J. Appl. Math. & Informatics Vol. **27**(2009), No. 3 - 4, pp. 757 - 762
Website: http://www.kcam.biz

SOME SYMMETRY PRESERVING TRANSFORMATION IN POPULATION GENETICS

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. We can also obtain a new diffusion operator L^* for diffusion coefficient and we prove that unique solution for L^*-martingale problem exists. In this note, we define new symmetric preserving transformation. Uniqueness for martingale problem and symmetric property will be proved.

AMS Mathematics Subject Classification: 92D25,60H30,47A05
Key words and phrases: Martingale problem, stochastic differential equation, diffusion operator, symmetry preserving transformation

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

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

Received July 25, 2008. Revised January 5, 2009. Accepted January 17, 2009. This research was supported by University of Incheon Research Grant, 2008-2009.

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 a measurable function defined on K.

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([2]) in case that the partition consists of two points. In this case, L is an 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([1]) 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.

The diffusion process with the generator L is easily shown to be ergodic since the matrix $\{q_{ij}\}$ generates an ergodic Markov chain.([3])

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

In this note, we obtain a new diffusion operator L^* for diffusion coefficient and we prove that unique solution for L^*-martingale problem exists. We also define new symmetric preserving transformation. Uniqueness for martingale problem and symmetric property will be proved.

2. Main results

We start with ;

Definition 1. 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

$$K^* = \left\{P = (P_i)_{i\in S} : \sum_{i\in S} P_i < +\infty\right\}.$$

Lemma 2. *Let L^* be a second order differential operator on K^**

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

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^*, L^*, 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

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

It is well-known that to show the existence and uniqueness of solutions for the (K^*, L^*, P_0)-martingale problem is equivalent to show that the stochastic differential equation has a unique solution. Therefore this result follows. □

By Lemma 1, we know that there exists a probability measure P^*(abbreviated by $P^* \sim L^*$) satisfying the following conditions ;

(1) $P^*(P(0) = P_0) = 1$ and

(2) denoting $M_f^*(t) = f(P(t)) - \int_0^t L^* f(P(t))ds$, $M_f^*(t)$ is a P^*-martingale for all $f \in C(K^*)$.

Defining

$$\langle f, g\rangle \equiv L^*(f \cdot g) - fL^*g - gL^*f \quad \text{for all } f, g \in C(K^*),$$

we know that

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

is a P^*-martingale.

Because L^* is local, we can extend the domain of L^*. Let K_{loc}^* denote the set of ξ for which there exists a compactly nested exhaustion $\{U_n\}_1^\infty$ and $\{\xi_n\}_1^\infty \subset K^*$ such that $\xi = \xi_n$ on U_n. Clearly K_{loc}^* is an algebra and since L^* is local, $L^*\xi$ is well defined by taking $L^*\xi(x) = L^*\xi_n(x)$ if $x \in U_n$. Thus $\langle \xi, \eta\rangle$ is also well defined for all $\xi, \eta \in K_{loc}^*$.

Given $\xi \in K_{loc}^*$, define L_ξ^* on K^* by

$$L_\xi^* f = L^* f + \langle \xi, f\rangle.$$

Theorem 3. *If $\xi \in K_{loc}^*$, then for each P_0 there is a precisely one $P_\xi^* \sim L_\xi^*$.*

Proof. Choose $\{U_n\}_1^\infty$ and $\{\xi\}_1^\infty$ and $\tau_n = \inf\Big\{t \geq 0 \ : \ P_0 \notin U_n\Big\}$. For $n \geq 1$, define

$$R^{(n)}(t) = exp\left[\xi_n(P(t)) - \xi_n(P(0)) - \int_0^t \left(L^*\xi_n + \frac{1}{2}\langle \xi_n, \xi_n\rangle\right)(P(s))ds\right].$$

Using Ito's formula for continuous semimartingales, we see that $dR^{(n)}(t) = R^{(n)}(t)dM_{\xi_n}^*(t)$. Hence, again by Ito's formula, for f, we have

$$\begin{aligned} d\Big(R^{(n)}(t)f(P(t))\Big) &= R^{(n)} f(P(t))dM_{\xi_n}^*(t) + R^{(n)}(t)dM_f^*(t) \\ &\quad + R^{(n)}(t)\Big(L^*f(P(t)) + \langle \xi, f\rangle(P(t))\Big)dt, \end{aligned}$$

since $dM_{\xi_n}^*(t)dM_f^*(t) = \langle \xi_n, f\rangle(P(t))dt$. In particular,

$$R^{(n)}(t)\left(f(P(t)) - \int_0^t L_{\xi_n}^* f(P(t))\right)$$

is a P^*-martingale for all f. Also, because $dR^{(n)}(t) = R^{(n)}(t)dM_{\xi_n}^*(t)$ and $R^{(n)}(\cdots)$ is bounded on finite time intervals, $R^{(n)}(t)$ is a P^*- martingale.

Because $R^{(n)}(t)$ is a P^*- martingale, we can find a unique measure $P^{(n)}$ such that $dP^{(n)} = R^{(n)}(t)dP^*$ for each $t \geq 0$. Moreover, it is easy to check that $P^{(n+1)} = P^{(n)}$ and therefore there is a unique $P^*_{(\xi)}$ satisfying $P^*_{(\xi)} = P^{(n)}$ for all n.

