

Блуждания по цепям

Александр Гиль и Антон Петрунин

Аннотация

Рассматривается пара примеров цепей Маркова. Заметка основана на лекции, прочитанной первым автором в школьном кружке в Сиэтле, www.assagames.com/nas/.

1 Вероятность и математические ожидания

Подкинув игральный кубик, мы можем выбросить с равными шансами 1, 2, 3, 4, 5 или 6 очков. *Исход* (результат) такого *испытания* (число очков, выпавшее на верхней грани кубика) является простым примером *случайной величины*. Естественно предположить, что исход одного испытания (подкидывания кубика) не зависит от исходов других таких же испытаний.

Давайте записывать значения нашей случайной величины, много раз подкидывая кубик, и вычислять долю испытаний, давших конкретный исход, например, 5 очков. При неограниченном увеличении числа испытаний эта доля стремится к числу, которое называется *вероятностью* этого исхода. Поскольку шансы каждого из шести исходов одинаковы, вероятность каждого исхода равна $\frac{1}{6}$ — сумма долей шести возможных исходов всегда равна 1, и в пределе все эти шесть долей уравниваются.

Давайте теперь не только записывать результат каждого испытания, но и считать среднее арифметическое всех записанных на данный момент результатов. Если эта последовательность стремится к определённому числу, это число называется *математическим ожиданием* или *средним значением* случайной величины (или просто *средним*).

В нашем примере искомое среднее значение числа очков можно посчитать по формуле

$$\frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 3\frac{1}{2}.$$

Авторы признательны Арсению Акопяну за помощь при составлении этой заметки.

Действительно, в пределе на долю каждого исхода 1, 2, 3, 4, 5 и 6 приходится $\frac{1}{6}$ количества всех исходов, и, значит, их среднее арифметическое должно стремиться к левой стороне равенства.

Иначе говоря, для вычисления среднего значения мы должны вычислить *взвешенную сумму* значений нашей величины для каждого исхода (1, 2, 3, 4, 5 и 6), взяв вероятность каждого исхода ($\frac{1}{6}$) в качестве его веса.

Хотя изложенные выше утверждения и кажутся очевидными, для того, чтобы придать им точный математический смысл, требуется развить теорию, которая останется за рамками нашей заметки. Мы обсудим только способы нахождения вероятностей и средних значений в чуть более сложных ситуациях.

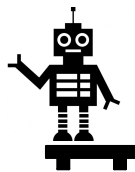
Вообще говоря, может оказаться, что среднее значение (математическое ожидание), вычисляемое как предел для неограниченно продлеваемой серии испытаний, не существует для данной схемы испытаний. Такое может случиться, поскольку у бесконечной последовательности чисел предельного значения может и не существовать.

Популярным примером ситуации без предельного значения является так называемый «Санкт-Петербургский парадокс». В наших задачах такого происходить не будет, но мы не будем это строго доказывать.

2 Робот Чебуратор на коротком столе

Электромеханическая игрушка «Робот Чебуратор», будучи включенной, через равные интервалы времени делает шаги одинаковой длины налево или направо, случайно выбирая между этими двумя направлениями с одинаковой вероятностью $\frac{1}{2}$.

Предположим, Чебуратор стоит на левом краю короткого стола: шагая влево, он чебурахнется со стола. Однако, он может шагнуть вправо и остаться на столе (на правом краю) — но чебурахнется после второго шага направо. Нас интересуют следующие две задачи.



- Какова вероятность того, что робот чебурахнется с левого края стола, и какова — что с правого?
- Сколько в среднем шагов сделает робот, чебурахнувшись на последнем шаге?

При этом мы полагаем, что существование требуемых чисел в обоих вопросах обосновано (это действительно верно, хотя требует доказательства); нам остаётся лишь найти значения этих чисел.

Решение первой задачи. Пусть p_1 и p_2 обозначают вероятности чебураханья с правого края стала из первого (на левом краю стола) и второго (на правом краю стола) положения Чебуратора. Заметим, что

$$p_1 = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot p_2.$$

Действительно, на первом шаге с вероятностью $\frac{1}{2}$ робот чебурахнется налево, и, значит, шансов чебурахнуться направо у него уже не будет; отсюда слагаемое $0 = \frac{1}{2} \cdot 0$. С той же вероятностью $\frac{1}{2}$ он перейдёт на второе место (на правом краю стола), где вероятность чебурахнуться направо мы обозначили p_2 ; отсюда слагаемое $\frac{1}{2} \cdot p_2$.

