Ann. Inst. Statist. Math.
36 (1984), Part B, 353-362

RANDOM COLLISION MODEL FOR RANDOM GENETIC DRIFT AND STOCHASTIC DIFFERENCE EQUATION

YOSHIAKI ITOH

(Received Oct. 28, 1983; revised Jan. 11, 1984)

Summary

At first we introduce a simple stochastic difference equation, to simulate random sampling drift in population genetics, which is naturally obtained from a random collision model. Next, we introduce a random collision model to simulate overdominance model in population genetics. We assume in a time interval Δt, a random collision of four particles, which represents overdominant selection, takes place at a certain probability, where a particle corresponds to a gene. We assume that mutation takes place by some rate and assume that every new mutation is different from extant alleles. We estimate mean heterozygosity by our simulation method and compare it with the result obtained by using a stochastic difference equation for overdominance model.

1. Introduction

Random collision model is useful to simulate various models in population genetics (see Itoh [6]). Recently, the amount of good data from natural population on the genetic variability is increasing, and the demand for analysis giving insight into molecular evolution and genetic structure of population is becoming increasingly important. But the problems which can be treated by analytical methods are restricted. Even to get the numerical result on Fokker-Planck equation is difficult when the dimension is more than two. Hence simulation methods are becoming important.

At first we introduce a symple difference scheme to simulate random sampling drift for genetic studies. A simulation method introduced by Pederson [14] is known and is a good numerical technique. But Pederson's difference equation scheme seems complicated. The method presented here is obtained by an approximate description of random collision model for random genetic drift (Itoh [3], [5], [6]), which is already used by Maruyama and Nei [10], Maruyama and Takahata [11], Takahata [15], Maruyama [9], Nei et al. [12], to discuss genetic variability maintained by mutation and overdominant selection in finite populations and is shown to be convenient. For the simulation studies, it is necessary to decompose a covariance matrix called drift matrix. For general covariance matrix, Cholesky decomposition is usually used. Pederson's method also gives a decomposition of the drift matrix implicitly. For our case an approximation of the random collision model automatically gives a decomposition of the drift matrix as shown in Section 2. There are several models for the studies of random sampling drift for example Wright's [16], Moran's [13] and others'. Our random collision model also seems to be one of the models which are simple and has clear image. In Section 3, we do not use approximation by stochastic differential equation. We simulate over-dominant selection model in population genetics directly by a four-particle random collision model. When we carry out computer simulations by stochastic differential equation we must descritize the time parameter. This makes the trajectory go across the boundary, which is not expected from the stochastic differential equation. Our random collision model for overdominant selection model gives a reasonable behaviour of trajectory at the boundary. We carried out a simulation study by our random collision model. We compared the heterozygosity obtained by our model with the result by Maruyama and Nei [10], and found these two results agree well with each other. Hence our result justifies the result by Maruyama and Nei, as well as the result by Maruyama and Nei justifies our result.

2. Simulation method by stochastic differential equation

In Wright's model it is supposed that each of the genes of the next generation is obtained by a random choice among the genes of the previous generation and that the whole population changes all at once. In Moran's model it is supposed that there are M individuals each formed from m alleles $A_1, A_2, \cdots, A_k$, and that at each instant at which the state of the model may change, one individual of the alleles, chosen at random, dies and is replaced by a new individual which is A_i with probability m_i/M, where m_i is the abundance of the allele A_i. It is supposed that the probability of any individual "dying" during an interval $(t, t+dt)$ and then being replaced by a new individual is λdt. Hence the mean number of such events in unit time is λM and the mean length of a generation is λ^{-1}. The following random collision model is another reasonable model.

Consider a population of M particles each of which is one of k

types, $A_1, A_2, \cdots, A_k$. The types may represent species, alleles, genotypes or other classification. We then consider random collisions between particles, which are assumed to occur at the rate λdt per time interval $(t, t+dt)$ for each particle. If a pair of particles of different types i and j collide, then after the collision the both particles are the type i with probability 1/2 and the type j with probability 1/2. If the type of the colliding particles are the same, no change occurs. In this random collision model two particles are chosen by random sampling without replacement at first, and from the two particles, two particles are chosen by random sampling with replacement.