Using the fact that $R^{(n)}(t)\left(f(P(t)) - \int_0^t L^*_{\xi_n} f(P(t))\right)$ is a martingale, we see that $P^{(n)} \sim L^*_{\xi_n}$. In fact $P^{(n)}$ is the only $P^* \sim L^*_{\xi_n}$. To see this, simply repeat the argument given above and thereby show that if $P^* \sim L^*_{\xi_n}$, then

$$(R^{(n)}(t))^{-1}\left(f(P(t)) - \int_0^t L^* f(P(t))\right)$$

is a martingale. Similarly, $(R^{(n)}(t))^{-1}$ is a martingale. Since $P^{(n)} \sim L^*_{\xi_n}$,

$$(f(P(t \wedge \tau_n) - \int_0^{t\wedge\tau_n} L^*_\xi f(P(s))ds$$

is a P^*_ξ-martingale. From this it is an elementary step to conclude that $P^*_\xi \sim L^*_\xi$. Conversely, if $P^* \sim L^*_\xi$, then

$$f(P(t \wedge \tau_n)) - \int_0^{t\wedge\tau_n} L^*_{\xi_n} f(P(s))ds$$

is a P^*-martingale. Thus

$$Q = P^* \bigotimes_{\tau_n(\cdot)} P^{(n)}_{P^0(\tau_n(\cdot))} \circ \theta^{-1}_{\tau_n(\cdot)} \sim L^*_{\xi_n}$$

([4]. Theorem 6.1.1, 6.1.2, 6.1.3) and so $Q = P^{(n)}$. Here $P^{(n)}_{P^0(\tau_n(\cdot))}$ is martingale measure at $\tau_n(\cdot)$ and θ_x is defined by $P(\cdot, \theta_s\omega) = P(\cdot + s, \omega)$. But this means that $P^* = P^*_\xi$. That is, P^*_ξ is the one and only $P^* \sim L^*_\xi$. □

Definition 4. We say that L^* is symmetric with respect to μ if

(1) $\int fL^*gd\mu = \int gL^*fd\mu, \quad f,g \in K^*$

(2) $\int L^*fd\mu = 0, \qquad f \in K^*.$

If L^* is symmetric, we know that μ is a reversible measure for the process determined by L^*. We are about to show that for any $\xi \in K^*_{loc}$, the measure μ_ξ given by

$$\mu_\xi(dx) = e^{2\xi(x)}\mu(dx)$$

is a reversible measure for the process determined by L^*_ξ. The archetypical example of this phenomenon is the case when $L^* = \frac{1}{2}d^2/dx^2$ and $\xi(x) = -x^2/2$. The corresponding processes are then Brownian motion and Ornstein-Uhlenbeck process and the measure are Lebesgue measure and Gauss measure.

Corollary 5. *If L^* is symmetric with respect to μ, then for each ξ,*

(1) $\int fL^*_\xi gd\mu_\xi = \int gL^*_\xi fd\mu_\xi, \quad f,g \in K^*$

(2) $\displaystyle\int L_\xi^* f d\mu_\xi = 0, \qquad f \in K^*.$

Proof. Set $\eta = e^{2\xi}$. Then $\eta \in K_{loc}^*$ and $\eta f \in K^*$ for all $f \in K^*$. Moreover, $\langle \eta, f\rangle = 2\eta\langle \eta, f\rangle$ for all $f \in K^*$. Hence

$$\begin{aligned} e^{2\xi} L_\xi^* f &= \eta L^* f + \frac{1}{2}\langle \eta, f\rangle = \frac{1}{2}\Big(2\eta L^* f + \langle \eta, f\rangle)\Big) \\ &= \frac{1}{2}\Big(\eta L^* f + L^*(\eta f) - f L^* \eta\Big) \end{aligned}$$

for all $f \in K^*$. We therefore have

$$\begin{aligned} \int g L_\xi^* f d\mu_\xi &= \frac{1}{2}\int (\eta g L^* f + g L^*(\eta f) - f g L^* \eta) d\mu \\ &= \frac{1}{2}\int (f L^*(\eta g) + \eta f L^* g - f g L^* \eta) d\mu \\ &= \int f L_\xi^* g d\mu_\xi. \end{aligned}$$

Also, by a simple approximation argument, $\displaystyle\int \eta L^* f d\mu = \int f L^* \eta d\mu$ and so $\displaystyle\int L_\xi^* f d\mu_\xi = 0.$ □

References

1. S.N.Either, *A class of degenerate diffusion processes occurring in population genetics,* Comm. Pure Appl. Math., **29** (1976), 483-493.
2. J.H.Gillespie, *Natural selection for within-generation variance in offspring number,* Genetics, **76** (1974), 601-606.
3. A.Shimizu, *Stationary distribution of a diffusion process taking values in probability distributions on the partitions,* Proceedings of a Workshop held in Nagoya, (1985), 100-114.
4. D. Stroock and S. Varadhan, *Multidimensional diffusion processes,* Springer-Verlag, (1979).

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 402-749, Korea
e-mail: choiwon@incheon.ac.kr