Bull. Korean Math. Soc. **41** (2004), No. 2, pp. 337–345

THE APPLICATION OF STOCHASTIC ANALYSIS TO COUNTABLE ALLELIC DIFFUSION MODEL

Won Choi

Abstract. In 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 can show existence and uniqueness of solution for stochastic differential equation and martingale problem associated with mean vector. Also, we examine that if the operator related to this martingale problem is connected with Markov processes under certain circumstance, then this operator must satisfy the maximum principle.

1. Introduction

Consider n locus model

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

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. If the partition X has α_i parts equal to i, then 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 chromosome of type X_i.

Let S be a countable set. In population genetics theory we often encounter diffusion process on the domain

$$K = \{p = (p_i)_{i \in S}\,;\, p_i \geq 0,\, \sum_{i \in S} p_i = 1\}$$

Received October 2, 2003.

2000 Mathematics Subject Classification: 92D25, 60H30, 47A05.

Key words and phrases: countable allelic model, martingale problem, stochastic differential equation, mean vector, second order differential operator.

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

Let L be a second order differential operator on K

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

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 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 ([5]) 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 ([3]) 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 ([6]) 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 ([4]) gave a partial result.

In this note, we try to apply diffusion processes for countable-allelic model. A key point is that the (K, L, p)-martingale problem in population genetics model is related to diffusion processes, so we can find existence and uniqueness of stochastic differential equation associated with mean vector. Also, we have to examine the relationship between martingale problem formulation and symmetric Markov processes.

2. Main results

The diffusion process with the generator L is easily shown to be ergodic since the matrix $\{q_{ij}\}$ generates an ergodic Markov chain. ([8]) Hence the diffusion has a unique stationary distribution $\nu(dp)$.

We begin with the following Lemma.

LEMMA 1. *The mean vector $\bar{P} = (\bar{P}_i)$ of stationary distribution $\nu(dp)$ satisfies the followings;*

(1) $\sum_j \bar{P}_j q_{ji} = 0$,
(2) $\sum_i \bar{P}_i = 1$.

Proof. See A. Shimizu ([8]). □

We are concerned with diffusion processes associated with second order differential operator L with random genetic drift

$$a_{ij} = p_i\beta_i\delta_{ij} + p_ip_j(\sum_{k\in S} p_i\beta_k - \beta_i - \beta_j).$$

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

In order to consider an stochastic differential equation for $p(t)$, we need boundary conditions and regularity condition on the drift coefficients b_i.

[Assumption for $b_i(p)$] : $\{b_i(p)\}_{i\in S}$ are real functions defined on K which satisfy the following conditions :

(i) $b_i(p) \geq 0$ if $p_i = 0$,
(ii) $\sum_{i\in S} b_i(p) = 0$ uniformly in $p \in K$,
(iii) there exists a matrix $\{c_{ij}\}_{i,j\in S}$ such that $c_{ij} \geq 0$ for every i and j of S, and

$$|b_i(p) - b_i(p')| \leq \sum_{j\in S} c_{ij}|p_j - p'_j|.$$

Suppose that $\{b_i(p)\}_{i\in S}$ satisfies the [Assumption for $b_i(p)$]. Then $p(t)$ is unique solution to stochastic differential equation

$$dp_i(t) = \sum_{k\in S} \alpha_{ik}(p(t))dB_k(t) + b_i(p(t))dt, \quad i \in S$$

where

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

and B_i are independent Brownian motions.

In order to construct the stochastic differential equation associated to mean vector, we need the following definition.

DEFINITION. A sequence $\{X_1, X_2, \cdots, X_K, \cdots\}$ 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 $i = 1, 2, \cdots,$.

The value

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

does not depend on the choice of (X_1, X_K)-chain.

Let X be any partition of n and let $\{X_1, X_2, \cdots, X_i, \cdots\}$ be a $((n), X_i)$-chain. Put

$$P_i = \prod_{k=1}^{i-1}\left(\frac{q_{j\ j+1}}{q_{j+1\ j}}\right), \quad P_{(n)} = 1.$$

Let

$$K_1 = \{P = (P_i)_{i\in S} : \sum_{i\in S} P_i < +\infty\}$$

and define a mapping $\bar{P}$ on K_1 called by mean vector

$$\bar{P}_i = \frac{P_i}{\sum_j P_j}.$$

Then $\bar{P}_i$ satisfies Lemma 1.

