Korean J. Math. **28** (2020), No. 1, pp. 1–7
http://dx.doi.org/10.11568/kjm.2020.28.1.1

ON THE PROBABILITY OF GENOTYPES IN POPULATION GENETICS

Won Choi

Abstract. A partition X describes that there exists α_i kinds of alleles occurring i loci for each i. All genes have multiple alleles, i.e., they exist in more than two allelic forms, although any one diploid organism can carry no more than two alleles. The number of possible genotypes in a multiple allel series depends on the number of alleles. We will deal with an n locus model in which mutation and gene conversion are taken into consideration. In this paper, we firstly find the probability $p_n(x)$ of genotype

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

with the rates of mutation and gene conversion. Also we find the probability of genotype without the rates of mutation and gene conversion and we apply this probability to two examples.

1. Introduction

Consider a partition to be a sequence

$$X = (x_1, x_2, \cdots, x_d) \in R^d.$$

Received August 13, 2019. Revised September 11, 2019. Accepted September 17, 2019.

2010 Mathematics Subject Classification: 92D10, 60H30, 60G44.

Key words and phrases: genotype, mutation and gene conversion rate. allele.

This research was supported by Incheon National University Research Grant, 2019-2020.

If the partition X has α_i parts equal to i, then we write

$$X = [1^{\alpha_1}, 2^{\alpha_2}, \cdots, n^{\alpha_n}].$$

We consider n locus model, so we find n genes on a chromosome. A partition X describes a state of a chromosome and X means that there exist d kinds of alleles which occupy x_1 loci, x_2 loci, $\cdots$, x_d loci. In other words, X describes that there exists α_i kinds of alleles occurring i loci for each i.
The concept of mutation follows the infinite allele model proposed by Kimura and Crow([3]) and the concept of gene conversion was introduced by Ohta([4]). The mutation and gene conversion was formulated by Shimizu ([5]). If a gene of the alleles occupying i loci is changed by mutation, the partition X is replaced by

$$X = [1^{\alpha_1+1}, 2^{\alpha_2}, \cdots, (i-1)^{\alpha_{i-1}+1}, \cdots n^{\alpha_n}].$$

If a gene conversion occurs between a gene of the alleles occupying i_1 loci and a gene of the alleles occupying i_2 loci, a partition X is changed to

$$X = [1^{\alpha_1}, 2^{\alpha_2}, \cdots, (i_1-1)^{\alpha_{i_1-1}+1}, i_1^{\alpha_{i_1}-1}, \cdots, i_2^{\alpha_{i_2}-1}, (i_2+1)^{\alpha_{i_2+1}+1}, n^{\alpha_n}].$$

Let q_{ij} denote "mutation rate" or "gene conversion rate" from a partition X_i to another partition X_j per generation measured on the t time scale and p_i denotes the frequency of type X_i.

Each individual is characterized by two elements called genes which represent two possible alternatives. However, many and possibly all genes have multiple alleles, i.e., they exist in more than two allelic forms, although any one diploid organism can carry no more than two alleles. The number of possible genotypes in a multiple allel series depends on the number of alleles. With one allele denoted by A, only one genotype is possible, AA. With two alleles denoted by A_1, A_2, three genotypes are possible; two homozygotes A_1A_1 and A_2A_2 and one heterozygote A_1A_2. With three alleles denoted by A_1, A_2, A_3, six genotypes are possible; three homozygotes A_1A_1, A_2A_2, A_3A_3 and three heterozygotes A_1A_2, A_1A_3, A_2A_3. In general, given n alleles, there are $n(n+1)/2$ genotypes, of which n are homozygotes and $n(n-1)/2$ are heterozygotes([1]). Since the martingale problem for diffusion operator is related to diffusion processes, W. Choi applied diffusion processes for countable-allelic model which is represented as the quotient of q_{ij} in population genetic model and he can define a new diffusion operator([2]).

We will deal with an n locus model in which mutation and gene conversion are taken into consideration. Also random partitions of the number n determined by chromosomes with n loci should be investigated. In this paper, we firstly describe the probability of genotype with the rates of mutation and gene conversion. Also we find the probability of genotype without the rates of mutation and gene conversion and we apply this probability to two examples.

2. Main Results

A finite sequence $\{X_1, X_2, \cdots, X_K\}$ of partitions is called (X_1, X_K)-*chain* if X_{i+1} is a consequent of X_i by mutation or gene conversion for each $i = 1, 2, \cdots, K-1$.

We begin with the following Theorem.

