Korean J. Math. **22** (2014), No. 1, pp. 29–36
http://dx.doi.org/10.11568/kjm.2014.22.1.29

# ON THE MARTINGALE EXTENSION OF LIMITING DIFFUSION IN POPULATION GENETICS

Won Choi

Abstract. The limiting diffusion of special diploid model can be defined as a discrete generator for the rescaled Markov chain. Choi( [2] ) defined the operator of projection $S_t$ on limiting diffusion and new measure $dQ = S_t dP$. and showed the martingale property on this operator and measure. Let $P_\rho$ be the unique solution of the martingale problem for $\mathcal{L}_0$ starting at $\rho$ and $\pi_1, \pi_2, \cdots, \pi_n$ the projection of $E^n$ on $x_1, x_2, \cdots, x_n$. In this note we define

$$dQ_\rho = S_t dP_\rho$$

and show that $Q_\rho$ solves the martingale problem for $\mathcal{L}_\pi$ starting at $\rho$.

## 1. Introduction

Let $E$ (a locally compact separable metric space) be the set of all possible alleles and $\nu_0$(in $\mathcal{P}(E)$ , the set of Borel probability measures on $E$) the distribution of the type of a new mutant. Suppose that $N$(a positive integer) is the diploid population size and $s(\mathbf{x})$ is the selection coefficient of allele $\mathbf{x}$.

Received July 26, 2013. Revised September 28, 2013. Accepted September 28, 2013.

2010 Mathematics Subject Classification: 92D10, 60H30, 60G44.

Key words and phrases: Diploid model, Limiting diffusion, Martingale problem, Population genetics.

This research was supported by Incheon National University Research Grant, 2013-2014.

We now consider the normal-selection model which define in W. Choi( [1]). The type space $E$ is unspecified. However, $\nu_0$ and the function $s$ must jointly satisfy the following condition; If $X$ is a random variable with distribution $\nu_0$, then $s(X)$ has the normal distribution with mean 0 and variance $\sigma^2$. Furthermore, $\sigma = \sigma_0/2N$ for an appropriate constant $\sigma_0$. There are therefore a number of possible choice for $E$, $\nu_0$, and $s$, including;

$$E = (0,1),\ \nu_0 = U(0,1),\ s(\mathbf{x}) = \sigma\mathbf{\Phi}^{-1}(\mathbf{x}),$$

where $\Phi$ is the standard normal distribution function,

$$E = R,\ \nu_0 = N(0,\sigma^2),\ s(\mathbf{x}) = \mathbf{x},$$

and

$$E = R,\ \nu_0 = N(0,\sigma_0^2),\ s(\mathbf{x}) = \mathbf{x}/2N.$$

For each positive integer $M$, let $\omega_M$ be a positive, symmetric, bounded, Borel function on $E^2$, let $R_M((p,q), dx \times dy)$ be an one-step transition function on $E^2 \times \mathcal{B}(E^2)$ satisfying

$$R_M((p,q), dx \times dy) = R_M((q,p), dy \times dx),$$

and $Q_M(p, dx)$ be an one-step transition function on $E \times \mathcal{B}(E)$.

Let $N$ be the diploid population size. We consider $M = 2N$ gametes and the mapping $\eta_M : E^M \to \mathcal{P}(E)$ by letting

$$\eta_M(p_1, p_2, \cdots, p_M) = \frac{1}{M}(\delta_{p_1} + \delta_{p_2} + \cdots + \delta_{p_M}).$$

