

# Метод стохастического уравнения обновления в определении статистики времени ожидания первого появления определенной последовательности состояний, последовательно посещенных чередующимся процессом обновления

С. А. Белан<sup>1)</sup>

*Институт теоретической физики им. Л. Д. Ландау РАН, 142432 Черноголовка, Россия*

*Национальный исследовательский университет Высшая школа экономики, факультет физики, 101000 Москва, Россия*

Поступила в редакцию 2 августа 2024 г.

После переработки 25 сентября 2024 г.

Принята к публикации 26 сентября 2024 г.

В работе демонстрируются преимущества метода стохастического уравнения обновления для описания статистики времени ожидания первого наблюдения периода времени фиксированной длины, в течение которого обновляемый процесс стохастических прыжков между двумя состояниями не покидает заданного состояния. Рассматриваются как марковские, так и произвольные распределения интервалов времени между прыжками. Сравнение аналитических предсказаний со случаем декоррелированного во времени процесса показывает, что корреляции могут как уменьшать, так и увеличивать соответствующее среднее время ожидания. Кроме того, сопоставляя экспоненциальные, субэкспоненциальные и модели с тяжелыми хвостами, характеризующиеся равными вероятностями наблюдения интересующего нас события, мы показываем, что более быстрое убывание плотности вероятности интервалов времени между прыжками подразумевает меньшее среднее время ожидания. Интересно, что независимо от деталей конкретной модели для рассматриваемых здесь прыжковых процессов как с дискретным, так и непрерывным временем, случайное время ожидания демонстрирует универсальность, становясь экспоненциально распределенным в пределе большой длительности ожидаемого события.

DOI: 10.31857/S0370274X24110124, EDN: RCRVEV

**Введение.** В стандартной формулировке задачи первого прохождения [1] событие прохождения связано с моментом первого входа процесса в определенное состояние или первого касания им границы некоторого множества состояний и, таким образом, характеризуется нулевой длительностью. Скажем, для процесса Бернулли событие первого прохождения соответствует появлению первого успеха (или неудачи) в серии испытаний, а плотность вероятности времени ожидания такого события описывается геометрическим распределением [2]. В качестве еще одного классического примера приведем одномерное броуновское движение, где событие первого прохождения связано с попаданием частицы в поглощающую точку, а результирующая статистика описывается распределением Леви–Смирнова [3, 4].

В более сложных модификациях обсуждаемой задачи рассматривается время ожидания некоторого “неточечного” события прохождения. Под последним понимается первое прохождение случайного процес-

са через заданную последовательность состояний за интервал времени фиксированной продолжительности. К примеру, автор работы [5] применил теорию мартингалов для вычисления среднего времени ожидания первого появления слова ABRACADABRA в случайной последовательности букв, генерируемой в мысленном эксперименте с печатающей обезьяной. Совсем недавно в работе [6] при помощи более прозрачного обратного метода Колмогорова была рассмотрена статистика времени ожидания некоторых простых серий исходов фиксированной длины в эксперименте с подбрасыванием симметричной монеты. В качестве схожего примера, но с непрерывным временем, можно вспомнить исследование статистики времени ожидания первого интервала заданной длины в пуассоновском потоке, что имеет значение для моделирования времени, требуемого пешеходам для пересечения транспортного потока, и для моделирования продолжительности мертвой эпохи в работе счетчиков Гейгера, см., например, классический учебник [7] и [8].

<sup>1)</sup>e-mail: sergb27@yandex.ru

Вышеупомянутые сценарии являются наиболее простыми с точки зрения аналитических вычислений, поскольку и испытания Бернулли, и поток Пуассона, представляют собой декоррелированные во времени процессы. Существует также ряд работ, посвященных времени ожидания первого наблюдения паттерна определенной продолжительности в моделях с временными корреляциями, простейшими примерами которых являются цепи Маркова первого порядка и системы с восстановлением. В частности, статистика времени ожидания некоторых достаточно простых цепочек исходов в последовательности, генерируемой цепью Маркова первого порядка с двумя состояниями, была рассмотрена в [9] путем построения аналогии с азартными играми. Более обще, работа [10] предлагает решение задачи определения статистического распределения времени ожидания любой наперед заданной цепочки исходов в дискретных коррелированных во времени данных, генерируемых цепями Маркова первого порядка и процессами восстановления (также известными как полумарковские процессы) с произвольным числом состояний. Основной областью применения этих результатов является анализ распределения определенных паттернов в длинных последовательностях ДНК [11]. Важным является также вклад работы [12], где исследовалась статистика времени ожидания достаточно длительного периода непрерывного пребывания в заданном состоянии чередующегося (т.е. обладающего двумя состояниями) процесса восстановления, представляющего популярную модель в анализе надежности.

Подводя итог, можно сказать, что различные версии задачи описания статистики времени ожидания события ненулевой длительности рассматривались разными методами в чисто математической литературе, в прикладной теории вероятностей и в физических журналах.

Целью данной статьи является демонстрация наиболее прямого и интуитивно понятного (по крайней мере для исследователей, работающих в области статистической физики) аналитического метода определения статистики времени ожидания в моделях прыжковых процессов со статистически независимыми интервалами между прыжками. Чтобы сделать изложение ключевых идей и вычислительных аспектов максимально простым, мы выбрали модель с всего двумя состояниями и рассматриваем самое простое возможное событие в качестве порогового – постоянное пребывание процесса в заданном состоянии в течение интервала времени фиксированной длины. Статистические характеристики времени ожидания в такой постановке задачи могут быть по-

лучены путем усреднения специфического представления случайной реализации этой величины – подход, который может быть назван методом стохастического уравнения обновления. Помимо методологических вопросов, особое внимание в данной работе уделяется пределу событий большой длительности, в котором, как будет показано ниже, возникает универсальное статистическое поведение. Другим важным аспектом предлагаемого анализа является сравнение результатов для различных типов распределений интервалов времени между прыжками. В заключительной части работы мы обсуждаем потенциальные приложения полученных результатов в областях исследований, где подобные вопросы, насколько нам известно, пока не рассматривались.

