One-dimensional h-path processes

Miyuki MAENO

Graduate School of Human Culture, Nara Women's University

1 One-dimensional generalized diffusion processes

Let $\overline{\mathbb{R}} = [-\infty, +\infty]$ and m be a nondecreasing right continuous function from $\overline{\mathbb{R}}$ into $\overline{\mathbb{R}}$. We set $\ell_1 = \inf\{x \in \overline{\mathbb{R}};\ m(x) > -\infty\}$, $\ell_2 = \sup\{x \in \overline{\mathbb{R}};\ m(x) < \infty\}$, $S_m = (\ell_1,\ \ell_2)$. Let s be a real valued continuous increasing function on S_m, and k be a real valued right continuous nondecreasing function on S_m. We set

$$S_*(\mu) = \{x \in S_m;\ \mu(x_1) < \mu(x_2) \text{ for } \ell_1 < {}^\forall x_1 < x < {}^\forall x_2 < \ell_2\}, \quad \mu = m \text{ or } k.$$

We assume $S_*(m) \neq \emptyset$ and $S_*(k) \subset S_*(m)$. Further we set

$$S_{**}(m) = S_*(m) \cup \{x;\ x = \ell_i \text{ with } |m(\ell_i)| + |s(\ell_i)| + |k(\ell_i)| < \infty,\ i = 1, 2\}.$$

We introduce the following two quantities.

$$J_{\mu,\nu}^{\ell_1} = \int_{(\ell_1,c]} d\mu(x) \int_{(x,c]} d\nu(y), \quad J_{\mu,\nu}^{\ell_2} = \int_{[c,\ell_2)} d\mu(x) \int_{[c,x)} d\nu(y),$$

where $d\mu$ and $d\nu$ are Borel measures on S_m.

Definition 1 For each $i = 1, 2$, ℓ_i is called to be

(s, m, k)-regular	if $J_{s,m+k}^{\ell_i} < \infty,\ J_{m+k,s}^{\ell_i} < \infty$,
(s, m, k)-exit	if $J_{s,m+k}^{\ell_i} < \infty,\ J_{m+k,s}^{\ell_i} = \infty$,
(s, m, k)-entrance	if $J_{s,m+k}^{\ell_i} = \infty,\ J_{m+k,s}^{\ell_i} < \infty$,
(s, m, k)-natural	if $J_{s,m+k}^{\ell_i} = \infty,\ J_{m+k,s}^{\ell_i} = \infty$.

Let $C_b(E)$ be the set of all bounded continuous functions on E, where E is a Borel set.

Definition 2 Let $D(\mathcal{G})$ be the space of all functions u in $C_b(S_m)$ satisfying the following conditions.

($\mathcal{G}$.1) There are a function $f \in C_b(S_*(m))$ and two constants A_1, A_2 such that

$$(1) \qquad u(x) = A_1 + A_2\{s(x) - s(c)\} + \int_{(c,x]} \{s(x) - s(y)\} f(y)\, dm(y) + \int_{(c,x]} \{s(x) - s(y)\} u(y)\, dk(y), \quad x \in S_m.$$

($\mathcal{G}$.2) For each $i = 1, 2$, if ℓ_i is regular, then $u(\ell_i) = 0$.

The operator $\mathcal{G}$ is defined by the mapping from $u \in D(\mathcal{G})$ to $f \in C_b(S_*(m))$ appeared in (1). The operator $\mathcal{G}$ is called a one-dimensional generalized diffusion operator (ODGDO for brief) with (s, m, k).

It is known that there exists a process $\mathbb{D} = [X(t);\, t \geq 0,\ P_x;\, x \in S_{**}(m)]$ with the generator $\mathcal{G}$, which is called a one-dimensional generalized diffusion process (ODGDP for brief) on S_m ([8], [13]). Further there exists a positive continuous function $p(t, x, y)$ satisfying

$$P_x(X(t) \in E) = \int_E p(t, x, y)\, dm(y), \quad t > 0,\ x \in S_{**}(m),\ E \in \mathcal{B}(S_m).$$

We call $p(t, x, y)$ the transition probability density function with respect to m.

2 One-dimensional generalized h-path processes

Let $\mathbb{D} = [X(t);\, t \geq 0,\ P_x;\, x \in S_{**}(m)]$ be the ODGDP with a generator $\mathcal{G}$ with $(s, m, 0)$. For $\alpha > 0$ and $i = 1, 2$, let $g_i(\cdot, \alpha)$ be a function on S_m satisfying the following properties.

(g.1) $g_i(x, \alpha)$ is positive and continuous in x.

