Stochastic Dynamics Associated with the Poisson-Dirichlet Distribution and the Dirichlet Process

Shui Feng
McMaster University

July 19-23, 2010

The Seventh Workshop on Markov Processes and Related Topics at Beijing Normal University

- GEM Distribution
- Poisson-Dirichlet Distribution
- Dirichlet Process
- GEM Process
- Infinite-Allele Model
- Fleming-Viot Process

1. GEM Distribution

For $0 \leq \alpha < 1, \theta > -\alpha$, let $U_1, U_2, \ldots$ be independent, and $U_i \sim Beta(1-\alpha, \theta + i\alpha)$. Set

$$V_1 = U_1, V_n = (1-U_1)\cdots(1-U_{n-1})U_n, n \geq 2.$$

The law of $(V_1, V_2, \ldots)$ is called the **GEM distribution**.

2. Poisson-Dirichlet Distribution

The law of the descending order statistics of $V_1, V_2, \ldots$ is called the **two-parameter Poisson-Dirichlet distribution**, denoted by $PD(\alpha, \theta)$.

The case $\alpha = 0$ corresponds to Kingman's Poisson-Dirichlet distribution.

3. Dirichlet Process

Let $\xi_1, \xi_2, \ldots$ be iid with common diffuse measure ν on $[0,1]$, and independently, $(P_1(\alpha,\theta), P_2(\alpha,\theta), \ldots)$ follows the two-parameter Poisson-Dirichlet distribution. The random measure on $[0,1]$

$$\Xi_{\alpha,\theta,\nu}(dx) = \sum_{i=1}^{\infty} P_i(\alpha,\theta)\delta_{\xi_i}(dx)$$

is called the **two-parameter Dirichlet process**, denoted by $\Pi_{\alpha,\theta,\nu}$.

Let

$$
\begin{aligned}
&[0,1]^\infty = \{(x_1, x_2, \ldots) : x_i \in [0,1], i = 1, 2, \ldots\},\\
&\triangle = \{\mathbf{p} = (p_1, p_2, \ldots) \in [0,1]^\infty : \sum_{i=1}^{\infty} p_i \leq 1\},\\
&\nabla = \{(p_1, p_2, \ldots) \in \triangle : p_1 \geq p_2 \geq \cdots \geq 0\},\\
&M_1([0,1]) = \text{the set of all probabilities on } [0,1].
\end{aligned}
$$

Then $PD(\alpha, \theta)$ is a probability on ∇, the GEM distribution is probability on $\triangle$, and the Dirichlet process is a probability on $M_1([0,1])$.

4. GEM Process

For $0 \leq \alpha < 1, \theta + \alpha \geq 1$ and any $i \geq 1$, set

$$c_i = \frac{1-\alpha}{2}, d_i = \frac{\theta + i\alpha}{2}.$$

For independent Brownian motions $B_1(t), B_2(t), \ldots$, consider the following SDE

$$dx_i(t) = (c_i - (c_i + d_i)x_i(t))dt + \sqrt{x_i(t)(1 - x_i(t))}dB_i(t).$$

The diffusion $x_i(t)$ takes values in $[0, 1)$ and is symmetric with reversible measure $Beta(1 - \alpha, \theta + i\alpha)$.

Consider the map

$$\Phi : [0,1]^\infty \to \triangle, \ (x_1, x_2, \ldots) \mapsto (p_1, p_2, \ldots),$$

where

$$p_1 = x_1, \ p_n = (1 - x_1) \cdots (1 - x_{n-1}) x_n, \ n \geq 2.$$

The **GEM process** is the $\triangle$-valued diffusion given by $\Phi(x_1(t), x_2(t) \ldots)$. Let $C^\infty_{cl}([0,1]^\infty)$ be the set of all bounded, C^∞ cylindrical functions on $[0,1]^\infty$ and

$$\mathcal{D} = \{ f|_\triangle : f \in C^\infty_{cl}([0,1]^\infty) \}.$$

For any f in $\mathcal{D}$, the generator of the GEM process is

$$\mathcal{L}f(\mathbf{p}) = \frac{1}{2}\sum_{i,j=1}^{\infty} a_{ij}(\mathbf{p})\frac{\partial^2 f}{\partial p_i \partial p_j} + \sum_{i=1}^{\infty} b_i(\mathbf{p})\frac{\partial f}{\partial p_i},$$

