

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

A Difusão com Reposicionamento Estocástico e o Modelo Contínuo de Ohta-Kimura

Telles Timóteo da Silva¹

Departamento de Física e Matemática, UFSJ, Ouro Branco, MG

Marcelo Dutra Fragoso²

Coordenação de Sistemas e Controle, LNCC, Petrópolis, RJ

Resumo. Neste trabalho analisamos a conexão entre a teoria dos processos de difusão com reposicionamento estocástico (*diffusion with stochastic resetting*) e a teoria de Ohta-Kimura da genética de populações. Essa conexão parte da equivalência entre os modelos verificada através das equações para distribuições de probabilidade associadas a esses modelos. Na dedução destas equações a teoria dos processos de Fleming-Viot desempenha um papel importante. Como consequência, isso permite que resultados relevantes obtidos para um modelo sejam transportados para o contexto do outro modelo. Isso é ilustrado no artigo através da prova de dois resultados.

Palavras-chave. Reposicionamento Estocástico, Modelo Contínuo de Ohta-Kimura, Processo de Fleming-Viot.

1 Introdução

O modelo escada de Ohta-Kimura (*Ohta-Kimura ladder model*) [15] surgiu no contexto da genética de populações, tendo por objetivo modelar a evolução de uma população para a qual reproduções aleatórias e mutações neutras se constituem nas forças evolutivas mais significativas. Considere uma população para a qual um locus gênico está sob observação. Durante sua evolução, a população acumula um grande número de genes alelos mutantes neutros. Estes distintos alelos gênicos podem ser detectados por experimentos de eletroforese [13–15]. Cada estado eletroforético (alelo gênico) é representado por um número inteiro e mutações ocorrem de um estado p para um estado vizinho $p - 1$ or $p + 1$.

Uma versão contínua do modelo escada de Ohta-Kimura foi introduzida por W. Fleming e M. Viot no âmbito do processo que ficou conhecido como processo de Fleming-Viot [6, 11, 12]. Os possíveis estados alélicos são representados num espaço de estado contínuo, por números reais. Fleming e Viot [11, p. 837] estudaram a distribuição da diferença $\xi = x_2 - x_1$ entre os alelos x_1 e x_2 , selecionados aleatoriamente da população. Considere $I(\xi, t)$ a função densidade dos alelos que diferem de ξ . Então $I(\xi, t)$ satisfaz

$$\frac{\partial I}{\partial t} = \frac{\partial^2 I}{\partial \xi^2} - 2I + 2\delta(\xi) \quad (1)$$

¹timoteo@ufsj.edu.br

²frag@lncc.br

e seu comportamento estacionário é $\lim_{t \rightarrow \infty} I(\xi, t) = \frac{1}{\sqrt{2}} \exp[-\sqrt{2}|\xi|]$.

Este modelo possui uma grande semelhança com um modelo de difusão com reposicionamento estocástico (*diffusion with stochastic resetting*) estudado por Evans e Majumdar [7,8], que representa uma coleção de partículas que se difundem pelo espaço \mathbb{R} , em tempos aleatórios retornam a uma posição pré-especificada x_0 , e a seguir continuam a difusão, repetindo o ciclo. Este processo modela, por exemplo, um indivíduo que procura por um objeto e, no caso de não obter sucesso, retorna à posição inicial para retomar a procura. A equação mestra para este processo é dada por

$$\frac{\partial p(x, t|x_0)}{\partial t} = D \frac{\partial^2 p(x, t|x_0)}{\partial x^2} - rp(x, t|x_0) + r\delta(x - x_0), \quad (2)$$

onde $p(x, t|x_0)$ é a probabilidade da partícula estar na posição x no tempo t [7, p.160601-2]. Evans e Majumdar obtêm várias propriedades do modelo, sendo a principal a distribuição estacionária $p_{st} = \frac{\alpha_0}{2} \exp[-\alpha_0|x - x_0|]$ (já esperada devido ao trabalho de Fleming e Viot). Aqui $\alpha_0 = \sqrt{r/D}$.

O objetivo deste trabalho é mostrar a equivalência entre os modelos de Ohta-Kimura e Evans-Majumdar e obter resultados para um modelo a partir de seu resultado equivalente no outro. Vê-se, por um lado, que a genética de populações pode se beneficiar da relação com modelos de reposicionamento estocástico (para verificar o estado da arte em modelos de reposicionamento estocástico, veja-se [9]). Por outro lado, para desenvolvimentos de modelos com reposicionamento estocástico, pode ser interessante tomar idéias emprestadas de modelos da genética de populações cuja área de pesquisa tem se desenvolvido muito desde o último século a partir de profundas formulações matemáticas. Fleming e Viot deduziram a equação (1) a partir de um formalismo matemático mais abstrato e, desde então, muitos avanços foram obtidos que complementam a formulação matemática [3,6,12] e dela extraem resultados aplicados [1,2,10].

