

Mathematical Properties of the Agent-based Model of Extinction — Recolonization for Population Genetics

N. V. Gaianov¹

DOI: [10.18255/1818-1015-2024-2-142-151](https://doi.org/10.18255/1818-1015-2024-2-142-151)

¹National Research University Higher School of Economics, Moscow, Russia

MSC2020: 60J70

Research article

Full text in Russian

Received April 26, 2024

Revised May 7, 2024

Accepted May 15, 2024

The individual-based model describes the dynamics of genetic diversity of a population scattered on a spatial continuum in the case of a finite number of individuals. During extinction events in a certain area, a portion of the population dies, after which new individuals with the genotype of the parent are born during recolonization event. In this paper we examine the model, as well as its modification, and derive properties related to population parameters. The study demonstrates that the lifespan of individuals follows an exponential distribution, allele probabilities remain constant over time, and the average heterozygosity, constrained by the number of individuals during extinction and recolonization, equals a similar quantity in the Moran model. The joint distribution of alleles is generalized for populations continuously scattered in space. Joint allele distribution and heterozygosity are computed through simulations.

Keywords: population model; recolonisation; spatial continuum; Moran model

INFORMATION ABOUT THE AUTHORS

Gaianov, Nikita V. (corresponding author)	ORCID iD: 0009-0001-0522-6856 . E-mail: nvgayanov@edu.hse.ru Research Assistant
--	--

Funding: Russian Science Foundation (Project no. 22-71-10056).

For citation: N. V. Gaianov, “Mathematical properties of the agent-based model of extinction — recolonization for population genetics”, *Modeling and Analysis of Information Systems*, vol. 31, no. 2, pp. 142–151, 2024. DOI: [10.18255/1818-1015-2024-2-142-151](https://doi.org/10.18255/1818-1015-2024-2-142-151).

Математические свойства агентной модели вымирания — реколонизации для популяционной генетики

Н. В. Гаянов¹

DOI: [10.18255/1818-1015-2024-2-142-151](https://doi.org/10.18255/1818-1015-2024-2-142-151)

¹Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

УДК 51-76

Научная статья

Полный текст на русском языке

Получена 26 апреля 2024 г.

После доработки 7 мая 2024 г.

Принята к публикации 15 мая 2024 г.

Агентная модель описывает динамику генетического разнообразия непрерывно распределенной популяции в случае конечного числа особей. В событии вымирания в некоторой области умирает часть популяции, после чего в ходе реколонизации рождаются новые особи с генотипом родителя. Мы рассматриваем модель, а также её модификацию, и получаем свойства, связанные с популяционными параметрами. В работе показано, что время жизни особей имеет экспоненциальное распределение, вероятности аллелей сохраняются во времени, средняя гетерозиготность при ограничении, связанном с числом особей при вымирании и реколонизации, равна аналогичной величине в модели Морана. Совместное распределение аллелей обобщено на случай популяций, непрерывно расположенных в пространстве. Совместное распределение аллелей и гетерозиготность посчитаны на симуляциях.

Ключевые слова: популяционная модель; реколонизация; непрерывное пространство; модель Морана

ИНФОРМАЦИЯ ОБ АВТОРАХ

Гаянов, Никита Владимирович
(автор для корреспонденции)

ORCID iD: [0009-0001-0522-6856](https://orcid.org/0009-0001-0522-6856). E-mail: nvgayanov@edu.hse.ru
Стажер-исследователь

Финансирование: Российский научный фонд (проект № 22-71-10056).

Для цитирования: N. V. Gaianov, “Mathematical properties of the agent-based model of extinction — recolonization for population genetics”, *Modeling and Analysis of Information Systems*, vol. 31, no. 2, pp. 142–151, 2024. DOI: [10.18255/1818-1015-2024-2-142-151](https://doi.org/10.18255/1818-1015-2024-2-142-151).

Введение

Популяционные модели описывают изменения генетического разнообразия популяции во времени. Такие модели используются во многих задачах популяционной генетики. Наиболее известные модели: модель Райта — Фишера, Морана, ступенька Кимуры (stepping stone) и др. [1] — либо не учитывают пространственную структуру популяции, либо разделяют её на подпопуляции, которые развиваются независимо друг от друга, но при этом между ними существует миграция. Такие модели не учитывают сценарии, в которых подпопуляции распределены непрерывно и граничат друг с другом.