То есть, зная две вероятности чебураханья направо после двух вариантов первого шага (0 и p_2) мы посчитали их взвешенную сумму, взяв в качестве весов вероятности соответствующих вариантов первого шага ($\frac{1}{2}$ и $\frac{1}{2}$).

Аналогично, рассмотрев два возможных исхода шага робота из второй позиции (на правом краю стола), получаем уравнение

$$p_2 = \frac{1}{2} \cdot p_1 + \frac{1}{2} \cdot 1.$$

Систему из этих двух уравнений с двумя неизвестными (p_1 и p_2) легко решить алгебраически, но мы сделаем это по-другому (этот другой способ пригодится нам позже при решении более сложной задачи).

Первое уравнение говорит нам, что p_1 равно среднему арифметическому 0 и p_2 ; второе уравнение говорит нам, что p_2 равно среднему арифметическому p_1 и 1 . То есть, в последовательности из четырёх чисел

$$0, p_1, p_2, 1$$

каждое из чисел, кроме 0 и 1 на краях, является средним арифметическим своих соседей, а значит, эта последовательность является арифметической последовательностью. Интервал от 0 до 1 разбит на три равных промежутка между членами последовательности; значит, $p_1 = \frac{1}{3}$ и $p_2 = \frac{2}{3}$.

Заметим, что если q_1 и q_2 — вероятности чебураханья налево (при первом и втором начальных положениях), то, проводя аналогичные вычисления, мы получим $q_1 = \frac{2}{3}$ и $q_2 = \frac{1}{3}$. Этот же результат можно получить, посмотрев на задачу через зеркало, меняющее местами право и лево.

Таким образом, из первой (левой) позиции робот чебурахается направо и налево с вероятностями $\frac{1}{3}$ и $\frac{2}{3}$. \square

Заметим, что

$$p_1 + q_1 = p_2 + q_2 = 1,$$

а, значит, вероятность того, что робот никогда не чебурахнется (строго чередуя шаги направо и налево), равна нулю. К этому же выводу мы бы пришли, вычислив вероятность того, что Чебуратор остаётся на столе после шага номер n , напрямую: эта вероятность равна $(\frac{1}{2})^n$ и приближается сколь угодно близко к нулю с возрастанием n — то есть, равна нулю в пределе при неограниченном росте n .

Аналогично решается и вторая задача.

Решение второй задачи. Обозначим s_1 и s_2 средние количества шагов, которые сделает робот, чебурахнувшись на последнем шаге, начиная с первой или второй позиции соответственно. Взглянув на робота через зеркало, меняющее местами право и лево, мы убеждаемся, что $s_1 = s_2$ — но мы и так это скоро увидим из вычислений.

Заметим, что

$$s_1 = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot (1 + s_2).$$

Действительно, на первом шаге с первой позиции с вероятностью $\frac{1}{2}$ робот чебурахнется налево, то есть он проделает всего один шаг; отсюда слагаемое $\frac{1}{2} \cdot 1$. С той же вероятностью он шагнёт вправо на вторую позицию, откуда в среднем он пройдёт ещё s_2 шагов. Учтя начальный шаг, получаем второе слагаемое $\frac{1}{2} \cdot (1 + s_2)$.

Аналогично получаем равенство

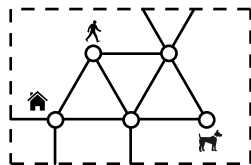
$$s_2 = \frac{1}{2} \cdot (1 + s_1) + \frac{1}{2} \cdot 1.$$

Мы опять получили систему из двух уравнений с двумя неизвестными s_1 и s_2 . Решив эту систему, получаем $s_1 = s_2 = 2$. \square

Попробуйте решить следующую задачу тем же способом.

Задача. На рисунке вы видите кусок, вырезанный из плана города, с отмеченными на нём собакой (\blacktriangle), её забывчивым хозяином (\blacktriangle) и их домом (\blacktriangle).

Хозяин ищет свой дом, выбирая с равной вероятности одну из сходящихся на перекрёстке дорог, и идёт до следующего перекрёстка ровно одну минуту. Дойдя до следующего перекрёстка, он опять выбирает с равной вероятностью одну из дорог (возможно, ту же, по которой только что шёл), и идёт по ней — и так далее, пока не попадёт на тот перекрёсток, на котором его дом.