Para sustentar nossa argumentação, vamos recapitular os principais aspectos do processo de Fleming-Viot (Seção 2), mostrando como derivar não apenas a equação (2), útil para um tipo de indivíduo (gene, partícula), mas também todo um conjunto de equações que fornecem densidades conjuntas para dois ou mais tipos (Seção 3). Na Seção 4.1, constatamos a equivalência entre os modelos. Aplicamos esta equivalência para mostrar um resultado de tempo médio de primeira passagem em genética de populações (Seção 4.2). Em seguida obtemos expressões para o cálculo de covariância e correlação entre diferenças de tipos (Seção 4.3), onde interpretamos o resultado em termos da diferença entre alelos genéticos e em termos do modelo de reposicionamento estocástico.

2 Processo de Fleming-Viot

Seja \mathcal{M} o conjunto das medidas de probabilidade em \mathbb{R} . Defina $\Omega := C([0, \infty[, \mathcal{M})$ como o conjunto de funções contínuas de $[0, \infty[$ em \mathcal{M} e $Y_t : \Omega \rightarrow \mathcal{M}$ o processo canônico, $Y_t(\omega) = \omega(t)$. Seja $\mathcal{F} = \mathcal{B}(\Omega)$ a σ -álgebra de subconjuntos de Borel e a filtração $\{\mathcal{F}_t\}$ em Ω dada por $\mathcal{F}_t := \bigcap_{\epsilon > 0} \mathcal{F}_{t+\epsilon}^0$, onde $\mathcal{F}_t^0 := \sigma(\{Y_u\} : 0 \leq u \leq t)$, e $\mathcal{F}_\infty := \bigvee_{n \in \mathbb{N}} \mathcal{F}_n$. Seja $\mu_0 \in \mathcal{M}$ o estado inicial de Y_t . Fixe $D, \gamma > 0$ e seja $\mathcal{L} \equiv \frac{d^2}{dx^2}$. Denotamos $\langle \beta, \mu \rangle = \int_{\mathbb{R}} \beta(x) \mu(dx)$.

Teorema 1 (Fleming & Viot (1979)). *Existe uma única medida de probabilidade \mathcal{P} sobre Ω satisfazendo:*

i) $\mathcal{P}[Y_0 = \mu_0] = 1$;

ii) para cada $\beta \in \mathcal{D}(\mathcal{L})$, $\langle \beta, Y_t \rangle = \langle \beta, Y_0 \rangle + M_t(\beta) + \frac{D}{2} \int_0^t \langle \mathcal{L}\beta, Y_s \rangle ds$ é uma \mathcal{F}_t -semimartingala, onde $M_t(\beta)$ é uma \mathcal{F}_t -martingala com variação quadrática predizível $\langle\langle M(\beta) \rangle\rangle_t = \gamma \int_0^t [\langle \beta^2, Y_s \rangle - \langle \beta, Y_s \rangle^2] ds$.

A demonstração pode ser vista em [5, 11, 12].

Do ponto de vista da genética populacional, \mathbb{R} é o espaço dos genes alelos para um locus. As proporções destes genes na população são modeladas por medidas de probabilidade em \mathcal{M} . O efeito da mutação é representado pelo operador \mathcal{L} . A deriva genética aleatória é interpretada como a *perda da diversidade genética devida à parte dos genes da população que não são passados para as gerações seguintes por meio da reprodução* e é representada pela martingala M ; γ é a taxa de amostragem dos genes. Veja [4, 11] para estas interpretações e [5, 6] para outros fatores que podem ser incluídos no modelo, tais como, seleção e conversão gênica.