We can approximate our random collision model by a stochastic difference equation (1). In it, the relative abundance of type i increase by $c\sqrt{x_i(t)x_j(t)}\,\Delta B_{ij}(t)$ by the interaction with j which results the decrease of the type j by $-c\sqrt{x_i(t)x_j(t)}\,\Delta B_{ij}(t)$, where $c=\sqrt{\lambda/M}$. Hence our random collision model automatically lead to the following equation (1), which has the drift matrix $c^2\{x_i(t)(\delta_{ij}-x_j(t))\}\Delta t$ as covariances.

For $i, j=1, 2, \cdots, m$, consider

$$\Delta x_i(t)=\sum_{\substack{j=1\\ i\neq j}}^{m} c\sigma_{ij}\sqrt{x_i(t)x_j(t)}\,\Delta B_{ij}(t)\ , \tag{1}$$

where

$$\sigma_{ij}=\begin{cases} 1 & \text{for } i>j \\ -1 & \text{for } i<j \end{cases}$$

$\Delta B_{ij}(t)$ $(i>j)$ are mutually independent one dimensional normal random variable with the mean 0 and the variance Δt. Let

$$x_i(t+\Delta t)=x_i(t)+\Delta x_i(t) \qquad \text{for } i=1, 2, \cdots, m\ .$$

Then this difference scheme represent the random sampling drift of m alleles, $1, 2, \cdots, m$ whose relative abundances at time t are $x_1(t), x_2(t), \cdots, x_m(t)$ respectively.

Cholesky decomposition is usually used for the decomposition of covariance matrix. For our case the method gives the representation

$$\Delta x_i(t)=c\sum_{j=1}^{m} d_{ij}(t)\Delta B_j(t) \tag{2}$$

where $\{d_{ij}(t)\}\{d_{ij}(t)\}^t=\{x_i(t)(\delta_{ij}-x_j(t))\}$ for a lower triangular matrix $\{d_{ij}(t)\}$ and its transpose $\{d_{ij}(t)\}^t$, and $\Delta B_j(t)$, $j=1, 2, \cdots, m$, are mutually independent normal random numbers with the expectation 0 and the variance Δt. This method is applicable to general covariance matrix and does not make use of the speciality of the drift matrix.

Pederson [14] gave a representation in which $x_i(t+\Delta t)$ is constructed from $x_i(t+\Delta t)$ for $j=1,2,\cdots,i-1$, and $x_j(x)$ for $j=1,2,\cdots,i$, as

$$x_i(t+\Delta t)=\frac{\left(1-\sum_{j=1}^{i-1} x_j(t+\Delta t)\right)x_i(t)}{1-\sum_{j=1}^{i-1} x_j(t)}+c\Delta B_i(t)\Bigg[x_i(t)(1-x_i(t)) - \left(\frac{x_i(t)}{1-\sum_{j=1}^{i-1} x_j(t)}\right)^2 \sum_{j,k=1}^{i-1} x_j(t)(\delta_{jk}-x_k(t))\Bigg]^{1/2}, \tag{3}$$

where $\Delta B_i(t)$, $i=1,2,\cdots,m$, are mutually independent normal random numbers.

Our method require ${}_mC_2$ mutually independent normal random numbers for each step, while by the above two methods mutually independent m normal random numbers are sufficient. But the equation (1) is very simple and the decomposition is explicitly given in it. In our representation, the decomposition is given by

$$\{x_i(t)(\delta_{ij}-x_j(t))\}=LL^t$$

by $m\times{}_mC_2$ matrix L in which a column vector with $\sqrt{x_ix_j}$, $(i>j)$ as ith component, $-\sqrt{x_ix_j}$ as jth component and 0 as the other component is arranged as the $\left(i-j+\sum_{k=0}^{i-1} k\right)$th column.

3. Simulation method by random collision model

We discuss the case in which mutation is absent. In the previous papers, the author introduced a three-particle random collision model for population biology (Itoh [4], [7]). Here we introduce four-particle random collision model to simulate overdominant selection model in population genetics.