Проблемы построения моделей с непрерывным расположением особей связаны с утверждением, доказанным в [2] о том, что не существует модели, в которой выполняются 3 свойства:

- 1) особи распределены с одинаковой плотностью;
- 2) особи размножаются независимо, со средним числом потомков равным единице;
- 3) новорождённые особи мигрируют согласно нормальному распределению с матожиданием, равным положению родителя.

Существуют модели, в которых существует локальная регуляризация плотности, например [3], однако для таких моделей сложно строить генеалогические деревья. В работе [4] был предложен новый класс моделей (пространственная Λ -Флеминг — Вайот модель), в которой изменение генетического разнообразия происходит благодаря событиям вымирания — реколонизации. Такой подход обеспечивает выполнение свойств (1) и (3), а события размножения становятся зависимыми. Для данной модели, а также для двойственной ей Λ -коалесценции было получено множество результатов [5–7], существуют работы связанные с изучением применимости модели [8, 9]

Пространственный Λ -Флеминг — Вайот процесс описывает случай бесконечной плотности и получается как предельный переход процесса с конечным числом особей, который будем называть агентным (individual-based) процессом вымирания — реколонизации, в дальнейшем просто процесс вымирания — реколонизации.

Наша цель заключается в изучении свойств процесса вымирания — реколонизации, обобщении популяционных параметров на случай непрерывно распределённой популяции и написании симулятора для данного процесса.

1. Описание процесса

Рассмотрим двумерный тор $\mathbb{T}$, реализованный как квадрат с ребром длины L , склеенный по противоположным сторонам. В момент времени $t = 0$ особи распределены согласно однородному пуассоновскому точечному процессу [10] на $\mathbb{T}$ с плотностью ρ . Особь описывается координатой на торе (x^1, x^2) и генотипом, представляющим собой вектор из нулей и единиц длины N_ℓ . В начальный момент времени все координаты векторов генотипа для всех особей независимо и одинаково распределены согласно распределению Бернулли с параметром p . Изменение популяции происходит благодаря событиям вымирания — реколонизации. Назовем временами событий вымирания — реколонизации последовательность $\{t_n, n \in \mathbb{N}\} \subset \mathbb{R}_+$ случайных величин с независимыми приращениями, такую что

$$t_1 \sim \text{Exp}(\lambda), \quad (1)$$

$$t_{n+1} - t_n \sim \text{Exp}(\lambda), n = 1, 2, \dots \quad (2)$$

Пусть $x = (x^1, x^2), z = (z^1, z^2) \in \mathbb{T}$. Расстояние на торе между x и z будем формально обозначать как $\|x - z\|$, положим его равным

$$\|x - z\| = \sqrt{\min(|x^1 - z^1|, L - |x^1 - z^1|)^2 + \min(|x^2 - z^2|, L - |x^2 - z^2|)^2}.$$

В момент времени t_n происходит событие вымирания — реколонизации.

1. На торе из равномерного распределения выбирается точка z — центр события. Каждой особи, живущей в момент времени до события назначается величина

$$v(x, z) = \exp\left(-\frac{\|x - z\|^2}{2\alpha^2\theta^2}\right),$$

где x — координаты особи. Согласно взвешенной вероятности, полученной по значениям $v(x, z)$, выбирается родитель.

2. Каждая особь умирает с вероятностью $u(x, z)$, где

$$u(x, z) = u_0 \exp\left(-\frac{\|x - z\|^2}{2\theta^2}\right).$$

3. Из обобщенного пуассоновского распределения на торе с плотностью $u(x, z)$ (z фиксировано) генерируются новые особи. Генотип новых особей равен генотипу родителя.

Параметр u_0 описывает интенсивность, с которой умирают особи, а θ и α описывают ширину области, в которой умирают особи и выбирается родитель, соответственно.

2. Теоретические свойства модели

2.1. Время жизни

Время жизни особи считается как разность момента времени события, в котором она умерла и момента времени события, в котором она родилась.