Собака двигается по тому же принципу, но бежит в полтора раза быстрее хозяина.

Докажите, что из данного начального положения собака в среднем прибежит домой на 20 секунд раньше хозяина.

3 Решения прямым подсчётом

Две задачи про Чебуратора, рассмотренные выше, допускают решения прямым подсчётом вероятностей. Мы опишем их, используя те же обозначения, что и раньше.

Решения. Решая вторую задачу, заметим, что робот чебурахается на n -ом шаге либо влево, либо вправо с вероятностью $(\frac{1}{2})^n$. Значит, среднее значение числа шагов можно найти, просуммировав ряд

$$s_1 = s_2 = \frac{1}{2} \cdot 1 + \left(\frac{1}{2}\right)^2 \cdot 2 + \left(\frac{1}{2}\right)^3 \cdot 3 + \dots$$

Для суммирования такого ряда требуется некоторое умение; однако, ряд действительно сходится к сумме, равной 2 — и неудивительно, ведь это же значение было получено выше.

Далее заметим, что если робот начинает движение с левой позиции, то он может чебурахнуться направо только на чётных шагах — а если он начинает с правой позиции, то только на нечётных. То есть,

$$\begin{aligned} p_1 &= \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^6 + \dots \\ p_2 &= \left(\frac{1}{2}\right)^1 + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^5 + \dots \end{aligned}$$

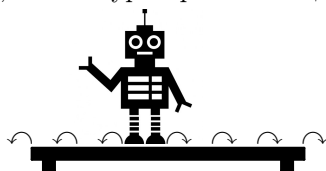
Применив формулу для суммы геометрической прогрессии, получаем те же результаты: $p_1 = \frac{1}{3}$ и $p_2 = \frac{2}{3}$. \square

Как видите, этот способ оказался сложнее. Кроме того, он не допускает лёгкого обобщения на случай длинного стола (на котором Чебуратор может сделать более одного шага). Как мы увидим ниже, в этом случае наше первое решение остаётся практически без изменений.

То, что среднее значение числа шагов с чебураханьем на последнем шаге равно двум, можно также увидеть, поставив следующий мысленный эксперимент. Представьте себе, что при каждом чебураханье мы ставим робота обратно на стол. Тогда вероятность того, что робот чебурахнется на шаге n равна $\frac{1}{2}$, то есть робот будет чебурахаться в среднем на каждом втором шаге. Значит, среднее количество шагов между чебураханьями равно 2.

4 Робот Чебуратор на длинном столе

Представим теперь, что Чебуратор стоит на длинном столе.



Предположим, что робот чебурахнется с левого края стола, сделав i шагов налево, и с правого края стола, сделав j шагов направо.

Нас интересуют те же задачи:

- Какова вероятность того, что робот чебурахнется с левого края стола, и какова — что с правого?
- Сколько шагов в среднем сделает робот, чебурахнувшись на последнем шаге?

Решение первой задачи. Заметим, что сумма $m = i + j$ одна и та же для всех позиций и число позиций равно $m - 1$. Пронумеруем возможные положения Чебуратора числами от 1 до $m - 1$, по числу шагов налево, которое Чебуратору нужно сделать, чтобы чебурахнуться.

Нам будет удобно добавить ещё два положения под номерами 0 и m : первое соответствует тому, что робот чебурахнулся налево, а второе соответствует тому, что робот чебурахнулся направо (из этих позиций выхода нет).

Обозначим p_i вероятность того, что Чебуратор чебурахнется направо, если начинается с позиции под номером i . Естественно, мы имеем

$$p_0 = 0 \text{ и } p_m = 1.$$

Применим тот же метод, что и раньше. Стоя на позиции номер i с вероятностью $\frac{1}{2}$ Чебуратор перейдёт на позицию номер $i + 1$. В этом случае вероятность того, что он чебурахнется вправо, будет p_{i+1} . С той же вероятностью $\frac{1}{2}$ он перейдёт на позицию номер $i - 1$. В последнем случае вероятность того, что он чебурахнется вправо, будет p_{i-1} . То есть,

$$p_i = \frac{1}{2} \cdot p_{i-1} + \frac{1}{2} \cdot p_{i+1}.$$

Иначе говоря, в последовательности из $m + 1$ числа

$$0 = p_0, p_1, p_2, \dots, p_m = 1$$

каждое число, кроме двух крайних, является средним арифметическим соседей; мы получили арифметическую последовательность, делящую интервал от 0 до 1 на m промежутков. Отсюда получаем $p_i = \frac{i}{i+j}$. \square

Решение второй задачи. Давайте использовать ту же нумерацию позиций, как и в решении первой задачи.