Consider a random mating population of effective size N, and assume that selection and mutation occur deterministically and that, after selection and mutation, $2N$ gametes are randomly chosen for the next generation. If we assume that the fitness of heterozygotes is 1 for all pairs of alleles and $1-s$ for all homozygotes, and that every new mutation is different from the extent alleles (infinite allele model by Wright [17], and Kimura and Crow [8]). Then we have

$$\begin{aligned} &\mathrm{E}(\Delta x_i(t))=2Nx_i(t)\{-v+s(J-x_i(t))/(1-sJ)\}\Delta t\\ &\mathrm{E}(\Delta x_i(t)\Delta x_j(t))=x_i(t)(\delta_{ij}-x_j(t))(\Delta t)\end{aligned} \tag{4}$$

by an appropriate scaling of time, where v is the mutation rate, $x_i(t)$ is the frequency of allele A_i at time t, $J=\sum_{i=1}^{n} x_i^2(t)$, and N is effective population size. The study of the allele frequency distribution in finite populations for overdominance selection was initiated by Wright [18]. Approximate formulas for the effective number of alleles that can be maintained in finite population are developed by Kimura and Crow [8], Ewens [1], [2] and others. Yokoyama and Nei [19] presented a general formula for the expected heterozygosity for the Wright model of overdominant selection when population size is larger than a certain level. Maruyama and Nei [10] used the equation (1) to study various properties of overdominant selection in a finite population by computer simulations, simulating the stochastic differential equation with expectations and covariances given by equation (4).

In a population there are n particles of m types, $A_1, A_2, \cdots, A_m$. Consider the following four-particle random collision. Four particles are chosen from the population by random sampling without replacement, and let the four particles be A_i, A_j, A_k, and A_l. A_i and A_j from an individual A_iA_j, and A_k and A_l from A_kA_l. The A_iA_j and the A_kA_l collide and produce two A_iA_js with probability $1/2+s_{ij,kl}$ and two A_kA_ls with probability $1/2+s_{kl,ij}$, where $s_{ij,kl}=-s_{kl,ij}$.

$$(5)\qquad s_{ij,kl}=\begin{cases} s/2 & \text{if } i\neq j \text{ and } k=l \\ -s/2 & \text{if } i=j \text{ and } k\neq l \\ 0 & \text{if } i\neq j \text{ and } k\neq l \text{ or } i=j \text{ and } k=l\,, \end{cases}$$

and then the two A_iA_js (or the two A_kA_ls), split into two A_is and two A_js, (or two A_ks and two A_ls). Hence by the above collision A_i, A_j, A_k and A_l become two A_is and two A_js or two A_ks and two A_ls. We assume that a collision takes place in a time interval $[t, t+dt]$, with probability Cdt. Let the array of alleles frequencies be $\vec{X}=(X_1, X_2, \cdots, X_n)$ at time t. We calculate the expectation $\mathrm{E}(\Delta X_i(t))$ and covariance $\mathrm{E}(\Delta X_i(t)\Delta X_j(t))$.

Let $\varepsilon_{\alpha_1\alpha_2\cdots\alpha_i}$ be the number of the elements of the set $\{\alpha_j|\alpha_j=\alpha_i$ for $1\leqq j\leqq i-1\}$, and $f(X_{\alpha_1}, X_{\alpha_2}, \cdots, X_{\alpha_i})=X_{\alpha_1}(X_{\alpha_2}-\varepsilon_{\alpha_1\alpha_2})(X_{\alpha_3}-\varepsilon_{\alpha_1\alpha_2\alpha_3})\cdots(X_{\alpha_i}-\varepsilon_{\alpha_1\alpha_2\cdots\alpha_i})$. For example,

$$f(X_1, X_2, X_2, X_1)=X_1X_2(X_2-1)(X_1-1)\,,$$

$$f(X_1, X_1, X_1, X_3)=X_1(X_1-1)(X_1-2)X_3\,.$$

We define random variables $X_{ij,kl}(\vec{X})$, for which

$$X_{ij,kl}(\vec{X})=-X_{kl,ij}(\vec{X})\,.$$

$$\Pr\left(X_{ij,kl}(\vec{X})=\frac{1}{2}\right)=C\left(\frac{1}{2}+s_{ij,kl}\right)\frac{f(X_i, X_j, X_k, X_l)}{n(n-1)(n-2)(n-3)}\Delta t$$

$$\Pr\left(X_{ij,kl}(\vec{X})=-\frac{1}{2}\right)=C\left(\frac{1}{2}-s_{ij,kl}\right)\frac{f(X_i, X_j, X_k, X_l)}{n(n-1)(n-2)(n-3)}\Delta t$$

$$\Pr\left(X_{ij,kl}(\vec{X})=0\right)=1-C\frac{f(X_i, X_j, X_k, X_l)}{n(n-1)(n-2)(n-3)}\Delta t\ .$$

