**KMITL Sci. Tech. J. Vol. 16 No. 2 Jul.-Dec. 2016**

# Analysis of Difference Equations for Population Genetics

Puntani Pongsumpun

Department of Mathematics, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand

## Abstract

The genetics of populations with discrete generations were studied in this paper. We explored the change of the relative frequencies of genes based on the *Hardy-Weinberg* law. The difference equations were formulated for describing their changes. The stability theorem was used for analysis. The numerical simulations were found with the different situations.

**Keywords:** difference equations, genetics, *Hardy-Weinberg* law, stability theorem

## 1. Introduction

Population genetics is the study of variation of genetics within population and with the examination and modelling of changes in the frequencies of genes and alleles in populations over space and time. Many genes found within a population, will occur in a number of different forms (or alleles). Difference equations were used for investigating and predicting the occurrence of specific alleles or combinations of alleles in populations, based on developments in the molecular understanding of genetics [1]. The development of population genetics with mathematical models was occurred long time ago. The early phase was discovered by Fisher, Haldane, and Wright. Their theories had relevant with medical genetics and evolutionary biology. They are worth for illustrating the translation of population genetics to mathematical models. The stability of equilibrium points was also relevant in their studies [2]. The gene pool of the population is the collection of all alleles of all genes found within a freely interbreeding population. Each member of the population receives its alleles from other members of the gene pool (its parent) and passes them on to other members of the gene pool (its offspring). We studied the variation in alleles and genotypes within the gene pool and how this variation changes from one generation to the next generation. The population size, mutation, genetic drift, natural selection, environmental diversity, migration and non-randommating patterns were influence to the genetic diversity within a gene pool [3]. The genetics problem is important because it can help human to face with current environmental variability as same as to reduce potentially effects of closed relative breeding. In 1999, Chen *et al.* [4] presented a differential equation model for gene expression and analyzed their model. In 2014, Sargolzaie and Miri-Moghaddam [5] analyzed regression equations

*Corresponding author: E-mail: kppuntan@kmitl.ac.th

of differential diagnosis of β-thalassemia trait and iron deficiency anemia in Southeast Iran. In the same year (2014), Sirachainan *et al.* [6] formulated the mathematical formula for differentiating Thalassemia trait and iron deficiency anemia in school age children. In this study, the population was supposed to satisfy the Hardy–Weinberg equilibrium model theorem, this stated that allele and genotype frequencies in a population remained constant from generation to generation in the absence of other evolutionary influences. The difference equations were used to explain this phenomena because it could predict the frequency of human genotype for the next generation.

## 2. Materials and Methods

In this study, there were two alleles, a and A. The population can pass from one generation to the next. Thus there are three combinations: AA, aa and aA. AA and aa were called homozygous. aA were called heterozygous. If mating is random and all genotypes were equally fit (had an equal probability of surviving to produce offspring). The gene frequencies did not change. This corresponded to the Hardy-Weinberg law. Let N be the size of population, p as the frequency of allele A and q as the frequency of allele a [7].

p = total size of A alleles / 2N

q = total size of a alleles / 2N

where p + q = 1. The equations based on the following assumptions [8-13]:

i) There were no mutations.

ii) Mating was random.

iii) There was no variation of the number of progeny from parents of different genotypes.

iv) Progeny had equally likely to survive.

Let x be the frequency of AA genotype,

y be the frequency of aA genotype,

z be the frequency of aa genotype then $x + y + z = 1$. It can be seen that

$p = x + 1/2y$

$q = 1/2y + z$.

Note that aA was equivalent to Aa.

The frequencies of each parent for particular genotypes were shown in Table 1.

**Table 1**. Mating table

| Father | Mother | Frequencies |
|---|---|---|
| AA | AA | $x^2$ |
| AA | Aa | xy |
| AA | aa | xz |
| Aa | AA | yx |
| Aa | Aa | $y^2$ |
| Aa | aa | yz |
| aa | AA | zx |
| aa | Aa | zy |
| aa | aa | $z^2$ |

When there was the mating between father and mother, there were four possible combinations. Cases of heterozygous parents and homozygous parents were shown in Figure 1.

