Korean J. Math. **29** (2021), No. 1, pp. 75–80
http://dx.doi.org/10.11568/kjm.2021.29.1.75

ON THE GENOTYPE FREQUENCIES AND GENERATING FUNCTION FOR FREQUENCIES IN A DYPLOID 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. So in this paper, we consider the method of whether some genotypes is in Hardy-Weinburg equilibrium. Also we calculate the probability generating function for the offspring number of genotype produced by a mating of the ith male and jth female under a diploid model of N population with N_1 males and N_2 females. Finally, we have conditional joint probability generating function of genotype frequencies.

1. Introduction

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.

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 $I^AI^A,\ I^AI^B$ 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([3]);

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

Received November 21, 2020. Accepted January 12, 2021. Published online March 30, 2021.
2010 Mathematics Subject Classification: 92D10, 60H30, 60G44.
Key words and phrases: genotype frequencies, probability generating function, allele.
This research was supported by Incheon National University Research Grant, 2020-2021.

For example, suppose $N_{AA} = 100,\ N_{AB} = 150$ and $N_{BB} = 50$. Then we have

$$f(I^AI^A) = 0.333,\ f(I^AI^B) = 0.5,\ f(I^BI^B) = 0.1667$$

with round off to the proper digit. 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;

$$p = f(I^A) = \frac{2N_{AA} + N_{AB}}{2N} = 0.583,\ q = f(I^B) = \frac{2N_{BB} + N_{AB}}{2N} = 0.4167$$

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

$$p = f(I^A) = f(I^AI^A) + \frac{1}{2}f(I^AI^B) = 0.583$$
$$q = f(I^B) = f(I^BI^B) + \frac{1}{2}f(I^AI^B) = 0.4167$$

The frequencies of the genotypes expected are calculated by using $p^2,\ 2pq$ and q^2. So in this paper, we consider the method of whether some genotypes is in these probabalities. Also we calculate the probability generating function for the offspring number of genotype under a diploid model of N population with N_1 males and N_2 females. Finally, we have conditional joint probability generating function of genotype frequencies.

2. Main Results

If a population is large, randomly mating and not affected by mutation, migration or natural selection, then the allelic frequencies of a population do not change and the genotypic frequencies will not change after one generation in the proportion of 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 genotypes are in the expected proportions of $p^2,\ 2pq,\ q^2$, the population is said to be in Hardy-Weinburg equilibrium([2], [3]).
The Hardy-Weinburg law indicated that when its conditions are satisfied, reproduction alone does not alter allelic or genotypic frequencies and the allelic frequencies determine the frequencies of genotypes.

The frequencies of the genotypes expected under Hardy-Weinberg equilibrium are calculated by using $p^2,\ 2pq$ and q^2.

Example 1. In the example of Introduction, the frequencies of the genotypes expected under Hardy-Weinberg equilibrium are

$$f(I^AI^A) = p^2 = 0.339889$$
$$f(I^AI^B) = 2pq = 0.4858722$$
$$f(I^BI^B) = q^2 = 0.17438976.$$

Multiplying each of these expected genotypic frequencies by the total number of observed genotypes in the population, we get the numbers expected for each genotypes;

$$I^AI^A = 0.339889 \times 300 = 101.9667$$
$$I^AI^B = 0.4858722 \times 300 = 145.76166$$
$$I^BI^B = 0.17438976 \times 300 = 52.316928.$$

We determine whether the differences between the observed and the expected numbers of each genotypes are due to chance.

$$\sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$
$$= \frac{(100 - 101.9667)^2}{101.9667} + \frac{(150 - 145.76166)^2}{145.76166} + \frac{(50 - 51.316928)^2}{51.316928}$$
$$= 0.0379330594 + 0.1232390325 + 0.033795853 = 0.1949679449.$$

The value of chi-square is about 0.19 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 1 degree of freedom has very large probability. Therefore the observed values differ from the expected value and genotypes observed are likely to be in Hardy-Weinberg proportions.

Example 2. The fitness is defined as the relative reproductive success of a genotype in case of natural selection([2], [3]). We fine the fitness of Example 1 for each genotype as following;

$$\text{the fitness of } I^AI^A = \frac{100}{150} = 0.6667$$
$$\text{the fitness of } I^AI^B = \frac{150}{150} = 1$$
$$\text{the fitness of } I^BI^B = \frac{50}{150} = 0.3333.$$

The selection coefficient is the relative intensity of selection against a genotype. Therefore the selection coefficients of I^AI^A, I^AI^B and I^BI^B are 0.3333, 0 and 0.6667, respectively.

Let us define a genotype by $w = (x; y)$. Denote $p_n(x)$ by the probability that in the n-th generation male (or female) individual transmits the gene x without any mutation or gene conversion.

