ON THE ADAPTED EQUATIONS FOR SEVERAL DYPLOID MODEL IN POPULATION GENETICS

Won Choi

ABSTRACT. For a locus with two alleles (I^A and I^B), the frequencies of the alleles are represented by

$$p = f(I^A) = \frac{2N_{AA} + N_{AB}}{2N}, \ q = f(I^B) = \frac{2N_{BB} + N_{AB}}{2N}$$

where N_{AA}, N_{AB} and N_{BB} are the numbers of $I^A I^A$, $I^A I^B$ and $I^B I^B$ respectively and N is the total number of populations. The frequencies of the genotypes expected are calculated by using p^2, $2pq$ and q^2. Choi showed the method of whether some genotypes is in these probabalities. Also he calculate the probability generating function for offspring number of genotype under a diploid model([1]). In this paper, let $x(t, p)$ be the probability that I^A become fixed in the population by time t-th generation, given that its initial frequency at time $t = 0$ is p. We find adapted equations for x using the mean change of frequence of alleles and fitness of genotype. Also we apply this adapted equations to several diploid model and it also will apply to actual examples.

1. Introduction

The gene pool of a population can be represented in terms of allelic frequencies. There are always fewer alleles than genotypes, so the gene pool of a population can be represented in fewer terms when allelic frequencies are used. Consider a population consisting of $2N$ genes at a single locus which are either A or B. The gene population can be represented in term of allelic frequencies. There are fewer alleles than genotypes, so the gene population can be represented in fewer term when allelic frequencies are used.

Consider a diploid model. For a locus with two alleles (I^A and I^B), the frequencies of the alleles are represented by the p and q and p, q can be calculated as follows;

$$p = f(I^A) = \frac{2N_{AA} + N_{AB}}{2N}$$
$$q = f(I^B) = \frac{2N_{BB} + N_{AB}}{2N}$$

where N_{AA}, N_{AB} and N_{BB} are the numbers of $I^A I^A$, $I^A I^B$ and $I^B I^B$ respectively and N is the total number of populations.

The alleles frequencies can be calculated from the genotype frequencies. To calculate allelic frequencies from genotypic frequencies, we add the frequency of the homozygote for each allele to half the frequency of the heterozygote([6]);

$$p = f(I^A) = f(I^A I^A) + \frac{1}{2} f(I^A I^B)$$

$$q = f(I^B) = f(I^B I^B) + \frac{1}{2} f(I^A I^B)$$

The frequencies of the genotypes expected are calculated by using p^2, $2pq$ and q^2. Choi showed the method of whether some genotypes is in these probabalities. Also he calculate the probability generating function for offspring number of genotype under a diploid model of N population with N_1 males and N_2 females and he find the conditional joint probability generating function of genotype frequencies([1]).

In this paper, we define the density and operator for the value of the frequency of one gene and find adapted equations as a follow-up for the frequency of alleles. Also we apply this adapted equations to several diploid model and it also will apply to actual examples.

2. The Main Results

Let $p(t)$ be the frequency of I^A-allele at time t. For the time being, assume that the trajectories of $p(t)$ can be approximated by paths which are continuous. Therefore we have a diffusion process.

Let L be one-dimensional diffusion operator

$$L(t, p) = a(t, p) \frac{\partial^2}{\partial p^2} + b(t, p) \frac{\partial}{\partial p}.$$

A probability P is called a solution of the *martingale problem* if it satisfies the following conditions,

  (1) $P(p(0) = p) = 1$.
  (2) denoting $M_f(t) = f(p(t)) - \int_0^t Lf(p(t)) ds$, $(M_f(t), \mathcal{F}_t)$ is a P-martingale.

The diffusion operator L was first introduced by Gillespie([3]). We define the one parameter family $\{T_t : t \geq 0\}$ of transformations by

$$T_t f = E[f(p(t))]$$

for an arbitrary $f(x)$. Then Ethier([2]) showed that there exists a strongly continuous non-negative contraction semigroup $\{T_t : t \geq 0\}$ and we have

$$T_t T_s = T_{t+s}.$$

On the other hand, let $f(t_1, p; t_2, y)$ $(t_1 < t_2)$ be the density that the value of the $p(t)$ is y at time t_2, given that it is p at time t_1 and

$$Q(t_1, t_2) g(p) = x(t_1, t_2, p) = E_f[g(y)]$$

for an arbitrary g. Here E_f means expectation with respect to density f. Since T_t is a strongly continuous non-negative contraction semigroup, we have

$$Q(t_1, t_2) = Q(t_2, s)Q(s, t_1)$$

for $t_1 < s < t_2$. Therefore

$$\frac{1}{\Delta t}\{Q(t_2 + \Delta t, t_1) - Q(t_2, t_1)\}$$
$$= \frac{1}{\Delta t}\{Q(t_2 + \Delta t, t_2)Q(t_2, t_1) - Q(t_2, t_1)\}$$
$$= \frac{1}{\Delta t}\{Q(t_2 + \Delta t, t_2) - I\}Q(t_2, t_1).$$