We assume all pairs $\{X_{ij,kl}, X_{kl,ij}\}$ are mutually independent. To simplify representations we consider an m dimensional linear space, $\left\{\sum_{i,j=1}^{m} x_{ij}A_iA_j \mid x_{ij}\in R,\ i, j=1, 2,\cdots, m\right\}$, over a field R which is generated by linearly independent elements A_iA_j, $i, j=1, 2,\cdots, m$. $(1+2X_{ij,kl})A_iA_j+(1+2X_{kl,ij})A_kA_l$ represents the outcome of a collision of A_iA_j and A_kA_l during $[t, t+\Delta t]$. We define an amount of α, $(\ )_\alpha$, as follows. For each A_iA_j, $i, j=1,\cdots, m$,

$$(A_iA_j)_\alpha=0 \qquad \text{if } \alpha\neq i, j\ ,$$

$$(A_iA_j)_\alpha=1 \qquad \text{if } \alpha=i,\ \alpha\neq j \text{ or } \alpha\neq i,\ \alpha=j\ ,$$

$$(A_iA_j)_\alpha=2 \qquad \text{if } i=j=\alpha\ .$$

For each element of the linear space,

$$\left(\sum_{i,j=1}^{m} x_{ij}A_iA_j\right)_\alpha=\sum_{i,j=1}^{m} x_{ij}(A_iA_j)_\alpha\ .$$

$$\begin{aligned}&(1+2X_{ij,kl})A_iA_j+(1+2X_{kl,ij})A_kA_l-(A_iA_j+A_kA_l)\\&=2X_{ij,kl}A_iA_j-2X_{ij,kl}A_kA_l\ ,\end{aligned}$$

is the outcome from a collision of A_iA_j and A_kA_l during $[t, t+dt]$. Hence, let $\{i, j, k, l\}$ be the set of i, j, k and l,

$$\begin{aligned}\mathrm{E}(\Delta X_\alpha)&=\sum_{\{i,j,k,l\}\ni\alpha}\mathrm{E}(2X_{ij,kl}A_iA_j-2X_{ij,kl}A_kA_l)_\alpha\\&=Cs\Delta t\sum_{i\neq j,k=l}\frac{f(X_i, X_j, X_k, X_l)}{n(n-1)(n-2)(n-3)}\{(A_iA_j)_\alpha-(A_kA_l)_\alpha\}\\&\quad-Cs\Delta t\sum_{i=j,k\neq l}\frac{f(X_i, X_j, X_k, X_l)}{n(n-1)(n-2)(n-3)}\{(A_iA_j)_\alpha-(A_kA_l)_\alpha\}\ .\end{aligned}$$

Making use of

$$\sum_{\substack{\{i,j,k,l\}\ni\alpha\\ i\neq j,k=l}}=\sum_{\substack{i\neq j,k=l\\ \alpha=i\\ \alpha\neq k}}+\sum_{\substack{i\neq j,k=l\\ \alpha=i\\ \alpha=k}}+\sum_{\substack{i\neq j,k=l\\ \alpha=j\\ \alpha\neq k}}+\sum_{\substack{i\neq j,k=l\\ \alpha=j\\ \alpha=k}}+\sum_{\substack{i\neq j,k=l\\ \alpha\neq i,j\\ \alpha=k}}\ ,$$

we have

$$\mathrm{E}(\Delta X_\alpha)=2Cs\Delta t\cdot\frac{2X_\alpha}{n}\left(\frac{\sum_k X_k(X_k-1)-n(X_\alpha-1)}{(n-1)(n-2)}\right)\ .$$

We calculate $\mathrm{E}(\Delta X_\alpha \Delta X_\beta)$, for $\alpha \neq \beta$.

$$\Delta X_\alpha \Delta X_\beta = \sum_{\{i,j,k,l\} \ni \alpha} (2X_{ij,kl}A_iA_j - 2X_{ij,kl}A_kA_l)_\alpha \times \sum_{\{i',j',k',l'\} \ni \beta} (2X_{i'j',k'l'}A_{i'}A_{j'} - 2X_{i'j',k'l'}A_{k'}A_{l'})_\beta \,.$$