Теорема 1. *Время жизни особи распределено экспоненциально с параметром $\frac{\lambda}{L^2} \iint_{\mathbb{T}} u_0 \exp -\frac{\|s\|^2}{2\theta^2} ds^1 ds^2$.*

Доказательство. Пусть $P(D|x, z)$ — вероятность, что в ходе события вымирания — реколонизации с центром в z особь, расположенная в x умрет. Эта вероятность равна $u(x, z)$. Так как события вымирания — реколонизации расположены равномерно на $\mathbb{T}$, то

$$P(D|x) = \frac{1}{L^2} \iint_{\mathbb{T}} u(x, z) dz = \frac{1}{L^2} \iint_{\mathbb{T}} u_0 \exp -\frac{\|x - z\|^2}{2\theta^2} dz^1 dz^2.$$

Интеграл берется по тору, так что, сделав замену $s = x - z$, мы получим, что

$$P(D|x) = \frac{1}{L^2} \iint_{\mathbb{T}} u_0 \exp -\frac{\|s\|^2}{2\theta^2} ds^1 ds^2.$$

Вероятность не зависит от x , и, так как особи распределены согласно однородному пуассоновскому процессу на торе, то положим

$$p_d = P(D) = \frac{1}{L^2} \iint_{\mathbb{T}} u_0 \exp -\frac{\|s\|^2}{2\theta^2} ds^1 ds^2.$$

Вероятности смерти особи в каждом событии вымирания — реколонизации равны, а сами события взаимоисключающие, значит, вероятность умереть через k событий вымирания — реколонизации равна

$$p_k = (1 - p_d)^{k-1} p_d, k = 1, 2, \dots$$

Пусть $f_k(t)$ — плотность вероятности суммы k независимых экспоненциально распределенных случайных величин с параметром λ . Такая сумма имеет гамма распределение с параметрами $(k, 1/\lambda)$.

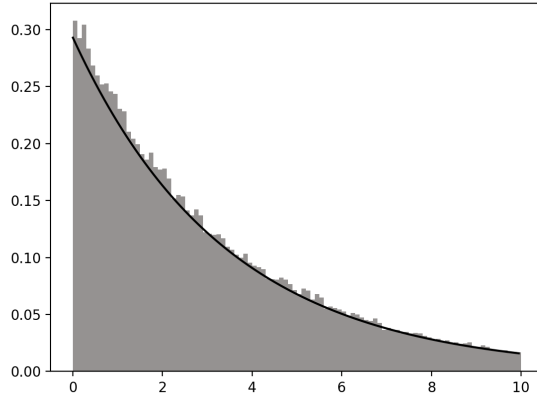


Fig. 1. Lifetime distribution based on simulations with parameters $u_0 = 0.4$, $\rho = 10000$, $\theta = 0.4$, and 10000 extinction — recolonization events. Bold line is a theoretical value.

Рис. 1. Распределение времени жизни в симуляции при параметрах $u_0 = 0.4$, $\rho = 10000$, $\theta = 0.4$, и 10000 событий вымирания — реколонизации. Выделенная линия — теоретическое значение.

Пусть $P(D_k)$ — вероятность умереть в k -ое событие вымирания — реколонизации, тогда плотность вероятности времени жизни $g(t)$ вычисляется по формуле полной вероятности:

$$g(t) = \sum_{k=1}^{\infty} P(D_k) f_k(t),$$

т. е.

$$g(t) = \sum_{k=1}^{\infty} (1 - p_d)^{k-1} p_d \frac{t^{k-1} \lambda^k e^{-\lambda t}}{\Gamma(k)} = \lambda p_d e^{-\lambda p_d t}.$$

Получается, что время жизни особи имеет экспоненциальное распределение с параметром λp_d . Среднее время жизни равно $\frac{1}{\lambda p_d}$. $\square$

Замечание. Частота смертей равна произведению частоты событий вымирания — реколонизации и вероятности умереть во время этого события, то есть λp_d , значит среднее время жизни равно $1/\lambda p_d$.