Seja a função $h = h(x_1, \dots, x_n)$ pertencente à álgebra gerada por $\{\beta_1(x_1), \dots, \beta_n(x_n) : \text{onde } \beta_i \in \mathcal{D}(\mathcal{L}), \text{ para } i = 1, \dots, n\}$. Dadas funções $\beta_i : \mathbb{R} \rightarrow \mathbb{R}, i = 1, \dots, n$, defina $\phi : \mathcal{M} \rightarrow \mathbb{R}$ por $\phi(\mu) = \langle h, \mu^n \rangle = \langle \beta_1, \mu \rangle \langle \beta_2, \mu \rangle \dots \langle \beta_n, \mu \rangle = \prod_{i=1}^n \langle \beta_i, \mu \rangle$, para cada $\mu \in \mathcal{M}(\mathbb{R})$. O processo Y_t é um processo de Markov e o gerador infinitesimal é dado por

$$[(\mathcal{G})(\phi)](\mu) = \frac{D}{2} \sum_{i=1}^n \left\langle \frac{\partial^2 h}{\partial x_i^2}, \mu^n \right\rangle + \frac{\gamma}{2} \sum_{j \neq k} (\langle \Theta_{j;k} h, \mu^{n-1} \rangle - \langle h, \mu^n \rangle), \quad (3)$$

onde $\Theta_{j;k} h = h(x_1, \dots, x_{k-1}, x_j, x_{k+1}, \dots, x_n)$. Temos também que

$$M_t(\phi) = \phi(Y_t) - \phi(Y_0) - \int_0^t (\mathcal{G}\phi)(Y_s) ds \quad (4)$$

é uma \mathcal{F}_t -martingala [5].

3 Densidades Conjuntas Médias

Vamos agora computar as funções $\Phi_m(x_1, x_2, \dots, x_m, t)$ que fornecem as densidades conjuntas médias dos tipos $x_1, \dots, x_m \in \mathbb{R}$ no tempo t [11]. Para isto defina Φ_m por

$$\int_{\mathbb{R}^m} \beta_1(x_1) \dots \beta_m(x_m) \Phi_m(x_1, \dots, x_m, t) dx_1 \dots dx_m = E[\langle \beta_1 \dots \beta_m, Y_t^m \rangle].$$

Pela propriedade de martingala (4) e usando a expressão (3) obtemos as seguintes equações:

$$\begin{aligned} \frac{\partial}{\partial t} \Phi_n(x_1, \dots, x_n, t) &= \frac{D}{2} \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2} \Phi_n(x_1, \dots, x_n, t) - \binom{n}{2} \gamma \Phi_n(x_1, \dots, x_n, t) \\ &+ \sum_{j \neq k} \frac{\gamma}{2} \delta(x_k - x_j) \Phi_{n-1}(\dots, x_{k-1}, x_{k+1}, \dots, t) \end{aligned} \quad (5)$$

para cada $n = 1, 2, \dots$

Estamos interessados no comportamento da distribuição das diferenças entre tipos e necessitamos de funções $\Psi_1(\xi, t)$ e $\Psi_2(\xi_1, \xi_2, t)$ que satisfazem as seguintes relações

$$\Psi_1(\xi, t) = \int_{x \in \mathbb{R}} \Phi_2(x, x + \xi, t) dx, \quad \Psi_2(\xi_1, \xi_2, t) = \int_{x_1 \in \mathbb{R}} \Phi_3(x_1, x_1 + \xi_1, x_1 + \xi_2, t) dx_1,$$

De (5) obtemos o seguinte sistema de equações para $\Psi_1, \tilde{\Psi}, \Psi_2$:

$$\frac{\partial}{\partial t} \Psi_1(\xi, t) = D \frac{\partial^2}{\partial \xi^2} \Psi_1(\xi, t) - \gamma \Psi_1(\xi, t) + \gamma \delta(\xi), \tag{6}$$

$$\begin{aligned} \frac{\partial}{\partial t} \Psi_2(\xi_1, \xi_2, t) = & D \left(\Delta_{\xi_1, \xi_2} + \frac{\partial^2}{\partial \xi_2 \partial \xi_1} \right) \Psi_2(\xi_1, \xi_2, t) - 3\gamma \Psi_2(\xi_1, \xi_2, t) \\ & + \gamma \left[\delta(\xi_1) \Psi_1(\xi_2, t) + \delta(\xi_2) \Psi_1(\xi_1, t) + \delta(\xi_1 - \xi_2) \Psi_1(\xi_1, t) \right], \end{aligned} \tag{7}$$

A solução deste sistema pode ser obtida por meio de transformada de Fourier.

4 Resultados Principais

4.1 Equivalência entre os Modelos

Constatamos, inicialmente, que o processo de Fleming-Viot representa tipos de indivíduos (genes, partículas) que sofrem difusão em \mathbb{R} , mas em tempos aleatórios interagem entre si: em termos de genética populacional, um gene é escolhido para morrer e outro é escolhido para ser duplicado. Em termos de reposicionamento estocástico, uma partícula é reposicionada na localização de outra partícula. As equações (1) ou (2) não capturam toda a riqueza da dinâmica desses tipos difundindo e interagindo, e a tarefa de extrair informações tendo apenas essas equações como ponto de partida parece ser uma tarefa árdua. Porém Y_t fornece uma visão completa da dinâmica, a partir de um modelo mais fundamental e geral, que tem recebido aprimoramentos desde 1979.