(g.2) $g_1(x, \alpha)$ is nondecreasing in x and $g_2(x, \alpha)$ is nonincreasing in x.

(g.3) If $|s(\ell_i)| < \infty$, then $g_i(\ell_i, \alpha) = 0$.

(g.4) $g_i(x, \alpha)$ satisfies

$$\begin{aligned} g_i(x, \alpha) =& g_i(c, \alpha) + (D_s g_i)(c, \alpha)\{s(x) - s(c)\} \\ &+ \alpha \int_{(c,x]} \{s(x) - s(y)\} g_i(y, \alpha)\, dm(y), \quad x \in S_m, \end{aligned}$$

where $D_s f(x) = \lim_{\varepsilon \downarrow 0} \{f(x+\varepsilon) - f(x)\}/\{s(x+\varepsilon) - s(x)\}$. We set $W(\alpha) = (D_s g_1)(x, \alpha) g_2(x, \alpha) - g_1(x, \alpha)(D_s g_2)(x, \alpha)$. Note that $W(\alpha)$ is a positive number independent of $x \in S_m$.

Definition 3 We set

$$G(\alpha, x, y) = G(\alpha, y, x) = W(\alpha)^{-1} g_1(x, \alpha) g_2(y, \alpha), \quad \alpha > 0,\ x, y \in S_m,\ x \leq y.$$

We call $G(\alpha, x, y)$ the Green function corresponding to the ODGDO $\mathcal{G}$ with $(s, m, 0)$.

It is known that the Green function $G(\alpha, x, y)$ corresponding to $\mathcal{G}$ coincides with the Laplace transform of $p(t, x, y)$, that is,

$$G(\alpha, x, y) = \int_0^\infty e^{-\alpha t} p(t, x, y)\, dt, \quad \alpha > 0,\ x, y \in S_{**}(m).$$

([8], [11]).

Definition 4 We call h a *superharmonic function on* S_m *with respect to* (s, m) if and only if h is positive and continuous on S_m and has the right derivative $D_s h$ which is right continuous and nonincreasing on S_m, and the set $\{x \in S_m;\ D_s h(x_1) > D_s h(x_2) \text{ for } \ell_1 < {}^\forall x_1 < x < {}^\forall x_2 < \ell_2\}$ is included in $S_*(m)$.

For a superharmonic function h on S_m with respect to (s, m), we set

$$m^h(x) = \begin{cases} -\infty & \text{if } x \in (-\infty, \ell_1), \\ \int_{(c,x]} h(y)^2\, dm(y) & \text{if } x \in [\ell_1, \ell_2), \\ \infty & \text{if } x \in [\ell_2, \infty), \end{cases} \qquad x \in \mathbb{R},$$

$$s^h(x) = \int_{(c,x]} h(y)^{-2}\, ds(y), \quad x \in S_m,$$

$$k^h(x) = -\int_{(c,x]} h(y)\, dD_s h(y), \quad x \in S_m.$$

Let $\mathcal{G}^h$ be the ODGDO with (s^h, m^h, k^h) and $\mathbb{D}^h$ be an ODGDP with the generator $\mathcal{G}^h$, which is called a one-dimensional h-path generalized diffusion process. We set

$$G^h(\alpha, x, y) = G^h(\alpha, y, x) = G(\alpha, x, y)/h(x)h(y), \quad \alpha > 0,\ x, y \in S_m,\ x \le y.$$

Theorem 5 The function $G^h(\alpha, x, y)$ is the Green function corresponding to the ODGDO $\mathcal{G}^h$ with (s^h, m^h, k^h).

By virtue of the uniqueness of Laplace transform, we obtain that

$$p^h(t, x, y) = p(t, x, y)/h(x)h(y), \quad t > 0,\ x, y \in S_{**}(m).$$

The states of the end point ℓ_1 are described in Tables 1, which suggests the behavior of sample paths of $\mathbb{D}^h$ near the end points of S_m.

