Poisson-Dirichlet 分布及可逆测度值过程

献给王梓坤教授 90 华诞

冯水

Department of Mathematics and Statistics, McMaster University, Hamilton L8S 4K1, Canada

E-mail: shuifeng@mcmaster.ca

收稿日期: 2017-11-25; 接受日期: 2018-02-13; 网络出版日期: 2019-01-18

Natural Sciences and Engineering Research Council of Canada 资助项目

摘要 Poisson-Dirichlet 分布是定义在无穷维单纯形上的概率. 它刻画了一个取值为离散概率的随机变量的分布. 与 Poisson-Dirichlet 分布密切相关的随机测度包括 GEM (Griffiths-Engen-McCloskey) 分布、Dirichlet 过程、两参数 Dirichlet 过程和两参数 Poisson-Dirichlet 分布等. 构造与这些分布相应的测度值过程是近些年比较活跃的研究课题. 本文介绍近年来这方面的发展状况, 并给出一些待研究的问题.

关键词 Poisson-Dirichlet 分布 Dirichlet 过程 两参数 Dirichlet 过程 测度值过程 Dirichlet 型 可逆过程

MSC (2010) 主题分类 60F10, 92D10

1 引言

现代概率论和随机过程的一个主要研究领域是无穷维随机测度和测度值随机过程. 这方面的研究, 既是数学理论本身的自然延伸, 也与其他学科的进步密不可分. 经过近二三十年的发展, 相应的理论已比较成熟. 许多开发出的模型被应用到生物学、通信、经济学、金融、遗传学、统计物理、机器学习、Bayes 非参数统计等领域. 本文中讨论的无穷维随机测度主要包括 Poisson-Dirichlet 分布、Dirichlet 过程、GEM 分布及这些分布的两参数推广. 相应的测度值随机过程都是以这些分布为平稳分布的可逆测度值过程. 研究这些模型主要动因来源于 Bayes 非参数统计和群体遗传学.

无穷维随机测度作为先验分布, 在 Bayes 非参数统计上的应用非常自然、活跃. 实际应用中的一个主要任务, 是通过已有的信息对一些未知量进行估计和预测. 在很多情形下, 这些未知量包括概率分布本身. 无穷维随机测度的使用将独立性抽样推广到了可交换性抽样. 这虽然在理论框架上更加一般, 但同时增加了模拟的难度. 随着计算技术的飞速发展, 许多曾经被认为是纯理论的模型变成了实际工作中不可缺少的工具.

英文引用格式: Feng S. Reversible measure-valued processes associated with the Poisson-Dirichlet distribution (in Chinese). Sci Sin Math, 2019, 49: 377–388, doi: 10.1360/N012017-00253

测度值过程的研究在很大程度上得益于概率论与遗传学的联系. 特别是群体遗传学对测度值过程的研究起到了非常大的促进作用. 考虑一个生物群体, 在各种生物力的驱动下演进. 如果将生物群体中的个体按不同的指标特征分类, 那么这些指标特征在该群体中的分布变化可以用一个概率值随机过程刻画. 理论研究的一条主线是从 Markov 链到扩散逼近, 再推广到无穷维空间. 这其中的模型包括 Hardy-Weinberg 律、Wright-Fisher 模型、Feller 扩散逼近、Kingman 的合并模型、Moran 模型、踩石 (stepping-stone) 模型、自然选择模型、GEM 表示、Kingman 的划分结构和 Fleming-Viot 无穷维扩散等.

本文第 2 节将给出 Poisson-Dirichlet 分布、Dirichlet 过程、GEM 分布及这些分布的两参数推广的定义及相关背景. 第 3 节介绍有关的可逆测度值过程模型. 在作为结尾的第 4 节, 我们会列出两个有待研究的问题.

2 Poisson-Dirichlet 分布及相关的随机测度

本节给出 Poisson-Dirichlet 分布、GEM 分布和 Dirichlet 过程的定义以及这些分布的两参数推广. 同时介绍一些与这些分布相关的实际背景. 令 S 为任意一个 Polish 空间, 距离为 d . 设定

$$\begin{aligned}\Delta &= \left\{ \mathbf{x} = (x_1, x_2, \dots) : x_i \geq 0, \sum_{i=1}^{\infty} x_i = 1 \right\}, \\ \nabla_{\infty} &= \left\{ \mathbf{p} = (p_1, p_2, \dots) : p_1 \geq p_2 \geq \dots \geq 0, \sum_{i=1}^{\infty} p_i = 1 \right\}, \\ M_1(S) &= S \text{ 上的所有概率的集合, 赋予弱收敛拓扑.}\end{aligned}$$

任意给定 $\theta > 0$, 考虑一系列独立同分布 $\text{Beta}(1, \theta)$ 随机变量 $U_1, U_2, \dots$, 定义

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

令 $\mathbf{P}(\theta) = (P_1(\theta), P_2(\theta), \dots)$ 代表 $\mathbf{V}(\theta) = (V_1, V_2, \dots)$ 从大到小的顺序统计量. 给定 $\nu_0 \in M_1(S)$, 设 $\xi_1, \xi_2, \dots$ 为一列独立于 $\mathbf{V}(\theta)$ 的具有同分布 ν_0 的独立随机变量.

定义 2.1 随机概率 $\mathbf{P}(\theta)$ 的分布称为参数为 θ 的 Poisson-Dirichlet 分布, 记为 $\text{PD}(\theta)$; 随机概率 $\mathbf{V}(\theta)$ 的分布称为 GEM 分布, 记作 $\text{GEM}(\theta)$. Dirichlet 过程, 记为 $\mathcal{D}(\theta, \nu_0)$, 是下面随机测度的分布:

$$\Xi_{\theta, \nu_0} = \sum_{i=1}^{\infty} P_i(\theta) \delta_{\xi_i} = \sum_{i=1}^{\infty} V_i(\theta) \delta_{\xi_i},$$

其中 δ_x 为集中在 x 点的 Dirac 测度. 为了简单起见, 有时也称 Ξ_{θ, ν_0} 为 Dirichlet 过程.

从上述定义可看出, $\text{PD}(\theta)$ 、 $\text{GEM}(\theta)$ 和 $\mathcal{D}(\theta, \nu_0)$ 分别是空间 ∇_{∞} 、 Δ 和 $M_1(S)$ 上的概率. Dirichlet 过程这个名字中的“过程”代表的是空间而不是时间. 它与下一节所要讨论的随机过程中的“过程”含义不太一样.