Hence we have

$$\begin{aligned}\mathrm{E}(\Delta X_\alpha \Delta X_\beta) = \sum_{\{i,j,k,l\} \ni \alpha,\beta} \mathrm{E}\,\{&2X_{ij,kl}(A_iA_j)_\alpha 2X_{ij,kl}(A_iA_j)_\beta \\ &-2X_{ij,kl}(A_kA_l)_\alpha 2X_{ij,kl}(A_iA_j)_\beta - 2X_{ij,kl}(A_iA_j)_\alpha 2X_{ij,kl}(A_kA_l)_\beta \\ &+2X_{ij,kl}(A_kA_l)_\alpha 2X_{ij,kl}(A_kA_l)_\beta\} \\ = \sum_{\{i,j,k,l\} \ni \alpha,\beta} \{&(A_iA_j)_\alpha(A_iA_j)_\beta - (A_kA_l)_\alpha(A_iA_j)_\beta - (A_iA_j)_\alpha(A_kA_l)_\beta \\ &+(A_kA_l)_\alpha(A_kA_l)_\beta\} C \frac{f(X_i, X_j, X_k, X_l)}{n(n-1)(n-2)(n-3)} \Delta t \,.\end{aligned}$$

We have

$$\begin{aligned}\sum_{\{i,j,k,l\} \ni \alpha,\beta} = &\sum_{\substack{i,j=\alpha \\ k=\alpha, l=\beta}} + \sum_{\substack{i,j=\alpha \\ k=\beta, l=\alpha}} + \sum_{\substack{i=\alpha, j=\beta \\ k,l=\alpha}} + \sum_{\substack{i=\beta, j=\alpha \\ k,l=\alpha}} + \sum_{\substack{i,j=\beta \\ k=\beta, l=\alpha}} \\ &+ \sum_{\substack{i,j=\beta \\ k=\alpha, l=\beta}} + \sum_{\substack{i=\beta, j=\alpha \\ k,l=\beta}} + \sum_{\substack{i=\alpha, j=\beta \\ k,l=\beta}} + \sum_{\substack{i,j=\alpha \\ k,l=\beta}} + \sum_{\substack{i,j=\beta \\ k,l=\alpha}} + \sum_{\substack{i,j=\alpha \\ k=\beta, l\neq\beta}} + \sum_{\substack{i,j=\alpha \\ k\neq\alpha,\beta, j=\beta}} \\ &+ \sum_{\substack{i=\beta, j\neq\alpha,\beta \\ k,l=\alpha}} + \sum_{\substack{i\neq\alpha,\beta, j=\beta \\ k,l=\alpha}} + \sum_{\substack{i,j=\beta \\ k=\alpha, l\neq\beta}} + \sum_{\substack{i,j=\beta \\ k\neq\alpha,\beta, l=\alpha}} + \sum_{\substack{k=\alpha, l\neq\alpha,\beta \\ i,j=\beta}} + \sum_{\substack{k\neq\alpha,\beta, l=\alpha \\ i,j=\beta}} \\ &+ \sum_{\substack{i=\beta, j\neq\alpha,\beta \\ k=\alpha, l\neq\alpha,\beta}} + \sum_{\substack{i=\beta, j\neq\alpha,\beta \\ k\neq\alpha,\beta, l=\alpha}} + \sum_{\substack{i\neq\alpha,\beta, j=\beta \\ k=\alpha, l\neq\alpha,\beta}} + \sum_{\substack{i\neq\alpha,\beta, j=\beta \\ k\neq\alpha,\beta, l=\alpha}} + \sum_{\substack{i=\alpha, j\neq\alpha,\beta \\ k=\beta, l\neq\alpha,\beta}} \\ &+ \sum_{\substack{i=\alpha, j\neq\alpha,\beta \\ k\neq\alpha,\beta, l=\beta}} + \sum_{\substack{i\neq\alpha,\beta, j=\alpha \\ k=\beta, l\neq\alpha,\beta}} + \sum_{\substack{i\neq\alpha,\beta, j=\alpha \\ k\neq\alpha,\beta, l=\beta}} + \sum_{\substack{i=\alpha, j=\beta \\ k,l\neq\alpha,\beta}} + \sum_{\substack{i=\beta, j=\alpha \\ k,l\neq\alpha,\beta}} \\ &+ \sum_{\substack{i,j\neq\alpha,\beta \\ k=\alpha, l=\beta}} + \sum_{\substack{i,j\neq\alpha,\beta \\ k=\beta, l=\alpha}} \,.\end{aligned}$$