Table 1

state of ℓ_1 \ $\inf S_*(k^h)$ \ $h(\ell_1)$	$= 0$		$\in (0,\infty)$		$= \infty$	
	$> \ell_1$	$= \ell_1$	$> \ell_1$	$= \ell_1$	$> \ell_1$	$= \ell_1$
$(s, m, 0)$-regular	(s^h, m^h, k^h)- entrance	(s^h, m^h, k^h)- entrance if $J^{\ell_1}_{k^h, s^h} < \infty$ (s^h, m^h, k^h)- natural if $J^{\ell_1}_{k^h, s^h} = \infty$	(s^h, m^h, k^h)- regular	(s^h, m^h, k^h)- regular if $D_s h(\ell_1) \in (-\infty, \infty)$ (s^h, m^h, k^h)- exit if $D_s h(\ell_1) = \infty$		
$(s, m, 0)$-exit	(s^h, m^h, k^h)- entrance	(s^h, m^h, k^h)- entrance if $J^{\ell_1}_{m^h+k^h, s^h} < \infty$ (s^h, m^h, k^h)- natural if $J^{\ell_1}_{m^h+k^h, s^h} = \infty$	(s^h, m^h, k^h)- exit	(s^h, m^h, k^h)-exit		
$(s, m, 0)$-entrance			(s^h, m^h, k^h)- entrance	(s^h, m^h, k^h)- entrance	(s^h, m^h, k^h)- regular if $J^{\ell_1}_{m^h, s^h} < \infty$ (s^h, m^h, k^h)- exit if $J^{\ell_1}_{m^h, s^h} = \infty$	(s^h, m^h, k^h)- regular if $J^{\ell_1}_{s^h, m^h+k^h} < \infty$ and $J^{\ell_1}_{m^h+k^h, s^h} < \infty$ (s^h, m^h, k^h)- exit if $J^{\ell_1}_{s^h, m^h+k^h} < \infty$ and $J^{\ell_1}_{m^h+k^h, s^h} = \infty$ (s^h, m^h, k^h)- natural if $J^{\ell_1}_{s^h, m^h+k^h} = \infty$ and $J^{\ell_1}_{m^h+k^h, s^h} = \infty$
$(s, m, 0)$-natural	(s^h, m^h, k^h)- natural	(s^h, m^h, k^h)- natural	(s^h, m^h, k^h)- natural	(s^h, m^h, k^h)- natural	(s^h, m^h, k^h)- natural	(s^h, m^h, k^h)- natural

3 Application to population genetics

We consider a locus with two alleles in a randomly mating population of N diploid individuals. We denote by A_1 the wide-type allele and by A_2 the mutant allele. Let $X(n)$ be the relative frequency (gene frequency) of A_1 at the n-th generation in the population ($n = 0, 1, 2, \cdots$). Mutation, selection and random genetic drift are the factors which change gene frequency $X(n)$. The Wright-Fisher model and the stochastic selection model are the fundamental stochastic models in population genetics. The Wright-Fisher model is a stochastic model due to random genetic drift and this stochastic force has no correlation between distinct generations. On the other hand, in the stochastic selection model stochastic force of selection has autocorrelation from generation to generation in general. These models are described by discrete time stochastic processes because we regard the generation as the time unit. It is difficult, however, to analyze these discrete time models. Then diffusion approximations are employed for the original discrete time models. In other words, we approximate a discrete time stochastic process in population genetics by an appropriate diffusion process by introducing a new time scaling ([2], [7]). A general stochastic model may be obtained by combining these diffusion models. We will deal with a diffusion process $\mathbb{D} = [X(t);\, t \geq 0,\ P_x;\, x \in I]$ that is the diffusion model with random genetic drift and stochastic selection, where $I = (0, 1)$. Further we introduce two deterministic factors of mutation and selection in this diffusion model.

It is known that the generator of the diffusion process $\mathbb{D}$ is given by

$$\begin{aligned} L &= \frac{1}{2}a(x)\frac{d^2}{dx^2} + b(x)\frac{d}{dx}, \\ a(x) &= \frac{1}{2N}x(1-x) + \gamma x^2(1-x)^2, \\ b(x) &= v - (u+v)x + \frac{\gamma}{2}\rho x(1-x)(1-2x) \\ &\qquad + \{(S_{11} - 2S_{12} + S_{22})x + S_{12} - S_{22}\}x(1-x) \end{aligned}$$

(see [6]). The meaning of each variable and parameter are as follows. The variable x is the gene frequency of A_1 ($0 \leq x \leq 1$). The parameter N is the population size ($1 \leq N \leq \infty$). Note that the case that $N = \infty$ corresponds to that without random genetic drift. Three genotypes A_1A_1, A_1A_2 and A_2A_2 have fitnesses $1 + w_n + S_{11}$, $1 + \frac{1}{2}w_n + S_{12}$ and $1 + S_{22}$ in the original discrete time model. Here w_n is the stochastic part of selection parameters at the n-th generation, and S_{11}, S_{12} and S_{22} are the deterministic part of selection parameters ($\min\{w_n + S_{11}, \frac{1}{2}w_n + S_{12}, S_{22}\} \geq -1$). It is assumed that stochastic selection has no dominance. We assume that $\{w_n :\ 0, \pm 1, \pm 2, \cdots\}$