Poisson-Dirichlet 分布是 Kingman 于 1975 年在文献 [1] 中引入的. 当时研究的一个背景来源于计算机科学中称作堆的一类数据结构. 其构造是通过 Gamma 过程 (增量服从 Gamma 分布的 Lévy 过程) 的跳跃进行归一化实现的. 参数 θ 代表 Gamma 过程的跳跃区间长度. 经过几十年的发展, Poisson-Dirichlet 分布已经成为许多领域中的基本模型. 下面是几个有代表性的例子.

例 2.2 (群体遗传学^[2]) 考虑一个大的生物群体及每一个个体染色体的一个固定位点. 假设遗传的过程只涉及中性的突变和随机漂移, 等位基因数无上界而且每一个突变都引入一个全新的基因, 那么当系统达到平衡时, 不同等位基因的比例从大到小排列则服从 Poisson-Dirichlet 分布. 参数 θ 对应于群体的突变速率.

例 2.3 (随机组合) 任意给定 $n \geq 1$, 集合 $[n] = \{1, \dots, n\}$ 的一个排列 σ 是 $[n]$ 到 $[n]$ 的 1-1 映射. 所有排列的集合记为 Σ_n . 对于每一个排列 σ , 定义

$$n_1 = \inf\{i \in [n] : \sigma^i(1) = 1\}.$$

集合 $\{1, \sigma(1), \dots, \sigma^{n_1-1}(1)\}$ 称为排列 σ 的从 1 开始的循环, 其长度为 n_1 . 类似地, 我们可以定义从集合 $[n] \setminus \{1, \sigma(1), \dots, \sigma^{n_1-1}(1)\}$ 最小数出发的循环, 并以此类推. 用 $K_n(\sigma)$ 代表 σ 所有循环的总数,

$$L_1(\sigma) \geq L_2(\sigma) \geq \dots \geq L_{K_n(\sigma)}(\sigma)$$

代表这些循环长度从大到小排序. 对于 $i > K_n(\sigma)$, 取 $L_i(\sigma) = 0$. 在 Σ_n 上先引入均匀分布, 然后对每一个 σ 用 $\theta^{K_n(\sigma)}$ 进行加权. 记由此所得 Σ_n 上的分布为 P_n . 那么可以证明文献 [3-5], 当 n 趋于无穷大时, $n^{-1}(L_1(\sigma), L_2(\sigma), \dots)$ 依分布收敛于 Poisson-Dirichlet 分布 $PD(\theta)$.

例 2.4 (随机数论) 任意给定 $N \geq 1$, 从 $[N] = \{1, \dots, N\}$ 中任意取一个数 X_N 并对其做如下素数分解:

$$X_N = \prod_q q^{C_q(X_N)},$$

其中 $C_q(X_N)$ 为素因子 q 的重数. 令 $q_1(X_N) \geq q_2(X_N) \geq \dots \geq q_N(X_N) \geq 1$ 为 X_N 的全部素因子 (可以重复), $q_i(X_N) = 1, i > \sum_q C_q(X_N)$. 定义 $A_i(X_N) = \log q_i(X_N)$. 可以证明 (参见文献 [6]) 当 N 趋于无穷大时, $(\log N)^{-1}(A_1(X_N), A_2(X_N), \dots)$ 依分布收敛于参数为 1 的 Poisson-Dirichlet 分布. 特别地, 当 N 很大时, 集合

$$\{1 \leq n \leq N : n \text{ 具有大于 } \sqrt{n} \text{ 的素因子}\}$$

中的元素总数大概是 $(\log 2)N$.

任意给定一列取值于 $[0, 1]$ 上的独立随机变量 $Z_1, Z_2, \dots$, 定义

$$Y_1 = Z_1, \quad Y_n = (1 - Z_1) \cdots (1 - Z_{n-1})Z_n, \quad n \geq 2.$$

随机向量 $\mathbf{Y} = (Y_1, Y_2, \dots)$ 及其分布称为残数配置模型 (residual allocation model, RAM).

显然, GEM 分布属于残数配置模型的一个子类. 它的构造可通过重复折断一根棍的过程来实现. 考虑一根长度为 1 的棍子, 赋予左端坐标为 0, 右端坐标为 1. 在棍子中间按 $\text{Beta}(1, \theta)$ 分布任取一点, 把它折断; 留下左边的一段, 将右边的一段归一化后再重复前面的折法. 这样下去, 该单位棍最终会被折成可数个小段. 把这些小段按折断的先后顺序排列, 所得向量则服从分布 $\text{GEM}(\theta)$.

McCloskey^[7] 给出了 GEM 分布的另一个特征. 考虑一个总体, 其中所有个体被分为可数类. 将这些类按比例从大到小进行标记. 比例最大的记为 1, 并以此类推. 假设这些类的比例从大到小的排列 $\mathbf{P}(\theta) = (P_1(\theta), P_2(\theta), \dots)$ 服从 Poisson-Dirichlet 分布 $PD(\theta)$. 从该总体中随机抽样, 并记 $\tau(k)$ 为第 k 次出现的类的标记, 那么,

$$\tau(\mathbf{P})(\theta) = (P_{\tau(1)}(\theta), P_{\tau(2)}(\theta), \dots)$$

称为以尺度偏置的排列 (size-biased permutation), 而且 $\tau(\mathbf{P})(\theta)$ 的分布为 GEM 分布.

如果定义

$$\bar{\Delta} = \left\{ \mathbf{x} = (x_1, x_2, \dots) : x_i \geq 0, \sum_{i=1}^{\infty} x_i \leq 1 \right\},$$

容易看出 GEM 分布可以自然地扩充到空间 $\bar{\Delta}$ 上.

Dirichlet 过程是 Ferguson^[8] 引入的. 它是有限维 Dirichlet 分布的无穷维推广. 我们知道, 对于任意 $m \geq 2$, $\alpha_1, \dots, \alpha_m > 0$, 有限维 Dirichlet 分布 $\mathcal{D}(\alpha_1, \dots, \alpha_m)$ 是定义在

$$\Delta_m = \left\{ (x_1, \dots, x_m) : x_i \geq 0, i = 1, \dots, m; \sum_{k=1}^m x_k = 1 \right\}$$

上具有密度

$$\frac{\Gamma(\alpha_1 + \dots + \alpha_m)}{\prod_{i=1}^m \Gamma(\alpha_i)} \prod_{i=1}^m x_i^{\alpha_i - 1}$$

的概率. 对于 S 的任意分割 $B_1, \dots, B_m$, 随机向量 $(\Xi_{\theta, \nu_0}(B_1), \dots, \Xi_{\theta, \nu_0}(B_m))$ 的分布为 $\mathcal{D}(\theta \nu_0(B_1), \dots, \theta \nu_0(B_m))$. 这个性质称为 Dirichlet 过程的划分性. 它在许多与 Dirichlet 过程有关的计算中起到关键作用. Dirichlet 过程的后验分布还是 Dirichlet 过程, 只改变了参数 (共轭性). 这一点使得它成为 Bayes 统计推断领域最重要的先验分布. 本文关于 Dirichlet 过程的定义来源于文献 [9].