**Марковская динамика.** Рассмотрим цепь Маркова с двумя состояниями, характеризующуюся вероятностями перехода  $p$  и  $q$  за один шаг дискретного времени, как показано на рис. 1а. Траектория такого прыжкового процесса может быть закодирована бинарной динамической переменной  $x_t$ . Более конкретно, мы будем говорить, что  $x_t = 1$ , где  $t = 0, 1, 2, \dots$ , если процесс находится в состоянии 1 на шаге  $t$ , и  $x_t = 0$ , если процесс в состоянии 2. Через  $k$  обозначим время, требуемое чтобы впервые встретить  $N$  единиц подряд в ряду значений переменной  $x_t$ . Среднее значение случайной величины  $k$  для произвольного начального состояния было вычислено в работе [9], в то время как полная статистика  $k$  может быть, в принципе, извлечена из более общих результатов, содержащихся в [10].

В этом разделе мы показываем, как получить статистику  $k$  на основе метода стохастического уравнения обновления. Кратко опишем суть данного метода. Вводя индикаторные функции, аргументы которых задают условия на начальное состояние и на момент первого перехода между парой состояний, мы можем разделить множество возможных наблюдений на взаимно исключающие сценарии, которые, будучи вместе взятыми, учитывают все возможные траектории процесса. Такое представление приводит к набору так называемых уравнений обновления на случайное время ожидания, а также на случайные переменные, входящие в декомпозицию этого времени по альтернативным историям динамики процесса. Благодаря марковскому свойству процесса, эти уравнения рекурсивны в том смысле, что они выражают некоторые случайные переменные через их статистически независимые копии. Строго говоря, число неизвестных при этом превосходит число уравнений, однако используя статистическую независимость можно получить замкнутые выражения

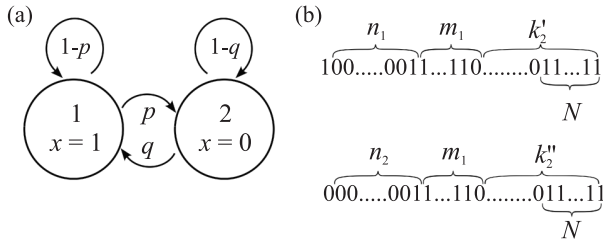


Рис. 1. (а) – Марковский процесс с двумя состояниями в дискретном времени с вероятностями перехода за отдельный шаг  $p$  и  $q$ . Состояния 1 и 2 соответствуют значениям  $x = 1$  и  $x = 0$  динамической переменной  $x_t$ . (б) – Представление случайного времени  $k$  ожидания первого появления серии из  $N$  единиц подряд в последовательности значений, принимаемых переменной  $x_t$ . Предположим, что в нулевой момент времени процесс находился в состоянии 0 (нижняя часть рис. (б)). Поскольку желаемая серия начинается с 1, процесс сперва должен оказаться в состоянии 1, что требует некоторого случайного числа шагов, которое на рисунке и в тексте обозначено как  $n_2$ . Далее, число шагов, спустя которое процесс затем покинет состояние 1, обозначено через  $m_1$ . Если  $m_1 > N - 1$ , то интересующая нас серия из  $N$  единиц подряд появится в течении этих  $m_1$  шагов, т.е. спустя  $n_1 + N - 1$  шагов спустя начала наблюдения. В противном случае, если  $m_1 \leq N - 1$ , процесс покидает состояние 1 слишком рано, так что необходимо будет подождать еще  $k_2''$  шагов, пока желаемая последовательность наконец не встретится. Благодаря марковскому свойству и статистической однородности, величины  $k_2$  и  $k_2''$  независимы и одинаково распределены. Аналогичное представление может быть построено, если процесс стартует из состояния 2 (верхняя часть рисунка (б)). Вышеописанные соображения формализованы в ур. (1), (2) и (3)

для усредненных характеристик случайного времени ожидания. Следуя этой логике ниже мы вычисляем характеристическую функцию, определенную как среднее  $\langle e^{isk} \rangle$ , и математическое ожидание  $\langle k \rangle$ .

Итак, очевидно,  $k$  может быть представлено как

$$k = k_1 I(x_0 = 1) + k_2 I(x_0 = 0), \quad (1)$$

где  $I(\dots)$  – это индикаторная функция, которая равна единице, если выполнено условие, заданное ее аргументом, и нулю в противоположном случае, в то время как  $k_1$  и  $k_2$  обозначают случайное время ожидания в эксперименте, где процесс исходно находился в состояниях 1 и 2 соответственно. Благодаря марковскому свойству процесса  $x_t$ , условные времена ожидания  $k_1$  и  $k_2$  удовлетворяют следующей системе сцепленных уравнений обновления

$$k_1 = n_1 + (N-1)I(m_1 > N-1) + (m_1 + k_2')I(m_1 \leq N-1), \quad (2)$$

$$k_2 = n_2 + (N-1)I(m_1 > N-1) + (m_1 + k_2'')I(m_1 \leq N-1), \quad (3)$$

где  $k_2'$  и  $k_2''$  обозначают статистически независимые копии  $k_2$ ;  $n_1 \geq 1$  – это момент времени, когда впервые наблюдается 1 при условии  $x_0 = 1$ ;  $m_1 \geq 1$  – это момент времени, когда впервые наблюдается 0, если  $x_0 = 1$ ;  $n_2 \geq 1$  – это момент времени, когда впервые наблюдается 1 при условии  $x_0 = 0$ , см. рис. 1b и подпись к нему.

