

Точные решения в математических моделях эволюции: формализм Гамильтона–Якоби

Якушкина Татьяна Сергеевна¹
Саакян Давид Багратович²
Ху Чин-Кун²

- 1) Московский государственный университет им. М.В. Ломоносова
Национальный исследовательский университет "Высшая школа экономики"
- 2) Institute of Physics, Academia Sinica, Taiwan

VII конференция по математическим моделям и численным методам в биологии и медицине

ИВМ РАН
2 ноября 2015

- Модели эволюции Эйгена и Кроу–Кимуры
- Аналитические выражения для усредненных характеристик модификации модели Кроу–Кимуры
- Однопиковый ландшафт приспособленности: решение и фазовый портрет
- Гладкий ландшафт приспособленности
- Многомерный ландшафт приспособленности
- Модификация модели Эйгена
- Аналитические выражения для динамики максимума

Модель Эйгена

Модель генома длины N (число нуклеотидов)

$$S_i = (s_1, \dots, s_N)$$

s_τ :

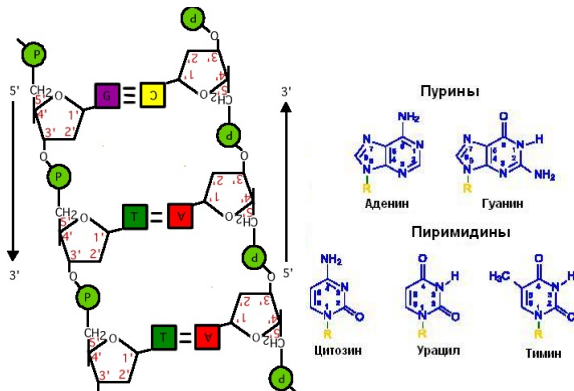
+1 — пурин (R)

-1 — пиримидин (Y)

$M = 2^N$ различных последовательностей S_i

d_{ij} — расстояние Хэмминга между S_i и S_j

$p_i(t)$ — вероятность последовательности S_i



Модель Эйгена: система уравнений

$$\frac{dp_i}{dt} = \{Q_{ii}r_i - \sum_j r_j p_j(t)\} p_i(t) + \sum_{j \neq i} Q_{ij} r_j p_j(t), \quad (1)$$
$$\sum_{i=0}^{M-1} p_i = 1$$

r_i — интенсивность воспроизводства S_i

Q_{ij} — интенсивность мутаций

$Q_{ii} = q^N \equiv Q$ — родительский тип ($S_i \rightarrow S_i$)

$Q_{ji} = q^{N-d_{ij}}(1-q)^{d_{ij}}$ — мутация ($S_j \rightarrow S_i$).

q — вероятность безошибочной репликации



- M. Eigen, Naturwissenschaften, 58(10), 1971: 465–523
- M. Eigen, J. McCaskill, P. Schuster, J Phys Chem, 92(24), 1988:181-1891
- M. Eigen, P. Schuster, The Hypercycle, Springer, 1979

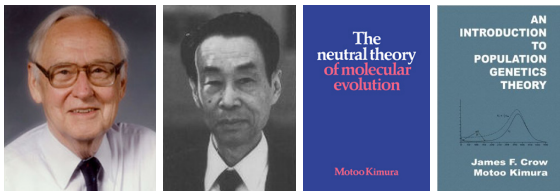
Модель Кроу–Кимуры

l – расстояние Хэмминга от опорной последовательности

$N_l = \frac{l!}{l!(L-l)!}$ разных последовательностей в одном классе

$$\frac{dP_l}{Ndt} = P_l[r_l - \mu] + \mu \frac{N-l+1}{N} P_{l-1} + \mu \frac{l+1}{N} P_{l+1} - P_l \sum_i r_i P_i \quad (2)$$

μ – интенсивность мутации, r_l – приспособленность



J. F. Crow and M. Kimura, An Introduction to Population Genetics Theory (Harper Row, NY, 1970).

Уравнение Гамильтона–Якоби

Схематически процесс мутации может быть представлен в виде цепочки:

$$\Leftrightarrow \square | -1 \square \Leftrightarrow \square | \square \Leftrightarrow \square | +1 \square \Leftrightarrow \quad (3)$$