1992 年, Pitman^[10] 把 Poisson-Dirichlet 分布推广到了两参数. 这个突破在随后的 20 多年中, 带来了很多相关的研究. 下面给出这类模型的定义.

任意给定 $0 \leq \alpha < 1$ 和 $\theta > -\alpha$, 令 $W_1, W_2, \dots$ 为一列独立随机变量. 对于每个 $k = 1, 2, \dots$, W_k 服从 $\text{Beta}(1 - \alpha, \theta + k\alpha)$ 分布. 设定

$$V_1(\alpha, \theta) = W_1, \quad V_n(\alpha, \theta) = (1 - W_1) \cdots (1 - W_{n-1}) W_n, \quad n \geq 2,$$

令 $\mathbf{P}(\alpha, \theta) = (P_1(\alpha, \theta), P_2(\alpha, \theta), \dots)$ 代表 $\mathbf{V}(\alpha, \theta) = (V_1(\alpha, \theta), V_2(\alpha, \theta), \dots)$ 从大到小的顺序统计量. 给定 $\nu_0 \in M_1(S)$, 设 $\xi_1, \xi_2, \dots$ 为一列独立于 $\mathbf{V}(\alpha, \theta)$ 的具有同分布 ν_0 的独立随机变量.

定义 2.5 随机概率 $\mathbf{P}(\alpha, \theta)$ 的分布称为两参数 Poisson-Dirichlet 分布, 记为 $\text{PD}(\alpha, \theta)$; 随机概率 $\mathbf{V}(\alpha, \theta)$ 的分布称为两参数 GEM 分布, 记作 $\text{GEM}(\alpha, \theta)$. 两参数 Dirichlet 或 Pitman-Yor 过程, 记为 $\mathcal{D}(\alpha, \theta, \nu_0)$, 是下面随机测度的分布:

$$\Xi_{\alpha, \theta, \nu_0} = \sum_{i=1}^{\infty} P_i(\alpha, \theta) \delta_{\xi_i} = \sum_{i=1}^{\infty} V_i(\alpha, \theta) \delta_{\xi_i}.$$

按惯例我们也称 $\Xi_{\alpha, \theta, \nu_0}$ 为两参数 Dirichlet 过程.

一个可数无穷随机离散分布称作具有 SBP (size-biased permutation) 不变性如果它的分布在以大小偏差的排列变换后保持不变. Pitman^[11] 证明了两参数 GEM 分布是满足 SBP 不变性的可数无穷随机离散分布的全体. 这充分反映出两参数模型的特殊性. 当然两参数模型不仅是数学结构上的推广, 更重要的是应用范围的扩大和延伸. 有关该领域近 20 年的发展, 读者可以参见文献 [12–14] 及其中的文献.

为了方便和清晰起见, 我们将本节讨论的 6 个分布及过程列表 (见表 1) 如下:

表 1 分布及过程

分布/过程	一参数	两参数
Poisson-Dirichlet	$PD(\theta)$	$PD(\alpha, \theta)$
GEM	$GEM(\theta)$	$GEM(\alpha, \theta)$
Dirichlet	$\mathcal{D}(\theta, \nu_0)$	$\mathcal{D}(\alpha, \theta, \nu_0)$

3 可逆测度值过程

本节介绍了 6 个分布的定义、背景、理论价值及一些应用. 在实际当中, 这些分布所刻画的往往是一些系统在平衡状态下的行为. 本节将介绍以这些分布为可逆测度的随机过程. 这些模型所刻画的则是相关系统在非平衡状态下的行为.

3.1 具有可逆分布 $PD(\theta)$ 的过程

定义 $\bar{\nabla}_\infty = \{\mathbf{p} = (p_1, p_2, \dots) : p_1 \geq p_2 \geq \dots \geq 0, \sum_{i=1}^\infty p_i \leq 1\}$. 以 $PD(\theta)$ 为平稳分布的过程通常取值于空间 ∇_∞ 或 $\bar{\nabla}_\infty$. 这就要求各个微观变化的结果仍然保持比例从大到小的排序.

3.1.1 CCF 过程

CCF (continuous coagulation-fragmentation) 过程的状态空间为 ∇_∞ . 给定 ∇_∞ 中任意元素 $\mathbf{p}$, 其一次大小偏差抽样 $\tilde{p}_1$ 满足

$$P\{\tilde{p}_1 = p_n\} = p_n, \quad n = 1, 2, \dots$$

CCF 过程的转移机制如下: 给定 $\mathbf{p}$, 做一次大小偏差抽样得 $\tilde{p}_1$, 还原后再做第二次大小偏差抽样得 $\tilde{p}_2$. 如果 $\tilde{p}_2 = \tilde{p}_1$, 那么在区间 $(0, \tilde{p}_1)$ 中任意去掉一点将 $\tilde{p}_1$ 分成两部分. 如果 $\tilde{p}_2 \neq \tilde{p}_1$, 那么将 $\tilde{p}_1$ 和 $\tilde{p}_2$ 合成一部分. 同时保持其他比例不变. 经过分裂或合并后, 与其他比例一起从大到小重新排列. 这样从 $\mathbf{p}$ 转移到 $\mathbf{p}'$. 由此得到的过程是取值于 ∇_∞ 的 Markov 过程.

Vershik 猜想^[15] CCF 过程的平稳分布唯一, 而且是参数为 1 的 Poisson-Dirichlet 分布. 文献 [16, 17] 证明了 $PD(1)$ 确实是 CCF 过程的平稳分布 (实际上是可逆的). 文献 [18] 给出了唯一性的证明.