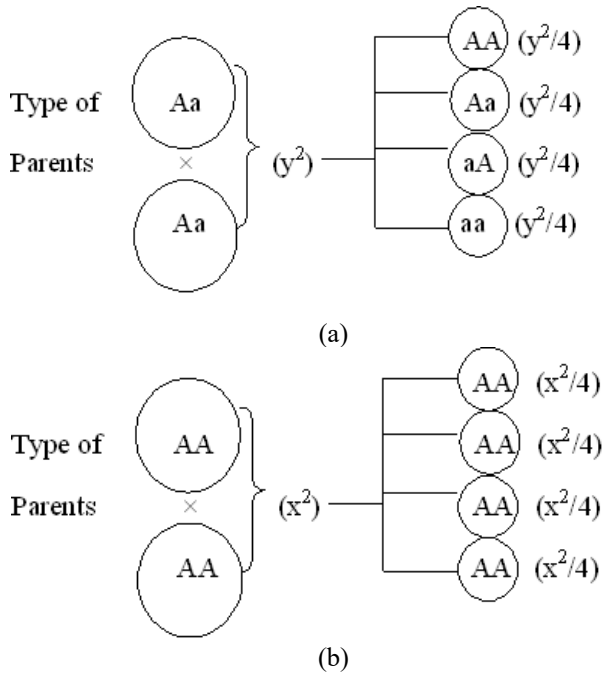


**Figure 1**. Offspring diagram
(a) Offspring of heterozygous parents (Aa×Aa)
(b) Offspring of homozygous parents (AA×AA)

The offspring table was shown in Table 2.

**Table 2**. Offspring table

| **Type of Parents** | **Offspring** | | | **Total frequencies** |
|---|---|---|---|---|
| | AA | aA | aa | |
| AA×AA | $x^2$ | 0 | 0 | $x^2$ |
| AA×Aa | xy | xy | 0 | 2xy |
| AA×aa | 0 | 2xz | 0 | 2xz |
| Aa×Aa | $y^2/4$ | $y^2/2$ | $y^2/4$ | $y^2$ |
| Aa×aa | 0 | yz | yz | 2yz |
| aa×aa | 0 | 0 | $z^2$ | $z^2$ |
| | $x^2$+ xy + $y^2/4$ | xy+2xz + $y^2/2$+ yz | $y^2/4$+ yz+$z^2$ | $x^2$+2xy +2xz+$y^2$ +2yz+$z^2$ |

The total frequency of offspring for each pair of parent can be shown in Table 2. They obtained from the possible combinations as shown in Figure 1. The n-th generation frequencies of each genotype (AA, aA and aa) were derived from the total frequencies of genotype for each pair of parent as shown in Table 2. The n-th generation frequency of each genotype can be shown in the form of difference equations (1):

$$\begin{aligned} x_{n+1} &= x_n^2 + x_n y_n + y_n^2/4 \\ y_{n+1} &= x_n y_n + 2x_n z_n + y_n^2/2 + y_n z_n \\ z_{n+1} &= y_n^2/4 + y_n z_n + z_n^2 \end{aligned} \tag{1}$$

By using standard dynamical modelling method, we need to find equilibrium points and determine their stabilities [5]. Equilibrium points (x*,y*,z*) of our equations (1) were found by setting $x_{n+1} = x_n = x^*, y_{n+1} = y_n = y^*, z_{n+1} = z_n = z^*$ into equations (1), then the possible non-zero equilibrium points were (0,0,1), (1,0,0), $(x_1^*, y_1^*, z_1^*)$ and $(x_2^*, y_2^*, z_2^*)$ where

$$\begin{aligned} x_1^* &= \frac{1}{2}(1-\sqrt{1-2v}-v), \\ z_1^* &= \frac{1}{2}(1+\sqrt{1-2v}-v), \\ y_1^* &= v, \\ x_2^* &= \frac{1}{2}(1+\sqrt{1-2v}-v), \\ z_2^* &= \frac{1}{2}(1-\sqrt{1-2v}-v), \\ y_2^* &= v \end{aligned} \tag{2}$$

and v was any real number.

The stability of each equilibrium point can be determined by checking the modulus of eigenvalues. If the modulus of each eigenvalue is not greater than 1, then that equilibrium point is stable [5].