A equação (6) é a mesma proposta em [7] para reposicionamento em 0. Além disso a equação (7), dependente da solução para (6), retorna a densidade conjunta dos alelos diferindo de ξ_1 e ξ_2 ou, equivalentemente, a densidade conjunta de duas partículas difundindo em \mathbb{R} e se reposicionando em 0.

Note que a equação (6) é igual à equação (1) com $D = 1$ e $\gamma = 2$. A equivalência entre os modelos de Ohta-Kimura e Evans-Majumdar está resumida na Tabela 1, sendo que uma mudança de variáveis, $\xi = x - x_0$, transforma a função $I(\xi, t)$ em $p(x, t|x_0)$.

4.2 Tempo Médio de Primeira Passagem

Vamos utilizar esta equivalência para obter a distribuição de probabilidade de sobrevivência e o tempo médio de primeira passagem no contexto de genética populacional. Evans e Majumdar tratam destes problemas no contexto de reposicionamento estocástico, enquanto, num contexto de genética populacional, até onde saibamos, eles ainda não foram formulados. Coloquemos as questões:

1. Enquanto os tipos genéticos se difundem e são amostrados, qual é a probabilidade $Q(\xi, t)$ de que sua diferença ξ se mantenha abaixo de uma diferença pré-estipulada $\bar{\xi}$ até o tempo t ?
2. Qual é o tempo médio $T(\bar{\xi})$ que ξ leva para atingir o valor $\bar{\xi}$ pela primeira vez?

Tais questões são interessantes num contexto biológico já que ξ é uma medida da diversidade genética, e é possível vislumbrar a necessidade de guiar a população para um valor alvo de diversidade. Pela equivalência entre os dois modelos $Q(\xi, t)$ satisfaz

$$\frac{\partial Q}{\partial t} = D \frac{\partial^2 Q}{\partial \xi^2} - \gamma Q(\xi, t) + \gamma Q(0, t)$$

onde $Q(\bar{\xi}, t) = 0$ e $Q(\xi, 0) = 1$. A transformada de Laplace $q(\xi, s) = \mathbb{L}\{Q(\xi, t)\}$ é

$$q(\xi, s) = \frac{1 - \exp\left[-\left(\frac{s+\gamma}{D}\right)^{\frac{1}{2}}(\bar{\xi} - \xi)\right]}{s + \gamma \exp\left[-\left(\frac{\gamma+s}{D}\right)^{\frac{1}{2}}\bar{\xi}\right]}.$$

Como explicado por Evans e Majumdar, embora não pareça haver como inverter esta transformada em geral, esta última equação é útil para resolver o problema do tempo médio de primeira passagem. Assim, pela equivalência entre os modelos temos o

Teorema 2. *O tempo médio de primeira passagem é*

$$T(\bar{\xi}) = q(0, 0) = \frac{1}{\gamma} \left\{ \exp\left[\left(\frac{\gamma}{D}\right)^{\frac{1}{2}}\bar{\xi}\right] - 1 \right\}.$$

Note que $T(\bar{\xi})$ cresce exponencialmente quando $\bar{\xi}$ se afasta de 0.

4.3 Covariância e Correlação

Para obter expressões para a variância, covariância e correlação, encontramos as funções características associadas às densidades nas equações (6) e (7).

Tabela 1: Equivalência entre os modelos

Ohta-Kimura: equação (1)	Evans-Majumdar: equação (2)
genes evoluem pelo espaço dos tipos gênicos	partículas se movimentam no espaço
distância orientada entre tipos: ξ	posição da partícula: x
diferença após re-amostragem: 0	posição de retorno: x_0
taxa de amostragem: 2	taxa de reposicionamento: r
constante de difusão: 1	constante de difusão: D
densidade média dos tipos diferindo de em ξ no tempo t : $I(\xi, t)$:	probabilidade da partícula estar em x no tempo t , tendo começado de x_0 no tempo 0: $p(x, t x_0)$

Teorema 3. *A variância, a covariância e a correlação são, respectivamente,*

$$\begin{aligned} \text{Var}(\xi) &= -\frac{\partial^2}{\partial k^2} \Psi_1(k, t) \Big|_{k=0} = \frac{2D}{\gamma} (1 - e^{-\gamma t}), \\ \text{Cov}(\xi_1, \xi_2) &= -\frac{\partial^2}{\partial k_1 \partial k_2} \Psi_2(k_1, k_2, t) \Big|_{k_1=k_2=0} = \frac{D}{\gamma} (1 - e^{-\gamma t}), \\ \rho(\xi_1, \xi_2, t) &= \frac{1}{2}. \end{aligned}$$

A demonstração detalhada aparecerá em trabalho futuro.