We begin with the following Lemma;

Lemma 1. *The probability $p_n(x)$ is the same for all n and forms a stationary state in the filial generation.*

Proof. In case neither mutation nor gene conversion, mutation or gene conversion rate is 1 for all $k = 1, 2, \cdots, r$. Therefore we have

$$p_{n+1}(x) = p_n(x) \sum_{k=1}^{r} q_{kx} p_n(k) = p_n(x) \sum_{k=1}^{r} p_n(k) = p_n(x)$$

for $n = 0, 1, 2, \cdots$ and $x = 1, 2, \cdots, r$ where q_{ij} is mutation or gene conversion rate from a partition i to another partition j([1]). This formula means that

$$p_n(x) = p_0(x)$$

and the probability $p_n(x)$ without mutation or gene conversion is the same for all n. See ([1]) for the detail proof. □

We consider a diploid model of N population with N_1 males and N_2 females. The alleles will be represented by I^A and I^B. At time n the individuals are distributed among the genotypes I^AI^A, I^AI^B and I^BI^B with $N_n^{(1)}, N_n^{(2)}, \cdots, N_n^{(6)}$, respectively. For example, $N_n^{(2)} = N_1 - N_n^{(1)} - N_n^{(3)}$ and $N_n^{(5)} = N_2 - N_n^{(4)} - N_n^{(6)}$.
Also, we consider the indepedent and identically distributed random variables $X_{ij}(i = 1, 2, \cdots, N_1; j = 1, 2, \cdots, N_2)$, the number of offsprings produced by a mating of the ith male and jth female. Let the probability generating function(p.g.f.) for each X_{ij} be $P(z)$ and let $p_{ij}^{(w)}$ be the probability that an offspring of a mating of a male of genotype x and a female of genotype y which is of w; $x = 1, 2, 3$; $y = 1, 2, 3$; $w = 1, 2, \cdots, 6$.

Then we have;

THEOREM 2. *Suppose that at $(n+1)$th generation, there are in all N_1N_2 matings which take the kth male and the lth female ($k = 1, 2, \cdots, N_1$; $l = 1, 2, \cdots, N_2$). Let $G_{kl}^{(w)}$ be the number of offsprings of type w where $w = 1, 2, \cdots, 6$ in the dyploid model and q_{kl} be the probability that a mating produces X_{kl} offspring.*
The p.g.f. of $G_{kl}^{(w)}$ is

$$\prod_{k=1}^{N_1}\prod_{l=1}^{N_2}\{\sum_{G_{kl}^{(w)}}\sum_{X_{kl}}\binom{X_{kl}}{G_{kl}^{(1)}, G_{kl}^{(2)}, G_{kl}^{(3)}, G_{kl}^{(4)}, G_{kl}^{(5)}, G_{kl}^{(6)}}[\prod_{w=1}^{6}(p_{kl}^{(w)}r_{kl}^{(w)})^{G_{kl}^{(w)}}]q_{kl}\}$$

Proof. Obviously X_{kl} are independent, $X_{kl} = \sum_{w=1}^{6} G_{kl}^{(w)}$ and

$$P(G_{kl}^{(w)}|X_{kl}) = \binom{X_{kl}}{G_{kl}^{(1)}, G_{kl}^{(2)}, G_{kl}^{(3)}, G_{kl}^{(4)}, G_{kl}^{(5)}, G_{kl}^{(6)}}\prod_{w=1}^{6}(p_{kl}^{(w)})^{G_{kl}^{(w)}}$$

for $w = 1, 2, \cdots, 6$. From the total probability theorem, we have

$$P(G_{kl}^{(w)}|X_{kl}) = \prod_{k=1}^{N_1}\prod_{l=1}^{N_2}\{\sum_{X_{kl}}P(G_{kl}^{(w)}|X_{kl})P(X_{kl})\}$$

for $w = 1, 2, \cdots, 6;\ k = 1, 2, \cdots, N_1;\ l = 1, 2, \cdots, N_2$. The p.g.f. of $G_{kl}^{(w)}$ is

$$E\{\prod_{k=1}^{N_1}\prod_{l=1}^{N_2}\prod_{w=1}^{6} {r_{kl}^{(w)}}^{G_{kl}^{(w)}}\}$$
$$= \prod_{k=1}^{N_1}\prod_{l=1}^{N_2}\{\sum_{G_{kl}^{(w)}}[(\prod_{w=1}^{6} {r_{kl}^{(w)}}^{G_{kl}^{(w)}})\sum_{X_{kl}} P(G_{kl}^{(w)}|X_{kl})P(X_{kl})]\}$$
$$= \prod_{k=1}^{N_1}\prod_{l=1}^{N_2}\{\sum_{G_{kl}^{(w)}}\sum_{X_{kl}}\binom{X_{kl}}{G_{kl}^{(1)}, G_{kl}^{(2)}, G_{kl}^{(3)}, G_{kl}^{(4)}, G_{kl}^{(5)}, G_{kl}^{(6)}}[\prod_{w=1}^{6}(p_{kl}^{(w)} r_{kl}^{(w)})^{G_{kl}^{(w)}}]q_{kl}\}.$$