THEOREM 1.

$$\rho = \left(\frac{q_{12}}{q_{21}}\right)\left(\frac{q_{23}}{q_{32}}\right)\cdots\left(\frac{q_{K-1\ K}}{q_{K\ K-1}}\right) = 1.$$

In other word, the value ρ does not depend on the choice of (X_1, X_K)-chain.

Proof. For simplicity, we consider only a case $K = 3$ and assume X_1 can be changed to X_2 and X_3 by mutation (i.e. X_1 is a consequent of X_2 and X_3 by gene conversion), besides X_2 and X_3 can be changed to each other by gene conversion. Putting

$$\begin{aligned}
X_1 &= [1^{\alpha_1}, \cdots, i^{\alpha_i}, \cdots, j^{\alpha_j}, \cdots, n^{\alpha_n}],\\
X_2 &= [1^{\alpha_1+1}, \cdots, (i-1)^{\alpha_{i-1}+1}, i^{\alpha_i-1}, \cdots, j^{\alpha_j}, \cdots, n^{\alpha_n}],\\
X_3 &= [1^{\alpha_1+1}, \cdots, i^{\alpha_i}, \cdots, (j-1)^{\alpha_{j-1}+1}, j^{\alpha_j-1}, \cdots, n^{\alpha_n}],
\end{aligned}$$

we have

$$\begin{aligned}
q_{12} &= i\alpha_i,\\
q_{21} &= (\alpha_1+1)(i-1)(\alpha_{i-1}+1),\\
q_{23} &= (i-1)(\alpha_{i-1}+1)j\alpha_j,\\
q_{32} &= i\alpha_i(j-1)(\alpha_{j-1}+1),\\
q_{31} &= (\alpha_1+1)(j-1)(\alpha_{j-1}+1),\\
q_{13} &= j\alpha_j.
\end{aligned}$$

Obviously, $\rho = 1$. □

Let us define a genotype by $(x;y)$, $(x = 1, 2, \cdots, r; y = 1, 2, \cdots, r)$. Denote by $b_n(x;y)$ the probability of the genotype $(x;y)$ which occurs mutation or gene conversion in the n-th generation. Define a_n as

$$a_n(x;x) = b_n(x;x),$$
$$a_n(x;y) = b_n(x;y) + b_n(y;x).$$

for $x = 1, 2, \cdots, r; y = 1, 2, \cdots, r$ and $p_n(x)$ as the probability that in the n-th generation individual object transmits the gene x by the mutation or gene conversion.

THEOREM 2.

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

Proof. Since there are $\frac{r(r+1)}{2}$ genotypes for r alleles, there are $\frac{r(r+1)}{2}$ such probabilities;

$$a_n(1;1),\ a_n(1;2),\ \cdots, a_n(1;r),\ a_n(2;2), \cdots, a_n(2;r), \cdots, a_n(r;r)$$

such that

$$\sum_{x \le y} a_n(x;y) = 1$$

for $n = 0, 1, 2, \cdots$.

Obviously we have

$$\begin{aligned} p_n(x) =& \frac{1}{2} q_{1x} a_n(1;x) + \frac{1}{2} q_{2x} a_n(2;x) + \cdots q_{xx} a_n(x;x) \\ &+ \frac{1}{2} q_{xx+1} a_n(x;x+1) + \cdots + \frac{1}{2} q_{xr} a_n(x;r) \end{aligned}$$

and

$$\sum_{x=1}^{r} p_n(x) = 1.$$

The individuals of genotype $(x;y)$ or $(y;x)$ all posses the gene x and transmit it with probability $\frac{1}{2}$ if $x \neq y$ and with probability 1 if $x = y$. Therefore the probability of the genotype $(x;y)$ in the $(n+1)$-st generation is obviously $p_n(x)p_n(y)$ and we have

$$a_{n+1}(x;x) = [p_n(x)]^2$$
$$a_{n+1}(x;y) = 2p_n(x)p_n(y).$$

a_{n+1} was presented in terms of a_n and hence the recurrence formula is obtained. Since the probability of the genotype $(x;y)$ in the $(n+1)$-st generation is $p_n(x)p_n(y)$, the distribution $b_{n+1}(x;y)$ is independent. Consequently we have

$$\begin{aligned} p_{n+1}(x) =& \frac{1}{2}\cdot 2q_{1x}p_n(1)p_n(x) + \frac{1}{2}\cdot 2q_{2x}p_n(2)p_n(x) + \cdots q_{xx}p_n(x)p_n(x) \\ &+ \frac{1}{2}\cdot 2q_{xx+1}p_n(x)p_n(x+1) + \cdots + \frac{1}{2}\cdot 2q_{xr}p_n(x)p_n(r) \end{aligned}$$

and

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

□

Corollary 3. *The probability $p_n(x)$ without mutation or gene conversion is the same for all n and stationary state in the filial generation.*