**Характеристическая функция.** Полная статистика случайной переменной  $T$  может быть извлечена из характеристической функции, которая определена как  $\langle e^{isT} \rangle$ . Уравнение (1) позволяет нам записать  $e^{isk} = e^{isk_1} I(x_0 = 1) + e^{isk_2} I(x_0 = 0)$ , и, следовательно,

$$\begin{aligned} \langle e^{isk} \rangle &= \langle e^{isk_1} I(x_0 = 1) \rangle + \langle e^{isk_2} I(x_0 = 0) \rangle = \\ &= \langle e^{isk_1} \rangle Pr[x_0 = 1] + \langle e^{isk_2} \rangle Pr[x_0 = 0] = \\ &= \frac{q}{p+q} \langle e^{isk_1} \rangle + \frac{p}{p+q} \langle e^{isk_2} \rangle. \end{aligned} \quad (4)$$

Мы предполагаем статистически-стационарное начальное условие, так что  $Pr[x_0=1] = \frac{q}{p+q}$  и  $Pr[x_0=0] = \frac{p}{p+q}$ .

Далее, из ур. (2) и (3) находим

$$\begin{aligned} e^{isk_1} &= e^{is(n_1+N-1)} I(m_1 > N-1) + \\ &+ e^{is(n_1+m_1+k_2')} I(m_1 \leq N-1), \end{aligned} \quad (5)$$

и

$$\begin{aligned} e^{isk_2} &= e^{is(n_2+N-1)} I(m_1 > N-1) + \\ &+ e^{is(n_2+m_1+k_2'')} I(m_1 \leq N-1). \end{aligned} \quad (6)$$

Используя статистическую независимость переменных  $n_1$ ,  $m_1$  и  $k_2$  в ур. (5), мы получаем

$$\begin{aligned} \langle e^{isk_1} \rangle &= e^{is(N-1)} \langle e^{isn_1} \rangle \langle I(m_1 > N-1) \rangle + \\ &+ \langle e^{isn_1} \rangle \langle e^{isk_2'} \rangle \langle e^{ism_1} I(m_1 \leq N-1) \rangle. \end{aligned} \quad (7)$$

Схожим образом, статистическая независимость переменных  $n_2$ ,  $m_1$  и  $k_2$  в ур. (6) позволяет нам записать

$$\begin{aligned} \langle e^{isk_2} \rangle &= \langle e^{is(N-1)} \rangle \langle e^{isn_2} \rangle \langle I(m_1 > N-1) \rangle + \\ &+ \langle e^{isn_2} \rangle \langle e^{isk_2''} \rangle \langle e^{ism_1} I(m_1 \leq N-1) \rangle. \end{aligned} \quad (8)$$

Очевидно,  $\langle e^{isk_2} \rangle = \langle e^{isk_2'} \rangle = \langle e^{isk_2''} \rangle$ . Тогда, решая систему сцепленных ур. (7) и (8) для  $\langle e^{isk_1} \rangle$  и  $\langle e^{isk_2} \rangle$ , имеем

$$\langle e^{isk_1} \rangle = \frac{e^{is(N-1)} \langle e^{isn_1} \rangle \langle I(m_1 > N-1) \rangle}{1 - \langle e^{isn_2} \rangle \langle e^{ism_1} I(m_1 \leq N-1) \rangle}, \quad (9)$$

$$\langle e^{isk_2} \rangle = \frac{e^{is(N-1)} \langle e^{isn_2} \rangle \langle I(m_1 > N-1) \rangle}{1 - \langle e^{isn_2} \rangle \langle e^{ism_1} I(m_1 \leq N-1) \rangle}. \quad (10)$$

Подстановка ур. (9) и (10) в ур. (4) дает

$$\langle e^{isk} \rangle = \frac{(q \langle e^{isn_1} \rangle + p \langle e^{isn_2} \rangle) e^{is(N-1)} \langle I(m_1 > N-1) \rangle}{(p+q)(1 - \langle e^{isn_2} \rangle \langle e^{ism_1} I(m_1 \leq N-1) \rangle)}. \quad (11)$$

Средние значения в правой части ур. (11) могут быть вычислены при помощи следующих вероятностных распределений

$$p_1(m_1) = p(1-p)^{m_1-1}, \quad (12)$$

и

$$p_2(n_2) = q(1-q)^{n_2-1}. \quad (13)$$

Тогда

$$\langle e^{ism_1} \rangle = \sum_{m_1=1}^{\infty} e^{ism_1} p_1(m_1) = \frac{pe^{is}}{1 - (1-p)e^{is}}, \quad (14)$$

$$\langle I(m_1 > N-1) \rangle = \sum_{m_1=N}^{\infty} p_1(m_1) = (1-p)^{N-1}, \quad (15)$$

$$\begin{aligned} \langle e^{ism_1} I(m_1 \leq N-1) \rangle &= \sum_{m_1=1}^{N-1} e^{ism_1} p_1(m_1) = \\ &= \frac{p((1-p)^N e^{iNs} - (1-p)e^{is})}{(p-1)(1 + (p-1)e^{is})}, \end{aligned} \quad (16)$$

и

$$\langle e^{isn_2} \rangle = \frac{qe^{is}}{1 - (1-q)e^{is}}. \quad (17)$$

Для выполнения усреднения по статистике случайной величины  $n_1$ , заметим что

$$n_1 = 1 \cdot I(b=1) + (1+n_2) \cdot I(b=0), \quad (18)$$

где  $b$  – это переменная Бернулли, которая равна единице с вероятностью  $1-p$ , и нулю с вероятностью  $p$ . Из ур. (18) следует, что

$$e^{isn_1} = I(x=1)e^{is} + I(x=0)e^{is(1+n_2)}, \quad (19)$$

и, поэтому,

$$\begin{aligned} \langle e^{isn_1} \rangle &= (1-p)e^{is} + pe^{is} \langle e^{isn_2} \rangle = \\ &= \frac{e^{is}(1-p + (p+q-1)e^{is})}{1 - (1-q)e^{is}}, \end{aligned} \quad (20)$$

где мы использовали ур. (17).