Here $\delta_p \in \mathcal{P}(E)$ denotes the unit mass at $p \in E$. The state space for this model is

$$\mathcal{K}_M(E) = \eta_M(E^M).$$

Given $\mu \in \mathcal{P}(E)$, we define $\mu_1 \in \mathcal{P}(E^2)$ and $\mu_2, \mu_3 \in \mathcal{P}(E)$ by

$$\mu_1(dp \times dq) = \omega_M(p,q)\mu^2(dp \times dq)/\langle \omega_M, \mu^2 \rangle,$$

$$\mu_2(dx) = \int_{E^2} R_M((p,q), dx \times E)\mu_1(dp \times dq),$$

$$\mu_3(dx) = \int_E Q_M(p, dx)\mu_2(dp).$$

The Markov chain has one-step transition function $P_M(\mu, d\theta)$ on $\mathcal{K}_M(E) \times (\mathcal{K}_M(E))$ defined by

$$P_M(\mu, \cdot) = \int_{E^M} (\mu_3)^M (dp_1 \times dp_2 \times \cdots \times dp_M)\delta_{\eta_M(p_1,p_2,\cdots,p_M)}(\cdot).$$

Choi( [1]) identified and characterized the limiting diffusion of this diploid model by defining discrete generator for the rescaled Markov chain. Also he defined the operator of projection $S_t$ on limiting diffusion and new measure $dQ = S_t dP$, and showed the martingale property on this operator and measure. ( [2]) Let $P_\rho$ be the unique solution of the martingale problem for $\mathcal{L}_0$ starting at $\rho$ and $\pi_1, \pi_2, \cdots, \pi_n$ the projections of $E^n$ on $x_1, x_2, \cdots, x_n$. In this note we define

$$dQ_\rho = S_t dP_\rho$$

and show that $Q_\rho$ solves the martingale problem for $\mathcal{L}_\pi$ starting at $\rho$.

## 2. Main Results

We define the discrete generator $\mathcal{L}_M$ for the $M$-the rescaled Markov chain and canonical coordinate process $\{\rho_t, t \geq 0\}$ :

$$(\mathcal{L}_M \phi)(\rho_t) = M \int_{\mathcal{P}_M} (\phi(\nu_t) - \phi(\rho_t)) P_M(\rho_t, \nu_t)$$

where $\mathcal{P}_M$ is given in the diploid models as described above.

We restrict our attention to test functions $\theta$ of the form

$$\theta(\nu_t) = \beta_1 \langle f_1, \nu_t \rangle \cdots \beta_k \langle f_k, \nu_t \rangle, \;\; \theta(\rho_t) = \langle f_1, \rho_t \rangle \cdots \langle f_k, \rho_t \rangle$$

where $f_1, \cdots, f_k \in \mathcal{B}(E)$ and $\{\beta_i\}$ is a set of non-negative constants satisfying that $\sup_i \beta_i < +\infty$. Assume that "mutation or gene conversion rate" is

$$\sum_{k \in S} \beta_k \langle f_i, \rho_t \rangle - \beta_i - \beta_j \;\; \text{for} \;\; \text{every} \;\; i < j,$$

in the diploid models as described above. This means the mutations or gene conversions occur with particular rate in case of $i < j$. See [3].

We start with;

Lemma 1. *Suppose that there exist a selection function $\sigma$ on $E^2$ and bounded linear operator $A, B$ on $\mathcal{B}(E)$ such that*

$$\omega_M(p,q) = 1 + \frac{1}{M}\sigma(p,q) + o(\frac{1}{M}),$$

$$\int_S f(x) R_M((p,q), dx \times S) = f(p) + \frac{1}{M}(Bf)(p,q) + o(\frac{1}{M}),$$

$$\int_S f(x) Q_M(p, dx) = f(x) + \frac{1}{M}(Af)(p) + o(\frac{1}{M}).$$

*Then there exist* $a_{f_i,f_j},\ b_{f_i} \in \mathcal{B}(\mathcal{P}(E))$ *such that*

$$\lim_{M \to \infty}(\mathcal{L}_M\theta)(\rho_t) = (\mathcal{L}_\pi\theta)(\rho_t) =$$

$$\sum_{1\le i<j\le k} a_{f_i,f_j} F_{z_i z_j}(\langle \mathbf{f}, \rho_t\rangle) + \sum_{i=1}^{k} b_{f_i} F_{z_i}(\langle \mathbf{f}, \rho_t\rangle)$$