If $x(t, p) = \int f(0, p; , t, y)g(y)dy$, there exists at least one diffusion process whose equation is

$$\frac{\partial x(t, p)}{\partial t} = L(t, p)x(t, p)$$

for each differential operator $L(t, p)$ and the converse is true.([2]) If $x(t, p) = x(0, t, p) = \int f(0, p; t, y)g(y)dy$, then we have

$$\frac{\partial x(t, p)}{\partial t} = a(t, p)\frac{\partial^2 x(t, p)}{\partial p^2} + b(t, p)\frac{\partial x(t, p)}{\partial p}.$$

we apply this result to dyploid model of population genetics;

LEMMA 1. *Denote $x(t, p)$ be the probability that I^A become fixed in the population by time t-th generation, given that its initial frequency at time $t = 0$ is p. Then we have adapted equation*

$$\frac{\partial x}{\partial t} = \frac{pq}{4N}\frac{\partial^2 x}{\partial p^2} + M(t)\frac{\partial x}{\partial p}$$

where $M(t)$ is mean change for the frequence of alleles.

Proof. This result follows easily from the properties of dyploid model and the result of M. Kimura.([4]) □

EXAMPLE 1. In case of Wright's model, $2N$ genes are drawn random from the gene cluster with replacement at each generation. If the population consists of j-th A genes and $(2N - j)$-th B genes, each drawing results in A or B with probabilities $j/2N$ or $(2N - j)/2N$ respectively. Since the mean change is zero obviously, we have following adapted equation for the frequence of gene A:

$$\frac{\partial x}{\partial t} = \frac{pq}{4N}\frac{\partial^2 x}{\partial p^2}.$$

Of course, this example is the case when there is no mutation. If there is a mutation, A genes mutate to B genes in the time interval Δt and B genes to A genes. Therefore the mean change cannot be zero.

The fitness is defined as the relative reproductive success of a genotype in case of natural selection. The natural selection takes place when individuals with adaptive traits produce a greater numbers of offspring than that produced by others in the population. If the adaptive traits have a genetic basis, they are inherited by the offspring and appear with greater frequency in the next generation ([5], [6]). The

fitness is the reproductive success of one genotype compared with the reproductive success of other genotypes in the population.

Then we have;

THEOREM 2. *Suppose that A and B genes are selectively different and that fitnesses are $1 + a$ and 1 for a constant a, respectively. If $|a| < 1$, we have adapted equation with respect to this model*

$$\frac{\partial x}{\partial t} = \frac{pq}{4N}\frac{\partial^2 x}{\partial p^2} + apq\frac{\partial x}{\partial p}$$

for the frequence of allele I^A.

Proof. The average fitness is $(1 + a)p + q$ and I^A frequency in next generation is

$$\frac{(1 + a)p}{(1 + a)p + q}.$$

Therefore the mean change is

$$\frac{(1 + a)p}{(1 + a)p + q} - p = \frac{apq}{1 + ap}.$$

The equatiion with respect to this model

$$\frac{\partial x}{\partial t} = \frac{pq}{4N}\frac{\partial^2 x}{\partial p^2} + \frac{apq}{1 + ap}\frac{\partial x}{\partial p}$$

and if $|a| < 1$, $a/(1 + ap) \to a$ and we have a result.   □

Suppose that the allelic frequencies of a population do not change and the genotypic frequencies will not change after one generation in the proportion p^2 (the frequency of $I^A I^A$), $2pq$ (the frequency of $I^A I^B$) and q^2 (the frequency of $I^B I^B$). Here p is the frequency of allele I^A and q is the frequency of allele I^B. When genotype are in the expected proportions of p^2, $2pq$, q^2, the population is said to be in Hardy-Weinburg equilibrium([5], [6]).

THEOREM 3. *Let the fitness for each genotype be $1 + a$, $1 + aw$, 1, respectively for a measure of the degree of dominance w. Then adapted equation with respect to this model is*

$$\frac{\partial x}{\partial t} = \frac{pq}{4N}\frac{\partial^2 x}{\partial p^2} + apq\frac{p + 2wq}{1 + ap^2 + 2awpq}\frac{\partial x}{\partial p}$$

for the frequence of allele I^A.