where

$$a_{ij}(\mathbf{p}) = p_i(\delta_{ij} - p_j) + p_i p_j \sum_{k=1}^{i-1} \frac{p_k}{\hat{p}_k} + \delta_{ij} p_i \hat{p}_{i-1},$$

$$b_i(\mathbf{p}) = p_i \sum_{k=1}^{i} \left(\frac{(\delta_{ki}\hat{p}_{k-1} - p_k)(c_k \hat{p}_{k-1} - (c_k + d_k)p_k)}{p_k \hat{p}_k} \right),$$

where $\hat{p}_k = (1 - \sum_{l=1}^{k} p_l)$

Properties

For simplicity, denote $PD(\alpha, \theta)$ by μ. Then the Dirichlet form associated with the GEM process is given by

$$\mathcal{E}(f, g) = \langle \mu, f(-\mathcal{L}g) \rangle$$

with domain $\mathcal{D}(\mathcal{E})$.

Theorem 1. (F and Wang (07)) (1) *The GEM process is the unique Feller process generated by* $\mathcal{L}$*;*

(2) *The GEM process is symmetric with GEM distribution as the reversible measure;*

(3) *There exists* $c > 0$ *such that for any* f *in* $\mathcal{D}$

$$\langle \mu, f^2 \log f^2 \rangle \leq c\mathcal{E}(f,f) + \langle \mu, f^2 \rangle \log \langle \mu, f^2 \rangle,$$

which is the Log-Sobolev inequality.

Denote the semigroup of the GEM process by P_t. Then it follows from the Log-Sobolev inequality that for any g in the domain of the Dirichlet form satisfying $g \geq 0, \langle \mu, g \rangle = 1$, the relative entropy of $P_t(g)$ with respect to μ converges to zero exponentially fast as t tends to infinity.

5. Infinite-Allele Models

In Pitman (02), the GEM distribution and the Poisson-Dirichlet distribution in the case of $\alpha = 0$ are shown to be invariant distribution of certain Markov chains involving the split-and-merge transformations of an interval-partition. The case of $\alpha > 0$ is studied in Bertoin (08) where the dynamics involve an exchangeable fragmentation-coagulation process.

Our focus here is on an infinite-dimensional diffusion process.

For any $n \geq 1$, let

$$\phi_1(\mathbf{p}) = 1, \ \phi_n(\mathbf{p}) = \sum_{i=1}^{\infty} p_i^n, \ n \geq 2, \ \mathbf{p} \in \nabla$$

and

$$\mathcal{D}_0 = \text{algebra generated by } \{\phi_n : n \geq 1\}.$$

For $f \in \mathcal{D}_0$, set

$$\mathcal{L}_{\alpha,\theta} f(\mathbf{p}) = \frac{1}{2} \left\{ \sum_{i,j=1}^{\infty} p_i(\delta_{ij} - p_j) \frac{\partial^2 f}{\partial p_i \partial p_j} - \sum_{i=1}^{\infty} (\theta p_i + \alpha) \frac{\partial f}{\partial p_i} \right\}.$$

Theorem 2. (Petrov(09)) (1) *The generator $\mathcal{L}_{\alpha,\theta}$ defined on $\mathcal{D}_0$ is closable in $C(\nabla)$. The closure, also denoted by $\mathcal{L}_{\alpha,\theta}$ for notational simplicity, generates a unique ∇-valued diffusion process $X_{\alpha,\theta}(t)$, the two-parameter infinite-allele diffusion process;*

(2) *The process $X_{\alpha,\theta}(t)$ is reversible with respect to $PD(\alpha,\theta)$;*

(3) *The spectrum of $\mathcal{L}_{\alpha,\theta}$ consists of the eigenvalues $\{0, -\lambda_2, -\lambda_3, \ldots\}$ with*

$$\lambda_n = \frac{n(n-1+\theta)}{2}, n \geq 0.$$

The eigenvalue 0 is simple, and the multiplicity of $-\lambda_n$ for $n \geq 2$ is $\pi(n) - \pi(n-1)$ with $\pi(n)$ denoting the total number of partitions of integer n.