From this we have

$$\mathrm{E}(\Delta X_\alpha \Delta X_\beta) = -\frac{4}{n(n-1)} C \Delta t X_\alpha X_\beta \,. \tag{6}$$

$$\Delta X_\alpha \Delta X_\alpha = \sum_{\{i,j,k,l\} \ni \alpha} \{(A_iA_j)_\alpha^2 - 2(A_iA_j)_\alpha(A_kA_l)_\alpha + (A_kA_l)_\alpha^2\} \,.$$

Using

$$\sum_{\{i,j,k,l\} \ni \alpha} = 4 \sum_{\substack{i,j=\alpha \\ k=\alpha, l\neq\alpha}} + 2 \sum_{\substack{i,j=\alpha \\ k,l\neq\alpha}} + 4 \sum_{\substack{i=\alpha \\ j,k,l\neq\alpha}} ,$$

we have

$$\mathrm{E}(\Delta X_\alpha \Delta X_\alpha) = \frac{4}{n(n-1)} C \Delta t X_\alpha (n - X_\alpha) \,. \tag{7}$$

Hence we have approximately the following

$$\mathrm{E}(\Delta x_\alpha)=\frac{4}{n}Csx_\alpha(\sum_k x_k^2-x_\alpha)\Delta t$$

$$\mathrm{E}(\Delta x_\alpha\Delta x_\beta)=\frac{4}{n^2}Cx_\alpha(\delta_{\alpha\beta}-x_\beta)\Delta t$$

where $x_\alpha=X_\alpha/n$, for $\alpha=1,2,\cdots,m$.

4. Simulation for infinite allele case

We assume that every new mutation is different from the extant alleles and in a time unit Δt each of the n alleles is replaced by a mutant with probability $(4v/n)C\Delta t$. Random collisions and mutations take place mutually independently. Hence we have

$$\mathrm{E}(\Delta x_\alpha)=\frac{4}{n}Cx_\alpha\{-v+s(\sum_k x_k^2-x_\alpha)\}\Delta t\ .$$

We choose $C=n^2/4$ and put $n=2N$. We have

$$\mathrm{E}(\Delta x_\alpha(t))=2N\{-v+sx_\alpha(\sum_k x_k^2-x_\alpha)\}\Delta t$$

$$\mathrm{E}(\Delta x_\alpha\Delta x_\beta)=x_\alpha(\delta_{\alpha\beta}-x_\beta)\Delta t$$

where $x_\alpha=X_\alpha/2N$ for $\alpha=1,2,\cdots,m$. The first equation is approximately equivalent with (4) when s is small. The variance caused by mutation is negligible. Hence we can use our random collision model as a simulation method for overdominance model when s is small. We can make the expectation and variance coincide to (4), if we choose, the state-dependent $s_{ij,kl}$,

$$s_{ij,kl}=\begin{cases}\dfrac{s}{2(1-sJ)} & \text{if } i\neq j \text{ and } k=l\\[2ex] -\dfrac{s}{2(1-sJ)} & \text{if } i=j \text{ and } k\neq l\\[2ex] 0 & \text{if } i=j \text{ and } k=l \text{ or } i\neq j \text{ and } k\neq l\,,\end{cases}$$

instead of (5). But here we discuss the case of small s and use the rule given by (5).

A step consists of the following successive two stages. In the first stage a random collision of four particles takes place and in the next stage a mutation takes place with probability $4v$, that is to say, one of n particles is randomly chosen and replaced by a mutant with probability $4v$. We repeat this step one by one and take the time average

$\bar{h}$ of heterozygosity $h=1-\sum x_j^2$. Initially we set all of the n particles are of one type. We take the time average of heterozygosity over the last half duration of the total steps, that is, we take the time average from time $T/2$ to T, to get the average heterozygosity of the stationary state. Our results are compared with the results by Maruyama and Nei [10] in Table 1

Table 1. Comparison of values of heterozygosity, obtained by the two methods, stochastic differential equation and random collision model

