Дискретная формула Фейнмана–Каца и законы геномики

С.В. Козырев

Законы Кунина (”взаимодействующий газ генов”) следуют из формулы Фейнмана–Каца для дискретного марковского процесса.

Аналог с действием в физике (сумма ”эволюционных усилий” по пути порождения генома).

Эволюционные уравнения, интегралы по траекториям

Ю. Л. Далецкий, Континуальные интегралы, связанные с операторными эволюционными уравнениями, УМН, 17 (5), 3–115 (1962).
M. Kac, On some connections between probability theory and integral equations, Proc. 2nd Berkeley Simpos. Math. Stat, and Probab., Berkeley, 189–215 (1951).

Обсуждается описание решения уравнения диффузии

$$\frac{\partial \psi}{\partial t} = a\Delta\psi + V(x,t)\psi$$

в виде интеграла по траекториям.

Марковский процесс с конечным числом состояний

Уравнение на (квадратных) матрицах в $\mathbb{R}^p$

$$\frac{dS(t,\tau)}{dt} = A(t)S(t,\tau),$$

то есть для матричных элементов

$$\frac{ds_{jk}(t,\tau)}{dt} = \sum_{r=1}^{p} a_{jr}(t)s_{rk}(t,\tau).$$

Тогда

$$S(t,\theta)S(\theta,\tau) = S(t,\tau), \quad \tau \le \theta \le t,$$

или

$$\sum_{r=1}^{p} s_{jr}(t,\theta)s_{rk}(\theta,\tau) = s_{jk}(t,\tau).$$

Пусть

$$a_{kk}(t) \le 0, \quad a_{jk}(t) \ge 0 (j \ne k), \quad \sum_{j=1}^{p} a_{jk}(t) = 0.$$

Тогда

$$s_{jk}(t, \tau) \ge 0,$$

$$\sum_{j=1}^{p} s_{jk}(t, \tau) = 1,$$

то есть функции $s_{jk}(t, \tau)$ суть вероятности перехода из состояния k в момент времени τ в состояние j в момент времени t.

Рассмотрим пространство M_l траекторий для такого марковского процесса — функций $x(t)$ на отрезке $[0, l]$ со значениями в конечном множестве $\{x_1, \dots, x_p\}$. Введём меру на пространстве M_l.

Обозначим $q(t_0, t_1, \dots, t_{n+1})$ разбиение интервала $0 = t_0 < t_1 < \dots < t_{n+1} = l$ и обозначим Γ набор $\Gamma = (\gamma_1, \dots, \gamma_n)$ подмножеств $\{x_1, \dots, x_p\}$.

Подмножество траекторий в M_l

$$Q_{ij}(q, \Gamma) = \{x(0) = x_i, \quad x(l) = x_j, \quad x(t_m) \in \gamma_m (m = 1, \dots, n)\},$$

мера такого подмножества

$$\mu_{ij}(q, \Gamma) = \sum_{x_{k_1} \in \gamma_1} \cdots \sum_{x_{k_n} \in \gamma_n} s_{jk_n}(l, t_n) s_{k_n k_{n-1}}(t_n, t_{n-1}) \dots$$

$$\dots s_{k_2 k_1}(t_2, t_1) s_{k_1 i}(t_1, 0).$$

Дискретный вариант формулы Фейнмана–Каца

Рассмотрим интеграл по траекториям

$$\psi_{ij}(t) = \int_{M_t} e^{\int_0^t V(x(s),s)ds} d\mu_{ij}.$$

Пусть $\Psi(t)$ есть матрица с элементами $\psi_{ij}(t)$. Тогда имеет место уравнение (Дайсона–Швингера)

$$\Psi(t) = S(t,0) + \int_0^t S(t,s)V(s)\Psi(s)ds,$$

где диагональная матрица $V_{rk}(s) = \delta_{rk} V(x_k, s)$,
и дифференциальное уравнение

$$\frac{d}{dt}\Psi(t) = [A(t) + V(t)]\,\Psi(t), \quad \Psi(0) = 1.$$

Таким образом, решение этих эволюционных уравнений задаётся интегралом по траекториям по мере μ.

Объекты, сгенерированные, а не просто наблюдаемые.
Геномика. Геномы — строки символов. Набор мутаций $E = [e_1, \ldots, e_n]$; например точечные замены, вставки, удаления, дупликации etc.
$w(e_s) > 0$ вес ("эволюционное усилие") для мутации e_s.
Аррениусовский фактор $e^{-\alpha w(e_s)}$ есть скорость мутации, $\alpha > 0$ обратная температура.

Рассмотрим уравнение эволюции в виде

$$\frac{df_j(t)}{dt} = \sum_i A_{ji} f_i(t),$$

f_i есть населённость генома i, A_{ji} скорость мутаций $A_{ji} = e^{-\alpha w(e_s)}$, если e_s переводит i в j.

Разложение оператора эволюции $Q(t) = e^{tA}$ в ряд теории возмущений будет содержать вклады, пропорциональные скоростям мутации $i \to j$ (генома в геном):

$$Q_{ji} = \sum_{p:i \to j} e^{-\alpha \sum_{k \in p} w(e_s(k))}, \tag{1}$$

сумма по k идёт по мутациям вдоль пути p порождения генома j из i (i.e. k-ая мутация на пути p есть e_s); сумма по путям $p : i \to j$ — порождение j из i мутациями в разном порядке.

Вариант формулы Фейнмана–Каца для дискретного марковского процесса.
Сумма весов мутаций по пути p есть взвешенная оценка колмогоровской сложности порождения генома — аналог действия $S(p)$ в физике. Сумма по путям в (1) есть аналог интеграла по путям $\int e^{S(p)} dp$ в физике.

Объяснение законов геномики "взаимодействующего газа генов"

Логнормальное распределение в эволюции белков.
Ортологичные белки в различных организмах связаны происхождением от единого предка. Распределение логарифма частот замен аминокислот в ортологичных белках близко к нормальному.

Эволюция белка случайными заменами аминокислот.
Скорость мутаций (1) даст распределение потомков

$$e^{-\alpha \sum_k E_k},$$

где E_k веса мутаций замен аминокислот.
Если веса независимы (нейтральная эволюция), это порождает логнормальное распределение частот белков в ортологичном семействе.

Степенное распределение размеров паралогичных семейств генов. Паралогичные гены в геноме порождаются дупликацией генов.

Вес эволюционного усилия дупликации E. Семейство N паралогичных генов отвечает N весов E. Нейтральная эволюция генной дупликацией даёт для семейств из N паралогичных генов

$$e^{-\alpha NE}$$

степенной закон распределения по размеру семейства.

Гиббсовское распределение "взаимодействующего газа генов"

E.V. Koonin, Are There Laws of Genome Evolution? PLoS Comput. Biol. 7(8): e1002173 (2011).
E.V. Koonin, The Logic of Chance: The Nature and Origin of Biological Evolution, FT Press, 2012.

Законы Кунина ("взаимодействующий газ генов") следуют из формулы Фейнмана–Каца для дискретного марковского процесса.
Связь с принципом наименьшего действия в физике (обсуждалась ещё Бюффоном и Мопертюи).

М.Громов, Кольцо тайн: вселенная, математика, мысль. МЦНМО, 2017.

M. Gromov, Ergo-brain, https://www.ihes.fr/~gromov/category/ergosystems/

M. Gromov, Great Circle of Mysteries. Mathematics, the World, the Mind, (Birkhauser Cham, 2018).