Предлагается подстановка:

$$p_l(t) \equiv p(x, t) = \exp Nu(x, t), \quad x = 1 - \frac{2l}{N} \quad (4)$$

с гладкой по x функцией $u(x, t)$, приводящая к уравнению Гамильтона–Якоби вида:

$$\begin{aligned} \frac{\partial u}{\partial t} + H(u', x) &= 0; \\ -H(u', x) &= f(x) + \mu \left(\frac{1+x}{2} e^{2u'} + \frac{1-x}{2} e^{-2u'} - 1 \right) \end{aligned} \quad (5)$$

Средняя приспособленность имеет вид:

$$\begin{aligned} R &= \max [f(x) + \mu(\sqrt{1-x^2} - 1)]|_x = \\ &f(x_0) + \mu(\sqrt{1-(x_0)^2} - 1) \end{aligned} \quad (6)$$

D.B.Saakian Journal of Stat. Physics, 128,(2007)781.

A. Martirosyan and D. B. Saakian, Phys. Rev. E 84, 021122 (2011).

Основные результаты для моделей Эйгена и Кроу–Кимуры

Основная задача — построить корректный непрерывный аналог моделей мутаций в дискретном пространстве.

- Стационарное распределение в моделях Эйгена и Кроу–Кимуры
- Поправки конечного объема в эволюционных моделях
- Средняя приспособленность и дисперсия приспособленности

E. Baake and H. Wagner, Genet. Res. 78:93 (2001).

E. Baake, M. Baake and A. Bover, J. Math. Biol. 50:83 (2005).

D.B.Saakian Journal of Stat. Physics, 128,(2007)781.

A. Martirosyan and D. B. Saakian, Phys. Rev. E 84, 021122 (2011).

Модели биологической эволюции с геном-мутатором

Гены-мутаторы (mutator genes) — гены, биохимическая активность которых может повышать частоту мутирования других генов.

Пример: *urvA*, *urvB*, *urvD* у *E. coli*.

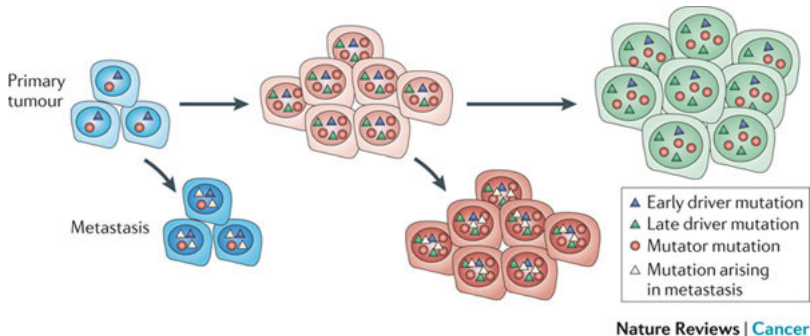


Рис.: Human cancers express mutator phenotypes: origin, consequences and targeting. Lawrence A. Loeb Nature Reviews Cancer 11, 450-457 (June 2011)

Модели биологической эволюции с геном-мутатором: модель Кроу–Кимуры

Геном — цепочка из $(N + 1)$ гена:

2 аллеля каждого гена: $s_\tau = \pm 1, \tau = 0, \dots, N$.

N генов образуют базовую часть генома

$S_i = (s_1, \dots, s_N), \quad i = 1, \dots, 2^N$.

1 ген-мутатор:

$s_0 = +1$ дикий тип, мутация с интенсивностью μ_1 ,

$s_0 = -1$ мутантный тип, мутация с интенсивностью μ_2 .

Мутация самого гена-мутатора:

$\alpha_1 s_0 : +1 \rightarrow -1$

$\alpha_2 s_0 : -1 \rightarrow +1$.

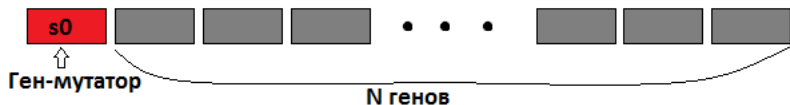


Рис.: Формализация структуры генома в модели

Модели биологической эволюции с геном-мутатором: модель Кроу–Кимуры

Опорная последовательность $S = (+1, \dots, +1)$

Расстояние Хэмминга между S и S_i :