$2Ns$	$4Nv$	$\bar{h}$		
		Stochastic differential equation Maruyama and Nei [10]	Random collision model $n=2N=50$ $T=1{,}000{,}000$	Random collision model $n=2N=100$ $T=4{,}000{,}000$
0	0.050	0.0476	0.0507	0.0457
5	0.0075	0.0450	0.0315	0.0295
0	0.100	0.0909	0.0838	0.0904
5	0.0175	0.0930	0.0879	0.1048
0	0.500	0.3333	0.3326	0.3276
5	0.150	0.3377	0.3642	0.3492
0	1.0	0.500	0.486	0.477
5	0.5	0.480	0.533	0.529
25	0.002	0.485	0.489	0.486
0	4.0	0.800	0.787	0.791
5	3.2	0.802	0.786	0.793
25	1.3	0.796	0.783	0.790

$\bar{h}$ for $2Ns=0$ by Maruyama and Nei [10] are obtained by an analytical formula.

Finally we briefly mention another possible applications of our model. The case with $s<0$ corresponds to negative overdominance model. Our random collision model may provide an effective method to study the effect of asymmetric overdominant selection. Another possible application is to physics. If we take into account the space parameter, our model may be a caricature for a kinetic Ising model.

Acknowledgement

The author is grateful to Prof. T. Maruyama and to Prof. H. Matsuda for fruitful discussion and comments. The author is also grateful to Ms. S. Ueda for the computer-program of the simulation.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Ewens, W. J. (1964). The maintenance of alleles by mutation, *Genetics*, **50**, 891-898.

[2] Ewens, W. J. (1979). *Mathematical Population Genetics*, Springer-Verlag, Berlin.

[3] Itoh, Y. (1973). On a ruin problem with interaction, *Ann. Inst. Statist. Math.*, **25**, 635-641.

[4] Itoh, Y. (1975). An *H*-theorem for a system of competing species, *Proc. Japan Acad.*, **51**, 374-379.

[5] Itoh, Y. (1979a). Random collision process on oriented graph, *Research Memorandum* No. 154, 1-20, The Institute of Statistical Mathematics, Tokyo.

[6] Itoh, Y. (1979b). Random collision models in oriented graphs, *J. Appl. Prob.*, **16**, 36-44.

[7] Itoh, Y. (1981). Non-associative algebra and Lotka-Volterra equation with ternary interaction, *Nonlinear Anal.*, **5**, 53-56.

[8] Kimura, M. and Crow, J. F. (1964). The number of alleles that can be maintained in a finite population, *Genetics*, **49**, 725-738.

[9] Maruyama, T. (1983). Stochastic theory of population genetics, *Bulletin of Mathematical Biology*, **45**, 521-554.

[10] Maruyama, T. and Nei, M. (1982). Genetic variability maintained by mutation and overdominant selection in finite populations, *Genetics*, **98**, 441-459.

[11] Maruyama, T. and Takahata, N. (1981). Numerical studies of the frequency trajectories in the process of fixation of null genes at duplicated loci, *Heredity*, **46**, 49-57.

[12] Nei, M., Maruyama, T. and Wu, C. I. (1983). Models of evolution of reproductive isolation, *Genetics*, **103**, 557-579.

[13] Moran, P. A. P. (1958). Random processes in genetics, *Proc. Phil. Soc.*, **54**, 60-71.

[14] Pederson, D. G. (1973). An approximate method of sampling a multinomial population, *Biometrics*, **29**, 814-821.

[15] Takahata, N. (1981). Genetic variability and rate of gene substitution in a finite population under mutation and fluctuating selection, *Genetics*, **98**, 427-440.

[16] Wright, S. (1939). Statistical genetics in relation to evolution, *Actualités Sci.*, No. 802.

[17] Wright, S. (1949a). Genetics of population, *Encyclopedia Britannica*, 14th ed., **10**, 111-112.

[18] Wright, S. (1949b). Adaptation and selection, *Genetics, Paleontology and Evolution*, (eds. G. L. Jepson, G. G. Simpson and T. Mayr), Princeton University Press, Princeton, New Jersey.

[19] Yokoyama, S. and Nei, M. (1979). Population dynamics of sex-determining alleles in honey bees and self-incompatibility alleles in plants, *Genetics*, **91**, 609-626.