Korean J. Math. **31** (2023), No. 1, pp. 17–23
http://dx.doi.org/10.11568/kjm.2023.31.1.17

ON THE ADAPTED EQUATIONS IN VARIOUS DYPLOID MODEL AND HARDY-WEINBURG EQUILIBRIUM IN A TRIPLOID MODEL

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^AI^A, I^AI^B and I^BI^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 defined the density and operator for the value of the frequency of one gene and found the adapted partial differential equation as a follow-up for the frequency of alleles and applied this adapted partial differential equation to several diploid model [1].

In this paper, we find adapted equations for the model for selection against recessive homozygotes and in case that the alley frequency changes after one generation of selection when there is no dominance. Also we consider the triploid model with three alleles I^A, I^B and i and determine whether six genotypes observed are in Hardy-Weinburg for equilibrium.

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.

Firstly, 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

---

Received November 7, 2022. Revised December 28, 2022. Accepted January 2, 2023.
2010 Mathematics Subject Classification: 92D10, 60H30, 60G44.
Key words and phrases: adapted equation, dyploid, triploid.
This research was supported by Incheon National University Research Grant, 2022-2023.

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^AI^A, I^AI^B and I^BI^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 [5];

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

In case of a locus with three alleles I^A, I^B and i, we have six genotypes and add the frequency of the homozygote to half the frequency of each heterozygous genotype that possesses the allels. So the grequencies of the alleles are

$$p = f(I^A) = f(I^AI^A) + \frac{1}{2}f(I^AI^B) + \frac{1}{2}f(I^Ai)$$
$$q = f(I^B) = f(I^BI^B) + \frac{1}{2}f(I^AI^B) + \frac{1}{2}f(I^Bi)$$
$$r = f(ii) = f(ii) + \frac{1}{2}f(I^AI^B) + \frac{1}{2}f(I^Bi).$$

The ABO blood groups can use as an example for three alleles. In a certain ABO blood population groups, the frequencies of the four blood groups are A (genotypes I^AI^A and I^Ai), B (genotypes I^BI^B and I^Bi), AB (genotype I^AI^B) and O (genotype ii).

Choi calculated 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 [2]. Also he defined the density and operator for the value of the frequency of one gene and found the adapted equations as a follow-up for the frequency of alleles and applied this adapted equations to several diploid model and it also applied to actual examples [1].

In this paper, we find adapted partial equations for the model of selection against recessive homozygotes and in case that the alley frequency changes after one generation of selection when there is no dominance. Also we consider the triploid with three alleles and determine whether six genotypes observed are in Hardy-Weinburg for equilibrium.

2. The Main Results

Let $p(t)$ be the frequency of I^A-gene 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.

We begin with the probability about the frequency of allele;

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 respectively for frequence of alleles.

Proof. This result follows easily from the property of dyploid model and the result of M. Kimura [3]. □

The selection coefficient is the relative intensity of selection against a genotype [4]. We usually note of selection for a special genotype. When selection is for one genotype, selection is automatically against at least one other genotype.
The related variable is the fitness. 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 [4,5]. The fitness is the reproductive success of one genotype compared with the reproductive success of other genotypes in the population.

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^AI^A), $2pq$ (the frequency of I^AI^B) and q^2 (the frequency of I^BI^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 [4,5].

Firstly, we consider the model for recessive homozygotes. Recessive homozygotes may have considerably reduced fitness. Selection will occur against the recessive homozygotes. Then we have;

Theorem 2. *Consider the model for selection against recessive homozygotes with selection coefficient α. The adapted equation*

$$\frac{\partial x}{\partial t} = \frac{pq}{4N}\frac{\partial^2 x}{\partial p^2} + \frac{\alpha pq^2}{1-\alpha q^2}\frac{\partial x}{\partial p}$$

for frequence of alleles I^A and

$$\frac{\partial x}{\partial t} = \frac{pq}{4N}\frac{\partial^2 x}{\partial p^2} - \frac{\alpha pq^2}{1-\alpha q^2}\frac{\partial x}{\partial p}$$

for frequence of alleles I^B.