Let

$$\begin{aligned} x_{n+1} &= f(x_n, y_n, z_n) \\ y_{n+1} &= g(x_n, y_n, z_n) \\ z_{n+1} &= h(x_n, y_n, z_n) \end{aligned}$$

and

$D = \begin{pmatrix} f_x(x^*,y^*,z^*) & f_y(x^*,y^*,z^*) & f_z(x^*,y^*,z^*) \\ g_x(x^*,y^*,z^*) & g_y(x^*,y^*,z^*) & g_z(x^*,y^*,z^*) \\ h_x(x^*,y^*,z^*) & h_y(x^*,y^*,z^*) & h_z(x^*,y^*,z^*) \end{pmatrix}$, then equilibrium point $(x^*, y^*, z^*)$ is stable if each eigenvalue $(\lambda)$ has modulus not greater than 1. Each eigenvalue can be calculated from solving

$$Det(D - \lambda I) = 0 .$$

From equations (1), we can have

$$D=\begin{pmatrix} 2x^*+y^* & x^*+y^*/2 & 0 \\ y^*+2z^* & x^*+y^*+z^* & 2x^*+y^* \\ 0 & y^*/2+z^* & y^*+2z^* \end{pmatrix}.$$

We check the modulus of eigenvalues for each equilibrium point which was defined in equations (2), we found that all eigenvalues had modulus not greater than 1 as we can see in Table 3. Therefore each equilibrium point was stable.

**Table 3**. Eigenvalues of each equilibrium point.

| **Equilibrium point (ep.)** | **Eigenvalues ($\lambda_i$ (i=1,2,3)** |
|---|---|
| First ep. (0,0,1) | $\lambda_1$= -1, $\lambda_2$=0, $\lambda_3$=1 |
| Second ep. (1,0,0) | $\lambda_1$= -1, $\lambda_2$=0, $\lambda_3$=1 |
| Third ep. $(x_1^*, y_1^*, z_1^*)$ | $\lambda_1 = \sqrt{1-2v}$, $\lambda_2 = -\frac{1}{2}(\sqrt{1+2v}+\sqrt{1-2v})$, $\lambda_3 = \frac{1}{2}(\sqrt{1+2v}-\sqrt{1-2v})$ |
| Fourth ep. $(x_2^*, y_2^*, z_2^*)$ | $\lambda_1 = -\sqrt{1-2v}$, $\lambda_2 = -\frac{1}{2}(\sqrt{1+2v}-\sqrt{1-2v})$, $\lambda_3 = \frac{1}{2}(\sqrt{1+2v}+\sqrt{1-2v})$ |

## 3. Results and Discussion

In this section, we simulated our difference equations (1) by using numerical method with the difference initial conditions. The program Mathematica (Wolfram Research, Champaign, IL) was used for simulating the numerical results. The initial conditions were randomly selected and it can be separated into 5 cases. The numerical results were shown in Figures 2-6.

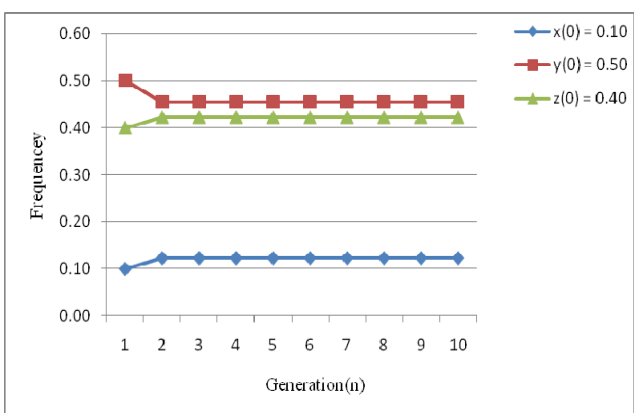


**Figure 2.** Case 1: numerical simulations from our difference equations (1) with x(0) = 0.10, y(0) = 0.50,z(0) = 0.40. The solutions converged to $(x_1^*, y_1^*, z_1^*)$,

where = 0.1225, = 0.455 and = 0.4225

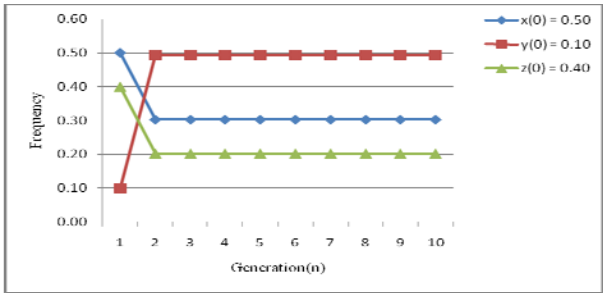


**Figure 3**. Case 2: numerical simulations from our difference equations (1) with x(0) = 0.50, y(0) = 0.10, z(0) = 0.40. The solutions converged to $(x_2^*, y_2^*, z_2^*)$, where $x_2^* = 0.3025$, $y_2^* = 0.495$ and $z_2^* = 0.2025$