$$l \equiv d_{1i} = (N - \sum_{\tau=1}^N s_{\tau})/2, \quad l = 0, 1, \dots, N$$

Усредненное состояние гена

$$x_l = \sum_{\tau=1}^N s_{\tau}/N, \quad x_l \in [-1, 1].$$

Распределение последовательностей по классам Хэмминга: $P_l(t)$, $Q_l(t)$

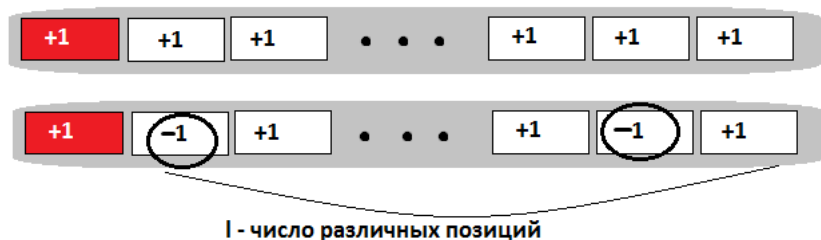


Рис.: Определение числа мутаций как количества измененных состояний генов l

Модели биологической эволюции с геном-мутатором: модель Кроу–Кимуры

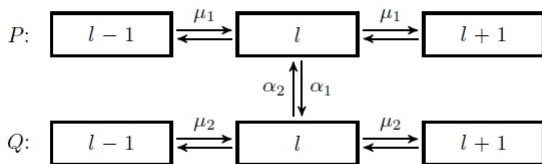


Рис.: Схема допустимых переходов между состояниями системы: верхняя цепочка соответствует дикому типу генома, нижняя — мутантному. l — число мутаций в основной части генома.

Модели биологической эволюции с геном-мутатором: система уравнений в постановке Кроу–Кимуры

$$\begin{aligned}\frac{dP_l(t)}{Ndt} &= \alpha_2 Q_l + P_l (f(x_l) - (\mu_1 + \alpha_1)) + \mu_1 \left(P_{l-1} \frac{N-l+1}{N} + P_{l+1} \frac{l+1}{N} \right) - P_l R, \\ \frac{dQ_l(t)}{Ndt} &= \alpha_1 P_l + Q_l (g(x_l) - (\mu_2 + \alpha_2)) + \mu_2 \left(Q_{l-1} \frac{N-l+1}{N} + Q_{l+1} \frac{l+1}{N} \right) - Q_l R, \\ R(t) &= \sum_l (P_l(t)f(x_l) + Q_l(t)g(x_l)), \\ x_l &= 1 - 2l/N, \quad 0 \leq l \leq N.\end{aligned}\tag{7}$$

$f(x_l)$ — функция приспособленности последовательностей дикого типа

$g(x_l)$ — функция приспособленности последовательностей с мутатор аллелем

Заметим, что функции приспособленности могут отличаться, а μ_2 в 10–100 раз больше, чем μ_1 .

Используем нелинейную подстановку

$$P_l \rightarrow \frac{P_l}{\sum_l (P_l + Q_l)}, \quad Q_l \rightarrow \frac{Q_l}{\sum_l (P_l + Q_l)}.$$

Усредненные характеристики модификации модели Кроу–Кимуры

- Средняя приспособленность популяции R
- Среднее состояние гена в популяции s
- Среднее состояние гена дикого типа s_1
- Среднее состояние гена мутантного типа s_2
- Вероятностное значение для подпопуляции мутантного типа q

$$R = \frac{\sum_l (P_l f(x_l) + Q_l g(x_l))}{\sum_l (P_l + Q_l)}, s = \frac{\sum_l (P_l + Q_l) x_l}{\sum_l (P_l + Q_l)},$$
$$s_1 = \frac{\sum_l P_l x_l}{\sum_l P_l}, s_2 = \frac{\sum_l Q_l x_l}{\sum_l Q_l}, q = \frac{\sum_l Q_l}{\sum_l (P_l + Q_l)}. \quad (8)$$

Вопросы для исследования в модифицированной модели Кроу–Кимуры с геном-мутатором

- A. Является ли среднее состояние гена в популяции положительным: $s > 0$?
- B. Совпадают ли средние состояния гена для дикого и мутантного типов $s_1 = s_2$?
- C. Является ли мутатор аллель строго доминантным в популяции или $q < 1$?
- D. Требуются ли поправки порядка N для q или R при $N \sim 10000$?

Аналитические выражения для усредненных характеристик модификации модели Кроу-Кимуры

Общий случай: $\alpha_1, \alpha_2, \mu_1, \mu_2$.

