J. Appl. Math. & Informatics Vol. 30(2012), No. 3 - 4, pp. 677 - 683
Website: http://www.kcam.biz

ON THE DIFFUSION OPERATOR IN POPULATION GENETICS†

WON CHOI

ABSTRACT. W.Choi([1]) obtains a complete description of ergodic property and several property by making use of the semigroup method. In this note, we shall consider separately the martingale problems for two operators A and B as a detail decomposition of operator L. A key point is that the (K, L, p)-martingale problem in population genetics model is related to diffusion processes, so we begin with some a priori estimates and we shall show existence of contraction semigroup $\{T_t\}$ associated with decomposition operator A.

AMS Mathematics Subject Classification : 92D10, 60H30, 60G44.
Key words and phrases : diffusion operator, martingale problem, contraction semigroup.

1. Introduction

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

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

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

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

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

Received August 1, 2011. Revised September 20, 2011. Accepted September 26, 2011.
†This research was supported by University of Incheon Research Grant, 2011-2012

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.

W.Choi([1]) obtains a complete description of ergodic property and several property by making use of the semigroup method. In this note, we shall consider separately the martingale problems for two operators A and B as a detail decomposition of operator L. A key point is that the (K, L, p)-martingale problem in population genetics model is related to diffusion processes, so we begin with some a priori estimates and we shall show existence of contraction semigroup $\{T_t\}$ associated with decomposition operator A.

2. Main results

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 $\mathrm{sup_i}\beta_\mathrm{i} < +\infty$, 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

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

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

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

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

$$dP_i(t) = \sqrt{\beta_i P_i(t)}dB_i(t) + \tilde{b}_i(P(t))dt, \;\; i \in S, \tag{1}$$

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 $c > (1/2)\mathrm{sup_{i\in S}}\beta_\mathrm{i}$.

It was shown easily that the existence and the uniqueness of solutions hold for the equation (1) when the drift coefficients $\{b_i(p)\}_{i\in S}$ satisfies the [Assumption for $b_i(p)$], not [Assumption for $\tilde{b}_i(P)$].([1]) So, we have the following result.

Lemma 1. *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

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

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

Proof. 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 (1) has a unique solution. Therefore this result follows from W. Choi([1]). □

Let S_d be the set of symmetric, non-negative definite, $d \times d$ matrices. To establish the main results, we shall consider separately the operators L_1 for

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

and for

$$B = \sum_{i\in S} \tilde{b}_i(P)\frac{\partial}{\partial P_i}. \tag{3}$$

We suppose that the norm on $C(K_1)$ is the supremum norm, denoted $||\cdot||_{K_1}$ and the seminorm $|\cdot|_{C^m(K_1)}$ on $C^m(K_1)$ is defined by

$$|f|_{C^m(K_1)} = \sum_{1\le|\alpha|\le m} ||D^\alpha f||_{K_1}.$$

Theorem 2. *For a positive integer m,and define the operators A and B by (2) and (3), where $\tilde{a} : K_1 \to S_d$ and $\tilde{b} \in C^m(K_1, R^d)$ satisfies $\langle \tilde{b}, \nabla\tilde{a}\rangle \ge 0$ on ∂K_1. If*

$$\frac{\partial}{\partial t}u = Au \tag{4}$$

$$\frac{\partial}{\partial t}v = Bv, \tag{5}$$

then

$$|u(t,\cdot)|_{c^m(K_1)} \le |u(0,\cdot)|_{C^m(K_1)} \tag{6}$$

$$|v(t,\cdot)|_{c^m(K_1)} \le e^{\lambda_m t}|v(0,\cdot)|_{C^m(K_1)}, \tag{7}$$

where λ_m is defined in the process of proof.

Proof. For each multi-index α, define the operator

$$A_\alpha = A + \sum_{i=1}^{d} (\frac{1}{2}\alpha_i - |\alpha| x_i) \frac{\partial}{\partial x_i},$$

and note that, since $\tilde{b}(x) = \frac{1}{2}\alpha - |\alpha|x$ satisfies the condition of Lemma 1, A_α satisfies the maximum principle in K_1, and that

$$D^\alpha A = A_\alpha D^\alpha - \frac{1}{2}|\alpha|(|\alpha| - 1)D^\alpha$$

on $C^{|\alpha|+2}(K_1)$. By differentiating (4), we therefore obtain

$$\frac{\partial}{\partial t} u^\alpha = A_\alpha u^\alpha - \frac{1}{2}|\alpha|(|\alpha| - 1)u^\alpha$$

for $1 \leq |\alpha| \leq m$, where $u^\alpha = D_x^\alpha u$, and these equations, together with the maximum principle, imply (6).