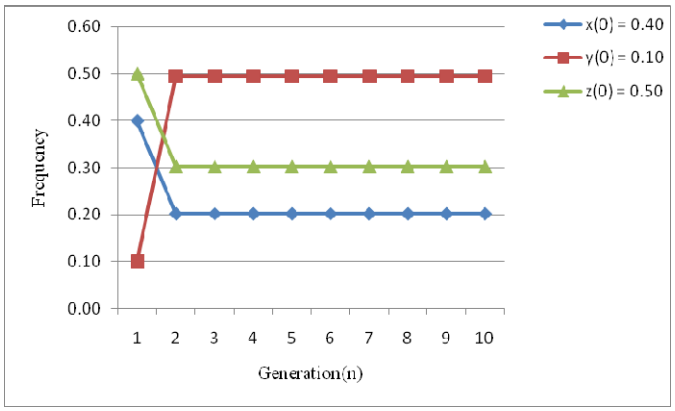


**Figure 4**. Case 3: numerical simulations from our difference equations (1) with x(0) = 0.40, y(0) = 0.10, z(0) = 0.50.The solutions converged to $(x_1^*, y_1^*, z_1^*)$, where $x_1^* = 0.2025$, $y_1^* = 0.495$ and $z_1^* = 0.3025$

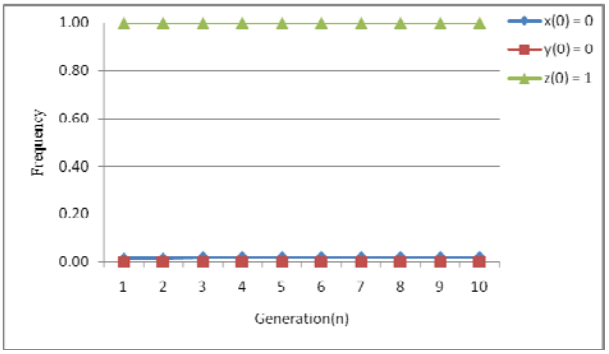


**Figure 5**. Case 4: numerical simulations from our difference equations (1) with x(0) = 0, y(0) = 0, z(0) = 1.The solutions converged to (0,0,1)

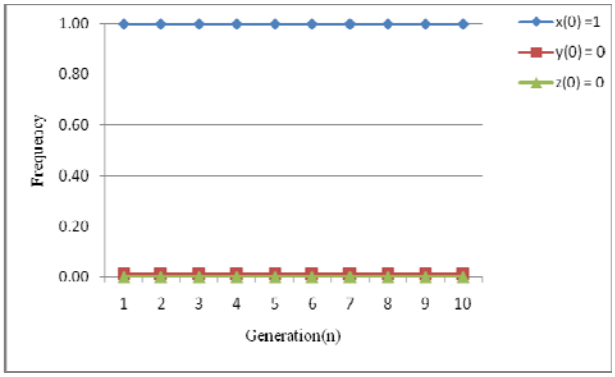


**Figure 6**. Case 5: numerical simulations from our difference equations (1) with x(0) = 1, y(0) = 0, z(0) = 0.The solutions converged to (1,0,0)

In this paper, we applied difference equations for describing the population genetics. The difference equations used in this study obtained from Keshet [7]. They were formed by using the idea of neutral Wright-Fisher model [14]. It stated that the genotype frequency of generation n+1 was formed from generation n. The numerical simulations were used to support the analytical results. The meaning of figure 2 to figure 6 can be explained as follows:

Case 1: if frequency of AA genotype, aA genotype, aa genotype of parent were 0.10, 0.50 and 0.40, respectively, then the frequencies of AA genotype, aA genotype, aa genotype for each offspring are 0.1225, 0.455, 0.4225, respectively, as shown in figure 2. The solutions converged to the third equilibrium point.

Case 2: if frequency of AA genotype, aA genotype, aa genotype of parent were 0.50, 0.10 and 0.40, respectively, then the frequencies of AA genotype, aA genotype, aa genotype for each offspring are 0.3025, 0.495, 0.2025, respectively, as shown in figure 3. The solutions converged to the fourth equilibrium point.

Case 3: if frequency of AA genotype, aA genotype, aa genotype of parent were 0.40, 0.10 and 0.50, respectively, then the frequencies of AA genotype, aA genotype, aa genotype for each offspring are 0.2025, 0.495, 0.3025, respectively, as shown in figure 4. The solutions converged to the third equilibrium point.

Case 4: if frequency of AA genotype, aA genotype, aa genotype of parent were 0, 0 and 1, respectively, then the frequencies of AA genotype, aA genotype, aa genotype for each offspring are 0, 0, 1, respectively, as shown in figure 5. The solutions converged to the first equilibrium point.