*uniformly in* $\rho_t \in \mathcal{K}_M(E)$, *where* $F_{z_i}$ *and* $F_{z_i z_j}$ *mean the partial derivative with respect to* $i$ *and* $i, j$, *respectively. Here*

$$\theta(\rho_t) = F(\langle f_1, \rho_t\rangle, \langle f_2, \rho_t\rangle, \cdots, \langle f_k, \rho_t\rangle) = F(\langle \mathbf{f}, \rho_t\rangle)$$

$$a_{f_i,f_j} = \beta_i\langle f_i f_j, \rho_t\rangle - \langle f_i, \rho_t\rangle\langle f_j, \rho_t\rangle(\sum_{k\in S}\beta_k\langle f_i, \rho_t\rangle - \beta_i - \beta_j)$$

$$b_{f_i} = \langle A f_i, \rho_t\rangle + \langle B f_i, \rho_t^2\rangle + \langle (f_i \circ \pi)\sigma, \rho_t^2\rangle - \langle f_i, \rho_t\rangle\langle \sigma, \rho_t^2\rangle,$$

*and* $\pi$ *is the projection of* $E^2$.

*Proof.* See [1]. □

In particular, the set of possible alleles, known as the type space, is a locally compact, separable metric space $E$ and the mutation operator $A$ is given

$$Af = \frac{1}{2}\theta(\langle f, \nu_0\rangle - f),$$

where $\theta > 0$.

Let $\pi_1, \pi_2, \cdots, \pi_n$ be the projection of $E^n$ on $x_1, x_2, \cdots, x_n$-coordinate, respectively. Define

$$S_t^{\pi_x, \pi_x+\pi_y} = \exp\{\langle \pi_y, \rho_t\rangle - \langle \pi_y, \rho_0\rangle - \int_0^t e^{-\langle \pi_y \cdot \rho_s\rangle}\mathcal{L}_{\pi_x} e^{\langle \pi_y, \rho_s\rangle} ds\}$$

where $x, y = 1, 2, \cdots, n,\ x \neq y$. For $\rho$, we denote by $P_\rho$ the unique solution of the martingale problem for $\mathcal{L}_0$ (i.e., the distribution of the neutral model) starting at $\rho_0$.

Theorem 2 allows us to define a mean one $\{\mathcal{F}_t\}$-martingale.

THEOREM 2. *Suppose* $\{\mathcal{F}_t\}$ *is corresponding filtration with respect to topology of uniform convergence on compact sets. For* $0 < \delta < \delta_0$, *there exists* $\delta_0$ *such that*

$$E^{P_\rho}\left[\frac{S_{t+\delta}^{0,\pi}}{S_t^{0,\pi}}|\mathcal{F}_t\right] = 1.$$

*Proof.* Define $\pi_K = (-K) \vee (\pi \wedge K),\ K = 1, 2, \cdots, n$ and note that $\{S_t^{0,\pi_K}\}$ is mean-one martingale from [2]. Hence

$$E^{P_\rho}\left[\frac{S_{t+\delta}^{0,\pi_K}}{S_t^{0,\pi_K}} e^{\langle \pi_K, \rho_t\rangle + \frac{1}{2}\theta\delta\langle \pi_K, \nu_0\rangle} | \mathcal{F}_t\right] = e^{\langle \pi_K, \rho_t\rangle + \frac{1}{2}\theta\delta\langle \pi_K, \nu_0\rangle}.$$

But,

$$\begin{aligned}\exp\left(\langle \pi_K, \rho_t\rangle + \frac{1}{2}\theta\delta\langle \pi_K, \nu_0\rangle\right) &\le \exp\left(\langle \pi_K, \rho_t\rangle + \frac{1}{2}\theta \int_t^{t+\delta} \langle \pi_K, \rho_s\rangle ds\right) \\ &\le \exp\left(\langle \pi_1, \rho_{t+\delta}\rangle + \frac{1}{2}\theta \int_t^{t+\delta} \langle \pi_1, \rho_s\rangle ds\right).\end{aligned}$$