3.1.2 IMNA 模型

IMNA (infinitely-many-neutral-alleles) 是文献 [19] 中引入的取值于 $\bar{\nabla}_\infty$ 的无穷维可逆扩散过程. 对任意整数 $n \geq 2$, 令 $\mathbf{X}_n = (X_1^n(t), \dots, X_n^n(t))$ 为 Δ_n 值的、具有无穷小算子

$$\mathcal{A}_{n,\theta} = \frac{1}{2} \left\{ \sum_{i,j=1}^n x_i(\delta_{ij} - x_j) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^n \left\{ -\theta x_i + \frac{\theta}{n-1}(1-x_i) \right\} \frac{\partial}{\partial x_i} \right\}$$

的 Wright-Fisher 扩散过程. 众所周知, 该过程是可逆的, 其可逆测度为 Dirichlet 分布 $\mathcal{D}(\frac{\theta}{n-1}, \dots, \frac{\theta}{n-1})$ (参见文献 [20]). 将 $\mathbf{X}_n = (X_1^n(t), \dots, X_n^n(t))$ 从大到小排序, 然后自然嵌入到空间 $\bar{\nabla}_\infty$. 文献 [19] 证明了当 n 趋于无穷时, 该过程序列依分布收敛到 IMNA 模型. 其可逆测度是 Poisson-Dirichlet 分布 $PD(0, \theta)$, 无穷小算子为

$$\mathcal{A}_\theta := \frac{1}{2} \left\{ \sum_{i,j=1}^\infty p_i(\delta_{ij} - p_j) \frac{\partial^2}{\partial p_i \partial p_j} - \sum_{i=1}^\infty \theta p_i \frac{\partial}{\partial p_i} \right\}.$$

令 $C(\bar{\nabla}_\infty)$ 代表 $\bar{\nabla}_\infty$ 上连续函数集合. 算子 $\mathcal{A}_\theta$ 的定义域是 $C(\bar{\nabla}_\infty)$ 中由函数集合 $\{1, \sum_{i=1}^\infty p_i^m, m = 2, 3, \dots\}$ 生成的子代数.

文献 [21] 利用 Dirichlet 型技巧给出了 IMNA 模型的另外一个构造, 同时刻画了过程轨道的行为.

3.2 具有可逆分布 $\text{PD}(\alpha, \theta)$ 的过程

本小节介绍的模型, 都与随机划分有关. 为此我们先给出有关的符号和定义. 令 $\mathbb{N}$ 代表自然数集合. $\mathbb{N}$ 上的一个划分 π 为 $\mathbb{N}$ 的一组至多可数不交非空子集 $\{\pi_i : i = 1, 2, \dots\}$ 满足 $\bigcup_i \pi_i = \mathbb{N}$. 每个集合 π_i 称为划分 π 的一个子块. 对任意自然数 n 和 $[n] = \{1, 2, \dots, n\}$, 划分 $\pi_n = \{[n] \cap \pi_i : i = 1, 2, \dots\}$ 表示 π 在 $[n]$ 上的限制. 用 $\mathcal{P}_\infty$ 和 $\mathcal{P}_n$ 分别代表 $\mathbb{N}$ 和 $[n]$ 上的所有划分构成的集合. $\mathbb{N}$ 上的一个随机划分 Π 是一个取值于 $\mathcal{P}_\infty$ 上的随机变量. Π 是可交换的, 如果它的分布在 $\mathbb{N}$ 上不同的排列下保持不变. 给定一个可交换的随机划分 $\Pi = \{\Pi_i : i = 1, 2, \dots\}$, 用 $|\Pi_i \cap [n]|$ 表示 $\Pi_i \cap [n]$ 中元素个数. 可以证明下列极限存在:

$$\lim_{n \rightarrow \infty} \frac{|\Pi_i \cap [n]|}{n} = X_i, \quad i = 1, 2, \dots$$

用 $\mathbf{P} = (P_1, P_2, \dots)$ 表示 $\{X_i : i = 1, 2, \dots\}$ 从大到小的排列. 这样每一个可交换的随机划分都对应于一个 $\bar{\nabla}_\infty$ 值的随机变量. 反过来对于 $\bar{\nabla}_\infty$ 中任意的 $\mathbf{p}$, 可通过 Kingman 的漆桶构造 [22] 得到一个可交换的随机划分 $\Pi(\mathbf{p})$. 更一般地, 所有可交换的随机划分都可以表示成漆桶构造的随机划分的凸组合.

3.2.1 EFC 过程

EFC (exchangeable fragmentation-coalecence) 过程是一类取可交换随机划分值的 Markov 过程. 这类过程的构造及性质可参见文献 [23]. 过程的演化涉及两个机制: 合并和分化. 这里的合并是指一个划分的两个或多个子块合并成一个大子块, 而分化则是指一个含有至少两个元素的子块被分为多个更小的子块. 合并和分化可通过两个分别定义在 $[0, 1]$ 和 $\bar{\nabla}_\infty$ 上的测度 Λ 和 ν 来刻画. 具体来讲, 给定一个 EFC 过程 Π_t 和任意 n , $\Pi_t^n = \Pi_t \cap [n]$ 是 $\mathcal{P}_n$ 值的连续时间 Markov 链, 而且 Π_t 由 $\{\Pi_t^n = \Pi_t \cap [n] : n = 1, 2, \dots\}$ 完全决定. 对于任意 $2 \leq k \leq m \leq n$ 和 $\mathcal{P}_n$ 中具有 m 个子块的划分 η , 令 η' 为合并 η 中 k 个子块成一个大子块后形成的划分. 从 η 到 η' 的转移速率为

$$\int_0^1 u^{k-2}(1-u)^{m-k} \Lambda(du).$$

另外, 设 $\{i_1, \dots, i_k\}$ 为划分 η 含 k 个元素的任意子块. 令 ζ 为 $\mathcal{P}_k$ 中含有至少两个子块的划分. 按 ζ 中的指标将 $\{i_1, \dots, i_k\}$ 分成更小的子块. 具体地讲, 对于 ζ 的每一个子块 A , $\{i_j : j \in A\}$ 为 $\{i_1, \dots, i_k\}$ 的一个更小的子块. 记由此分化而得的划分为 η' . 相应的转移速率为 $\int_{\bar{\nabla}_\infty} \nu(d\mathbf{p}) \text{P}\{\Pi^k(\mathbf{p}) = \zeta\}$. 由此得到的 EFC 过程由参数测度 Λ 和 ν 完全决定.

给定一系列独立同分布标准指数随机变量 $E_1, E_2, \dots$ 由文献 [24, 命题 10] 可知, 下述极限几乎处处存在:

$$L = \lim_{n \rightarrow \infty} \frac{n(\sum_{i=1}^n E_i)^{-1}}{(\sum_{k=1}^\infty (\sum_{j=1}^k E_j)^{-1/\alpha})^\alpha}.$$

下面取 $\Lambda(du) = (1-u)^{\theta/\alpha} du$, $\nu(d\mathbf{p}) = L^{-1} \text{PD}(\alpha, 0)(d\mathbf{p})$.

定理 3.1 [12] 以上述 Λ 和 ν 为参数测度的 EFC 过程是具有可逆分布 $\text{PD}(\alpha, \theta)$ 的 Markov 过程.

