability theorem

$$P[n \, events \, in \, (0, t + \tau)]$$

$$= \sum_{k=0}^{n} P[\, k \, events \, in \, (0,t) and \, n-k \, events \, in \, (t,t + \tau)]$$

$$= \sum_{k=0}^{n} P[\, k \, events \, in \, (0,t)] P[n-k \, events \, in \, (t,t + \tau)]$$

$$P_n(t+\tau) = \sum_{k=0}^{n} P_k(t) P_{n-k}(\tau); \; f \; or \; n > 0 \; and \; when \; \tau = h$$

$$P_n(t+h) = \sum_{k=0}^{n} P_k(t) P_{n-k}(h)$$

$$P_n(t+h) = P_0(t)P_n(h) + P_1(t)P_{n-1}(h) + \sum_{i=2}^{n} P_i(t)P_{n-i}(h)$$

$$P_n(t+h) = P[no \, event \, in \, (t,t,+h)] \, P[N(t) = n]$$

$$+ \, P[one \, event \, in \, (t,t,+h)] \, P[N(t) = n - 1] + \sum_{i=2}^{n} P_i(t)P_{n-i}(h)$$

$$P_n(t+h) = P_n(t)[1 - \lambda h + 0(h)] + P_{n-1}(t)[\lambda h + 0(h)] \sum_{i=2}^{n} P_i(t) 0(h)$$

$$P_n(t+h) = P_n(t) - \lambda h P_n(t) + \lambda h P_{n-1}(t) + 0(h)$$

$$P_n(t+h) - P_n(t) = -\lambda h P_n(t) + \lambda h P_{n-1}(t) + 0(h)$$

$$\frac{P_n(t+h)}{h} - \frac{P_n(t)}{h} = -\lambda P_n(t) + \lambda P_{n-1}(t) + \frac{0(h)}{h}$$

$$\lim_{h\to 0}\frac{P_n(t+h)}{h}-\frac{P_n(t)}{h}=-\lambda P_n(t)+\lambda P_{n-1}(t)+\lim_{h\to 0}\frac{0(h)}{h}$$

$$\frac{dP_n(t)}{d(t)}=-\lambda P_n(t)+\lambda P_{n-1}(t)$$

$$P_n^{'}(t)=-\lambda P_n(t)+\lambda P_{n-1}(t) \qquad (6)$$

multiply both side of equation (6) by e^{yt}, we have

$$e^{\lambda t}P_n^{'}(t)=-\lambda P_n(t)e^{\lambda t}+\lambda P_{n-1}(t)e^{yt} \qquad (7)$$

let $Q_n(t)=P_n(t)e^{\lambda t}$ and $Q_{n-1}(t)=P_{n-1}(t)e^{\lambda t}$

$$Q_n^{'}(t)=\lambda e^{\lambda t}P_n(t)+P_n^{'}(t)e^{\lambda t}$$

$$=\lambda e^{\lambda t}P_n(t)+\left(-\lambda P_n(t)e^{\lambda t}+\lambda P_{n-1}(t)e^{\lambda t}\right)$$

$$Q_n^{'}(t)=\lambda P_{n-1}(t)e^{\lambda t}$$

$$Q_n^{'}(t)=\lambda Q_{n-1}(t) \qquad (8)$$

Using the boundary conditions that $Q_0^{'}(t)=1$ and $Q_n^{'}(t)=0$

From equation (8)

When n = 1

$$Q_1^{'}(t)=\lambda$$

Integrating both side with respect to t

$$\int Q_1^{'}(t)dt=\int \lambda dt$$

$$Q_1(t)=\lambda t$$

When n = 2

$$Q_2'(t) = \lambda Q_1(t)$$

$$Q_2'(t) = \lambda^2 t$$

$$\int Q_2'(t)dt = \int \lambda^2 t dt$$

$$Q_2(t) = \frac{\lambda^2 t^2}{2} = \frac{(\lambda t)^2}{2!}$$

When n = 3

$$Q_3'(t) = \lambda Q_2(t)$$

$$Q_3'(t) = \lambda \frac{(\lambda t)^2}{2} = \frac{\lambda^3 t^2}{2}$$

$$\int Q_3'(t)dt = \frac{\lambda^3}{2} \int t^2 dt$$

$$Q_3(t) = \frac{\lambda^3 t^3}{2 * 3} = \frac{(\lambda t)^3}{3!}$$

In general

$$Q_n(t) = \frac{(\lambda t)^n}{n!}$$

$$e^{\lambda t} P_n(t) = \frac{(\lambda t)^n}{n!}$$

$$P_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

This shows that considering birth of offspring in branching process as occurrence of event has a Poisson probability density function. Now at steady state and when $n = j$.