Consider the solution to stochastic differential equation for $P_i(t)$

$$dP_i(t) = \sqrt{\beta_i P_i(t)}dB_i(t) + \tilde{b}_i(P(t))dt, \ \ i \in S \tag{2.1}$$

where

$$\tilde{b}_i(P(t)) = b_i(\bar{P}(t)) + c\bar{P}_i(t) + \bar{P}_i(t)(\beta_i - \sum_{k\in S}\bar{P}_k(t)\beta_k)$$

for a constant $c > 0$ satisfying $c > (1/2)sup_{i\in S}\beta_i$.

It will be shown that the existence and the uniqueness of solutions hold for the equation (2.1) when the drift coefficients $\{b_i(p)\}_{i\in S}$ satisfies the [Assumption for $b_i(p)$], not [Assumption for $\tilde{b}_i(P)$].

THEOREM 2. *Suppose that $\{b_i(p)\}_{i\in S}$ satisfies the [Assumption for $b_i(p)$]. Then the existence and the uniqueness of solutions hold for the equation* (2.1).

Proof. For any ε, let

$$X_\varepsilon = \{P = (P_i)_{i\in S} : \sum_{i\in S} P_i \geq \varepsilon\}.$$

Note that $\{\tilde{b}(P)\}_{i\in S}$ are continuous on P_ε.

If $P(0) \in X_\varepsilon$ and $\tau_\varepsilon = inf\{t > 0 : \sum_{i\in S} P_i(t) = \varepsilon\}$, we have

$$P(t) \in X_\varepsilon \text{ for } 0 \le t \le \tau_\varepsilon$$

and

$$P_i(t \wedge \tau_\varepsilon) = P_i(0) + \int_0^{t\wedge\tau_\varepsilon} \sqrt{\beta_i P_i(s)} dB_i(s) + \int_0^{t\wedge\tau_\varepsilon} \tilde{b}_i(P(s))ds.$$

Therefore, there exists a X-valued solution $P(t)$ of (2.1) up to $\tau = \lim_{\varepsilon\downarrow 0} \tau_\varepsilon$. Since $P(\tau = +\infty\} = 1$, there exists a solution of the equation (2.1) taking values in K_1.

In order to prove the uniqueness of solutions, suppose that $\{P(t)\}$ and $\{P'(t)\}$ be two solutions of (2.1) taking values in K_1. For any $\varepsilon > 0$ and $M > 0$ define

$$\eta = \inf\{t > 0 \ : \ \sum_{i\in S} P_i(t) \notin (\varepsilon, M) \text{ or } \sum_{i\in S} P_i'(t) \notin (\varepsilon, M)\}.$$

From the [Assumption for $b_i(p)$], it follows that there exists a constant $C > 0$ such that for every P and P' of X_ε,

$$\begin{aligned}
&\sum_{i\in S} E\{|P_i(t\wedge\eta) - P_i'(t\wedge\eta)|\} \\
&\le \int_0^t \sum_{i\in S} E\{|\tilde{b}_i(P(s\wedge\eta)) - \tilde{b}_i(P'(s\wedge\eta))|\}ds \\
&\le C\int_0^t \sum_{i\in S} E\{|P_i(s\wedge\eta) - P_i'(s\wedge\eta)|\}ds.
\end{aligned}$$

Hence, by Gronwall's inequality ([7]) we have

$$P\{P(t) = P'(t) \text{ for } 0 \le t \le \eta\} = 1,$$

which implies the uniqueness of solutions for (2.1) because of

$$\lim_{M\to\infty,\, \varepsilon\downarrow 0} \eta = +\infty \text{ a.s.}$$

□

COROLLARY 3. *Let L_1 be a second order differential operator on K_1*

$$L_1 = \sum_{i,j\in S} \tilde{a}_{ij}(P)\frac{\partial^2}{\partial P_i \partial P_j} + \sum_{i\in S} \tilde{b}_i(P)\frac{\partial}{\partial P_i}$$

where

$$\tilde{a}_{ij} = \begin{cases} (\textit{number of elements } S) \times \sqrt{\beta_i \beta_j P_i(t) P_j(t)} & \textit{if } S \textit{ is finite} \\ 0 & \textit{if } S \textit{ is infinite.} \end{cases}$$

Then the uniqueness of solution for the (K_1, L_1, P_0)-martingale problem holds.