Theorem 3. (F and Sun(09)) (1) *The bilinear form*

$$\mathcal{E}_{\alpha,\theta}(f,g) = \frac{1}{2}\int_{\nabla} \sum_{i,j=1}^{\infty} p_i(\delta_{ij} - p_j)\frac{\partial f}{\partial p_i}\frac{\partial g}{\partial p_j}(\mathbf{p})dPD(\alpha,\theta)$$

is closable on $L^2(\nabla; PD(\alpha,\theta))$ *and the closure is a regular Dirichlet form. The symmetric Hunt process associated with the Dirichlet form coincides with the process in Theorem 2.*

(2). *For any* $k \geq 1$, *let*

$$C_k = \nabla \cap \{\sum_{i=1}^{k} p_i = 1\}, D_k = C_k \cap \{p_k > 0\}.$$

If $\theta + k\alpha < 1$, *then the process* $X_{\alpha,\theta}(t)$ *will hit any subset of* C_k *with* *non-zero* $(k-1)$*-dimensional Lebesgue measure; if* $\theta + k\alpha \geq 1$*, then* D_k *is not hit by* $X_{\alpha,\theta}$*.*

Remarks. Theorems 2 and 3 are generalizations of the results in Ethier and Kurtz (81), Ethier (92), and Schmuland (91). There is a fundamental change in boundary behavior when α changes from zero to a positive number. In Schmuland (91), all finite dimensional simplex will be hit as long as $\theta < 1$; for $\alpha > 0$, the condition $\theta + k\alpha$ puts additional restrictions on the dimension of the simplex that can be hit.

6. Fleming-Viot Process

Let S be a compact metric space, $C(S)$ be the set of continuous functions on S, $M_1(S)$ the space of probability measures on S equipped with the usual weak topology, and ν a diffuse probability in $M_1(S)$. Consider operator A of the form

$$Af(x) = \frac{\theta}{2} \int (f(y) - f(x))\nu(dy), \ f \in C(S).$$

Define

$$\mathcal{D} = \{u : u(\mu) = f(\langle \phi, \mu \rangle), \ f \in C_b^{\infty}(\mathbf{R}), \phi \in C(S), \mu \in M_1(S)\},$$

where $\langle \phi, \mu \rangle$ is the integration of ϕ with respect to μ and $C_b^{\infty}(\mathbf{R})$ denotes the set of all bounded, infinitely differentiable functions on $\mathbf{R}$.

Then the **Fleming-Viot process with neutral parent independent mutation** (FV process) is a pure atomic measure-valued Markov process with generator

$$\mathcal{A}u(\mu) = \langle A\delta u(\mu)/\delta\mu(\cdot), \mu\rangle + \frac{f''(\langle\phi,\mu\rangle)}{2}\langle\phi,\phi\rangle_\mu, \;\; u \in \mathcal{D},$$

where

$$\delta u(\mu)/\delta\mu(x) = \lim_{\varepsilon\to 0+} \varepsilon^{-1}\{u((1-\varepsilon)\mu + \varepsilon\delta_x) - u(\mu)\},$$
$$\langle\phi,\psi\rangle_\mu = \langle\phi\psi,\mu\rangle - \langle\phi,\mu\rangle\langle\psi,\mu\rangle,$$

and δ_x stands for the Dirac measure at $x \in S$. It is known (Ethier (90)) that the Fleming-Viot process with parent independent mutation is reversible with the Dirichlet process $\Pi_{0,\theta,\nu}$ as the reversible measure.

Dirichlet Form Formulation

Set

$$\mathcal{F} := \mathrm{Span}\{\langle f_1, \mu\rangle \cdots \langle f_k, \mu\rangle : f_1, \ldots, f_k \in C(S), k \geq 1\}.$$

Consider the following symmetric bilinear form

$$\mathcal{E}_{FV}(u, v) = \frac{1}{2} \int \langle \frac{\delta u}{\delta \mu(\cdot)}, \frac{\delta v}{\delta \mu(\cdot)} \rangle_\mu \Pi_{0,\theta,\nu_0}(d\mu), \quad u, v \in \mathcal{F}.$$

The Fleming-Viot process is the symmetric Hunt process associated with this form.