$$P_j = \frac{(\lambda t)^j e^{-\lambda t}}{j!}$$

The probability generating function of the process is

$$G_p(S) = \sum_{j=0}^{\infty} P_j S^j$$

$$G_p(S) = \sum_{j=0}^{\infty} \frac{(\lambda t)^j e^{-\lambda t} S^j}{j!}$$

$$G_p(S) = \left[\frac{(\lambda t)^0 e^{-\lambda t} S^0}{0!} + \frac{(\lambda t)^1 e^{-\lambda t} S^1}{1!} + \frac{(\lambda t)^2 e^{-\lambda t} S^2}{2!} + \ldots\right]$$

$$G_p(S) = \left[e^{-\lambda t} + \frac{\lambda t e^{-\lambda t} S}{1!} + \frac{(\lambda t)^2 e^{-\lambda t} S^2}{2!} + \ldots\right]$$

$$G_p(S) = e^{-\lambda t}\left[1 + \frac{\lambda t S}{1!} + \frac{(\lambda t S)^2}{2!} + \ldots\right]$$

From Taylor series expansion of $e^{-\lambda t}\left[1 + \frac{\lambda t S}{1!} + \frac{(\lambda t S)^2}{2!} + \ldots\right] = e^{\lambda t S}$

$$G_p(S) = e^{-\lambda t} e^{\lambda t S}$$

$$G_p(S) = e^{\lambda t(S-1)}$$

In a process where branching occurs from a Poisson process to logarithmic distribution, the most elegant way to determine the resulting distribution is by use of probability generating function. (Christian Walck, 2007).

The probability generating function of logarithmic distribution is

$$G_l(S) = \frac{ln(1-sq)}{ln(1-q)} = \alpha\, ln(1-sq)$$

Where

$0 \leq q \leq 1$ and $\alpha = \dfrac{1}{ln(1-q)}$

For branching process in n-steps

$G(S) = G_1(G_2(\ldots G_{n-1}(G_n(S))\ldots))$

The resulting probability model is obtained by nesting the logarithmic probability generating function in Poisson probability generating function.

$$G(S) = G_p(G_l(S))$$

$$G(S) = exp\{\lambda t[\alpha\, ln(1-sq)-1]\}$$

$$G(S) = exp\{\lambda t\alpha\, ln(1-sq)-\lambda t\}$$

$$G(S) = exp\ ln(1-sq)^{\lambda t\alpha}\ exp^{-\lambda t}$$

$$G(S) = (1-sq)^{\lambda t\alpha}\ exp^{-\lambda t}$$

let $\lambda t\alpha = -j$

From this we have $\lambda t = -\dfrac{j}{\alpha}$

$$\lambda t = \frac{-j}{[ln(1-q)]^{-1}} = -j\, ln(1-q)$$

$$G(S) = (1-sq)^{-j}\ exp^{-[-j\, ln(1-q)]}$$

$$G(S)=(1-sq)^{-j}\ exp^{\,j\,ln(1-q)}$$

$$G(S)=(1-sq)^{-j}\ exp\ ln(1-q)^{j}$$

$$G(S)=(1-sq)^{-j}(1-q)^{j}$$

$$G(S)=P^{j}(1-sq)^{-j}$$

$$G(S)=\left(\frac{P}{1-sq}\right)^{j} \tag{9}$$

The resulting model is recognized as probability generating function of negative binomial distribution with parameter *j* and *p*.

RESULTS AND DISCUSSION

In generation 1:

j = 10

X_1 = 19

$$P=\frac{j}{X_1}=\frac{10}{19}=0.5263$$

$$q=1-p=1-0.5263=0.4737$$

Substituting in equation (4)

$$G_1(S)=\left(\frac{0.5263}{1-0.4737s}\right)^{10}=0.00163(1-0.4737s)^{-10}$$

$$\mu_1 = \left.\frac{dG_1(s)}{ds}\right|_{s=1} = 0.00772(1-0.4737)^{-11}$$

$$= \left.\frac{dG_1(s)}{ds}\right|_{s=1} = \frac{0.00772}{0.00086} = 9 \textit{ offspring}$$

$$\mu_2 = \left.\frac{dG_2(s)}{ds}\right|_{s=1} = \mu_1 \times \mu_1 = \mu_1^2$$

$$\mu_2 = 9 \times 9 = 81 \textit{offspring}$$

The model can also be used in predicting number of offspring at higher generation, the expected number of offspring in generation *n* can be obtain using the derived model below.

$$\mu_n = \left.\frac{dG_n(s)}{ds}\right|_{s=1} = \left(\frac{10 \times 0.4737}{0.5263}\right)^n \qquad (10)$$

Applying equation 2 to equation 9

$$G(S) = \sum_{j=0}^{\infty} P_j S^j$$

$$G(S) = \sum_{j=0}^{\infty} P^j \left(\frac{1}{1-sq}\right)^j$$

$$G(S) = 1 + P\left(\frac{1}{1-sq}\right) + P^2\left(\frac{1}{1-sq}\right)^2 + P^3\left(\frac{1}{1-sq}\right)^3 + \ldots$$

$$G(S) = 1 + \sum_{j=1}^{\infty} P^j \left(\frac{1}{1-sq}\right)^j \qquad (11)$$

From the equation above when $j = 0$, $G(S) = 1$. This means the probability of extinction is 1 if there are 0 offspring in any generation.