Quando γ vai a 0, a variância tende a $2Dt$, como esperado, e a covariância vai a Dt . A correlação é $1/2$ e surpreendentemente não é influenciada pelos valores de γ ou D .

Do ponto de vista da genética, ξ_1 e ξ_2 representam as diferenças genéticas $x_2 - x_1$ e $x_3 - x_1$, e a re-amostragem, quando altera o valor de x_1 , influencia ξ_1 e ξ_2 simultaneamente. Daí a correlação ser positiva. Isso também explica a derivada mista na equação (7).

Do ponto de vista do modelo com reposicionamento ξ_1 e ξ_2 representam as posições de duas partículas correlacionadas (com correlação $1/2$), e que devido ao mecanismo de reposicionamento, em tempos aleatórios retornam ao 0 ou saltam uma para a posição da outra.

5 Conclusões

Concluimos que há uma equivalência entre o modelo contínuo de Ohta-Kimura e a difusão com reposicionamento estocástico de Evans-Majumdar. Esta equivalência permitiu que dois resultados relevantes fossem obtidos. O processo de Fleming-Viot se constitui no formalismo matemático abstrato ideal para deduzir as equações para ambos os modelos e ressaltar esta equivalência. Trabalhos futuros serão organizados de forma a mostrar esta equivalência em contextos mais gerais.

Agradecimentos

Esta pesquisa foi parcialmente financiada pelo CNPq via Projetos de Pesquisa No. 304801/2016-1 e 421486/2016-3, pelo FACC/PCI/LNCC e pela FAPERJ Projeto No. 201.696/2018. Pesquisa realizada enquanto o primeiro autor foi visitante no LNCC.

Referências

- [1] M. Bieniek, K. Burdzy, and S. Finch. Non-extinction of a Fleming-Viot particle model. *Probab. Theory Relat. Fields*, 153:293—332, 2012.
- [2] M. Bieniek, K. Burdzy, and S. Pal. Extinction of Fleming-Viot-type particle systems with strong drift. *Electronic Journal of Probability*, 17(11):1–15, 2012.

- [3] T. T. Da Silva and M. D. Fragoso. A note on jump-type Fleming-Viot processes. *Statistics & Probability Letters*, 76:821–830, 2006.
- [4] T. T. Da Silva and M. D. Fragoso. Sample paths of jump-type Fleming-Viot processes with bounded mutation operators. *Statistics & Probability Letters*, 78:1784–1791, 2008.
- [5] D. A. Dawson. Measure-valued Markov processes. In P. L. Hennequin, editor, *École d'Été de Probabilités de Saint-Flour XXI*, volume 1541 of *Lecture Notes in Math.*, pages 1–260. Springer-Verlag, Berlin, 1993.
- [6] S. N. Ethier and T. G. Kurtz. Fleming-Viot processes in population genetics. *SIAM J. Control and Optimization*, 31(2):345–386, 1993.
- [7] M. R. Evans and S. N. Majumdar. Diffusion with stochastic resetting. *Physical Review Letters*, 106:160601, 2011.
- [8] M. R. Evans and S. N. Majumdar. Diffusion with resetting in arbitrary spatial dimension. *J. Phys. A: Math. Theor.*, 47(285001):1–19, 2014.
- [9] R. C. Falcao and M. R. Evans. Interacting Brownian motion with resetting. *J. Stat. Mech.*, 023024:1–16, 2017.
- [10] P. Fatheddin and J. Xiong. Large Deviation Principle for Some Measure-Valued Processes. *Stochastic Process. Appl.*, 125(3):970–993, 2015.
- [11] W. Fleming and M. Viot. Some measure-valued Markov processes in population genetics theory. *Indiana Univ. Math. J.*, 28(5):817–843, 1979.
- [12] S. Hiraba. Jump-Type Fleming-Viot Processes. *Adv. Appl. Prob.*, 32:140–158, 2000.
- [13] P. A. P. Moran. Wandering distribution and the electrophoretic profile. *Theoretical Population Biology*, 8:318–330, 1975.
- [14] P. A. P. Moran. Wandering distribution and the electrophoretic profile II. *Theoretical Population Biology*, 10:145–149, 1976.
- [15] T. Ohta and M. Kimura. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res. Camb.*, 22:201–204, 1973.