Получено выражение для **среднего фитнеса** в популяции

$$R = \max_x [V_+(x)],$$

$$V_{\pm}(x) = \frac{f(x) + g(x)}{2} - \frac{1 + \mu}{2} - \frac{\alpha_1 + \alpha_2}{2} + \frac{1 + \mu}{2} \sqrt{1 - x^2} \pm \frac{1}{2} \sqrt{A(x)^2 + 4\alpha_1\alpha_2},$$

где $\mu_1 = 1$, $\mu_2 = \mu$, $A(x) = f(x) - g(x) - \alpha_1 + \alpha_2 + (1 - \mu)(\sqrt{1 - x^2} - 1)$.

Уравнение связи со средним состоянием гена

$$R = \frac{f(s) + g(s) - \alpha_1 - \alpha_2}{2} + \frac{1}{2} \sqrt{(f(s) - g(s) - \alpha_1 + \alpha_2)^2 + 4\alpha_1\alpha_2}. \quad (9)$$

Модель с пренебрежимо малыми интенсивностями обратных мутаций гена-мутатора

Пусть $\alpha_2 = 0$, $\alpha_1 = a$, $\mu_1 = 1$, $\mu_2 = \mu$, тогда обе ветви решения $V_{\pm}(x)$ допустимы.
Фаза мутатора $V_-(x)$

$$R = \max_x [g(x) + \mu(\sqrt{1-x^2} - 1)],$$
$$g(s) = R.$$

Смешанная фаза $V_+(x)$

$$R = \max_x [f(x) - a + \sqrt{1-x^2} - 1],$$
$$g(s) = R.$$

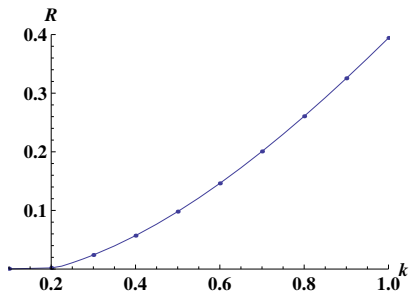


Рис.: Зависимость среднего фитнеса R от k в модели с линейным ландшафтом приспособленности $f(x) = kx$, при $\mu_1 = 1$, $\mu_2 = 10$, $\alpha_2 = 0$, $\alpha_1 = 0.02$. Две фазы: смешанная $R = \sqrt{k^2 + 1} - \alpha_1 - 1$ и фаза мутатора $R = \sqrt{k^2 + (\mu_2)^2} - \mu_2$, $q = 1$. Граница задана условием $\alpha_1 = \mu_2 - 1 + \sqrt{k^2 + 1} - \sqrt{k^2 + (\mu_2)^2}$. Здесь $k_c \approx 0.212$.

Однопиковый ландшафт приспособленности: решение и фазовый портрет

Рассмотрим функцию фитнеса, равную нулю всюду, кроме $l = 0$: $f(x_0) = J$. Обозначим R_{mu} — средний фитнес для мутатор фазы, R_{mix} — смешанной фазы и R_{ns} — неселективной фазы.

$$R_{mix} = J - 1 - a, \quad R_{mu} = J - \mu, \quad R_{ns} = 0.$$

Получено уравнение для вероятности мутантного типа в популяции:

$$P_0 = \frac{(J - a - 1)(\mu - a - 1)}{J(\mu - 1)}.$$

$$q = 1 - P_0 \frac{J}{J - 1}. \quad (10)$$

Фазовый портрет системы с однопиковой функцией фитнеса

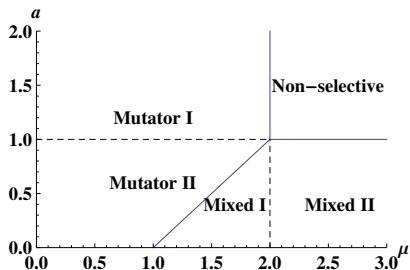


Рис.: Система задана при параметрах $\alpha_1 = a, \alpha_2 = 0, \mu_1 = 1, \mu_2 = \mu$. Три возможных фазы: смешанная $0 < s < 1, 0 < q < 1$, неселективная $s = 0, 0 < q \leq 1$ и мутатор фаза $0 < s, q = 1$. Граница между неселективной и мутатор фазами задана условием $\mu = J$, между неселективной и смешанной $a = J - 1$, смешанной и мутатор фазами $a + 1 = \mu$.