Подстановка выражений (14), (15), (16), (17) и (20) в ур. (11) дает окончательный ответ для характеристической функции времени ожидания  $k$

$$\langle e^{isk} \rangle = \frac{q(1-p)^N e^{iNs} (1 - (1-p-q)e^{is})(1 - (1-p)e^{is})}{(p+q)[(1-p)(1 - (1-q)e^{is})(1 - (1-p)e^{is}) - qpe^{is}((1-p)e^{is} - (1-p)^N e^{iNs})]}. \quad (21)$$

*Среднее время ожидания.* Точное распределение  $p(k)$ , являющееся обратным преобразованием Фурье от  $\langle e^{isk} \rangle$ , не может быть получено аналитически. Тем не менее ур. (21) может быть использовано для вычисления среднего значения  $\langle k \rangle$ , поскольку статистические моменты  $k$  определяются коэффициентами разложения характеристической функции  $\langle e^{isk} \rangle$  в ряд по степеням  $s$ .

Из разложения до линейных членов  $\langle e^{isk} \rangle = 1 + is\langle k \rangle + O(s^2)$ , мы находим следующее достаточно простое выражение

$$\langle k \rangle = \frac{p+q}{pq}(1-p)^{-N+1} - \frac{1}{p+q} - \frac{1-p}{p}, \quad (22)$$

которое в пределе больших  $N$  принимает вид

$$\langle k \rangle_{\infty} = \frac{p+q}{pq}(1-p)^{-N+1}. \quad (23)$$

Исследуем, как общий ответ, заданный ур. (22), упрощается при  $q = 1-p$ . В этом случае исходы

$x_0, x_1, x_2, \dots$  становятся статистически независимыми, так что процесс  $x_t$  сводится к простому процессу Бернулли. Из ур. (22) получаем

$$\langle k \rangle^{\text{Bern}} = \frac{1}{p(1-p)^N} - \frac{1}{p}. \quad (24)$$

Отметим, что после подстановки  $p = q = 1/2$ , которая соответствует эксперименту с подбрасыванием симметричной монеты, ур. (24) воспроизводит результат

$$\langle k \rangle^{f.c.} = 2^{N+1} - 2, \quad (25)$$

ранее известный из работы [6].

Уравнение (22) показывает, что корреляции могут как увеличивать, так и уменьшать среднее время ожидания в сравнении со случаем отсутствия корреляций. Действительно, давайте сравним эксперимент по подкидыванию честной монеты и марковский процесс с двумя состояниями, характеризующиеся одинаковой вероятностью получения  $N$  единиц подряд

в  $N \gg 1$  последовательных исходах. Приравнявая соответствующие вероятности в этих двух моделях,  $\frac{1}{2^N}$  и  $\frac{q}{p+q}(1-p)^{N-1}$ , и, учитывая ур. (23), мы получаем условие  $\langle k \rangle_\infty = \frac{2^N}{p}$ . Вместе с ур. (25) это дает

$$\frac{\langle k \rangle_\infty}{\langle k \rangle_\infty^{f.c.}} \approx \frac{1}{2p}, \quad (26)$$

так что  $\langle k \rangle_\infty / \langle k \rangle_\infty^{f.c.} < 1$ , где  $p > 1/2$ , и  $\langle k \rangle_\infty / \langle k \rangle_\infty^{f.c.} > 1$  при  $p < 1/2$ . Интересно заметить, что, как следует из (26) марковские корреляции первого порядка не могут уменьшить среднее время ожидания более чем в два раза.

*Предел больших  $N$ .* Используя  $\langle k \rangle$  как оценку для характерной ширины функции распределения  $p(k)$ , мы можем заключить, что для описания “тела” этой функции достаточно знать ее характеристическую функцию в области  $s \lesssim \langle k \rangle^{-1}$ . Примечательно, ур. (22) показывает, что для достаточно больших  $N$  имеет место разделение масштабов времени  $\langle k \rangle \gg N$ . Следовательно, для указанных выше релевантных значений  $s$  мы можем упростить ур. (21), раскладывая экспоненты в его правой части. Это приводит к следующему приближению для характеристической функции

$$\langle e^{isk} \rangle \approx \frac{pq(1-p)^N}{pq(1-p)^N - i(1-p)(p+q)s}, \quad (27)$$

откуда сразу же следует экспоненциальная предельная форма распределения вероятностей  $p(k) \approx \frac{1}{\langle k \rangle_\infty} \exp(-\frac{k}{\langle k \rangle_\infty})$  со средним значением  $\langle k \rangle_\infty = \frac{p+q}{pq}(1-p)^{-N+1}$  (см. ур. (23)), представляющим асимптотику общего выражения (22) при  $N \gg 1$ .

*Немарковская динамика.* Перейдем теперь к случаю бинарного случайного процесса с немарковскими корреляциями. А именно, рассмотрим стохастические прыжки между двумя состояниями, см. рис. 2а, согласно следующим простым правилам. Если процесс перешел в состояние 1, то он остается там случайное время  $t_1$ , имеющее плотность распределения общего вида  $\rho_1(t_1)$ . Схожим образом, время пребывания  $t_2$  процесса в состоянии 2 также представляет собой случайную переменную с произвольной плотностью распределения  $\rho_2(t_2)$ . Случайные длительности интервалов времени между прыжками являются статистически независимыми. Прыжковые процессы с такими свойствами известны в литературе как полумарковские процессы или марковские процессы обновления.

Пусть  $x(t)$  – это бинарная функция, для которой значение  $x(t) = 1$  означает, что процесс находится

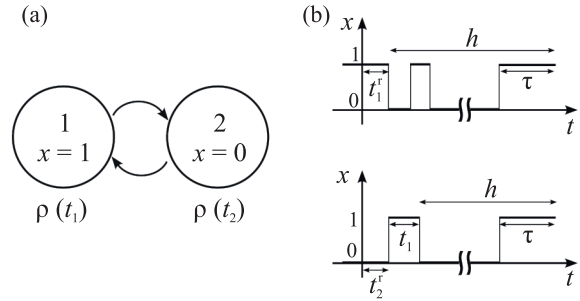


Рис. 2. Прыжковый процесс с двумя состояниями в непрерывном времени с произвольными плотностями вероятности  $\rho_1(t_1)$  и  $\rho_2(t_2)$  интервалов  $t_1$  и  $t_2$  между прыжками

в состоянии 1 в момент времени  $t$ , в то время как значение  $x(t) = 0$  соответствует состоянию 2. Мы начинаем наблюдать за этим процессом в случайный момент времени и хотим описать статистику времени ожидания  $T$  первого промежутка времени длины  $\tau$ , в течение которого процесс не покидает состояния 1. Для упрощения анализа, в этом разделе будет предполагаться, что  $\tau$  – это дискретная переменная. Отметим, что если исходно имеется дискретная модель, то переход к непрерывному пределу оправдан, если все релевантные масштабы времени велики в сравнении с дискретным шагом, т.е.  $\langle t_1 \rangle, \langle t_2 \rangle, \tau \gg 1$ .

