Bull. Korean Math. Soc. **40** (2003), No. 4, pp. 677–683

THE APPLICATION OF STOCHASTIC DIFFERENTIAL EQUATIONS TO POPULATION GENETIC MODEL

Won Choi and Dug-Hwan Choi

Abstract. In multi-allelic model $X = (x_1, x_2, \cdots, x_d)$,

$$M_f(t) = f(p(t)) - \int_0^t Lf(p(t))ds$$

is a P-martingale for diffusion operator L under the certain conditions. In this note, we examine the stochastic differential equation for model X and find the properties using stochastic differential equation.

1. Introduction

Consider a partition to be a sequence

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

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

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

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. 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.

In population genetics theory we often encounter diffusion process on the compact domain

$$K = \{(p_1, p_2, \cdots, p_d); p_1 \geq 0, \cdots, p_d \geq 0, p_1 + p_2 + \cdots + p_d = 1\}$$

Received March 10, 2003.
2000 Mathematics Subject Classification: 92D10, 60H30, 60G44.
Key words and phrases: allelic model, martingale problem, stochastic differential equation, random genetic drift, mutation rate, natural selection.

We suppose that the vector $p(t) = (p_1, p_2, \cdots, p_d)$ of gene frequencies varies with time t.

Let L be a second order differential operator on K

$$L = \sum_{i,j=1}^{d} a_{ij}(p) \frac{\partial^2}{\partial p_i \partial p_j} + \sum_{i=1}^{d} b_i(p) \frac{\partial}{\partial p_i} \tag{1.1}$$

with domain $C^2(K)$, where $\{a_{ij}\}$ is a real symmetric and non-negative definite matrix defined on K and $\{b_i\}$ is an R^d-valued measurable function defined on K. The coefficient $\{a_{ij}\}$ comes from chance replacement of individuals by new ones after random mating and $\{b_i\}$ is represented by the addition of "mutation or gene conversion rate" and the effect of natural selection. The operator L has the same form as the generator of the diffusion describing a $p(t)$-allele model incorporating mutation and random drift with single locus, but we could give a remark that the matrix q_{ij} depends on the combinatorial structure of the partitions.

We assume that $\{a_{ij}\}$ and $\{b_i\}$ are continuous on K. Let $\Omega = C([0,\infty) : K)$ be the space of all K-valued continuous function defined on $[0,\infty)$. A probability P on $(\Omega, \mathcal{F})$ is called a solution of the (K, L, p)-*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 for each $f \in C^2(K)$.

The diffusion operator L was first introduced by Gillespie ([4]) in case that the partition consists of two points. In this case, L is a one-dimensional diffusion operator. However, the uniqueness of solutions of the (K, L, p)-martingale problem has not been generally established. For this problem, Either ([2]) proved that if $\{a_{ij}(p)\} = \{p_i(\delta_{ij} - p_j)\}$ for Kronecker symbol δ_{ij} and $\{b_i(p)\}$ are C^4-functions satisfying a certain condition, then the uniqueness of the (K, L, p)-martingale problem holds. Also, Okada ([5]) showed that the uniqueness holds for a rather general class in two dimension. In case that L reduces to an infinite allelic diffusion model of the Wright-Fisher type, Either ([3]) gave a partial result.

In this note, we try to apply the stochastic differential equation for multi-allelic model. A key point is that the (K, L, p)-martingale problem in population genetics model is related to simpler stochastic differential equation, so we can find various properties for multi-allelic model. Also, we find a solution of stochastic differential equation for simple-allelic model.

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

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

we have

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

Obviously, $\rho = 1$. □

Now, we are concerned with diffusion processes associated with second order differential operator L with random genetic drift

$$a_{ij} = p_i\beta_i\delta_{ij} + p_ip_j\left(\sum_{k=1}^{d} p_i\beta_k - \beta_i - \beta_j\right).$$

Here $\{\beta_i\}$ is non-negative constant satisfying that $sup_i\beta_i < +\infty$, and δ_{ij} stands for the Kronecker symbol.

We used instead of K the simplex

$K_1 = \{(p_1, p_2, \cdots, p_{d-1}); p_1 \geq 0, \cdots, p_{d-1} \geq 0, p_1 + p_2 + \cdots + p_{d-1} \leq 1\}$.