3.2.2 Petrov 扩散

给定任意 $n \geq 1$ 及 $\mathcal{P}_n$ 中的划分 η . 如果 η 的所有子块为 $\{\eta_1, \dots, \eta_k\}$ (显然 $k \leq n$), 那么, $\{|\eta_1|, \dots, |\eta_k|\}$ 是 n 的一个整数划分. 用 $|\eta|^\uparrow$ 表示这个划分从大到小的排序. 定义集合

$$\mathbb{K}_n = \{(|\eta|^\uparrow, 0, \dots) : \eta \in \mathcal{P}_n\}, \quad \mathbb{K} = \bigcup_n \mathbb{K}_n.$$

Petrov^[25] 研究了一列 $\mathbb{K}_n$ 值的 Markov 链 $\mathbf{X}_n(t)$, $n = 2, \dots$. 其转移机制涉及从 $\mathbb{K}_n$ 到 $\mathbb{K}_{n+1}$ 的上转移, 然后再从 $\mathbb{K}_{n+1}$ 回到 $\mathbb{K}_n$ 的下转移. 具体的转移速率依赖于 Ewens-Pitman 划分结构 (参见文献 [13, 14]). 由于 $\mathbb{K}_n/n$ 是 ∇_∞ 的子集, 当 n 趋于无穷大时 Markov 链 $\mathbf{X}_n(t)$ 经过扩散标度变换后收敛到一个无穷维扩散过程, 称为 Petrov 扩散. 该过程的状态空间是 $\overline{\nabla}_\infty$, 可逆测度为 $\text{PD}(\alpha, \theta)$, 其无穷小算子具有形式

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

与 IMNA 模型一样, 算子 $\mathcal{A}_{\alpha, \theta}$ 的定义域也是 $C(\overline{\nabla}_\infty)$ 中由函数集合 $\{1, \sum_{i=1}^{\infty} p_i^m, m = 2, 3, \dots\}$ 生成的子代数. 当 $\alpha = 0$ 时, 算子 $\mathcal{A}_{\alpha, \theta}$ 与 $\mathcal{A}_\theta$ 等同. Petrov 扩散的另外一个构造参见文献 [26], 使用的是 Dirichlet 型技巧. 值得指出的是, 对于每个 i , 函数 p_i 都不属于算子 $\mathcal{A}_{\alpha, \theta}$ 的定义域, 但属于 Dirichlet 型的定义域.

Petrov 扩散和 IMNA 模型与 Bayes 统计有着很自然的联系. 这方面的结果可参见文献 [27, 28].

IMNA 模型的自然背景是群体遗传学. 文献 [29] 通过引入一种特殊的选择机制, 第一次给出了 Petrov 扩散与群体遗传学的联系.

最近, 文献 [30] 引入了两类状态空间为区间划分的扩散过程. 其中两个过程分别以 $\text{PD}(1/2, 0)$ 和 $\text{PD}(1/2, 1/2)$ 为平稳分布. 这两种情形之所以特殊, 主要是因为相应的分布刻画的是 Brown 运动和 Brown 桥的游弋区间的分布行为 (参见文献 [31, 32]).

3.3 GEM 过程

任意给定两列正数 $\mathbf{c} = (c_i)_{i \geq 1}$, $\mathbf{d} = (d_i)_{i \geq 1}$, 令 $(B_t^i)_{i \geq 1}$ 为一列独立标准 Brown 运动. 对每一个 i , 考虑下列随机微分方程:

$$dX_t^i = \frac{1}{2}(c_i - (c_i + d_i)X_t^i)dt + \sqrt{X_t^i(1 - X_t^i)}dB_t^i.$$

我们知道 $\text{Beta}(c_i, d_i)$ 分布是上述扩散过程的可逆分布. 定义

$$Y_t^1 = X_t^1, \quad Y_t^n = (1 - X_t^1) \cdots (1 - X_t^{n-1})X_t^n, \quad n \geq 2$$

和

$$V^1 = U^1, \quad V^n = (1 - U^1) \cdots (1 - U^{n-1})U^n, \quad n \geq 2,$$

其中 $(U^i)_{i \geq 1}$ 为一列独立 $\text{Beta}(c_i, d_i)$ 随机变量.

文献 [33] 证明了, 如果 $\inf_i d_i \geq 1$, 那么过程 $(Y_t^1, Y_t^2, \dots)$ 是一个无穷维可逆扩散, 称为 GEM 过程. 其可逆测度为 RAM 模型 $(V^1, V^2, \dots)$ 的分布, 其无穷小算子为

$$\mathcal{A}^{\mathbf{c}, \mathbf{d}} = \frac{1}{2} \left[\sum_{i,j=1}^{\infty} a_{ij}(\mathbf{y}) \frac{\partial^2}{\partial y_i \partial y_j} + \sum_{i=1}^{\infty} b_i(\mathbf{y}) \frac{\partial}{\partial y_i} \right],$$

其中

$$a_{ij}(\mathbf{y}) := y_i y_j \sum_{k=1}^{i \wedge j} \frac{(\delta_{ki}(1 - \sum_{l=1}^{k-1} y_l) - y_k)(\delta_{kj}(1 - \sum_{l=1}^{k-1} y_l) - y_k)}{y_k(1 - \sum_{l=1}^k y_l)},$$

$$b_i(\mathbf{y}) := y_i \sum_{k=1}^i \frac{(\delta_{ik}(1 - \sum_{l=1}^{k-1} y_l) - y_k)(c_k(1 - \sum_{l=1}^{k-1} y_l) - (c_k + d_k)y_k)}{y_k(1 - \sum_{l=1}^k y_l)},$$

这里约定 $\sum_{i=1}^0 = 0$ 和 $\prod_{i=1}^0 = 1$. 如果进一步约定 $\frac{0}{0} = 1$, 那么 $\mathcal{A}^{c,d}$ 的定义域可以扩展到空间 $\bar{\Delta}$ 上.

特别地, 当 $c_i = 1 - \alpha$ 和 $d_i = \theta + i\alpha$ 时, 与此相应的 GEM 过程的可逆测度为 $\text{GEM}(\alpha, \theta)$. 这类模型在文献 [34] 中被用来研究依赖于时间的概率密度估计. 文献 [33] 还建立了 Poincaré 和 log-Sobolev 不等式. 另外, 文献 [35] 进一步得到了 GEM 过程的 Harnack 不等式.

3.4 可逆分布为 $\mathcal{D}(\theta, \nu_0)$ 的过程

本小节介绍可逆分布为 $\mathcal{D}(\theta, \nu_0)$ 的随机过程. 与上一节不同的是, 这类过程不仅涉及比例的演化, 更重要的是要考虑每个个体的指标或型的变化. 加入个体指标信息后, 过程的结构变得更加复杂而且增加了奇异性.