We apply the dominated convergence theorem to show that the right hand side of above inequality is integrable. Let $q = \frac{p}{p-1}$ and define $\delta_0 = 2p/\theta q$. By the Hölder and Jensen inequalities,

$$\begin{aligned}&E^{P_\rho}[\exp\left(\langle \pi_1, \rho_{t+\delta}\rangle + \frac{1}{2}\theta \int_t^{t+\delta} \langle \pi_1, \rho_s\rangle ds\right)] \\ &\le (E^{P_\rho}[\exp(p\langle \pi_1, \rho_{t+\delta}\rangle)])^{1/p} (E^{P_\rho}\left[\exp\left(\frac{1}{2}\theta q \int_t^{t+\delta} \langle \pi_1, \rho_s\rangle ds\right)\right])^{1/q} \\ &\le E^{P_\rho}[\langle e^{p\pi_1}, \rho_{t+\delta}\rangle]^{1/p} \left(\frac{1}{\delta}\int_t^{t+\delta} E^{P_\rho}[\langle e^{\theta q\delta\pi_1/2}, \rho_s\rangle] ds\right)^{1/q}\end{aligned}$$

Let $\{U(t)\}$ be the semigroup on $\mathcal{B}(E)$ with generator $A$. Since

$$E^{P_\rho}[\langle g, \rho_t\rangle] = \langle U(t)g, \rho\rangle \le \langle g, \rho\rangle \vee \langle g, \nu_0\rangle$$

for $g \in \mathcal{B}(E)$, we have

$$\begin{aligned}&E^{P_\rho}[\exp\left(\langle \pi_1, \rho_{t+\delta}\rangle + \frac{1}{2}\theta \int_t^{t+\delta} \langle \pi_1, \rho_s\rangle ds\right)] \\ &\le [\langle e^{p\pi_1}, \rho\rangle \vee \langle e^{p\pi_1}, \nu_0\rangle]^{1/p} [\langle e^{\theta q\delta\pi_1/2}, \rho\rangle \vee \langle e^{\theta q\delta\pi_1/2}, \nu_0\rangle]^{1/q} \\ &\le \langle e^{p\pi_1}, \rho\rangle \vee \langle e^{p\pi_1}, \nu_0\rangle\end{aligned}$$

if $0 < \delta < \delta_0$, and the proof is complete. □

Theorem 2 allows us to define $Q_\rho$ by

$$dQ_\rho = S_t^{0,\pi} dP_\rho.$$

Choi( [2]) proved that $Q_\rho^K$ solves the martingale problem for $\mathcal{L}_{\pi_K}$ starting at $\rho_0$. In advance, we now show that $Q_\rho$ solve the martingale problem for $\mathcal{L}_\pi$ starting at $\rho_0$.

THEOREM 3. *The measure $Q_\rho$ is a solution of the $(E^n, \mathcal{L}_\pi, \rho_0)$-martingale problem.*

*Proof.* Define

$$M_t^\pi = \phi(\rho_t) - \phi(\rho_0) - \int_0^t (\mathcal{L}_\pi \phi)(\rho_s) ds$$

and

$$M_t^{\pi_K} = \phi(\rho_t) - \phi(\rho_0) - \int_0^t (\mathcal{L}_{\pi_K} \phi)(\rho_s) ds$$

for $\phi \in \mathcal{D}(\mathcal{L}_\pi)$. Then $M_t^{\pi_K}$ is $Q_\rho^K$-martingale for $Q_\rho^K$ from the result of [2] with $\pi$ replaced by $\pi_K$ and we have

$$E^{Q_\rho^K}[M_{t+\delta}^{\pi_K} - M_t^{\pi_K} | \mathcal{F}_t] = 0.$$

Hence

$$E^{P_\rho}[(M_{t+\delta}^{\pi_K} - M_t^{\pi_K}) S_{t+\delta}^{0,\pi_K} | \mathcal{F}_t] = 0$$

and

$$E^{P_\rho}[(M_{t+\delta}^{\pi_K} - M_t^{\pi_K}) \frac{S_{t+\delta}^{0,\pi_K}}{S_t^{0,\pi_K}} e^{\langle \pi_K, \rho_t \rangle + \frac{1}{2}\theta\delta\langle \pi_K, \nu_0 \rangle} | \mathcal{F}_t] = 0.$$