is a discrete time stationary process with the mean $E[w_n] = 0$. The parameter $\gamma = \sum_{k=-\infty}^{\infty} E[w_o w_k]/4$ is a degree of autocorrelated stochastic selection ($0 \leq \gamma < \infty$). The parameter ρ denotes the type of stochastic selection ($\rho \geq 1$). The case that $\rho = 1$ with $N < \infty$ is called the TIM model ([12]) and the case that $\rho > 1$ with $N = \infty$ is called the SAS-CFF model ([5]). The mutation rate per generation from A_1 to A_2 [resp. from A_2 to A_1] is denoted by u [resp. v] ($u,\ v \geq 0$). Here we set

$$s_o(x) = \exp\left\{-2\int_c^x \frac{b(y)}{a(y)}\,dy\right\}, \qquad m_o(x) = \frac{2}{a(x)}\exp\left\{2\int_c^x \frac{b(y)}{a(y)}\,dy\right\},$$
$$s(x) = \int_c^x s_o(y)\,dy, \qquad m(x) = \int_c^x m_o(y)\,dy.$$

The states of the end points 0 and 1 are described in Tables 2 and 3.

Table 2 The state of the end point 0

	$\lvert s(0+)\rvert$	$\lvert m(0+)\rvert$	state
$N < \infty,\ \ v = 0$	$< \infty$	$= \infty$	exit
$N < \infty,\ \ 0 < 4Nv < 1$	$< \infty$	$< \infty$	regular
$N < \infty,\ \ 4Nv \geq 1$	$= \infty$	$< \infty$	entrance
$N = \infty,\ \ v > 0$	$= \infty$	$< \infty$	entrance
$N = \infty,\ \ v = 0,\ \ u < S_{12} - S_{22} + \gamma(\rho - 1)/2$	$= \infty$	$< \infty$	natural
$N = \infty,\ \ v = 0,\ \ u = S_{12} - S_{22} + \gamma(\rho - 1)/2$	$= \infty$	$= \infty$	natural
$N = \infty,\ \ v = 0,\ \ u > S_{12} - S_{22} + \gamma(\rho - 1)/2$	$< \infty$	$= \infty$	natural

Table 3 The state of the end point 1

	$s(1-)$	$m(1-)$	state
$N < \infty,\ \ u = 0$	$< \infty$	$= \infty$	exit
$N < \infty,\ \ 0 < 4Nu < 1$	$< \infty$	$< \infty$	regular
$N < \infty,\ \ 4Nu \geq 1$	$= \infty$	$< \infty$	entrance
$N = \infty,\ \ u > 0$	$= \infty$	$< \infty$	entrance
$N = \infty,\ \ u = 0,\ \ v < S_{12} - S_{11} + \gamma(\rho - 1)/2$	$= \infty$	$< \infty$	natural
$N = \infty,\ \ u = 0,\ \ v = S_{12} - S_{11} + \gamma(\rho - 1)/2$	$= \infty$	$= \infty$	natural
$N = \infty,\ \ u = 0,\ \ v > S_{12} - S_{11} + \gamma(\rho - 1)/2$	$< \infty$	$= \infty$	natural

Let us consider the following five conditions.

(i) $N < \infty, \quad 0 \leq 4Nu < 1, \quad 0 \leq 4Nv < 1.$
In the case that $0 < 4Nu < 1$, the end point 1 is assumed to be absorbing.

(ii) $N < \infty, \quad 0 \leq 4Nu < 1, \quad 4Nv \geq 1.$
In the case that $0 < 4Nu < 1$, the end point 1 is assumed to be absorbing.

(iii) $N = \infty, \quad u = 0, \quad v > \max\{0, S_{12} - S_{11} + \gamma(\rho - 1)/2\}$,

(iv) $N = \infty, \quad u = v = 0, \quad S_{11} > S_{12} + \gamma(\rho - 1)/2 \geq S_{22}$.

(v) $N = \infty, \quad u = v = 0, \quad \min\{S_{11}, S_{12}\} > S_{12} + \gamma(\rho - 1)/2$.