3.4.1 弱交互作用过程

这类模型的指标或型空间为 $S = \{1, 2, \dots, d\}$, 其中 d 可以有限也可以是无穷大. 有限指标空间的模型最早见于文献 [36]. 这里只考虑 $d = \infty$ 的情形. 这个模型来自文献 [37]. 记 $a_i = \theta \nu_0(\{i\})$, $i = 1, \dots, +\infty$. 考虑下列随机微分方程:

$$dX_i(t) = \frac{1}{2}(a_i(1 - |X(t)|_1) - a_\infty X_i(t))dt + \sqrt{X_i(t)(1 - |X(t)|_1)}dB^i(t), \quad i = 1, 2, \dots,$$

这里 $|X(t)|_1 = X_1(t) + X_2(t) + \dots$.

文献 [37] 证明了无穷维扩散 $(X_1(t), X_2(t), \dots)$ 是可逆的, 其可逆测度为 $\mathcal{D}(\theta, \nu_0)$. 同时刻画了生成元的谱, 得到了精确的谱空隙.

我们称这个模型为弱交互作用过程是相对于 Wright-Fisher 扩散而言. 在 Wright-Fisher 扩散中, 每两个个体之间都发生交互作用. 但是在弱交互作用过程中, 所有个体都只与指标为无穷大的个体 $X_\infty(t) = 1 - |X(t)|_1$ 发生交互作用.

3.4.2 Fleming-Viot 过程

本节假设 ν_0 没有原子. 也就是说指标空间 S 是不可数的, 而且对 S 中每一个元素 x , $\nu_0(\{x\}) = 0$. 为了简单起见, 我们进一步假设 S 是紧空间. 令 $C^\infty(\mathbb{R}^d)$ 代表 $\mathbb{R}^d$ 上无穷可微函数集. 算子 A 为一个 Markov 过程的无穷小生成元, 其定义域为 $D(A)$.

对任意 $\phi \in D(A)$, $f \in C^\infty(\mathbb{R})$, $\mu \in M_1(S)$, 令 $\langle \mu, \phi \rangle$ 表示函数 ϕ 关于 μ 的积分, 并定义 $M_1(S)$ 上的函数 $F(\mu) = f(\langle \mu, \phi \rangle)$. 我们考虑的 Fleming-Viot 过程 (参见文献 [38, 39]) 是一个取值于 $M_1(S)$ 的测度值过程. 其无穷小算子为

$$\mathcal{L}F(\mu) = f'(\langle \mu, \phi \rangle)\langle \mu, A\phi \rangle + \frac{1}{2} \int_S \int_S f''(\langle \mu, \phi \rangle)\phi(x)\phi(y)Q(\mu; dx, dy),$$

其中

$$Q(\mu; dx, dy) = \mu(dx)\delta_x(dy) - \mu(dx)\mu(dy).$$

算子的第一项代表突变, 第二项代表再抽样或随机漂移.

用 $B_b(S)$ 和 $B_b(S^2)$ 分别表示 S 上和乘积空间 S^2 上的有界可测函数集. 对任意的对称函数 $V(\cdot, \cdot) \in B_b(S^2)$, 定义

$$R(\mu, dx) := \int_S \left[\int_S V(y, z) \mu(dz) \right] Q(\mu; dx, dy).$$

带有选择的 Fleming-Viot 过程具有无穷小算子

$$\mathcal{L}_V F(\mu) = \mathcal{L}F(\mu) + f'(\langle \mu, \phi \rangle) \langle R(\mu), \phi \rangle.$$

称 $V(x, y)$ 为适应函数.

突变算子有许多选择. 其中与上一代独立的突变算子 (简称 PI (parent independent) 算子) 具有如下形式:

$$Ag(x) = \frac{\theta}{2} \int (g(y) - g(x)) \nu_0(dy), \quad g \in B_b(S).$$

下面的定理反应出 PI 算子的特殊性.

定理 3.2 假设突变算子 A 生成的半群是不可约的.

(1) (参见文献 [40, 41]) 突变算子为 PI 算子的 Fleming-Viot 过程是可逆的, 其唯一可逆测度为 $\mathcal{D}(\theta, \nu_0)$.

(2) (参见文献 [42]) 带有选择的 Fleming-Viot 过程可逆的充分必要条件是突变算子为 PI 算子. 相应的可逆测度为

$$Z^{-1} \exp\{V(\mu)\} \mathcal{D}(\theta, \nu_0)(d\mu),$$

其中 Z 为归一化常数, 同时, $V(\mu) = \int_S \int_S V(x, y) \mu(dx) \mu(dy)$.

在文献 [42] 中, 状态空间可以是局部紧的. 另外通过对拟不变性的研究, 文献 [43] 给出了上述结果不同的证明, 并将模型推广到带有重组的情形.

3.4.3 Wasserstein 扩散

当状态空间 $S = [0, 1]$ 时, Dirichlet 过程与 Wasserstein 扩散有非常紧密的关系. 定义

$$\mathcal{C} = \{h : h \text{ 是从 } [0, 1] \text{ 到 } [0, 1] \text{ 的非降函数, 在 } [0, 1] \text{ 上右连续同时满足 } h(1) = 1\}.$$

令 m 代表 $[0, 1]$ 上的均匀分布. 引入映射 $\chi : \mathcal{C} \rightarrow M_1([0, 1])$, $h \rightarrow m \circ \chi^{-1}$. 不难看出, 该映射为满单映射. 其逆映射为

$$\chi^{-1} : M_1([0, 1]) \rightarrow \mathcal{C}, \quad \mu \rightarrow h_\mu(t) = \inf\{s \in [0, 1] : \mu([0, s]) > t\}.$$

如果在 $\mathcal{C}$ 上赋予 $L^2(m)$ 范数, 在 $M_1(S)$ 上使用二阶 Wasserstein 距离 (生成的拓扑与弱拓扑相同), 那么映射 χ 是等距的. 这样, $\mathcal{C}$ 和 $M_1([0, 1])$ 上的结果基本等价.

考虑映射 $\iota : M_1([0, 1]) \rightarrow \mathcal{C}$, $\mu \rightarrow \tilde{h}_\mu(t) = \mu([0, t])$, 并在 $\mathcal{C}$ 上引入概率 $\mathcal{D}_\theta = \mathcal{D}(\theta, m) \circ \iota^{-1}$.