Proof. The average fitness is

$$(1 + a)p^2 + (1 + aw)2pq + q^2 = 1 + ap^2 + 2awpq.$$

Therefore, since I^A frequency in next generation is

$$\frac{(1 + a)p^2 + (1 + aw)pq}{1 + ap^2 + 2awpq}$$

and the mean change is

$$\frac{(1 + a)p^2 + (1 + aw)pq}{1 + ap^2 + 2awpq} - p = apq\frac{p + 2wq}{1 + ap^2 + 2awpq},$$

we have the result.   □

EXAMPLE 2. Suppose $N_{AA} = 20$, $N_{AB} = 50$ and $N_{BB} = 100$ which means the number of zygotes in one generation. The allelic frequencies can be calculated from either the numbers or the frequencies of the genotypes. To calculate allelic frequencies from the numbers of genotypes, we try following calculations with round off to the proper digit;

$$p = f(I^A) = \frac{2N_{AA} + N_{AB}}{2N} = 0.2647, \ q = f(I^B) = \frac{2N_{BB} + N_{AB}}{2N} = 0.7353.$$

To calculate the allelic frequencies from genotypic frequencies, we try following calculations;

$$p = f(I^A) = f(I^A I^A) + \frac{1}{2}f(I^A I^B) = 0.2647$$

$$q = f(I^B) = f(I^B I^B) + \frac{1}{2}f(I^A I^B) = 0.7353.$$

The frequencies of the genotypes expected under Hardy-Weinberg equilibrium are

$$I^A I^A = p^2 = 0.0701$$
$$I^A I^B = 2pq = 0.3893$$
$$I^B I^B = q^2 = 0.5407.$$

Suppose that the average number of viable offspring produced by three genotype $I^A I^A$, $I^A I^B$ and $I^B I^B$ are $2, 5$ and 10, respectively. This means the average number of progeny per individual in next generation. We find the fitness for each genotype as following;

$$\text{the fitness of } I^A I^A = \frac{2}{10} = 0.2$$

$$\text{the fitness of } I^A I^B = \frac{5}{10} = 0.5$$

$$\text{the fitness of } I^B I^B = \frac{10}{10} = 1.$$

Since the degree of dominance is $w = 0.625$, we have adapted equation

$$\frac{\partial x}{\partial t} = 0.0029\frac{\partial^2 x}{\partial p^2} + 0.1920\frac{\partial x}{\partial p}$$

for the frequence of allele I^A.

REMARK. The selection coefficient is the relative intensity of selection against a genotype([5]). We usually note of selection for a special genotype. When selection is for one genotype, selection is automatically against at least one other genotype. In Theorem 3, The average selection coefficient is

$$ap^2 + 2awpq.$$

So, we can have same adapted equation

$$\frac{\partial x}{\partial t} = \frac{pq}{4N}\frac{\partial^2 x}{\partial p^2} + apq\frac{p + 2wq}{1 + ap^2 + 2awpq}\frac{\partial x}{\partial p}$$

using selection coefficient, without the fitness. In case of Example 2, selection coefficients for three genotypes are 0.8, 0.5 and 0, respectively.

COROLLARY 4. *Suppose that the allele frequency changes after one generation of selection against genotypes carrying the dominant allele. Then we have adapted equation with respect to this model*

$$\frac{\partial x}{\partial t} = \frac{pq}{4N}\frac{\partial^2 x}{\partial p^2} + \left[\frac{p(1-b)(2-p)}{1-b+bq^2} - p\right]\frac{\partial x}{\partial p}$$

for the frequence of allele I^A. Here b is selection coefficient.

Proof. The fitness are $1-b, 1-b$ and 1 for genotypes I^AI^A, I^AI^B and I^BI^B, respectively. The average fitness is $1-b+bq^2$ and I^A frequency in next generation is

$$\frac{p(1-b)(2-p)}{1-b+bq^2}.$$

Therefore the mean change is

$$\frac{p(1-b)(2-p)}{1-b+bq^2} - p$$

and we have result.   □

References

[1] W.Choi, *On the genotype frequencies and generating function for frequencies in a dyploid model,* Korean J. Mathematics. **29** (1) (2021).
[2] S.N.Ethier *A class of degenerate diffusion processes occurring in population genetics* , Communications on Pure and Applied Mathematics. Vol XXIX (1976).
[3] J.H.Gillespie, *Natural selection for within-generation variance in offspring number,* Genetics, **76** (1974), 601–606.
[4] M.Kimura, *A Stochastic model of Compensatory Neutral Evolution* , Proceedings of a Workshop held in Nagoya, July 8-12. Stochastic Methods in Biology (1985)
[5] R. Lewis, *Human Genetics : Concepts and Applications,* McGraw-Hill Education (2016).
[6] B.A. Pierce, *Genetics Essentials : Concepts and Connections,* W.H.Freeman and Company (2014), 216–240.

Won Choi
Department of Mathematics, Incheon National University, Incheon Korea.
E-mail: choiwon@inu.ac.kr