Case 5: if frequency of AA genotype, aA genotype, aa genotype of parent were 1, 0 and 0, respectively, then the frequencies of AA genotype, aA genotype, aa genotype for each off spring are 1, 0, 0, respectively, as shown in figure 6. The solutions converged to the second equilibrium point.

The organism had two copies of the same allele for a gene. If it carried two copies of the same dominant allele, it can be homozygous dominant. If it carried two copies of the same recessive allele, it can be homozygous recessive. Heterozygous means that an organism had two different alleles of a gene. For case 1, 2 and 3, the frequency of each genotype for offspring was depend on the frequency of his/her parent. It can be seen that for case 4 and case 5, the frequency of homozygous dominant for parent was 1, and then the frequency of homozygous dominant for offspring is also 1. If the frequency of homozygous recessive for parent was 1, then the frequency of homozygous recessive for offspring was also 1. We can see that in each case, it converged to

the equilibrium solutions. Thus, the numerical results corresponded to the standard dynamical modeling theorem [7].

## 4. Conclusions

Genetics was the science of heredity which originated in the 19th century. Using the difference equations for population genetics, we described how organism population reproduce and propagate various properties to their offspring. The changes in the gene pool through the generations can be found from iterating difference equations (1) with the different initial values. Convergence of the n-th generation frequency of each genotype $(x_n, y_n, z_n)$ to either (0,0,1) or (1,0,0) was based on the initial value $(x_0, y_0, z_0)$. If $(x_0, y_0, z_0)$ was closer to (0,0,1), then $(x_n, y_n, z_n)$ converged to the equilibrium point (0,0,1). If $(x_0, y_0, z_0)$ was closer to (1,0,0), then $(x_n, y_n, z_n)$ converged to the equilibrium point (1,0,0). Based on the likelihood of mating, we can determine the probability of a given match resulting in offspring of a given genotype. The results of our difference equations could predict the occurrence of the frequency genotype for the next generations.

## References


[1] Crow, J.F. and Kimura, M., **1970**. *An Introduction to population genetics theory*. Harper and Row, New York.

[2] Mathematical Genetics, [online] Avaiable at: http://www.genetics.ucla.edu/courses/hg236b/Lange_Chapter_PopGenetics.pdf.

[3] Population genetics, [online] Avaiable at: http://www2.le.ac.uk/departments/genetics/vgec/schoolscolleges/topics/population-genetics.

[4] Chen, T., He, H.L. and Church, G.M., **1999**. Modeling gene expression with differential equations, *Pacific Symposium on Biocomputing*, 4, 29–40.

[5] Sargolzaie, N. and Miri-Moghaddam, E., **2014**. A local equation for differential diagnosis of β-thalassemia trait and iron deficiency anemia by logistic regression analysis in Southeast Iran. *Hemoglobin*, 38(5), 355–358.

[6] Sirachainan, N., Iamsirirak, P., Charoenkwan, P., Kadegasem, P., Wongwerawattanakoon, P., Sasanakul, W., Chansatitporn, N. and Chuansumrit, A., **2014**. New Mathematical Formula for Differentiating Thalassemia Trait and Iron Deficiency Anemia in Thalassemia Prevalent Area: A study in Healthy School-Age Children, *The Southeast Asian Journal of Tropical Medicine and Public Health*, 45(1), 174–182.

[7] Keshet, E.L., **1989**. *Mathematical models in biology*. 1st ed. New York: Random House.

[8] Ewens, W.J., **1979**. *Mathematical Population Genetics*. Springer–Verlag,New York.

[9] Li, C.C., **1976**. *A First Course in Population Genetics*. Boxwood Press, Pacific Grove, Calif.

[10] Roughgarden, J., **1979**. *Theory of Population Genetics and Evolution Ecology: An Introduction*. Macmillan, New York.

[11] Segel, L.A., **1984**. *Modeling Dynamic Phenomena in Molecular and Cellular Biology*. Cambridge University Press, Cambridge.

[12] Smith, J.M., **1968**. *Mathematical Ideas in Biology*. Cambridge University Press, Cambridge.

[13] Allman, E.S. and Rhodes, J.A., **2004**. *Mathematical models in Biology: An Introduction*, Cambridge University Press.

[14] Etheridge, A., **2009**. Some Mathematical Models from Population Genetics, [online] Avaiable at: http://www.stats.ox.ac.uk/~etheridg/orsay/notes1.pdf.