对于任意 $d \geq 1$, $x \in [0, 1]$, $f \in C_b^1(\mathbb{R}^d)$, $\phi_1, \dots, \phi_d \in L^2([0, 1])$, 定义 $\mathcal{C}$ 上函数

$$F(h) = f(\langle \phi_1, h \rangle, \dots, \langle \phi_d, h \rangle)$$

和梯度 (Fréchet 导数)

$$\mathbb{D}F(h)(x) := \sum_{i=1}^d \partial_i f(\langle \phi_1, h \rangle, \dots, \langle \phi_d, h \rangle) \phi_i(x),$$

其中 $\langle \cdot, \cdot \rangle$ 表示 $L^2([0, 1])$ 中的内积.

定理 3.3 ^[44] 存在取值于空间 $\mathcal{C}$ 的可逆扩散过程. 其可逆测度和 Dirichlet 型分别为 $\mathcal{D}_\theta$ 和

$$\mathbb{E}(F) = \int_{\mathcal{C}} \|\mathbb{D}F(h)\|_{L^2}^2 \mathcal{D}_\theta(dh).$$

由于 $\mathcal{C}$ 和 $M_1(S)$ 在映射 χ 下是等距的, 上述结果可自然地通过 χ 而得到空间 $M_1(S)$ 上的可逆扩散过程. 但是这样构造的过程的可逆测度并不是 Dirichlet 过程 $\mathcal{D}(\theta, m)$. 如果我们用 $\iota(M_1([0, 1]))$ 表示映射 ι 的值域, 那么在 $\iota(M_1([0, 1]))$ 上取映射 ι 的逆可以得到 $M_1(S)$ 上另一个可逆扩散过程 (参见文献 [45]). 这时的可逆测度确是 Dirichlet 过程 $\mathcal{D}(\theta, m)$.

3.5 可逆分布为 $\mathcal{D}(\alpha, \theta, \nu_0)$ 的扩散过程

前面我们看到 $\mathcal{D}(\alpha, \theta, \nu_0)$ 是 $\mathcal{D}(\theta, \nu_0)$ 的一个自然推广. 由于 $\mathcal{D}(\theta, \nu_0)$ 是 Fleming-Viot 过程的可逆测度, 因此, 很自然地我们要考虑能否构造可逆分布为 $\mathcal{D}(\alpha, \theta, \nu_0)$ 的扩散过程, 从而得到 Fleming-Viot 过程的两参数推广.

当状态空间 S 只含有两个元素时, 文献 [26] 给出了完整构造. 一般情形的研究, 进展很缓慢. 最近, 文献 [46] 考虑了特殊的参数区域 $\alpha = 1/2, \theta > -1/2$, 并通过 Dirichlet 型方法构造了相应的扩散过程. 具体来讲, 令 $\mathcal{F} = \{F(\mu) = f(\langle \mu, \phi_1 \rangle, \dots, \langle \mu, \phi_d \rangle) : \phi_i \in B_b(S), 1 \leq i \leq d, f \in C^\infty(\mathbb{R}^d), d \geq 1\}$, 定义梯度算子

$$\nabla_x F(\mu) := \left. \frac{dF}{ds}(\mu + s\delta_x) \right|_{s=0} = \sum_{i=1}^d \partial_i f(\langle \mu, \phi_1 \rangle, \dots, \langle \mu, \phi_d \rangle) \phi_i(x),$$

并用 $\nabla F(\mu)$ 表示函数 $x \rightarrow \nabla_x F(\mu)$. 引入下列 Dirichlet 型:

$$\mathcal{E}(F, G) = \frac{1}{2} \int_{\mathcal{M}_1(S)} \text{Cov}_\mu(\nabla F(\mu), \nabla G(\mu)) \mathcal{D}(\alpha, \theta, \nu_0)(d\mu), \quad F, G \in \mathcal{F},$$

其中 Cov_μ 表示关于 μ 的协方差.

如果 ν_0 的支撑是状态空间 S 的可数子集, 那么上述 Dirichlet 型决定一个可逆扩散. 可逆测度为 $\mathcal{D}(1/2, \theta, \nu_0)$. 如果 ν_0 没有原子, 那么可逆扩散可以通过对上述 Dirichlet 型进行放松后取 Mosco 极限 ^[47] 而得到.

4 两个问题

Dirichlet 过程、Poisson-Dirichlet 分布与 GEM 分布之间有着非常紧密的联系. 一个很自然的问题是, 类似的关系在相应的测度值过程之间是否也存在? 在 Fleming-Viot 过程与 IMNA 模型之间, 这方面的结果确实存在. 具体来讲 (参见文献 [13, 41]), Fleming-Viot 过程不同指标的比例从大到小排序后所得的过程就是 IMNA 模型. 但下面的问题还没有具体的答案.

问题 4.1 GEM 过程与 IMNA 模型的边缘分布之间的关系是什么?

上一节介绍了几种情形下, 以 $\mathcal{D}(\alpha, \theta, \nu_0)$ 为可逆测度的扩散过程. 这里包括状态空间只有两个元素的情形和状态空间一般但 $\alpha = 1/2$ 的情形. 下面的问题可参见文献 [13].

问题 4.2 对于一般的参数区域 $0 < \alpha < 1$, $\theta + \alpha > 0$, 以 $\mathcal{D}(\alpha, \theta, \nu_0)$ 为可逆测度的非平凡可逆 Markov 过程是否存在? 如何构造?

致谢 感谢王梓坤教授的启蒙和多年的帮助, 衷心祝贺先生 90 华诞.