The resulting model obtained is recognized as probability generating function of negative binomial distribution with parameter j and p. This shows that when birth pattern in branching process is considered as an occurrence of event in Poisson process, the result leads to logarithmic distribution producing a negative binomial probability generating function. The result gives a description of the branching system of a family as a negative binomial distribution.

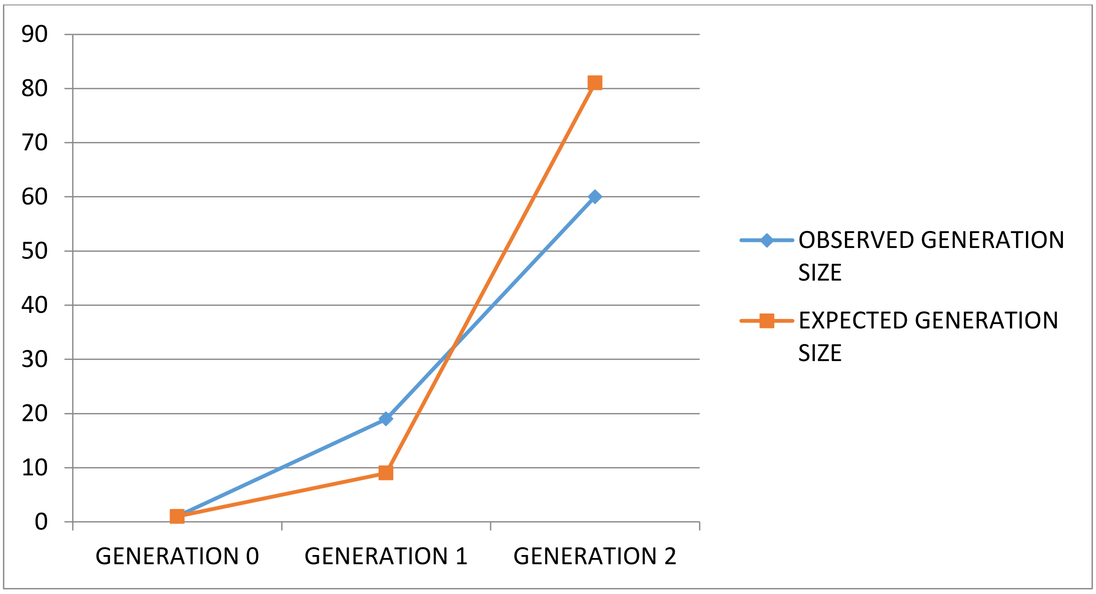

Figure 2: Graphical Representation of the Generation Size

CONCLUSION

A Probability Model of a Family Using Branching Process in Relation to Poisson, Logarithmic and Negative Binomial Distribution was designed. The resulting model is recognized as probability generating function of negative binomial distribution with parameter j and p. The model can also be used in predicting number of offspring at higher generation, the expected number of offspring in generation n can be obtain using the derived model. From the data collected $X_0 = 1$, $X_1 = 19$ and $X_2 = 60$ are the population sizes of generation 0, generation 1 and generation 2 respectively. The expected generation sizes obtained are: $X_0 = 1$, $X_1 = 9$ and $X_2 = 81$. The result shows that the higher the generation size the larger the population size. The population size of X_n generation agrees with the derived model below.

$$X_n = \left(\frac{10 \times 0.4737}{0.5263} \right)^n$$

showing the probability of extinction of the family is 1 when the generation size is zero. Hence, we conclude that the resulting probability model of branching system of the family is negative binomial or branching system of the family follows negative binomial distribution.

REFERENCES

Bashir, M. Y. (2015). Stochastic Process. Lecture Note Book. Bayero University Kano, Nigeria.

Gonzalez, M. and Puerto, I. (2010). Branching process: Genealogy and evolution. *Estadistica*, 26, (2), 109

Haccou P., Jagers P. and Vatutin V. (2005). *Branching processes: variation, growth and extinction of populations*. U.K, Cambridge: University Press.

Joice, D. F. (2014). *The Poisson process: probability and statistics.* Kiagbodo, Delta, Nigeria: Clark University.

Walck, C. (2007). *Handbook on statistical distributions for experimentalists*. University of Stockholm press.