Обозначим s_i среднее количество шагов робота с чебураханьем на последнем шаге, если он начинает с позиции под номером i . Естественнo предположить, что $s_0 = s_m = 0$ — ведь попадание на позиции 0 и m означает, что робот уже чебурахнулся.

Попробуем, как и раньше, посчитать значение s_i новым способом. После одного шага с i -ой позиции Чебуратор окажется на позиции $i - 1$ или $i + 1$ с равными вероятностями $\frac{1}{2}$. После этого ему останется в среднем пройти s_{i-1} и s_{i+1} шагов соответственно. Не забыв учесть уже пройденный шаг, получаем

$$s_i = \frac{1}{2} \cdot (s_{i-1} + 1) + \frac{1}{2} \cdot (s_{i+1} + 1),$$

или

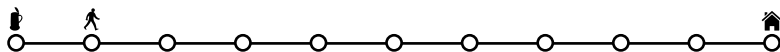
$$s_{i+1} = 2 \cdot s_i - s_{i-1} - 2.$$

Применяя эти равенства, последовательно двигаясь по позициям слева направо, получаем $s_i = i \cdot (s_1 - i + 1)$ для любого i . Поскольку $s_m = 0$, получаем $s_1 = m - 1$ и, значит, $s_i = i \cdot j$. \square

Заметим, что если Чебуратор начинает на середине стола, то есть $i = j$, то он чебурахнется в среднем за i^2 шагов.

Для закрепления материала мы советуем решить следующую задачу.

Задача. Пьяница вышел из бара, расположенного в 10 кварталах от своего дома на той же улице и идёт домой. Чтобы сделать дорогу веселее, он разнообразит её следующим образом. Дойдя до перекрёстка, он подкидывает монетку — и, если она выпадает орлом, он продолжает путь в том же направлении; если же она выпадает решкой, он разворачивается и идёт в противоположном направлении. Если ему случается вернуться к бару, он всегда разворачивается в сторону дома; если же он дошёл до своего дома, то завершает свой путь.



Докажите, что в среднем за такую прогулку пьяница проходит 100 кварталов, а среднее число его возвращений к бару равно 9.

5 Бактерии в пробирке

Задача. В пробирке живут 10 бактерий: 3 зелёных и 7 жёлтых. Каждую минуту происходит следующее: равновероятно выбирается пара из 10 бактерий, после чего первая бактерия пары погибает, а вторая делится на две своих точных копии. Таким образом, в конце минуты в пробирке снова остаётся ровно 10 бактерий.

Какова вероятность того, что через некоторое время все 10 будут зелёными?

Решение. Мы сведём задачу к уже решённой.

Предположим, что через несколько минут число зелёных бактерий стало i (где $0 < i < 10$). Тогда спустя минуту их число может стать $i - 1$, i или $i + 1$. При этом вероятности первого и последнего исходов равны. Действительно, исход с $i - 1$ зелёной бактерией происходит, когда первая бактерия случайно выбранной пары оказалась зелёной, а вторая — жёлтой. Исход с $i + 1$ зелёной бактерией происходит, наоборот, когда первая бактерия случайно выбранной пары — жёлтая, вторая — зелёная. Эти две ситуации симметричны и случаются с одинаковыми вероятностями. Исход i можно рассматривать «пропуском хода».

Таким образом, наша задача становится похожей на задачу про робота Чебуратора — только теперь Чебуратор ходит направо и налево с равными положительными вероятностями, меньшими, чем $\frac{1}{2}$, и с оставшейся вероятностью пропускает ход. Однако заметим, что пропускание хода можно вовсе не учитывать. То есть, ответ в задаче можно получить, подставив $i = 3$ и $j = 7$ в задаче про Чебуратора на длинном столе. Два состояния одноцветности бактерий — состояния под номером 0 и под номером 10 — будут соответствовать состояниям робота, чебурахнувшегося с левого и правого края стола; состояния под номерами от 1 до 9 будут соответствовать всем 9-ти позициям робота на столе. Таким образом, с вероятностью $\frac{3}{10}$ все бактерии станут зелёными, и с вероятностью $\frac{7}{10}$ все станут жёлтыми. \square

Эту же задачу (а значит, и первую задачу про Чебуратора) можно решить без вычислений.