Note that the integrand in above equation is bounded by the argument used for Theorem 2. For such $\delta$ used at Theorem 2, we conclude that

$$E^{P_\rho}[(M_{t+\delta}^{\pi} - M_t^{\pi}) \frac{S_{t+\delta}^{0,\pi}}{S_t^{0,\pi}} e^{\langle \pi, \rho_t \rangle + \frac{1}{2}\theta\delta\langle \pi, \nu_0 \rangle} | \mathcal{F}_t] = 0.$$

On the other hand,

$$E^{P_\rho}[(M_t^{\pi_K})^2 S_t^{0,\pi_K}] = E^{Q_\rho^K}[(M_t^{\pi_K})^2] = E^{Q_\rho^K}[\langle\langle M\pi_K \rangle\rangle_t] \le Ct$$

where $\langle\langle M\pi_K \rangle\rangle_t = \int_0^t \psi(\pi_s) ds$ is increasing process and $C$ is a constant with

$$\psi(\pi_s) = \sum_{i,j=1}^k (\beta_i \langle f_i f_j, \rho_t \rangle - \langle f_i, \rho_t \rangle \langle f_j, \rho_t \rangle (\sum_k \beta_k \langle f_i, \rho_t \rangle - \beta_i - \beta_j)) F_{z_i z_j}(\langle \mathbf{f}, \rho_t \rangle) \le C.$$

From the Fatou's lemma, we have

$$E^{P_\rho}[(M_t^\pi)^2 S_t^{0,\pi}] \le Ct$$

and

$$E^{Q_\rho}[(M_t^\pi)^2] \le Ct.$$

Therefore the integrands in below equations are integrable and we conclude that

$$E^{P_\rho}[(M_{t+\delta}^\pi - M_t^\pi) S_{t+\delta}^{0,\pi} | \mathcal{F}_t] = 0$$

and

$$E^{Q_\rho^K}[M_{t+\delta}^\pi - M_t^\pi | \mathcal{F}_t] = 0.$$

□

By Theorem 3, we know that there exists a probability measure $Q_\rho$ satisfying the following conditions ;

(1) $Q_\rho(\rho(0) = \rho_0) = 1$ and
(2) denoting $M_{\phi_1}(t) = \phi_1(\rho(t)) - \int_0^t \mathcal{L}_\pi \phi_1(\rho(s))ds$, $M_{\phi_1}(t)$ is a $Q_\rho$-martingale.

Therefore we conclude with;

Corollary 4. *Defining*

$$\langle \phi_1, \phi_2 \rangle \equiv \mathcal{L}_\pi(\phi_1 \cdot \phi_2) - \phi_1 \mathcal{L}_\pi \phi_2 - \phi_2 \mathcal{L}_\pi \phi_1,$$

$$(M_{\phi_1}(t))^2 - \int_0^t \langle \phi_1, \phi_2 \rangle (\rho(s))ds$$

*is a $Q_\rho$-martingale.*

*Proof.* Since the measure $Q_\rho$ is a solution of $L_\pi$-martingale problem, the result directly follows from quadratic covariation process. □

## References

[1] W.Choi, *On the limiting diffusion of special diploid model in population genetics*, Bull. Korean Math. Soc. **42** (2) (2005), 397–404.
[2] W.Choi, *On the martingale property of limiting diffusion in special diploid model*, J. Appl. Math. info. **31** (1) (2013), 241–246.
[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.

Won Choi
Department of Mathematics
Incheon National University
Incheon 406-772, Korea
*E-mail*: `choiwon@incheon.ac.kr`