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

$$\tilde{a}_{ij}(P) = \sum_{k \in S} \tilde{\alpha}_{ik}(P)\alpha_{jk}(P), \quad \tilde{\alpha}_{ij}(P) = \sqrt{\beta_i P_i(t)}.$$

Then $P_i(t)$ is a solution to stochastic differential equation

$$dP_i(t) = \tilde{\alpha}_{ij}(P(t))dB_i(t) + \tilde{b}_i(P(t))dt, \quad i \in S$$

It is well-known that to show the existence and uniqueness of solutions for the (K_1, L_1, P_0)-martingale problem is equivalent to show that the stochastic differential equation (2.1) has a unique solution. Therefore this result follows immediately from the Theorem 2. □

Let $C(K_1)$ be the Banach space of all continuous functions on K_1 with the uniform norm. Suppose that $\{b_i(p)\}_{i \in S}$ satisfies the [Assumption for $b_i(p)$]. Then Choi and Lee ([1]) show that there exists a unique strongly continuous contraction semigroup $\{T_t\}$ on $C(K_1)$ such that

(1) $T_t f \geq 0$ for any $f \in C(K_1)$ and $T_t 1 = 1$,
(2) $T_t f - f = \int_0^t T_s L_1 f ds$.

If μ is a measure on state space of T_t for which

$$\int fAgd\mu = \int gAfd\mu, \quad f, g \in C(K_1), \tag{2.2}$$

where A denotes the generator of T_t, then under mild conditions it is possible to show that

$$\int fT_t gd\mu = \int gT_t fd\mu, \quad f, g \in C(K_1). \tag{2.3}$$

But, since in practice we seldom knows enough about A to check a relation like (2.2), the observation that (2.2) usually implies (2.3) hardly can be considered a very useful one. Instead of (2.2), what we have a chance of testing in many practical circumstance is the following result very valuable to ours.

THEOREM 4. *Suppose that $\{b_i(p)\}_{i\in S}$ satisfies the [Assumption for $b_i(p)$]. Then there exist a measure μ such that the formula*

$$\int fL_1 g d\mu = \int gL_1 f d\mu \tag{2.4}$$

is satisfied.

Proof. Setting

$$\mu(dP) = \exp\left[\int \frac{\tilde{b}(y) - \tilde{a}\,'(y)}{\tilde{a}(y)} dy\right] dP,$$

we have easily result. □

REMARK. In Theorem 4, L_1 is not A but is more tractable operator that determines T_t in some weaker sense than does A.

By Corollary 3, we know that there exists a probability measure P^* satisfying the following conditions ;

(1) $P^*(P(0) = P_0) = 1$ and
(2) denoting $M_f^*(t) = f(P(t)) - \int_0^t L_1 f(P(t))ds$, $M_f^*(t)$ is a P^*-martingale for all $f \in C(K_1)$.

Defining

$$\langle f, g\rangle \equiv L_1(f\cdot g) - fL_1 g - gL_1 f \quad \text{for all } f, g \in C(K_1),$$

we meet with ;

THEOREM 5.

$$(M_f^*(t))^2 - \int_0^t \langle f, f\rangle (P(s))ds$$

is a P^*-martingale.

Proof.

$$\begin{aligned}
(M_f^*(t))^2 &= f^2(P(t)) - 2M_f^*(t)\int_0^t L_1 f(P(s))ds - \left(\int_0^t L_1 f(P(s))ds\right)^2 \\
&\cong \int_0^t L_1 f^2(P(s))ds - 2\int_0^t M_f^*(t) L_1 f(P(s))ds \\
&\quad - \left(\int_0^t L_1 f(P(s))ds\right)^2 \\
&= \int_0^t \langle f, f\rangle (P(s))ds,
\end{aligned}$$

where $X(\cdot) \cong Y(\cdot)$ means that $X(t) - Y(t)$ is a P^*-martingale and we have used the "integration by parts" lemma ([9]) for martingales to get from the second expression to the third. □

Define Υ on $C(K_1) \times C(K_1)$ by $\Upsilon(f,g) = -\int fL_1 g d\mu$.

The following result tell us that if operator L_1 satisfying the martingale problem and (2.4) is connected with Markov processes, then this operator L_1 must satisfy the maximum principle.