Proof. The fitness for each genotype I^AI^A, I^AI^B and I^BI^B are $1,\ 1,\ 1-\alpha$ respectively, therefore the average fitness is

$$p^2 + 2pq + q^2(1-\alpha) = 1 - \alpha q^2.$$

The frequency of the I^A allele after one generation of selection is calculated by adding the frequency of the I^AI^A genotype and half of the frequency of the I^AI^B genotype. Therefore we have

$$\frac{p^2 + pq}{1 - \alpha q^2} = \frac{p}{1 - \alpha q^2}$$

and the mean change is

$$\frac{p}{\alpha q^2} - p = \frac{\alpha p q^2}{1 - \alpha q^2}$$

we have the first result from Lemma 1.

The frequency of the I^B allele after one generation of selection is calculated by adding the frequency of the I^BI^B genotype and half of the frequency of the I^AI^B genotype. Therefore we have

$$\frac{pq + q^2(1-\alpha)}{1 - \alpha q^2} = \frac{q - \alpha q^2}{1 - \alpha q^2}$$

and the mean change is

$$\frac{q - \alpha q^2}{1 - \alpha q^2} - q = \frac{-\alpha p q^2}{1 - \alpha q^2}$$

we have the second result from Lemma 1. □

Example 1. Suppose $N_{AA} = 100,\ N_{AB} = 50$ and $N_{BB} = 20$ which means the number of zygotes in one generation. The allelic frequencies was calculated from either the numbers or the frequencies of the genotypes. In this case, we have with round off to the proper digit

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

Suppose that the average number of viable offspring produced by three genotype I^AI^A, I^AI^B and I^BI^B are $40, 40$ and 10, respectively. This means the average number of progeny per individual in next generation. We find the fitness for each genotype as following;

$$\begin{aligned} \text{the fitness of } I^AI^A &= \frac{40}{40} = 1 \\ \text{the fitness of } I^AI^B &= \frac{40}{40} = 1 \\ \text{the fitness of } I^BI^B &= \frac{10}{40} = 0.25 \end{aligned}$$

and this model is obviously for selection against recessive homozygotes with selection coefficient $\alpha = 0.75$ in Theorem 2. Therefore from Theorem 2, we have adapted equations

$$\frac{\partial x}{\partial t} = 0.00027\frac{\partial^2 x}{\partial p^2} + 0.04821\frac{\partial x}{\partial p}$$

for frequence of alleles I^A and

$$\frac{\partial x}{\partial t}=0.00027\frac{\partial^2 x}{\partial p^2}-0.04821\frac{\partial x}{\partial p}$$

for frequence of alleles I^B. Here the coefficients were rounded to five decimal places.

Assume that the fitness of the heterozygotes is between the fitnesses of the two homozygotes. Consider that the selection coefficient against the heterozygotes is exactly half the selection coefficient against the disfavored homozygotes.

Then we have;

THEOREM 3. *Assume that the allele frequency changes after one generation of selection when there is no dominance. The adapted equations*

$$\frac{\partial x}{\partial t}=\frac{pq}{4N}\frac{\partial^2 x}{\partial p^2}+\frac{pq}{2}\cdot\frac{2\alpha-p}{1-\alpha q}\frac{\partial x}{\partial p}$$

for frequence of alleles I^A *and*

$$\frac{\partial x}{\partial t}=\frac{pq}{4N}\frac{\partial^2 x}{\partial p^2}+\frac{1}{2}\alpha pq\frac{p}{1-\alpha q}\frac{\partial x}{\partial p}$$

for frequence of alleles I^B.

Proof. The fitness for each genotype I^AI^A, I^AI^B and I^BI^B are $1,\ 1-\alpha/2,\ 1-\alpha$ respectively, therefore the average fitness is

$$p^2+2pq(1-\alpha/2)+q^2(1-\alpha)=1-\alpha q.$$

The frequency of the I^A allele after one generation of selection is calculated by adding the frequency of the I^AI^A genotype and half of the frequency of the I^AI^B genotype. Therefore we have

$$\frac{p^2+pq(1-\alpha/2)}{1-\alpha q}=\frac{p}{2}\cdot\frac{2-\alpha q}{1-\alpha q}$$

and the mean change is

$$\frac{p}{2}\cdot\frac{2-\alpha q}{1-\alpha q}-p=\frac{pq}{2}\cdot\frac{2\alpha-p}{1-\alpha q}$$