2.2. Сохранение вероятностей аллелей

Будем называть координаты генотипа аллелями. Говорим что аллель производная, если соответствующая координата равна единице.

Лемма 1. Вероятность производного аллеля у особей не зависит от времени.

Доказательство. При $0 \leq t < t_1$ вероятность что аллель производная равна p по начальному условию. При $t_1 \leq t < t_2$ для особей, переживших событие вымирания — реколонизации, вероятность производной аллели равна p по начальному условию, а для особей, родившихся в момент вымирания — реколонизации вероятность равна вероятности производного аллеля у родителя, то есть тоже равна p . Аналогично рассматривая остальные события вымирания — реколонизации, получаем, что вероятности не зависят от времени для любого t . $\square$

Лемма 2. Пусть в момент времени t в популяции существует ровно N индивидуумов, обозначим за G_l^i значение l -ой координаты генотипа i особи тогда $P(G_l^{i_0} = 1 | \sum_{i=1}^N G_l^i = k) = \frac{k}{N}$.

Доказательство.

$$\begin{aligned}
 P\left(G_l^{i_0} = 1 \mid \sum_{i=1}^N G_l^i = k\right) &= \frac{P(G_l^{i_0} = 1)P(\sum_{i=1, i \neq i_0}^N G_l^i = k-1)}{P(\sum_{i=1}^N G_l^i = k)} = \\
 &= \frac{p \binom{N-1}{k-1} p^{k-1} (1-p)^{N-k}}{\binom{N}{k} p^k (1-p)^{N-k}} = \frac{\binom{N-1}{k-1}}{\binom{N}{k}} = \frac{k}{N}. \quad (3)
 \end{aligned}$$

□

2.3. Связь с моделью Морана

Зафиксируем один аллель (координату в генотипе). Пусть X_t — число особей с производным аллелем в момент времени t , тогда назовем величины $p_{ml}(n) = P(X_{t_{n+1}} = m | X_{t_n} = l)$ переходными вероятностями числа особей с производным аллелем. Рассмотрим следующую (неэквивалентную) модификацию. Пусть $t_n = n$, а в событиях вымирания — реколонизации умирает и рождается ровно одна особь.

Теорема 2. *Переходные вероятности для модификации процесса равны переходным вероятностям для модели Морана.*

Доказательство. Зафиксируем число особей N . Согласно лемме 2, вероятность выбрать особь с производным аллелем при условии того, что в популяции их k равна k/N . Пусть X_t — число особей с производным аллелем в момент времени t , тогда

$$P(X_{t+1} = k+1 | X_t = k) = \frac{k}{N} \left(1 - \frac{k}{N}\right), \quad (4)$$

$$P(X_{t+1} = k-1 | X_t = k) = \frac{k}{N} \left(1 - \frac{k}{N}\right), \quad (5)$$

$$P(X_{t+1} = k | X_t = k) = 1 - 2 \frac{k}{N} \left(1 - \frac{k}{N}\right). \quad (6)$$

В остальных случаях вероятность равна нулю.

□

Так как вероятности равны вероятностям в модели Морана, то все свойства, такие как время фиксации, гетерозиготность и т. д. переносятся на модель вымирания — реколонизации с этими ограничениями.

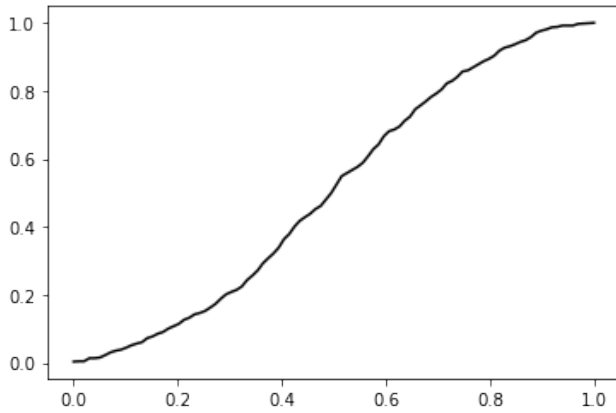
3. Популяционные характеристики для анализа непрерывно расположенных популяций

Существует множество характеристик для анализа популяций, таких как распределение аллелей, совместное распределение аллелей, гетерозиготность и т. д. Эти характеристики не учитывают пространственное расположение: в случае нескольких популяций они жестко разделяют их. Для модели вымирания — реколонизации и других пространственных моделей необходимо разработать обобщения характеристик, учитывающих пространственную структуру популяций.