Under these conditions, there exists the following limit distribution,

$$(2) \quad P^*(t, x, E) = \lim_{n\to\infty} P_x\left(X(t) \in E \mid \sigma_{\eta_n} < \sigma_{\xi_n}\right), \quad t > 0, \ x \in I, \ E \in \mathcal{B}(I),$$

for any sequences $\{\xi_n\}$, $\{\eta_n\}$ satisfying $\xi_n \downarrow 0$, $\eta_n \uparrow 1$ as $n \to \infty$, where σ_y be the first hitting time at $y \in I$. We set

$$\kappa(x) = \begin{cases} \dfrac{s(x) - s(0+)}{s(1-) - s(0+)}, & \text{if (i) or (v) is satisfied,} \\ 1, & \text{otherwise,} \end{cases} \qquad 0 \leq x \leq 1.$$

Theorem 6 Assume one of (i), (ii), (iii), (iv), (v). Then $P^*(t, x, E)$, $t > 0$, $x \in I$, $E \in \mathcal{B}(I)$, is the transition probability of $\mathbb{D}^\kappa$ whose generator is given by $\mathcal{G}^\kappa$. If (i) or (v) is satisfied, then $\mathcal{G}^\kappa$ is given by

$$\mathcal{G}^\kappa = \frac{1}{2}a(x)\frac{d^2}{dx^2} + b^\kappa(x)\frac{d}{dx},$$

where $b^\kappa(x) = b(x) + s_o(x)\{s(x) - s(0)\}^{-1}a(x)$. If one of (ii), (iii) or (iv) is satisfied, then $\mathcal{G}^\kappa = L$.

Proposition 7 If (i), (ii) or (iii) is satisfied, the end point 0 is $(s^\kappa, m^\kappa, 0)$-entrance. If (iv) or (v) is satisfied, the end point 0 is $(s^\kappa, m^\kappa, 0)$-natural. If (i) or (ii) is satisfied and $u = 0$, the end point 1 is $(s^\kappa, m^\kappa, 0)$-exit. If (i) or (ii) is satisfied and $u > 0$, the end point 1 is $(s^\kappa, m^\kappa, 0)$-regular. If (iii), (iv) or (v) is satisfied, the end point 1 is $(s^\kappa, m^\kappa, 0)$-natural.

The distributions like (2) sometimes appear in population genetics. Namely, let $\mathbb{D} = [X(t);\, t \geq 0,\ P_x;\, x \in I]$ be a diffusion approximation of gene frequency process. If the end points 0 and 1 are accessible, then the conditional distribution $P_x\left(X(t) \in E \mid \sigma_1 < \sigma_0\right)$ stands for the distribution of frequency of a certain allele under the condition that it reaches fixation before it disappears from the population. Such distributions and related topics are studied by Ewens [3], [4] and Karlin and Taylor [9] in some cases.

References

[1] J. L. Doob, Conditional Brownian motion and the boundary limits of harmonic functions, *Bull. Soc. Math. France* **85** (1957), 431–458.

[2] Ethier, S. N. and Kurtz, T. G., Markov Processes: Characterization and Convergence, John Wiley & Sons, New York, 1986.

[3] Ewens, W. J., Conditional diffusion processes in population genetics, *Theor. Pop. Biol.* 4 (1973), 21–30.

[4] Ewens, W. J., Mathematical Population Genetics, Springer-Verlag, New York, 1979.

[5] Gillespie, J. H., A general model to account for enzyme variation in natural populations V. The SAS-CFF model, Theor. Pop. Biol. **14** (1978), 1–45.

[6] Gillespie, J. H., The Causes of Molecular Evolution, Oxford Univ. Press, 1991.

[7] Iizuka, M. and Matsuda, H., Weak convergence of discrete time non-Markovian processes related to selection models in population genetics, J. Math. Biology **15** (1982), 107–127.

[8] K. Itô and H. P. McKean, Jr., *Diffusion Processes and their Sample Paths*, Springer-Verlag, New York, 1974.

[9] Karlin, S. and Taylor, H. M., *A Second Course in Stochastic Processes*, Academic Press, New York, 1981.

[10] H. P. McKean, Jr., Elementary solutions for certain parabolic differential equations, *Trans. Amer. Math. Soc.* **82** (1956), 519–548.

[11] Minami, N., Ogura, Y. and Tomisaki, M., Asymptotic behavior of elementary solutions of one-dimensional generalized diffusion equations, *Ann. Probab.* **13** (1985), 698–715.

[12] Takahata, N., Ishii, K. and Matsuda, H., Effect of temporal fluctuation of selection coefficient on gene frequency in a population, Proc. Natl. Acad. Sci. USA **72** (1975), 4541–454.

[13] S. Watanabe, On time inversion of one-dimensional diffusion processes, *Z. Wahrsch. verw. Geb.* **31** (1975), 115–124.