Гладкий ландшафт приспособленности: решение

$$\alpha_1 = a, \alpha_2 = 0.$$

Для вероятности мутатного типа в популяции получено

$$1 - q \sim \frac{a}{\mu} \exp \left[N \left(\int_{s_1}^{s_3} u'(x) - \int_{s_2}^{s_3} \hat{u}'(x) \right) \right], \quad (11)$$

Линейный ландшафт приспособленности $f(x) = kx$ в случае малого параметра a :

$$1 - q = \exp \left[\frac{-Na^2}{2(\sqrt{k^2 + 1} - 1)\mu} \right]. \quad (12)$$

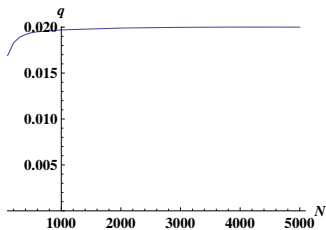


Рис.: Зависимость доли мутантного типа в популяции q от длины генома N . Однопиковый ландшафт приспособленности с параметрами: $J = 1.05$, $\mu_1 = 1$, $\mu_2 = 10$, $\alpha_1 = 0.001$.

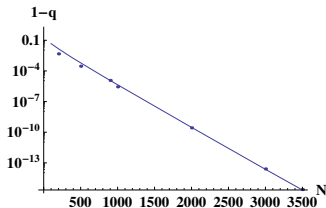


Рис.: Зависимость доли мутантного типа в популяции q от длины генома N . Линейный ландшафт приспособленности с параметрами $f(x) = kx$, $k = 1$, $\mu_1 = 1$, $\mu_2 = 10$, $\alpha_2 = 0$, $\alpha_1 = 0.3$.

Многомерный ландшафт приспособленности

Геном разделен на d длиной Ny_i каждая.

Фитнес для последовательностей дикого типа равен $\sum_i f(x_i)$, мутантного $\sum_i g(x_i)$, где $x_i = 1 - 2l_i/(Ny_i)$ и l_i — число мутаций в i -м участке.

Интенсивности мутаций равны μ_i для дикого типа и ν_i для мутантного.

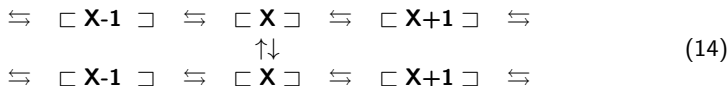
Средний фитнес вычисляется как максимум функции

$$V_{\pm}(x) = \frac{\sum_i (f_i(x_i) + g_i(x_i) - y_i(\mu_i + \nu_i + \alpha_1 + \alpha_2))}{2} + \sum_i y_i \frac{\mu_i + \nu_i}{2} \sqrt{1 - x_i^2} \\ \pm \frac{1}{2} \sqrt{(A - \alpha_1 + \alpha_2)^2 + 4\alpha_1\alpha_2}, \\ A = \sum_i \left(f_i(x_i) - g_i(x_i) + y_i(\mu_i - \nu_i) \left(\sqrt{1 - x_i^2} - 1 \right) \right). \quad (13)$$

- Получены выражения для среднего фитнеса в популяции
- При пренебрежимо малых обратных мутациях наблюдается смешанная и мутатор фазы (неселективной может не быть)
- В модели с однопиковым ландшафтом построен фазовый портрет
- Для гладких функций получено выражение для q
- Показана зависимость q от величины N

Эволюционная модель с теоретико-игровой динамикой

Основное уравнение химической кинетики с двумя цепочками :



$$\begin{aligned}
 \frac{dP(X, t)}{Ndt} = & P(X-1, t)R_{+1}^A(X-1) + P(X+1, t)R_{-1}^A(X+1) \\
 & + P(X, t)R_0^A(X) + Q(X, t)R^{BA}(X);
 \end{aligned} \tag{15}$$

$$\begin{aligned}
 \frac{dQ(X, t)}{Ndt} = & Q(X-1, t)R_{+1}^B(X-1) + Q(X+1, t)R_{-1}^B(X+1) \\
 & + Q(X, t)R_0^B(X) + P(X, t)R^{AB}(X).
 \end{aligned}$$

$$\sum_{0 \leq X \leq N} (P(X, \tau) + Q(X, \tau)) = 1, \quad \forall \tau \tag{16}$$

$$\begin{aligned}
 R_{-1}^A(X) + R_{+1}^A(X) + R_0^A(X) + R^{AB}(X) &= 0, \\
 R_{-1}^B(X) + R_{+1}^B(X) + R_0^B(X) + R^{BA}(X) &= 0.
 \end{aligned} \tag{17}$$

V.Galstyan, D.B.Saakian, Phys.Rev.E. 86,(2012)011125.