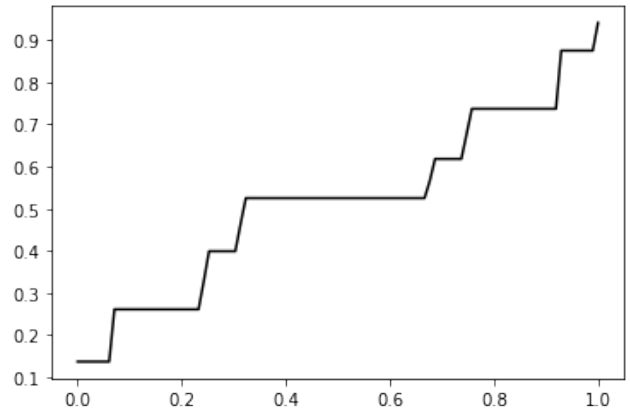
3.1. Распределение аллелей

Предположим что особи в популяции расположены согласно нормальному распределению с центром в точке z . Рассмотрим функцию

$$h(z, x; \beta) = \exp\left(-\frac{\|x - z\|^2}{2\beta^2}\right).$$

**Fig. 2.** Allele distribution

with $u_0 = 0.1, \rho = 10000, \theta = 0.15, \alpha = 1, N_\ell = 1000$. Left is distribution after 1000 events of extinction — recolonization, right is distribution after 5000 events of extinction — recolonization.

**Рис. 2.** Распределение аллелей

при $u_0 = 0.1, \rho = 10000, \theta = 0.15, \alpha = 1, N_\ell = 1000$. Слева — после 1000 событий вымирания реколонизации, справа — после 5000 событий вымирания — реколонизации.

Такая функция будет придавать больший вес особям, находящимся ближе к центру, и чем дальше особь от центра популяции, тем меньше её вклад.

Мы определим частоту аллеля l в точке x как

$$\theta^l = \frac{\sum_{x_k^{(t)} \sim l} h(x, x_k^{(t)})}{\sum_{x_k^{(t)}} h(x, x_k^{(t)})},$$

где сумма в числителе берется по всем живым особям с производным аллелем в позиции l в момент времени t , а знаменатель берется по всем живым особям в момент времени t .

Затем распределение аллелей определяется как

$$D^t(x, \theta) = \#\{l : \theta_l \leq \theta\} / N_\ell,$$

где $\#$ обозначает количество элементов в множестве.

Замечание. Классическое распределение аллелей ξ_i , как правило, вводится как число сайтов (координат), которые являются производными ровно в i особях. Т.е, стандартное определение будет представлять гистограмму, а нововведённое — функцию распределения. Ясно, что

$$\frac{\xi_i}{N_\ell} = \lim_{\beta \rightarrow \infty, \varepsilon \rightarrow 0+} D(x, i/N) - D(x, i/N - \varepsilon).$$

Мы также вводим совместное распределение аллелей для двух популяций, расположенных в точках x и y :

$$\theta_x^l = \frac{\sum_{x_k^{(t)} \sim l} h(x, x_k^{(t)})}{\sum_{x_k^{(t)}} h(x, x_k^{(t)})},$$

$$\theta_y^l = \frac{\sum_{x_k^{(t)} \sim l} h(y, x_k^{(t)})}{\sum_{x_k^{(t)}} h(y, x_k^{(t)})},$$

$$D^t(x, y, \theta_1, \theta_2) = \#\{l : \theta_x^l \leq \theta_1, \theta_y^l \leq \theta_2\} / N_{al}.$$