Proof. In case neither mutation nor gene conversion, we have $q_{kx} = 1$ 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 $x = 1, 2, \cdots, r$. 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. □

In case of m characters a genotype is described by $2m$ numbers

$$(x_1, x_2, \cdots, x_m; y_1, y_2, \cdots, y_m).$$

With respect to m characters the genotype of an individual is characterized by m pairs of number. Two individuals are of the same type if each of the m characters corresponds the same pair. Hence there are

$$\left(\frac{r(r+1)}{2}\right)^m$$

genotypes. In the formulation of a new individual a parent of genotype

$$(x_1x_2, \cdots, x_m; y_1, y_2, \cdots, y_m)$$

transmits to the offspring, corresponding to each of the m characters, one of the two genes which the individual posseses with respect to this character. The probability of transmitting any of these combinations is the same and therefore equal to $1/2^m$.

In case of individual $(1,2,3;1,5,9)$, the pair $1,1$ corresponds to the first character the pair $2,5$ to the second and $3,9$ to the third. Under the assumptions of Mendel's theory this individual is of the same genotype with $(1,5,3;1,2,9)$ and $(1,2,9;1,5,3)$, and of course with $(1,5,9;1,2,3)$, etc. In case of $m=3$, it may transmit eight combinations which in the proceeding example reduce to four, because the individual is homozygous in the first character. These four combinations are $1,2,3$ or $1,5,3$ or $1,2,9$ or $1,5,9$ each with probability $\frac{1}{4}$.

EXAMPLE 1. Consider $m=r=2$ without mutatiion or gene conversion and let $p_n(x_1,x_2)$ be the probability that in the n-th generation an individual transmits the genes x_1,x_2. Then we have

$$p_n(1,1)=a_n(1,1;1,1)+\frac{1}{2}a_n(1,1;1,2)+\frac{1}{2}a_n(1,1;2,1)+\frac{1}{2}a_n(1,1;2,2).$$

Since

$$a_{n+1}(x;x)=[p_n(x)]^2$$
$$a_{n+1}(x;y)=2p_n(x)p_n(y),$$

we get recurrence formula

$$p_{n+1}(1,1)=[p_n(1,1)]^2+\frac{1}{2}\cdot 2p_n(1,1)p_n(1,2)+\frac{1}{2}\cdot 2p_n(1,1)p_n(2,1)+\frac{1}{2}\cdot 2p_n(1,1)p_n(2,2).$$

Therefore

$$p_{n+1}(1,1)=p_n(1,1).$$

EXAMPLE 2. Consider a diploid model. The alleles will be represented by a and A and the population follows by random mating with non-overlapping generation. At time t the individuals are distributed among the genotypes aa, Aa and AA. Denote $P_{ij}^{(c)}$ as the probability that an offspring of the $i\times j$ mating is of type c. In this case

$$P_{ij}^{(1)}=P_iP_jP_k,\ P_{ij}^{(3)}=Q_iQ_jP_k$$

where $P_1 = P_4,\ P_2 = P_5$ and $P_3 = P_6$ are the probabilities that a parent of type $aa,\ Aa,\ AA$ transmits the first gene a, while $Q_1 = Q_4,\ Q_2 = Q_5$ and $Q_3 = Q_6$ are the probabilities that a individual transmits the second gene A. Of course $P_i + Q_i = 1$ for $i = 1, 2, \cdots, 6$

If it is allowed mutation or gene conversion, we get

$$P_{11}^{(1)} = (1 - q_{11})^2 P_k.$$

References

[1] Francisco J.Ayala, *Population and Evolutionary Genetics; A Primer*, The Benjamin/Cummings Publishing Company, (1982).

[2] W.Choi, *The application of stochastic analysis to population genetics model*, J. App. Math. Info. **23** (2007), 455–460.

[3] M.Kimura and J.F.Crow, *The number of alleles that can be maintained in a finite population*, Genetics **49** (1964).

[4] T.Ohta, *On the evolution of multigene families*, Theor. Por. Biol. **23** (1983) 216–240.

[5] A.Shimizu, *Stationary distribution of a diffusion process taking values in probability distributions on the partitions*, Proceeding of a Workshop held in Nagoya, Japan. (1985).

Won Choi
Department of Mathematics
Incheon National University
Incheon 406-772, Republic of Korea
E-mail: choiwon@inu.ac.kr