Two-Parameter FV

For $0 < \alpha < 1$, set

$$\mathcal{E}^{\alpha}_{FV}(u,v) = \frac{1}{2}\int \langle \frac{\delta u}{\delta\mu(\cdot)}, \frac{\delta v}{\delta\mu(\cdot)}\rangle_{\mu} \Pi_{\alpha,\theta,\nu_0}(d\mu), \quad u,v \in \mathcal{F},$$

Natural Questions: Is the bilinear form $\mathcal{E}^{\alpha}_{FV}$ closable? Does it give rise to a regular Dirichlet form?

Positive answers to these questions will lead to a two-parameter Fleming-Viot process.

Two special cases are confirmed in F and Sun (09).

Special Case I

Assume that $\theta = 0, 0 < \alpha < 1, S = \{0, 1\}$ with $\beta = \nu(\{0\}), \hat{\beta} = 1 - \beta$. Then we have

Theorem 4. *The form $\mathcal{E}^{\alpha}_{FV}$ is closable and the closure is a regular Dirichlet form.*

The generator in this case has the form

$$\mathcal{A}_{\alpha,0} f(p) = \frac{1}{2} p(1-p) f^{''}(p) + \frac{\alpha}{2} f'(p) \left[(1-2p) - \right.$$
$$\left. \frac{2\hat{\beta}^2 p^{2\alpha}(1-p) - 2\beta^2(1-p)^{2\alpha} p + 2\beta\hat{\beta}(1-2p)p^{\alpha}(1-p)^{\alpha}\cos(\alpha\pi)}{\hat{\beta}^2 p^{2\alpha} + p^2(1-p)^{2\alpha} + 2\beta\hat{\beta} p^{\alpha}(1-p)^{\alpha}\cos(\alpha\pi)} \right]$$

Special Case II

Assume that $\theta > 0, 0 < \alpha < 1, S = \{0, 1\}$.

Theorem 5. *The form $\mathcal{E}_{FV}^{\alpha}$ is closable and the closure is a regular Dirichlet form.*

The generator in this case has the form

$$\begin{aligned}\mathcal{A}_{\alpha,\theta} f(p) &= \frac{1}{2} p(1-p) f''(p) + \frac{1}{2} f'(p) \left[(1-2p) \right. \\ &\quad \left. + p(1-p) g'_{\alpha,\theta}(p) / g_{\alpha,\theta}(p) \right],\end{aligned}$$

where $g_{\alpha,\theta}(p)$ is the density function of $\Xi_{\alpha,\theta}(\{0\})$.

General Case

The problem remains open! The results so far seem to suggest that the parameter α will only change the structure of the drift.

References

1. J. Bertoin (2008). Two-parameter Poisson–Dirichlet measures and reversible exchangeable fragmentation-coalescence processes. *Combin. Probab. Comput.* **17**, No. 3, 329–337.

2. S.N. Ethier (1990). The infinitely-many-neutral-alleles diffusion model with ages. *Adv. Appl. Probab.* **22**, 1–24.

3. S.N. Ethier (1992). Eigenstructure of the infinitely-many-neutral-alleles diffusion model. *J. Appl. Probab.* **29**, 487-498.

4. S.N. Ethier and T.G. Kurtz (1981). The infinitely-many-neutral-alleles diffusion model. *Adv. Appl. Probab.*, **13**, 429-452.

5. S. Feng and F.Y. Wang (2007). A class of infinite-dimensional diffusion processes with connection to population genetics. *J. Appl. Prob.* **44**, 938-949.

6. S. Feng and W. Sun (2009). Some diffusion processes associated with two parameter Poisson-Dirichlet distribution and Dirichlet process. *Probab. Theory Relat. Fields*, DOI 10.1007/s00440-009-0238-2.

7. L.A. Petrov (2009). Two-parameter family of infinite-dimensional diffusions on the Kingman simplex. *Funct. Anal. Appl.* **43**, No. 4, 279–296.

8. J. Pitman (2002). Poisson-Dirichlet and the GEM invariant distributions for the split-and-merge transformation of an interval-

partition. *Combin. Probab. Comput.* **11**, 501514.

9. B. Schmuland (1991). A result on the infinitely many neutral alleles diffusion model. *J. Appl. Probab.* **28**, 253-267.

THANK YOU!