As for (7), there exist functions $c_{\alpha\gamma}$, defined for each pair of multi-indices α and γ with $1 \leq |\gamma| \leq |\alpha| \leq m$, such that

$$D^\alpha B = B D^\alpha + \sum_{1 \leq |\gamma| \leq |\alpha|} c_{\alpha\gamma} D^\gamma$$

on $C^{|\alpha|+1}(K_1)$ for $1 \leq |\alpha| \leq m$. By differentiating (5), we therefore obtain

$$\frac{\partial}{\partial t} v^\alpha = B v^\alpha + \sum_{1 \leq |\gamma| \leq |\alpha|} c_{\alpha\gamma} v^\gamma$$

for $1 \leq |\alpha| \leq m$. Here $v^\gamma = D_x^\gamma v$. These equations, together with the maximum principle for B, imply (7) with

$$\lambda_m = \max_{1 \leq |\gamma| \leq \mathrm{m}} \sum_{1 \leq |\gamma| \leq |\alpha| \leq \mathrm{m}} ||\mathrm{c}_{\alpha\gamma}||_{\mathrm{K}_1},$$

where the norm on $C(K_1)$ is the supremum norm. □

For each $P \in K_1$, let $\mathfrak{M}$ be the set of solutions to the martingale problem for A starting at P. Then martingale implies

$$E_P^Q[f(P(t))] = f(P) + \int_0^t E_P^Q[Af(P(s))]ds \tag{8}$$

for each $f \in C^2(K_1)$, $Q \in \mathfrak{M}$. We define the one parameter family $\{T_t : t \geq 0\}$ of transformations from $C(K_1)$ to the space of bounded functions on K_1 by

$$T_t f = E_P^Q[f(P(t))].$$

Then we have;

Theorem 3. *There exists a strongly continuous non-negative contraction semigroup $\{T_t : t \geq 0\}$ associated with decomposition operator A on $C(K_1)$.*

Proof. Let a be an arbitrary integer, a_n be the number of multi-indices α with $|\alpha| \leq n$. Suppose the coordinates of R^{a_n} are to be indexed by these multi-indices. Define $g^n : K_1 \to R^{a_n}$ by $g^n_\alpha(P) = P^\alpha$ and

$$Ag^n_\alpha = (\Pi_n g^n)_\alpha \tag{9}$$

where $\Pi_n \in R^{a_n} \bigotimes R^{a_n}$.

Choose $u^n : [0, \infty) \times K_1 \to R^{a_n}$ such that $u^n_\alpha(t, \cdot) = T_t g^n_\alpha$. By (8) and (9), we get

$$u^n(t, \cdot) = g^n + \int_0^t \Gamma_n u^n(s, \cdot) ds.$$

By solving this equation, we have

$$u^n(t, \cdot) = e^{t\Pi_n} g^n$$

for each $t \geq 0$. Hence

$$T_t\langle \theta, g^n \rangle = \langle e^{t\Pi_n \theta}, g^n \rangle, \tag{10}$$

for $\theta \in R^{a_n}$.

Let $\mathfrak{P}(K_1)$ denote the subspace of $C(K_1)$ consisting of all polynomials in $P_1, P_2, \cdots, P_d$. By (10),

$$T_t T_s = T_{t+s}$$

and

$$\mathrm{lim_{t\to 0}||T_t f - f||_{K_1} = 0}.$$

Therefore, since $\mathfrak{P}(K_1)$ is dense in $C(K_1)$, and since $||T_t f||_{K_1} \leq ||f||_{K_1}$ for each $f \in C(K_1)$, $\{T_t : t \geq 0\}$ is strongly continuous non-negative contraction semigroup on $C(K_1)$. □

Corollary 4. *For each positive integer m and $f \in C^m(K_1)$,*

$$|T_t f|_{C^m(K_1)} \leq |f|_{C^m(K_1)}.$$

Proof. Define $u(t, \cdot) = T_t f,\ t \geq 0$. By (9) and (10), hypothesis (4) of Theorem 2 is satisfied. Therefore the result follows easily. □

References

1. W.Choi and B.K.Lee *On the diffusion processes and their applications in population genetics*, J. Applied Mathematics and Computing, Vol 15, No 1-2 (2004), 415-423
2. S.N.Either, *A class of degenerate diffusion processes occurring in population genetics*, Comm. Pure Appl. Math., **29** (1976), 483-493.
3. S.N.Either, *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.

Won Choi received his Ph.D at Sung Kyun Kwan University under the direction of Yeong Don Kim. Since 1993 he has been at the University of Incheon. In 1996 and 2003, he was in Steklov Mathematical Institute of Russia and University of Iowa, respectively. His research interests center on stochastic processes and Bio-mathematics.

Department of Mathematics, University of Incheon, Incheon 406-772, Korea.
e-mail:choiwon@incheon.ac.kr