Случайное время ожидания  $T$  может быть представлено следующим образом

$$T = T_1 I(x(0) = 1) + T_2 I(x(0) = 0), \quad (28)$$

где  $T_1$  и  $T_2$  обозначают случайные времена ожидания при условии, что в момент, когда мы начинаем наблюдать за процессом, он был в состоянии 1 и 2 соответственно. Эти случайные величины, в свою очередь, удовлетворяют следующей системе уравнений

$$T_1 = \tau I(t_1^r \geq \tau) + (t_1^r + t) I(t_1^r < \tau), \quad (29)$$

$$T_2 = t_2^r + \tau I(t_1 \geq \tau) + (t_1 + t) I(t_1 < \tau), \quad (30)$$

$$t = t_2 + \tau I(t_1' \geq \tau) + (t_1' + t') I(t_1' < \tau). \quad (31)$$

Здесь  $t_1^r$  ( $t_2^r$ ) обозначает остаточное время пребывания, т.е. время, оставшееся до ухода процесса из состояния 1 (2), если он был в этом состоянии в случайный момент, когда мы начали его наблюдать;  $t$  – это оставшееся время ожидания в ситуации, если процесс покидает состояние 1 до того, как произошло интересующее нас событие первого прохождения;  $t_1'$  и  $t'$  – это статистически-независимые копии  $t_1$  и  $t$  соответственно. Посредством введения индикаторных функций и переменной  $t$  мы учитываем в ур. (29),

(30) и (31) все возможные варианты динамики процесса, см. также рис. 2б.

*Среднее время ожидания.* Выполняя усреднение по статистике прыжков в ур. (28), мы получаем

$$\langle T \rangle = \langle T_1 \rangle \text{Pr}[x(0) = 1] + \langle T_2 \rangle \text{Pr}[x(0) = 0] = \frac{\langle t_1 \rangle \langle T_1 \rangle + \langle t_2 \rangle \langle T_2 \rangle}{\langle t_1 \rangle + \langle t_2 \rangle}, \quad (32)$$

где учтено, что в статистически-стационарном режиме вероятность найти процесс в состоянии 1 (2) дается отношением  $\frac{\langle t_1 \rangle}{\langle t_1 \rangle + \langle t_2 \rangle}$  ( $\frac{\langle t_2 \rangle}{\langle t_1 \rangle + \langle t_2 \rangle}$ ), где  $\langle t_1 \rangle = \int_0^\infty dt_1 \rho_1(t_1) t_1$  и  $\langle t_2 \rangle = \int_0^\infty dt_2 \rho_2(t_2) t_2$  [13].

Далее, используя статистическую независимость случайных величин  $t_1$ ,  $t_1^r$ ,  $t_1'$ ,  $t$  и  $t'$ , входящих в ур. (29), (30) и (31), мы находим замкнутую систему уравнений для  $\langle T_1 \rangle$ ,  $\langle T_2 \rangle$  и  $\langle t \rangle$ . Решая эту систему, получаем

$$\langle T_1 \rangle = \tau + \langle t_1^r I(t_1^r < \tau) \rangle + [\langle t_2 \rangle + \langle t_1 I(t_1 < \tau) \rangle] \frac{\langle I(t_1^r < \tau) \rangle}{\langle I(t_1 \geq \tau) \rangle}, \quad (33)$$

и

$$\langle T_2 \rangle = \langle t_2^r \rangle + \tau + \langle t_1 I(t_1 < \tau) \rangle + [\langle t_2 \rangle + \langle t_1 I(t_1 < \tau) \rangle] \frac{\langle I(t_1 < \tau) \rangle}{\langle I(t_1 \geq \tau) \rangle}. \quad (34)$$

Подстановка ур. (33) и (34) в ур. (32) дает

$$\begin{aligned} \langle T \rangle &= \tau + \frac{\langle t_1 \rangle}{\langle t_1 \rangle + \langle t_2 \rangle} \times \\ &\times \left[ \langle t_1^r I(t_1^r < \tau) \rangle + (\langle t_2 \rangle + \langle t_1 I(t_1 < \tau) \rangle) \frac{\langle I(t_1^r < \tau) \rangle}{\langle I(t_1 \geq \tau) \rangle} \right] + \\ &+ \frac{\langle t_2 \rangle}{\langle t_1 \rangle + \langle t_2 \rangle} \left[ \langle t_2^r \rangle + \frac{\langle t_1 I(t_1 < \tau) \rangle}{\langle I(t_1 \geq \tau) \rangle} + \langle t_2 \rangle \frac{\langle I(t_1 < \tau) \rangle}{\langle I(t_1 \geq \tau) \rangle} \right]. \end{aligned} \quad (35)$$

Для средних значений в правой части вышестоящего уравнения, включающих в себя случайный интервал  $t_1$ , мы легко находим

$$\langle I(t_1 < \tau) \rangle = \int_0^\tau \rho_1(t_1) dt_1, \quad (36)$$

$$\langle I(t_1 \geq \tau) \rangle = \int_\tau^\infty \rho_1(t_1) dt_1, \quad (37)$$

и

$$\langle t_1 I(t_1 < \tau) \rangle = \int_0^\tau t_1 \rho_1(t_1) dt_1. \quad (38)$$