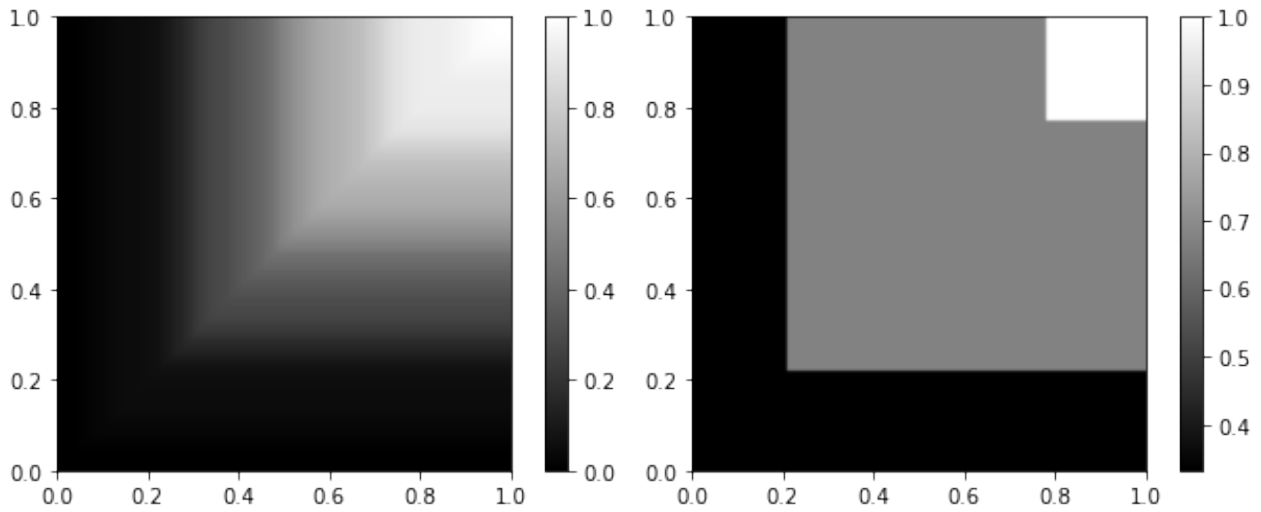


Fig. 3. Joint allele distribution for $\rho = 10000$, $\theta = 0.15$, $u_0 = 0.1$. On the left — after 1000 extinction — recolonization events, on the right — after 5000 extinction — recolonization events.

Рис. 3. Совместное распределение аллелей для $\rho = 10000$, $\theta = 0.15$, $u_0 = 0.1$. Слева — после 1000 событий вымирания — реколонизация, справа — после 5000 событий вымирания — реколонизации.

3.2. Гетерозиготность

Пусть X_t — количество особей с производным аллелем, а N_t — количество живых особей в момент времени t . Тогда величина $H_t = \frac{X_t(N_t - X_t)}{N^2}$ называется гетерозиготностью. Нулевая гетерозиготность указывает на фиксацию варианта аллеля. Вычисление средней гетерозиготности для популяции осложняется случайным размером популяции и случайным числом новых индивидуумов на каждом шаге. Однако, при наложенных ограничениях в 2.3, средняя гетерозиготность вычисляется аналогично модели Морана (см. например [11]):

$$\mathbb{E}(H_t) = H_0(1 - 2/N^2)^{\lfloor t \rfloor},$$

где $\lfloor \cdot \rfloor$ — функция округления в меньшую сторону. В общем случае были проведены симуляции с различными параметрами u_0 и θ , показывающие уменьшение гетерозиготности.

4. Программный симулятор

Был разработан симулятор для модели вымирания — реколонизации. Код написан на языке Python с использованием Cython для оптимизации. Для генерации индивидуумов с использованием неоднородного обобщенного пуассоновского процесса используется метод выборки с отклонением (acceptance-rejection) [10]: количество новых индивидуумов получается из распределения Пуассона с параметром $\Lambda = \iint_{\mathbb{T}} u(x, z) dx^1 dx^2 = L^2 p_d$. Новые индивидуумы генерируются равномерно на торе, а затем либо принимаются с вероятностью $u(z, x)/\Lambda$, либо отклоняются и генерируются заново.

Код симулятора доступен на GitHub по ссылке: <https://github.com/Genomics-HSE/slfvp>.