Сначала докажем, что с вероятностью 1 через некоторое время все бактерии в пробирке будут потомками одной.

Для начала заметим, что с положительной вероятностью это может случиться за первые 10 минут. Вероятность этого события (обозначим это число p) мала, но вероятность того, что за достаточно длинный промежуток времени найдутся такие 10 минут, в пределе равна 1.

Действительно, вероятность противоположного исхода (того, что после 10-ти минут в пробирке будут потомки более чем одной из первоначально находившихся там бактерий) равна $1 - p$, и это число меньше 1. Значит, вероятность того, что никакой из n последовательных 10-минутных интервалов не завершился тем, что в пробирке остались только потомки одной из бактерий, бывших в пробирке в начале этого интервала, равна $(1 - p)^n$. При увеличивающемся n это число приближается сколь угодно близко к нулю, значит, предельная вероятность того, что рано или поздно такой 10-минутный интервал случится — при неограниченном увеличении времени ожидания такого интервала — равна 1.

При этом в самом начале каждая из десяти бактерий имеет равные шансы на то, чтобы стать прародителем всех выживших. Поэтому в трёх из десяти случаев все станут зелёными, и в семи из десяти все станут жёлтыми.

6 Броуновское движение

Рассмотрим случай, когда Чебуратор стоит на середине длинного стола с чётным $m = 2 \cdot n$ — чтобы чебурахнуться, ему необходимо сделать или n шагов вправо, или столько же влево. Как мы выяснили, среднее количество шагов, которое он делает, чебурахаясь на последнем, равно n^2 . Это наблюдение можно использовать в обратном направлении: повторив это испытание много раз и оценив среднее значение числа шагов до чебураханья мы, зная размер стола, сможем оценить длину шага Чебуратора. Последнее может оказаться полезным, если шаги Чебуратора настолько мелки, что не поддаются прямому измерению.

Поведение Чебуратора на столе мало чем отличается от хаотического движения малых твёрдых частиц, плавающих в жидкости, под действием ударов молекул жидкости. Это хаотическое движение открыл Роберт Браун в начале XIX-го века, а в начале XX-го века удалось оценить среднее значение числа молекул в единице объёма по параметрам броуновского движения. Этот эксперимент послужил убедительным аргументом в утверждении атомной теории строения вещества; при этом для его проведения оказалось достаточным наблюдений с помощью обычного микроскопа.

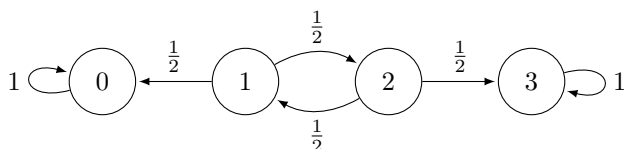
Конечно, эта физическая задача сложнее наших задач о Чебураторе. Тем не менее, идея решения у этих задач одна и та же.

7 Да, а о чём мы говорили?

Мы рассмотрели частный случай *цепей Маркова*, названных так в честь Андрея Андреевича Маркова-старшего (его сына тоже звали Андрей Андреевичем и он тоже занимался математикой).

А именно, мы рассматривали *конечные цепи Маркова с поглощающими состояниями*. Каждое *поглощающее состояние* соответствует чебураханью нашего Чебуратора — с левого или с правого края стола.

Цепь Маркова можно задать ориентированным графом, на каждом ребре которого написана вероятность перехода из одного состояния в другое. Например, задача о Чебураторе на коротком столе может быть описана таким графом:



Состояниям 0, 1, 2 и 3 соответствуют состояния Чебуратора «чебурахнулся налево», «стоит на левом краю стола», «стоит на правом краю стола» и «чебурахнулся направо». Заметим, что сумма вероятностей на стрелках, исходящих из какого-то состояния, всегда равна единице.

Цепи Маркова оказались удобной и очень востребованной моделью. Например, ранжирование веб-страниц, используемое поисковыми системами, основано на цепях Маркова. С помощью цепей Маркова создаются алгоритмы для автоматического написания музыки. Есть важные приложения в химии, биологии, лингвистике, экономике — а одно из применений этой теории в физике мы вкратце описали в предыдущем разделе.