Касательно усреднения по  $t_1^r$  and  $t_2^r$ , заметим, что функции плотности вероятности этих случайных величин имеют вид  $p_1(t_1^r) = \frac{1}{\langle t_1^r \rangle} \int_{t_1^r}^\infty dt_1 \rho_1(t_1)$  и  $p_2(t_2^r) = \frac{1}{\langle t_2^r \rangle} \int_{t_2^r}^\infty dt_2 \rho_2(t_2)$  соответственно (см. [1]). Следовательно

$$\langle I(t_1^r < \tau) \rangle = \frac{\int_0^\tau dt_1^r \int_{t_1^r}^\infty dt_1 \rho_1(t_1)}{\int_0^\infty dt_1 \rho_1(t_1) t_1}, \quad (39)$$

$$\langle t_1^r I(t_1^r < \tau) \rangle = \frac{\int_0^\tau dt_1^r t_1^r \int_{t_1^r}^\infty dt_1 \rho_1(t_1)}{\int_0^\infty dt_1 \rho_1(t_1) t_1}, \quad (40)$$

и

$$\langle t_2^r \rangle = \frac{\int_0^\infty dt_2^r t_2^r \int_{t_2^r}^\infty dt_2 \rho_2(t_2)}{\int_0^\infty dt_2 \rho_2(t_2) t_2}. \quad (41)$$

Это означает, в частности, что  $\langle t_2^r \rangle = \langle t_2 \rangle / (2 \langle t \rangle)$  в противоположность наивной оценке  $\langle t_1 \rangle / 2$  [14]. Несколько неожиданное поведение статистики оставшегося до очередного прыжка времени лежит в основе хорошо известного парадокса инспекции в теории обновляемых процессов [15].

Выражение (35) совместно с ур. (36)–(41) дает точный общий ответ для среднего времени ожидания  $\langle T \rangle$ . Важно отметить, этот результат отличается от представленного для аналогичной модели с двумя состояниями в работе [12]. Это объясняется тем, что наш анализ предполагает случайность выбора начального момента наблюдения (см. ур. (32)), в то время как автор работы [12] выполнил свои вычисления в предположении, что наблюдение за процессом начинается в момент его очередного входа в состояние 1.

Как видно из приведенных выше вычислений, в принятой нами формулировке задачи среднее  $\langle T \rangle$  конечно, если существуют статистические моменты  $\langle t_1 \rangle$ ,  $\langle t_2 \rangle$  и  $\langle t_2^2 \rangle$ . Заметьте, однако, что хотя среднее время ожидания расходится для распределений с достаточно тяжелыми хвостами, метод стохастического уравнения обновления тем не менее может быть применен для определения сходящихся статистических характеристик, например характеристической функции и производящей моменты функции.

*Предел больших  $\tau$ .* Сфокусируемся на случае экстремального события, когда процесс тратит в состоянии 1 время  $\tau$ , которое велико в сравнении со средним интервалом времени  $\langle t_1 \rangle$  между прыжками и средним временем  $\langle t_1^r \rangle = \langle t_1^2 \rangle / (2 \langle t_1 \rangle)$ , оставшимся до очередного прыжка (предполагаем, что соответствующее среднее конечно). Тогда ведущие вклады в  $\langle T_1 \rangle$  и  $\langle T_2 \rangle$  связаны с последними членами в правых частях ур. (33) и (34). То есть  $\langle T_1 \rangle \approx \langle T_2 \rangle \approx \frac{\langle t_1 \rangle + \langle t_2 \rangle}{\langle I(t_1 \geq \tau) \rangle}$ , и, следовательно, среднее время ожидания оценивается как

$$\langle T \rangle_\infty \approx \frac{\int_0^\infty dt_1 \rho_1(t_1) t_1 + \int_0^\infty dt_2 \rho_2(t_2) t_2}{\int_\tau^\infty dt_1 \rho_1(t_1)}, \quad (42)$$

где мы учли ур. (32).

Информативно сравнить среднее время, предсказываемое ур. (42) для экспоненциального, субэкспоненциального и распределения с тяжелыми хвостами. В экспоненциальном случае,  $\rho_1(t_1) = \lambda_1 e^{-\lambda_1 t_1}$  и  $\rho_2(t_2) = \lambda_2 e^{-\lambda_2 t_2}$ , мы находим из ур. (42)

$$\langle T \rangle_{\infty}^{\text{exp}} \approx \frac{\lambda_1 + \lambda_2}{\lambda_1 \lambda_2} e^{\lambda_1 \tau}. \quad (43)$$

Как простой пример субэкспоненциальной статистики мы выбираем полу-гауссово распределение  $\rho_1(t_1) = \sqrt{\frac{2}{\pi \sigma_1^2}} \exp(-\frac{t_1^2}{2\sigma_1^2})$  и  $\rho_2(t_2) = \sqrt{\frac{2}{\pi \sigma_2^2}} \exp(-\frac{t_2^2}{2\sigma_2^2})$ , получая

$$\langle T \rangle_{\infty}^{\text{norm}} \approx (1 + \frac{\sigma_2}{\sigma_1}) \tau e^{\frac{\tau^2}{2\sigma_1^2}}. \quad (44)$$

Наконец, используя распределение Парето,  $\rho_1(t_1) = \frac{\alpha_1 t_{01}^{\alpha_1}}{t_1^{\alpha_1+1}} I(t_1 \geq t_{01})$  и  $\rho_2(t_2) = \frac{\alpha_2 t_{02}^{\alpha_2}}{t_2^{\alpha_2+1}} I(t_2 \geq t_{02})$  с  $\alpha_1, \alpha_2 > 3$ , как основополагающий пример распределения с тяжелыми хвостами, мы находим

$$\langle T \rangle_{\infty}^{\text{pow.law}} \approx (\frac{\alpha_1}{\alpha_1 - 1} t_{01} + \frac{\alpha_2}{\alpha_2 - 1} t_{02}) \frac{\tau_1^{\alpha_1}}{t_{01}^{\alpha_1 - 1}}. \quad (45)$$