we have the first result from Lemma 1.
The frequency of the I^B allele after one generation of selection is calculated by adding the frequency of the I^BI^B genotype and half of the frequency of the I^AI^B genotype. Therefore the mean change is

$$\frac{pq(1-\alpha/2)+q^2(1-\alpha)}{1-\alpha q}-q=\frac{1}{2}\alpha pq\frac{p}{1-\alpha q}$$

we have the second result from Lemma 1. □

Consider the triploid with three alleles. If there are three genes A_1, A_2 and A_3 with frequencies p, q and r, respectively, then the genotype divided into six groups and frequencies of phenotype A_1, A_2, A_1A_2 and A_3 are $p^2+2pr, q^2+2qr, 2pq$ and r^2, respectively. With three genes A_1, A_2 and A_3, six genotypes are possible, of which three are homozygotes and three are heterozygotes. In general, given n genes or

n alleles, there are $\frac{n(n+1)}{2}$ genotypes, of which n are homozygotes and $\frac{n(n-1)}{2}$ are heterozygotes.

Example 2. Consider the ABO blood group. There are determined by three alleles I^A, I^B and i. To calculate a genotype frequency, we simply add up the number of individuals possessing a genotype and divide by the total number of individuals in the population. For example

$$f(I^AI^A) = \frac{\text{number of } I^AI^A \text{ individuals}}{N}.$$

Assume that in a certain population the frequencies of the four blood groups are

$$\begin{aligned}&\text{phenotype } A \text{ (genotypes } I^AI^A,\ I^Ai) = 0.4\\ &\text{phenotype } B \text{ (genotypes } I^BI^B,\ I^Bi) = 0.2\\ &\text{phenotype } AB \text{ (genotype } I^AI^B) = 0.1\\ &\text{phenotype } O \text{ (genotype } ii) = 0.3.\end{aligned}$$

Suppose that the frequencies of I^A, I^B and i is p, q and r, respectively. This means for example, the frequency of the genotype ii is r^2 and therefore about $r = 0.55$ with round off to the proper digit, which is the frequency of the i allele. The joint frequency of phenotypoes B and O is $(q + r)^2 = 0.5$ and therefore the frequency of allele I^B is about $q = 0.16$. The frequency of allele I^A is available in a similar way and so, $p = 0.29$ about. Multiplying each of these expected genotypic frequencies by the total number of observed genotypes in the population, we can get the numbers expected for each genotypes.

If $N = 100$, then we have the number expected for genotype

$$\begin{aligned}I^AI^A &= p^2 \times 100 = 8.41,\ I^Ai = 2pr \times 100 = 31.9\\ I^AI^B &= 2pq \times 100 = 9.28,\ I^Bi = 2qr \times 100 = 17.6\\ I^BI^B &= q^2 \times 100 = 2.56,\ ii = r^2 \times 100 = 30.25.\end{aligned}$$

We can determine the value of chi-square whether the differences between the observed and the expected numbers of each genotypes are due to chance.

$$\begin{aligned}\sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} &= \frac{(20-8.41)^2}{8.41} + \frac{(20-31.9)^2}{31.9} + \frac{(10-9.28)^2}{9.28}\\ &\quad + \frac{(10-17.6)^2}{17.6} + \frac{(10-2.56)^2}{2.56} + \frac{(30-30.25)^2}{30.25}.\end{aligned}$$

The value of chi-square is about 45.37 and the degree of freedom for Hardy-Weinberg equilibrium is the number of expected genotypes classes minus the number of associated alleles. The chi-square value with 3 degree of freedom has very small probability. Therefore the observed values differ from the expected value and genotypes observed are not likely to be in Hardy-Weinberg proportions.

References

[1] W. Choi, *On the adapted equations for several dyploid model in population genetics*, Korean J. Math. **30** (1) (2022), 62–72.

[2] W. Choi, *On the genotype frequencies and generating function for frequencies in a dyploid model*, Korean J. Math. **29** (1) (2021), 75–80.
[3] M. Kimura, *A Stochastic model of Compensatory Neutral Evolution* , Proceedings of a Workshop held in Nagoya, July 8-12, Stochastic Methods in Biology (1985).
[4] R. Lewis, *Human Genetics : Concepts and Applications*, McGraw-Hill Education (2016).
[5] 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 22012, Korea
E-mail: `choiwon@inu.ac.kr`