5. Численные результаты

Модель вымирания — реколонизации представляет новый способ исследования генетического разнообразия популяции, учитывая пространственное расположение её особей. Для модели вымирания-реколонизации были получены некоторые свойства. Было обнаружено, что продолжительность жизни индивидуумов имеет экспоненциальное распределение с средним, равным $1/\lambda p_d$. Вероятности производных аллелей у индивидуумов сохраняются со временем. Было доказа-

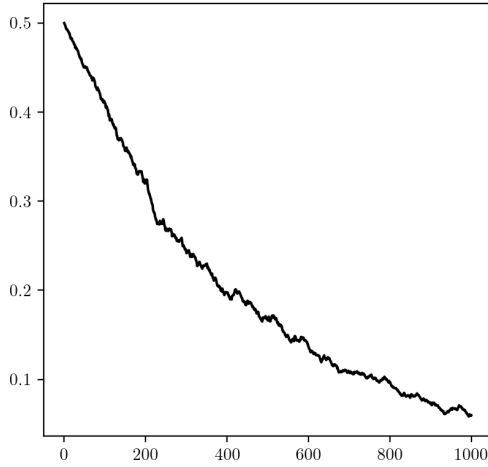


Fig. 4. Average heterozygosity at parameters $\theta = 0.6$, $\rho = 2000$, and 1000 extinction — recolonization events. Averaging was performed over 100 runs. On the left, $u_0 = 0.2$, on the right, $u_0 = 0.4$.

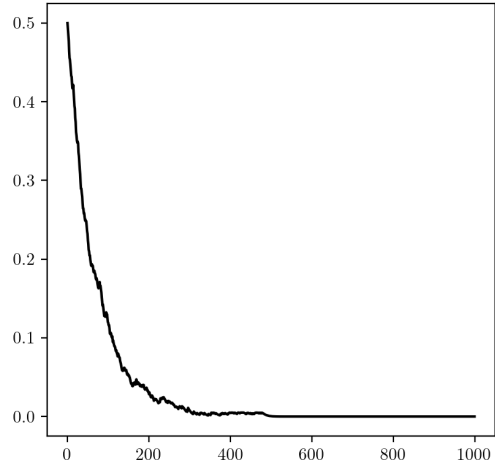


Рис. 4. Усредненная гетерозиготность при параметрах $\theta = 0.6$, $\rho = 2000$ и 1000 событиям вымирания — реколонизации. Усреднение проводилось по 100 запускам. Слева $u_0 = 0.2$, справа $u_0 = 0.4$

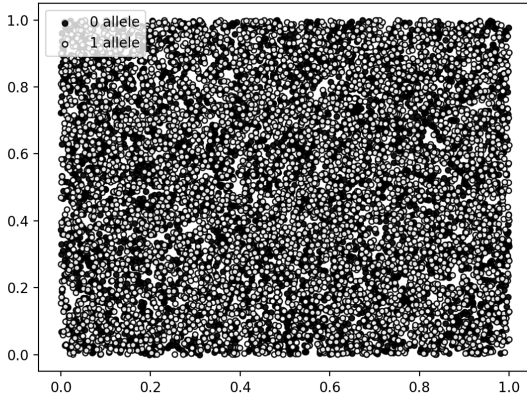


Fig. 5. Individuals with their alleles after one event of extinction — recolonization with parameters $\theta = 0.1$, $u_0 = 0.9$, $\rho = 10000$.

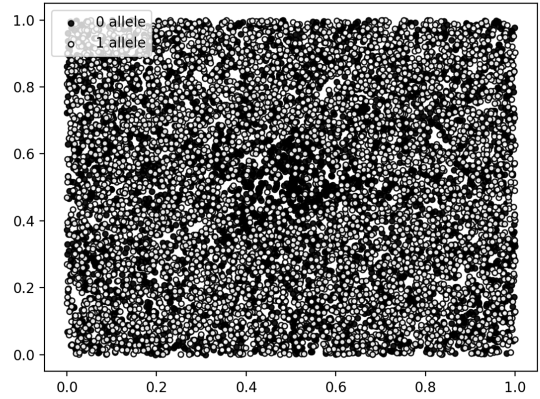


Рис. 5. Особи и их аллели до и после одного события вымирания — реколонизации с различными $\theta = 0.1$, $u_0 = 0.9$, $\rho = 10000$.