Вероятность  $\pi$  наблюдения непрерывного пребывания процесса в состоянии 1 в течении интервала времени  $\tau$ , начиная с какого-то случайно выбранного момента, равна

$$\pi = \frac{\langle t_1 \rangle}{\langle t_1 \rangle + \langle t_2 \rangle} Pr[t_1^r \geq \tau] = \quad (46)$$

$$= \frac{1}{\langle t_1 \rangle + \langle t_2 \rangle} \int_{\tau}^{\infty} dt_1^r \int_{t_1^r}^{\infty} dt_1 \rho_1(t_1). \quad (47)$$

Тогда получаем

$$\pi_{\text{exp}} = \frac{\lambda_2}{\lambda_1 + \lambda_2} \exp(-\lambda_1 \tau),$$

$$\pi_{\text{norm.}} = \frac{\sigma_1^3}{(\sigma_1 + \sigma_2) \tau^2} \exp(-\frac{\tau^2}{2\sigma_1^2})$$

и

$$\pi_{\text{pow.law}} = \frac{\alpha_1 t_{01}^{\alpha_1} \tau^{2-\alpha_1} + \alpha_2 t_{02}^{\alpha_2} \tau^{2-\alpha_2}}{(\alpha_2 - 1) \alpha_1 t_{01} + (\alpha_1 - 1) \alpha_2 t_{02}}$$

для экспоненциального, полу-гауссова и степенного распределений соответственно. В целях сравнения рассмотрим статистически симметричную ситуацию, с одинаково распределенными интервалами  $t_1$  и  $t_2$  и выберем параметры распределений (т.е. величины  $\lambda$ ,  $\sigma$ ,  $t_0$  и  $\alpha$ ) таким образом, что  $\pi_{\text{exp}} = \pi_{\text{norm.}} = \pi_{\text{pow.law}}$ . Легко тогда проверить, используя ур. (43)–(45), что в пределе больших  $\tau$  имеет место следующее упорядочение:  $\langle T \rangle_{\infty}^{\text{norm}} < \langle T \rangle_{\infty}^{\text{exp}} < \langle T \rangle_{\infty}^{\text{pow.law}}$ . Таким образом, более быстрое затухание плотности распределения

влечет за собой меньшее значение среднего времени ожидания периода экстремально затянувшегося пребывания процесса в заданном состоянии.

Помимо среднего значения  $\langle T \rangle_{\infty}$ , определенно-го ур. (42), выведем также предельную форму распределения вероятностей. Поскольку в рассматриваемом пределе как  $T_1$ , так и  $T_2$  определены преимущественно слагаемым  $t$  в ур. (29) и (30), то для времени ожидания  $T$  мы можем приближенно записать

$$T \approx t_2 + \tau I(t'_1 \geq \tau) + (t'_1 + T') I(t'_1 < \tau), \quad (48)$$

и, следовательно,

$$e^{isT} \approx e^{is(t_2 + \tau)} I(t'_1 \geq \tau) + e^{is(t_2 + t'_1 + T')} I(t'_1 < \tau). \quad (49)$$

Статистическая независимость переменных  $t_2$ ,  $t'_1$  и  $T'$  в ур. (49) позволяет прийти к следующему замкнутому результату для производящей моменты функции

$$\langle e^{isT} \rangle \approx \frac{e^{is\tau} \langle e^{ist_2} \rangle \langle I(t_1 \geq \tau) \rangle}{1 - \langle e^{ist_2} \rangle \langle e^{ist_1} I(t_1 < \tau) \rangle}. \quad (50)$$

“Тело” характеристической функции соответствует области  $s < \langle T \rangle^{-1}$ . Благодаря разделению масштабов времени  $\langle T \rangle_{\infty} \gg \tau$ ,  $\langle t_1 \rangle$ ,  $\langle t_2 \rangle$ , которое следует из ур. (42) (см. также ур. (43), (44) и (45)), мы можем упростить ур. (51), используя малость безразмерных параметров  $s\tau$ ,  $st_1$ ,  $st_2 \ll 1$  для типичных значений случайных переменных  $t_1$  и  $t_2$  и релевантных значений переменной Лапласа  $s$ . Оставляя члены нулевого порядка в числителе и первого порядка малости в знаменателе, получаем

$$\langle e^{isT} \rangle \approx \frac{\langle I(t_1 \geq \tau) \rangle}{\langle I(t_1 \geq \tau) \rangle - i(\langle t_1 \rangle + \langle t_2 \rangle)s}. \quad (51)$$

Наконец, выполняя обратное преобразование Фурье от  $\langle e^{isT} \rangle$  в рамках аппроксимации (51) и учитывая ур. (37), мы находим экспоненциальное распределение  $P(T) \approx \frac{1}{\langle T \rangle_{\infty}} \exp(-\frac{T}{\langle T \rangle_{\infty}})$  времени ожидания  $T$  в пределе больших значений  $\tau$ , с соответствующим средним значением  $\langle T \rangle_{\infty}$ , которое дается ур. (42).

**Закключение.** В заключении, мы хотели бы обсудить некоторые потенциальные применения результатов, полученных в ходе многолетних теоретических исследований статистики времени ожидания определенных паттернов в стохастической динамике.

Прежде всего, марковская цепь с двумя состояниями представляет собой полезный инструмент теоретического моделирования паттернов выпадения осадков [16–24]. Если, скажем, состояние 1 соответствует сухому дню в некоторой местности (отсутствие осадков), а состояние 2 – дождливому дню

(ненулевое количество выпавших осадков), то событие прохождения в контексте принятой нами здесь модели будет представлять собой засухливый период определенной длительности. Таким образом, представленный анализ вместе с предшествующими исследованиями аналогичной тематики (в особенности ссылкой [9], рассматривающей случай произвольного распределения начального состояния процесса) может дать предсказания касательно статистики времени ожидания периодов персистентных метеорологических условий, включая сценарий экстремально долгих засух, что представляет интерес для оценки рисков в агропромышленном планировании.