Переход к системе уравнений непрерывного параметра

Предлагается замена

$$P(X, t) = v_1 \exp[Nu(x, t)]; \quad Q(X, t) = v_2 \exp[Nu(x, t)]. \quad (18)$$

где $x = X/N$.

$$\begin{aligned} v_1 q &= v_1(r_+^A(x)e^{-u'} + r_-^A(x)e^{u'} + r_0^A(x)) + v_2 r^{BA}(x), \\ v_2 q &= v_2(r_+^B(x)e^{-u'} + r_-^B(x)e^{u'} + r_0^B(x)) + v_1 r^{AB}(x). \end{aligned} \quad (19)$$

$$\begin{aligned} r_0^A(x) &= -r_+^A(x) - r_-^A(x) - r^{AB}(x); \\ r_0^B(x) &= -r_+^B(x) - r_-^B(x) - r^{BA}(x). \end{aligned} \quad (20)$$

Уравнение Гамильтона-Якоби и динамика максимума

Разложение функции в окрестности максимума $y(t)$

$$u(x, t) = -\frac{V(t)}{2}[x - y(t)]^2. \quad (21)$$

позволяет вычислить динамику

$$\frac{dy(t)}{dt} \equiv H'_p(y, 0) = \sigma \Delta r^A(y) + (1 - \sigma) \Delta r^B(y); \quad (22)$$

$$\sigma = \frac{r^{BA}(y)}{r^{BA}(y) + r^{AB}(y)}. \quad (23)$$

Биологические модели в литературе для одной цепочки (игровая форма постановки)

Динамика системы, определяемой репликаторным уравнением

$$\dot{x}_i = x_i((A\bar{x})_i - (x, A\bar{x})) \quad (24)$$

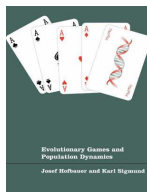
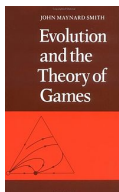
для системы с матрицей выплат A и распределением $\bar{x} = (x_1, \dots, x_n)$ в двумерном

случае определяется положением равновесия: $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$;

$$(a - c + d - b)x_1 = d - b. \quad (25)$$

Классы систем:

- "Дилемма заключенного": матрица $(a - c)(d - b) \leq 0$. Для любого начального положения траектория сходится к 0 (или 1).
- "Координационная игра": $a > c, d > b, (a - c)(d - b) > 0$. Внутреннее положение равновесия неустойчиво.
- "Ястребы-голуби": $a < c, d < b$. Устойчивое положение равновесия x_1 .



Подходы к вычислению интенсивностей перехода в модели с двумя цепочками

Процесс Морана. Для двух матриц A и B (2×2) средние выплаты по двум стратегиям:

$$\pi_1^A(X) = \frac{a(X-1) + b(N-X)}{N-1}; \pi_2^A(X) = \frac{cX + d(N-X-1)}{N-1} \quad (26)$$

Интенсивности перехода определяются уравнениями:

$$R_+^A(X) = \frac{1-w + w\pi_1^A(X)}{1-w + w\langle \pi^A(X) \rangle} \cdot \frac{X}{N} \cdot \frac{N-X}{N}; \quad (27)$$

$$R_-^A(X) = \frac{1-w + w\pi_2^A(X)}{1-w + w\langle \pi^A(X) \rangle} \cdot \frac{X}{N} \cdot \frac{N-X}{N}; \quad (28)$$

где средние выплаты по всей популяции имеют вид:

$$\langle \pi^A(X) \rangle = \frac{\pi_1^A(X)X + \pi_2^A(X)(N-X)}{N}. \quad (29)$$

$$x = \frac{X}{N}:$$

$$\begin{aligned} r_+^A(x) &= \lim_{N \rightarrow \infty} \left(\frac{\pi_1^A - \pi_2^A}{\Gamma + \langle \pi^A \rangle} \cdot \frac{x}{N} \cdot \frac{N-x}{N} \right) = \\ &= \frac{x}{\Gamma + \langle \pi^A(x) \rangle} (\pi_1^A(x) - \langle \pi^A(x) \rangle); \end{aligned} \quad (30)$$