We will restate in the stochastic differential equation, by projecting K onto K_1.

THEOREM 2. *Suppose that* $b_i(p) \geq 0$ *whenever* $p_i = 0$, *and* $\sum_{i=1}^{d} b_i(p) = 0$ *for all* $p \in K$. *Then*

$$\begin{aligned} \langle M_f(t) \rangle &= \sum_{i,j=1}^{d} \int_0^t (p_i(v)\beta_i(v)\delta_{ij} \\ &\quad + p_i(v)p_j(v)(\sum_{k=1}^{d} p_i(v)\beta_k(v) - \beta_i(v) - \beta_j(v))) \\ &\times \frac{\partial p(v)}{\partial p_i}\frac{\partial p(v)}{\partial p_j} dv \end{aligned}$$

where $\langle \cdot \rangle$ *denotes a quadratic variation process.*

Proof. We first choose $\{\alpha_{ij}(p)\}$ as follows :

$$\alpha_{ij}(p) = (\delta_{ij} - p_i)\sqrt{\beta_j p_j} \ .$$

Then it satisfies

$$\sum_{k=1}^{d-1} \alpha_{ik}(p)\alpha_{jk}(p) = p_i\beta_i\delta_{ij} + p_i p_j(\sum_{k=1}^{d-1} p_i\beta_k - \beta_i - \beta_j). \tag{2.1}$$

Let $p(t)$ be the solution to stochastic differential equation

$$dp_i(t) = \sum_{k=1}^{d-1} \alpha_{ik}(p(t))dB_k(t) + b_i(p(t))dt, \ \ i = 1, 2, \cdots, d-1 \tag{2.2}$$

where $p_d = 1-(p_1+p_2+\cdots+p_{d-1})$ and $B_1, B_2, \cdots, B_{d-1}$ are independent Brownian motions.

It is well-known that to show the existence and uniqueness of solutions of the (K_1, L, p)-martingale problem is equivalent to show that the stochastic differential equation has a unique solution ([1], [6]).

Let

$$\alpha_{dk} = -\sum_{i=1}^{d-1} \alpha_{ik}.$$

Since $b_d = -\sum_{i=1}^{d-1} b_i$, stochastic differential equation (2.2) holds for $i = d$. Moreover (2.1) holds when $i = d$ or $j = d$. From the Ito stochastic differential rule ([1], [6]),

$$M_f(t) = \sum_{k=1}^{d-1} \int_0^t \theta_k dB_k, \qquad \theta_k = \sum_{j=1}^{d} \frac{\partial \alpha_{jk}}{\partial p_j}.$$

Let $\langle M_f(t) \rangle$ denote the increasing process associated with the martingale $M_f(t)$. Since it has the property that $\langle M_f(0) \rangle = 0$ and $M_f^2(t) - \langle M_f(t) \rangle$ is a martingale [1], [6], we have

$$\begin{aligned}
\langle M_f(t) \rangle &= \sum_{k=1}^{d-1} \int_0^t \theta_k^2(v) dv, \\
\langle M_f(t) \rangle &= \sum_{i,j=1}^{d} \int_0^t \Big(p_i(v)\beta_i(v)\delta_{ij} \\
&\qquad + p_i(v)p_j(v)\big(\sum_{k=1}^{d} p_i(v)\beta_k(v) - \beta_i(v) - \beta_j(v) \big) \Big) \\
&\qquad \times \frac{\partial p(v)}{\partial p_i} \frac{\partial p(v)}{\partial p_j} dv.
\end{aligned}$$

□

COROLLARY 3. *Let $a_{ij}(p)$ be any polynomial satisfying* (1.1). *Suppose that $a_{ii}(p) = 0$ if $p_i = 0$, and $\sum_{i,j=1}^{d} a_{ij}(p) = 0$ if $\sum_{i=1}^{d} p_i = 1$. Then, there exists a positive constant ε such that $\langle M_f^{\varepsilon}(0) \rangle = 0$ and $\{M_f^{\varepsilon}(t)\}^2 - \langle M_f^{\varepsilon}(t) \rangle$ is a martingale. Here*

$$\begin{aligned}
\langle M_f^{\varepsilon}(t) \rangle &= \sum_{i,j=1}^{d} \int_0^t (p_i(v)\beta_i(v)\delta_{ij} \\
&\qquad + p_i(v)p_j(v)(\sum_{k=1}^{d} p_i(v)\beta_k(v) - \beta_i(v) - \beta_j(v)) \\
&\qquad + \varepsilon a_{ij}(p)) \frac{\partial p(v)}{\partial p_i} \frac{\partial p(v)}{\partial p_j} dv.
\end{aligned}$$

Proof. This result follows directly from Theorem 2 and T. Shiga ([7]).