THEOREM 6. *Assume that*

$$\int L_1 f d\mu = 0. \tag{2.5}$$

Then we have the following properties

(1) $\Upsilon(f,f) \geq 0$ *and*

(2) $\Upsilon(\varphi \circ f, \varphi \circ f) \leq ||\varphi' \circ f||^2 \Upsilon(f,f)$, *where* $\Phi = \{\varphi \in C^\infty(R) : \varphi(0) = 0\}$ *and* $\varphi \circ f \in C(K_1)$ *for all* $\varphi \in \Phi$.

Proof. Simply integrating the equation

$$L_1(f \cdot g) = \langle f, g\rangle + fL_1 g + gL_1 f,$$

we have

$$\Upsilon(f,g) = \frac{1}{2}\langle f, g\rangle d\mu$$

from the Theorem 5 and (2.5). Hence condition (1) follows easily.

Next, define

$$[M(\cdot), M(\cdot)](t) = \lim_{n\to\infty} \sum_{k=0}^{[2^n t]-1} \left(M\left(\frac{k+1}{2^n}\right) - M\left(\frac{k}{2^n}\right)\right)^2$$

for square-integrable semimartingales $M(\cdot)$. Then from the Theorem 4 and theory of martingales ([2], [7]), we can identify $\int_0^t \langle f, f\rangle(P(s))ds$ as the dual previsable projection of $[M_f^*(\cdot), M_f^*(\cdot)](\cdot)$. In particular, $\int_0^t \langle f, f\rangle(P(s))ds$ is nondecreasing, and so it is clear that $\langle f, f\rangle(P_0) \geq 0$. Furthermore, since

$$[M_f^*(\cdot), M_f^*(\cdot)] = [f(P(\cdot)), f(P(\cdot))],$$

it is easy to see that

$$||\varphi' \circ f||^2 [M_f^*(\cdot), M_f^*(\cdot)] - [M_{\varphi\circ f}^*(\cdot), M_{\varphi\circ f}^*(\cdot)]$$

is nondecreasing.

Therefore, since projection is a linear operation,

$$||\varphi' \circ f||^2 \int_0^t \langle f, f\rangle(P(s))ds - \int_0^t \langle \varphi\circ f, \varphi\circ f\rangle(P(s))ds$$

is nondecreasing.

Hence we have

$$\langle \varphi \circ f, \varphi \circ f \rangle (P_0) \leq ||\varphi' \circ f||^2 \langle f, f \rangle (P_0),$$

and condition (2) follows directly. □

REMARK. Actually, L_1 is at worst a "second order" operator. Therefore, if it is satisfied (2.4), (2.5), its square root can be at worst a "first order" operator, and so condition (1) of Theorem 6 is somewhat natural. However, even though we accept the argument that L_1 is second order, it is difficult to see how to make the reasoning about the square root of L_1 rigorous.

ACKNOWLEDGEMENT. This paper was written during the author's staying at University of Iowa, U.S.A. I wish to thank the University of Iowa, in particular professor Jian Huang, for his support, friendship and hospitality.

References

[1] W. Choi and B. K. Lee, *On the diffusion processes and their applications in population genetics*, In press, 2004.
[2] C. Dellacherie and P. A. Meyer, *Probabikites et potentiel*, Part B, Hermann, Paris, 1980.
[3] S. N. Either, *A class of degenerate diffusion processes occurring in population genetics*, Comm. Pure Appl. Math. **29** (1976), no. 5, 483–493.
[4] ———, *A class of infinite-dimensional diffusions occurring in population genetics*, Indiana Univ. Math. J. **30** (1981), no. 6, 925–935.
[5] J. H. Gillespie, *Natural selection for within-generation variance in offspring number,* Genetics **76** (1974), no. 3, 601–606.
[6] N. Okada, *On the uniqueness problem of two dimensional diffusion processes occurring in population genetics*, Z. Wahrsch. Verw. Gebiete **56** (1981), no. 1, 63–74.
[7] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, Springer-Verlag, 1994.
[8] A. Shimizu, *Stationary distribution of a diffusion process taking values in probability distributions on the partitions*, Stochastic methods in biology (Nagoya, 1985), 100–114, Lecture Notes in Biomath., 70, Springer, Berlin, 1987.
[9] D. Stroock and S. Varadhan, *Multidimensional diffusion processes*, Grundlehren 233, Springer-Verlag, Berlin-New York, 1979.

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