где $\langle \pi^A(x) \rangle = x\pi_1^A(x) + \pi_2^A(x)(1-x)$; $\Gamma = \frac{1-w}{w}$ – базовый фитнес.
 Что приводит к т.н. скорректированной репликаторной динамике

$$r_+^A(x) = x \left(\frac{(A\bar{x})_1}{(\bar{x}, A\bar{x})} - 1 \right); \quad r_-^A(x) = x \left(\frac{(A\bar{x})_2}{(\bar{x}, A\bar{x})} - 1 \right) \quad (31)$$

Т.о. получены необходимые для вычислений

$$\begin{aligned} \Delta r^A(x) &= x \left(\frac{(A\bar{x})_1 - (A\bar{x})_2}{(\bar{x}, A\bar{x})} \right); \\ \Delta r^B(x) &= x \left(\frac{(B\bar{x})_1 - (B\bar{x})_2}{(\bar{x}, B\bar{x})} \right). \end{aligned} \quad (32)$$

Механизм локального регулирования.

$$R_+^A(X) = \left(\frac{1}{2} + \frac{w}{2} \frac{\pi_1^A - \pi_2^A}{\Delta\pi_{\max}^A} \cdot \frac{X}{N} \cdot \frac{N-X}{N} \right); \quad (33)$$

$$R_-^A(X) = \left(\frac{1}{2} + \frac{w}{2} \frac{\pi_2^A - \pi_1^A}{\Delta\pi_{\max}^A} \cdot \frac{X}{N} \cdot \frac{N-X}{N} \right). \quad (34)$$

($\tau = A, B$):

$$\begin{aligned} r_+^\tau(x) &= \lim_{N \rightarrow \infty} \left(\frac{1}{2} + \frac{w}{2} \frac{\pi_1^A - \pi_2^A}{\Delta\pi_{\max}^A} \right) \cdot \frac{X}{N} \cdot \frac{N-X}{N} = \\ &= \kappa X (\pi_1^A - \langle \pi^A(x) \rangle); \end{aligned} \quad (35)$$

где $\kappa = \frac{w}{\Delta\pi_{\max}}$.

$$\begin{aligned} \Delta r^A(x) &= x\kappa_1 ((A\bar{x})_1 - (A\bar{x})_2); \\ \Delta r^B(x) &= x\kappa_2 ((B\bar{x})_1 - (B\bar{x})_2). \end{aligned} \quad (36)$$

Сравнение с аналитических результатов с численным моделированием

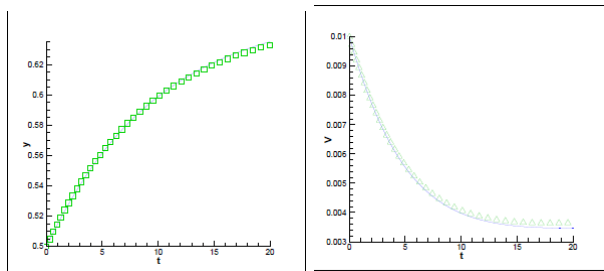





Рис.: $N = 1000$: Maximum dynamics via Moran process for $y = \sum_X [P(X, t) + Q(X, t)] \frac{X}{N}$. The numerical solution of y calculated by Eq.(3) (smooth line) is represented versus our analytical results by the HJE method (squares). We take transition rates $r^{AB} = 0.5$; $r^{BA} = 1$; $A = [3 \ 1; 3.2 \ 1.5]$; $B = [7 \ 0.1; 7.5 \ 0.3]$ Variance $Var = 1/Q$ dynamics, $Var = \sum_X [P(X, t) + Q(X, t)] (\frac{X}{N} - 1)^2$.

-  *Yakushkina T., Saakian D.B., Hu C.-K.* Evolutionary Games with Randomly Changing Payoff Matrices // Journal of the Physical Society of Japan. – 2015. – T. 84. – №. 6. – С. 064802.
-  *Yakushkina T., Saakian D.B., Hu C.-K.* Exact Dynamics for a Mutator Gene Model // Chinese Journal of Physics.-2015 - 53, DOI: 10.6122/CJP.20150910A.
-  *Saakian D.B., Yakushkina T., Hu C.-K.* The Rich Phase Structure of a Mutator Model // Preprint, 2015.

DBS and CKH thank Taiwan-Russia collaborative research grant with Grant Number 101-2923-M-001 -003 -MY3

Спасибо за внимание!