Во-вторых, мы ожидаем, что рассмотренный здесь прыжковый процесс с двумя состояниями в непрерывном времени может воспроизводить свойства статистически-стационарного турбулентного сигнала в рамках телеграфного приближения [25–28], представляющего собой разновидность огрубленного описания данных. При таком описании, из сигнала исключают информацию об его амплитуде, оставляя только информацию о знаке флуктуирующей переменной. Проще говоря, мы приписываем значение 1 телеграфной динамической переменной, если интересующая нас компонента поля скорости турбулентной жидкости в рассматриваемой точке положительна в заданный момент времени, и значение 0 – в противоположном случае. Пересечение сигналом нуля тогда соответствует переходу между парой огрубленных состояний турбулентной скорости. Аналогично, можно ввести телеграфное приближение для флуктуаций температуры в турбулентной конвекции и для флуктуаций концентрации пассивного скаляра, переносимого турбулентным течением [29]. Если известны плотности распределения интервалов времени между последовательными пересечениями нуля турбулентным сигналом, то результаты представленного анализа могут быть использованы для извлечения теоретических предсказаний касательно статистики времени ожидания периодов персистентности сигнала.

Следует иметь в виду, однако, что в обоих из вышеуказанных приложений для достижения более точного количественного соответствия между измеряемой статистикой времени ожидания и ее теоретическим описанием могут потребоваться более сложные статистические модели. В частности, марковские модели с не зависящими от времени вероятностями переходов не подходят для моделирования метеорологических наблюдений в местностях с выраженной сезонной зависимостью интенсивности осад-

ков (см. работу [30] и приведенные в ней ссылки). Что же касается телеграфного приближения турбулентных сигналов, то требуется предварительный анализ экспериментальных данных, чтобы проверить, является ли достаточно точным, существенное для использованной здесь модели предположение о статистической независимости интервалов времени между моментами смены знака компонентами флуктуационной составляющей поля скорости.

**Финансирование работы.** Работа была поддержана Министерством науки и высшего образования Российской Федерации (проект FFWR-2024-0017.)

**Конфликт интересов.** Автор данной работы заявляет, что у него нет конфликта интересов.

1. S. Redner, *A guide to first-passage processes*, Cambridge university press, Cambridge (2001).
2. R. G. Gallager, *Stochastic processes: theory for applications*, Cambridge University Press, Cambridge (2013), p. 17.
3. E. Schrödinger, *Physikalische Zeitschrift* **16**, 289 (1915).
4. S. Chandrasekhar, *Rev. Mod. Phys.* **15**(1), 1 (1943).
5. D. Ai, *Martingales and the ABRACADABRA problem*, (2011); <http://www.math.uchicago.edu/~may/VIGRE/VIGRE2011/REUPapers/Ai.pdf>.
6. S. Redner, *Physica D: Nonlinear Phenomena*, **442**, 133551 (2022).
7. W. Feller, *Stochastic processes: theory for applications*, John Wiley & Sons, N.Y. (1991), v. 2 (v. 81), VI. 7, p. 189.
8. J. C. Tanner, *Biometrika*, **40**(1/2), 58 (1953).
9. J. Glaz, M. Kulldorff, V. Pozdnyakov, and J. M. Steele, *Journal of Applied Probability* **43**(1), 127 (2006).
10. V. T. Stefanov, *Journal of Applied Probability* **40**(4), 881 (2003).
11. V. T. Stefanov, S. Robin, and S. Schbath, *Discrete Appl. Math.* **155**(6–7), 868 (2007).
12. T. Haynes and E. A. Davis, *Technometrics* **12**(3), 697 (1970).
13. W. Feller, *Stochastic processes: theory for applications*, John Wiley & Sons, N.Y. (1991), v. 2 (v. 81), XI. 9, p. 380.
14. P. J. Aston and P. K. Marriott, *Phys. Rev. E* **57**(1), 1181 (1998).
15. W. Feller, *Stochastic processes: theory for applications*, John Wiley & Sons, N.Y. (1991), v. 2 (v. 81), VI. 7, p. 187.
16. K. R. Gabriel and J. Neumann, *Q. J. R. Meteorol. Soc.* **88**(375), 90 (1962).
17. J. R. Green, *Journal of the Royal Statistical Society Series B: Statistical Methodology* **26**(2), 345 (1964).
18. R. W. Katz, *Journal of Applied Meteorology* **671**, 1962 (1977).



19. X. Lana and A. Burgueño, International Journal of Climatology: A Journal of the Royal Meteorological Society **18**(7), 793 (1998).
20. D.S. Wilks and R.L. Wilby, Prog. Phys. Geogr. **23**(3), 329 (1999).
21. W.O. Ochola and P. Kerkides, Irrigation and drainage: the journal of the International Commission on Irrigation and Drainage **52**(4), 327 (2003).
22. W.W. Ng and U.S. Panu, J. Hydrol. Journal of hydrology **380**(1–2), 222 (2010).
23. S.N. Majumdar, P. von Bomhard, and J. Krug, Phys. Rev. Lett. **122**(15), 158702 (2019).
24. R.D. Stern and R. Coe, Journal of the Royal Statistical Society Series A: Statistics in Society **147**(1), 1 (1984).
25. K.R. Sreenivasan and A. Bershadskii, J. Stat. Phys. **125**(5), 1141 (2006).
26. P. Kailasnath and K. R. Sreenivasan, Physics of Fluids A: Fluid Dynamics **5**(11), 2879 (1993).
27. D. Cava, G.G. Katul, A. Molini, and C. Elefante, Journal of Geophysical Research: Atmospheres **117**(D1), 1 (2012).
28. M. Heisel, Phys. Rev. Fluids **7**(1), 014604 (2022).
29. A. Bershadskii, J.J. Niemela, A. Praskovsky, and K.R. Sreenivasan, Phys. Rev. E **69**(5), 056314 (2004).
30. P.V. Bomhard, *Development of a statistical daily precipitation model and its application to precipitation records*, Doctoral dissertation, Universität zu Köln, Köln, Köln (2017).