но, что процесс вымирания — реколонизации, при дополнительных ограничениях, сводится к модели Морана. Были определены пространственные аналоги распределения аллелей и совместного распределения аллелей.

Для анализа модели вымирания — реколонизации был разработан симулятор. Теоретические результаты были сопоставлены с результатами симуляции. Время симуляции (в среднем) масштабируется линейно с числом событий, плотностями и квадратично с длиной тора. Значительный вклад в оптимизацию времени симуляции внес выбор тора в качестве пространства, на котором живут особи, что делает интеграл $\iint_{\mathbb{T}} v(z, x) dz$ независимым от x . Для различных параметров модели были получены графики гетерозиготности.

Выводы

Популяционные модели, учитывающие пространственное расположение, представляют собой новое направление в популяционной генетике. Несмотря на ограничения, выявленные в [2], существуют процессы, успешно моделирующие генетическое разнообразие в пространстве. Рассмотренная модель вымирания — реколонизации относится к этому классу. Она, некоторой модификацией, рассмотренной в 2.3, сводится к модели Морана, и может рассматриваться как её обобщение. Всё ещё остаются вопросы по оценке средней гетерозиготности в общем случае. Неизвестно, насколько структура тора применима к реальным сценариям. В качестве дальнейшего изучения пространственных популяционных моделей интересно привлечение мутаций, отбора, рекомбинации для моделирования популяции. Интересно, как наши новые пространственные распределения частот аллелей совпадают с исходными.

References

- [1] R. Durrett and R. Durrett, *Probability models for DNA sequence evolution*. Springer, 2008. doi: [10.1007/978-0-387-78168-6](https://doi.org/10.1007/978-0-387-78168-6).
- [2] J. Felsenstein, “A pain in the torus: Some difficulties with models of isolation by distance”, *The American Naturalist*, vol. 109, no. 967, pp. 359–368, 1975.
- [3] A. M. Etheridge, “Survival and extinction in a locally regulated population”, *The Annals of Applied Probability*, vol. 14, no. 1, pp. 188–214, 2004.
- [4] A. Etheridge, *Some Mathematical Models from Population Genetics: École D’Été de Probabilités de Saint-Flour XXXIX-2009*. Springer Science & Business Media, 2011. doi: [10.1007/978-3-642-16632-7](https://doi.org/10.1007/978-3-642-16632-7).
- [5] N. Barton, A. Etheridge, and A. Véber, “A new model for evolution in a spatial continuum”, *Electronic Journal of Probability*, vol. 15, pp. 162–216, 2010. doi: [10.1214/EJP.v15-741](https://doi.org/10.1214/EJP.v15-741).
- [6] N. Biswas, A. Etheridge, and A. Klimek, “The spatial Lambda-Fleming-Viot process with fluctuating selection”, *Electronic Journal of Probability*, vol. 26, pp. 1–51, 2021. doi: [10.1214/21-EJP593](https://doi.org/10.1214/21-EJP593).
- [7] A. M. Etheridge, A. Véber, and F. Yu, “Rescaling limits of the spatial Lambda-Fleming-Viot process with selection”, *Electronic Journal of Probability*, vol. 25, pp. 1–89, 2020. doi: [10.1214/20-EJP523](https://doi.org/10.1214/20-EJP523).
- [8] S. Guindon, H. Guo, and D. Welch, “Demographic inference under the coalescent in a spatial continuum”, *Theoretical population biology*, vol. 111, pp. 43–50, 2016.
- [9] T. Joseph, M. Hickerson, and D. Alvarado-Serrano, “Demographic inference under a spatially continuous coalescent model”, *Heredity*, vol. 117, no. 2, pp. 94–99, 2016.
- [10] R. L. Streit and R. L. Streit, *The Poisson point process*. Springer, 2010. doi: [10.1007/978-1-4419-6923-1_2](https://doi.org/10.1007/978-1-4419-6923-1_2).
- [11] J. Wakely, *Coalescent Theory: An Introduction*. Macmillan Learning, 2016, ISBN: 9780974707754.