参考文献

- 1 Kingman J F C. Random discrete distributions. J R Stat Soc Ser B Stat Methodol, 1975, 37: 1–22
- 2 Kingman J C F. Mathematics of Genetic Diversity. Philadelphia: SIAM, 1980
- 3 Arratia R, Barbour A D, Tavaré S. Logarithmic Combinatorial Structures: A Probabilistic Approach. EMS Monographs in Mathematics. Zürich: European Mathematical Society, 2003
- 4 Kingman J F C. The population structure associated with the Ewens sampling formula. Theor Popul Biol, 1977, 11: 274–283
- 5 Vershik A M, Shimdt A A. Limit measures arising in the asymptotic theory of symmetric groups, I. Theory Probab Appl, 1977, 22: 70–85
- 6 Billingsley P. On the distribution of large prime divisors. Period Math Hungar, 1972, 2: 283–289
- 7 McCloskey J W. A model for the distribution of individuals by species in an environment. PhD Thesis. East Lansing: Michigan State University, 1965
- 8 Ferguson T S. A Bayesian analysis of some nonparametric problems. Ann Statist, 1973, 1: 209–230
- 9 Sethuraman J A. A constructive definition of Dirichlet priors. Statist Sinica, 1994, 4: 639–650
- 10 Pitman J. The two-parameter generalization of Ewens' random partition structure. Technical Report 345. Berkeley: University of California, 1992
- 11 Pitman J. Random discrete distributions invariant under size-biased permutation. Adv in Appl Probab, 1996, 102: 525–939
- 12 Bertoin J. Two-parameter Poisson-Dirichlet measures and reversible exchangeable fragmentation-coalescence processes. Combin Probab Comput, 2008, 17: 329–337
- 13 Feng S. The Poisson-Dirichlet Distribution and Related Topics. Heidelberg: Springer, 2010
- 14 Pitman J. Combinatorial Stochastic Processes. New York: Springer-Verlag, 2006
- 15 Tsilevich N V. Stationary random partitions of positive integers. Theory Probab Appl, 2000, 44: 60–74
- 16 Gnedin A, Kerov S. A characterization of GEM distributions. Combin Probab Comput, 2001, 10: 213–217
- 17 Pitman J. Poisson-Dirichlet and GEM invariant distributions for split-merge transformations of an interval partition. Combin Probab Comput, 2002, 11: 501–514
- 18 Zerner M P W, Zeitouni O, Mayer-Wolf E, et al. The Poisson-Dirichlet law is the unique invariant distribution for uniform split-merge transformations. Ann Probab, 2004, 32: 915–938
- 19 Ethier S N, Kurtz T G. The infinitely-many-neutral-alleles diffusion model. Adv in Appl Probab, 1981, 13: 429–452
- 20 Wright S. Adaptation and selection. In: Genetics, Paleontology, and Evolution. Princeton: Princeton University Press, 1949, 365–389
- 21 Schmuland B. A result on the infinitely many neutral alleles diffusion model. J Appl Probab, 1991, 28: 253–267
- 22 Kingman J F C. The coalescent. Stochastic Process Appl, 1982, 13: 235–248
- 23 Berestycki J. Exchangeable fragmentation-coalescence processes and their equilibrium measures. Electron J Probab, 2004, 9: 770–824
- 24 Pitman J, Yor M. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. Ann Probab, 1997, 25: 855–900
- 25 Petrov L A. Two-parameter family of infinite-dimensional diffusions on the Kingman simplex. Funct Anal Appl, 2009, 43: 279–296
- 26 Feng S, Sun W. Some diffusion processes associated with two parameter Poisson-Dirichlet distribution and Dirichlet process. Probab Theory Related Fields, 2010, 148: 501–525
- 27 Ruggiero M. Species dynamics in the two-parameter Poisson-Dirichlet diffusion model. J Appl Probab, 2014, 51: 174–190
- 28 Ruggiero M, Walker S. Countable representation for infinite dimensional diffusions derived from the two-parameter Poisson-Dirichlet process. Electron Comm Probab, 2009, 14: 501–517
- 29 Costantini C, De Blasi P, Ethier S N, et al. Wright-Fisher construction of the two-parameter Poisson-Dirichlet diffusion.

- Ann Appl Probab, 2017, 27: 1923–1950
- 30 Forman N, Pal S, Rizzolo D, et al. Diffusions on a space of interval partitions with Poisson-Dirichlet stationary distributions. ArXiv:1609.06706, 2017
 - 31 Perman M, Pitman J, Yor M. Size-biased sampling of Poisson point processes and excursions. Probab Theory Related Fields, 1992, 92: 21–39
 - 32 Pitman J, Yor M. Arcsine laws and interval partitions derived from a stable subordinator. Proc Lond Math Soc (3), 1992, 3: 326–356
 - 33 Feng S, Wang F Y. A class of infinite-dimensional diffusion processes with connection to population genetics. J Appl Probab, 2007, 44: 938–949
 - 34 Mena R H, Ruggiero M. Dynamic density estimation with diffusive Dirichlet mixtures. Bernoulli, 2016, 22: 901–926
 - 35 Feng S, Wang F Y. Harnack inequality and applications for infinite-dimensional GEM processes. Potential Anal, 2016, 44: 137–153
 - 36 Jacobsen M. Examples of multivariate diffusions: Time-reversibility; a Cox-Ingersoll-Ross type process. Department of Theoretical Statistics, Preprint 6, University of Copenhagen, 2001
 - 37 Feng S, Miclo L, Wang F Y. Poincaré inequality for Dirichlet distributions and infinite-dimensional generalization. ALEA Lat Am J Probab Math Stat, 2017, 14: 361–380
 - 38 Ethier S N, Kurtz T G. Fleming-Viot processes in population genetics. SIAM J Control Optim, 1993, 31: 345–386
 - 39 Fleming W, Viot M. Some measure-valued Markov processes in population genetics theory. Indiana Univ Math J, 1979, 28: 817–843
 - 40 Ethier S N, Kurtz T G. Markov Processes: Characterization and Convergence. New York: John Wiley & Sons, 1986
 - 41 Ethier S N, Kurtz T G. Convergence to Fleming-Viot processes in the weak atomic topology. Stochastic Process Appl, 1994, 54: 1–27
 - 42 Li Z H, Shiga T, Yao L. A reversible problem for Fleming-Viot processes. Electron Comm Probab, 1999, 4: 71–82
 - 43 Handa K. Quasi-invariance and reversibility in the Fleming-Viot process. Probab Theory Related Fields, 2002, 122: 545–566
 - 44 von Renesse M K, Sturm K T. Entropic measure and Wasserstein diffusion. Ann Probab, 2009, 37: 1114–1191
 - 45 Shao J H. A new probability measure-valued stochastic process with Ferguson-Dirichlet process as reversible measure. Electron J Probab, 2011, 16: 271–292
 - 46 Feng S, Sun S. A dynamic model for the two-parameter Dirichlet process. ArXiv:1706.06146, 2017
 - 47 Mosco U. Composite media and asymptotic Dirichlet forms. J Funct Anal, 1994, 123: 368–421

Reversible measure-valued processes associated with the Poisson-Dirichlet distribution

Shui Feng

Abstract The Poisson-Dirichlet distribution is a probability on the space of infinite dimensional simplex. It is the distribution of a discrete-probability-valued random variable. The closely associated random measures include the GEM distribution, the Dirichlet process and their two-parameter generalizations. One of the active areas for research in modern probability is concerned with the construction of reversible measure-valued processes associated with these random measures. This article will survey recent development in this area and list a few interesting problems for further consideration.

Keywords Poisson-Dirichlet distribution, Dirichlet process, two-parameter Dirichlet process, measure-valued process, Dirichlet form, reversible process

MSC(2010) 60F10, 92D10

doi: 10.1360/N012017-00253