□

EXAMPLE. (A model with natural selection) Let the coefficients $\{b_i(p)\}$ have the form

$$b_i(p) = g_i(p) + h_i(p)$$

where

$$g_i(p) = \sum_{k=1}^{d} p_k q_{ki}$$

$$q_{ij} \geq 0 \text{ for } i \neq j, \quad q_{ii} = -\sum_{j \neq i} q_{ij}.$$

For $i \neq j$, q_{ij} represents a rate of change of type x_i to type x_j.

For a finite number of types $x_1, x_2, \cdots, x_d$, a standard model for incorporating natural selection is to take

$$h_i(p) = p_i \left(\sum_{k=1}^{d} p_k m_{ki} - \sum_{j,l=1}^{d} p_j p_l m_{jl} \right).$$

The interpretation is that $x_1, x_2, \cdots, x_d$ are possible alleles carried by a gamete at some gene locus, and m_{ij} is a fitness coefficient of the genotype (x_i, x_j) such that $m_{ji} = m_{ij}$. If p_i is the frequency of type i, then $p_i p_j$ represents the frequency of (x_i, x_j).

Consider a model with natural selection and let $d = 1$. Suppose that $\alpha' = 0$. Then the existence and uniqueness of solutions of the (K, L, p)-martingale problem is equivalent to show that the stochastic differential equation

$$dp_i(t) = \alpha(p_i(t)) \circ dB(t) + [g(p_i(t)) + h(p_i(t))]\, dt, \quad p_i(0) = p_0$$

has a unique solution. Here operation $\circ$ denotes the symmetric $\mathcal{Q}$-*multiplicaton.*

Let $u(z_1, z_2)$ be the solution of

$$\frac{\partial u}{\partial z_2} = a(u), \quad u(z_1, 0) = z_1,$$

and A_t be the solution of

$$\frac{dA_t}{dt} = g(u(A_t, B_t)) + h(u(A_t, B_t)), \quad A_0 = p_0.$$

Then the solution $p_i(t)$ is given by

$$p_i(t) = u(A_t, B_t).$$

Indeed, by the chain rule

$$dp_i(t) = a(u(A_t, B_t)) \circ dB_t + \left(\frac{\partial u}{\partial z_1}\right)(A_t, B_t)\,(g(u(A_t, B_t)) + h(u(A_t, B_t)))\,dt,$$

and hence

$$dp_i(t) = \alpha(p_i(t)) \circ dB(t) + [g(p_i(t)) + h(p_i(t))]\,dt.$$

ACKNOWLEDGEMENT. The first author is grateful to Professor Jian Huang for several helpful suggestions concerning biological ideas during my staying at University of Iowa.

References

[1] W. Choi and V. I. Shin, *Exact closed form equations of optimal nonlinear filtering for a class of bilinear systems*, Auomat. and Remote Control **60** (1999), 1151–1161.

[2] S. N. Ethier, *A class of degenerate diffusion processes occurring in population genetics*, Comm. Pure Appl. Math. **29** (1976), 483–493.

[3] ———, *A class of infinite dimensional diffusions occurring in population genetics*, Indiana Univ. J. **30** (1981), 925–935.

[4] J. H. Gillespie, *Natural selection for within-generation variance in offspring number*, Genetics **76** (1974), 601–606.

[5] N. Okada, *On the uniqueness problem of two dimensional diffusion processes occurring in population genetics*, Z. Wahr. Verw. Geb. **56** (1981), 63–74.

[6] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, Springer-Verlag, 1994.

[7] T. Shiga, *Diffusion processes in population genetics*, J. Math. Kyoto Univ. **21-1** (1981), 133–151.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF INCHEON, INCHEON 402-749, KOREA
E-mail: choiwon@incheon.ac.kr
ameschoi59@yahoo.co.kr