□

Suppose that the generation $n+1$ has N_1 males and $N_2 = N - N_1$ females. If the parbability of an offspring being male is p_1 and female p_2, the probability of having N_1 males and N_2 females from $N_1 N_2$ matings of parents in the ith generation is

$$\text{thecoefficient of } z_1^{N_1} z_2^{N_2} \text{ in } [P(p_1 z_1 + p_2 z_2)]^{N_1 N_2}.$$

Then we meet with;

THEOREM 3. *The conditional expection of*

$$\prod_{w=1}^{6} {r^{(w)}}^{N_{n+1}^{(w)}},$$

given $N_n^{(1)}, N_n^{(2)}, \cdots, N_n^{(6)}$ *is the product of*

$$\{\text{the coefficient of } z_1^{N_1} z_2^{N_2} \text{ in } [P(p_1 z_1 + p_2 z_2)]^{N_1 N_2}\}^{-1}$$

and the coefficient of $z_1^{N_1} z_2^{N_2}$ *in*

$$\prod_{i=1}^{3}\prod_{j=4}^{6}[P\{z_1(\sum_{w=1}^{3} p_{ij}^{(w)} r^{(w)}) + z_2(\sum_{w=4}^{6} p_{ij}^{(w)} r^{(w)})\}]^{N_n^{(i)} N_n^{(j)}}.$$

Proof. Theorem 2 can be represented by

$$\prod_{k=1}^{N_1}\prod_{l=1}^{N_2}\{\sum_{X_{kl}}(p_{kl}^{(1)} r_{kl}^{(1)} + p_{kl}^{(2)} r_{kl}^{(2)} + \cdots + p_{kl}^{(6)} r_{kl})^{(6)})^{X_{kl}} q_{kl}\}$$
$$\prod_{k=1}^{N_1}\prod_{l=1}^{N_2} P(p_{kl}^{(1)} r_{kl}^{(1)} + p_{kl}^{(2)} r_{kl}^{(2)} + \cdots + p_{kl}^{(6)} r_{kl}^{(6)}).$$

We know that there are $N_n^{(i)} N_n^{(j)}$ matings of a male of genotype i and a female of genotype j $(i, j = 1, 2, 3)$. Therefore the population at generation $n+1$ is

$$N_{n+1}^{(w)} = \sum_{k=1}^{N_1}\sum_{l=1}^{N_2} G_{kl}^{(w)},\ w = 1, 2, \cdots, 6$$

and we have

$$E[\prod_{w=1}^{6} r^{(w)N_{n+1}^{(w)}}|N_n^{(1)}, N_n^{(2)}, \cdots, N_n^{(6)}] = \prod_{i=1}^{3}\prod_{j=4}^{6}[P(p_{ij}^{(1)}r^{(1)}+p_{ij}^{(2)}r^{(2)}+\cdots+p_{ij}^{(6)}r^{(6)})]^{N_n^{(i)}N_n^{(j)}}$$

Therefore we get the joint probability generating function for $N_{n+1}^{(1)}, N_{n+1}^{(2)}, \cdots, N_{n+1}^{(6)}$ conditional on $N_1 = N_{n+1}^{(1)} + N_{n+1}^{(2)} + N_{n+1}^{(3)}$ and $N_2 = N_{n+1}^{(4)} + N_{n+1}^{(5)} + N_{n+1}^{(6)}$ and we have the result, given the population at the ith generation. □

REMARK. In Theorem 2 and Theorem 3, $p_{ij}^{(w)}$ can be represented by $p_n(x)$. By Lemma 1, $p_1(A) = p_4(A),\ p_2(A) = p_5(A)$ and $p_1(B) = p_4(B),\ p_2(B) = p_5(B)$. So in case of a model allowing mutation, we have

$$p_{11}^{(1)} = p_1^2(A)p_k = (1-\gamma_1)^2 p_k$$

where γ_1 is the probability of a mutation from A to B.

References

[1] W. Choi, *On the probability of genotypes in population genetics* , Korean J. Mathematics **28** (1) (2020).

[2] R. Lewis, *Human Genetics : Concepts and Applications*, McGraw-Hill Education (2016).

[3] 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 406-772, Republic of Korea
E-mail: choiwon@inu.ac.kr