**ГБОУ «Лицей №1553 им. В. И. Вернадского»**

**Курсовая работа**

Изучение времени появления данной последовательности символов в случайно генерируемом ряду.

**Автор:** Бахтин Александр Андреевич

**Науч. рук.:** к.ф.-м.н. Ногин Дмитрий Юрьевич

**МОСКВА 2016**

Целью нашей работы было решение такой задачи:

*Монету подкидывают каждую секунду. Выигрыш наступает при выпадении определенной последовательности орлов и решек. Когда, в зависимости от последовательности и вероятностей выпадения орла и решки (считая их сумму равной 1) в среднем наступает выигрыш?*

На математическом языке эту задачу можно сформулировать так:

*При создании (генерации) некоего ряда чисел в него на каждой секунде (на каждом шаге) с известными вероятностями P и Q (P+Q=1) добавляются, соответственно, 1 или 0. После начала генерации она может длиться бесконечно. Имеется заранее заданная последовательность единиц и нолей. Когда она в среднем появится в генерируемом ряду первый раз?*

Для решения этой задачи нужно ответить на два вопроса:

1. Как составить уравнение, в котором искомое (время/ шаг появления последовательности) будет выражено через данные вероятности и, возможно, другие необходимые данные?
2. Как, в свою очередь, рассчитать эти данные?
3. Составление уравнения.

Пусть искомое время равно х, рассмотрим пошагово процесс генерации последовательности: пусть $P\_{n}$ и $Q\_{n}$ это, соответственно, вероятность выпадения верного и неверного символа на шаге n. В «идеальной» ситуации каждый раз выпадает верный символ, тогда искомое х равно длине N данной последовательности, но пусть на первом шаге с вероятностью $Q\_{1}$ выпадает неверный символ, тогда искомое время увеличилось, ведь мы потратили на 1 секунду больше. И такое может произойти на каждом шаге, следовательно, в общем виде это уравнение выглядит так:

*х* = $Q\_{1}\left(x+n\_{1}\right)+P\_{1}\left(Q\_{2}\left(x+n\_{2}\right)+P\_{2}\left(…Q\_{n}\left(x+n\_{n}\right)+P\_{n}N\right)…\right)$

Т.е. с вероятностью $Q\_{n}$ мы увеличили время, нужное для выпадения этой последовательности на некое n (назовем это *потерей*), а с вероятностью $P\_{n}$ продвинулись дальше. Собственно, вычислению потерь и посвящен раздел II данной работы.

Теперь попробуем выразить из данного уравнения х. Для этого сначала раскроем скобки:

$$x=Q\_{1}x+Q\_{1}n\_{1}+P\_{1}Q\_{2}x+P\_{1}Q\_{2}n\_{2}+P\_{1}P\_{2}Q\_{3}x+P\_{1}P\_{2}Q\_{3}n\_{3}+P\_{1}P\_{2}P\_{3}...Q\_{N}x+P\_{1}P\_{2}P\_{3}...Q\_{N}n\_{N}+P\_{1}P\_{2}P\_{3}...P\_{N}N$$

Перенесем все слагаемые с х в левую часть уравнения и вынесем х за скобку, получим:

$x×\left(1-Q\_{1}-P\_{1}Q\_{2}-P\_{1}P\_{2}Q\_{3}-...-P\_{1}P\_{2}P\_{3}...P\_{N-1}Q\_{N}\right)$=$Q\_{1}n\_{1}+P\_{1}Q\_{2}n\_{2}+P\_{1}P\_{2}Q\_{3}n\_{3}+P\_{1}P\_{2}P\_{3}P\_{N-1}Q\_{N}n\_{4}+P\_{1}P\_{2}P\_{3}...P\_{N}N$.

Рассмотрим получившуюся скобку отдельно:

$$1-Q\_{1}-P\_{1}Q\_{2}-P\_{1}P\_{2}Q\_{3}-...-P\_{1}P\_{2}P\_{3}...P\_{N-1}Q\_{N}$$

По условию задачи $Q\_{n}+P\_{n}=1$.

Получаем $P\_{1}-P\_{1}Q\_{2}-P\_{1}P\_{2}Q\_{3}-...-P\_{1}P\_{2}P\_{3}P\_{N-1}Q\_{N}$.

Вынесем $P\_{1}$ за скобку: $P\_{1}\left(1-Q\_{2}-P\_{2}Q\_{3}-...-P\_{2}P\_{3}...P\_{N-1}Q\_{N}\right)$.

Это равно $P\_{1}\left(P\_{2}-P\_{2}Q\_{3}-...-P\_{2}P\_{3}...P\_{N-1}Q\_{N}\right)$. Теперь выносим $P\_{2}$ за скобку и так далее. В итоге получится $P\_{1}P\_{2}P\_{3}...P\_{N}$. Это можно записать в общем виде:

$1-Q\_{1}-P\_{1}Q\_{2}-...-P\_{1}P\_{2}…Q\_{n}$=$P\_{1}P\_{2}…P\_{n}$.

Вернемся к нашему уравнению:

$x×\left(P\_{1}P\_{2}P\_{3}...P\_{N}\right)=Q\_{1}n\_{1}+P\_{1}Q\_{2}n\_{2}+P\_{1}P\_{2}Q\_{3}n\_{3}+P\_{1}P\_{2}P\_{3}...Q\_{N}n\_{N}+P\_{1}P\_{2}P\_{3}...P\_{N}N$.

Разделим обе части на $P\_{1}P\_{2}P\_{3}P\_{4}:$

$x=\frac{Q\_{1}n\_{1}}{P\_{1}P\_{2}P\_{3}…P\_{N}}+\frac{P\_{1}Q\_{2}n\_{2}}{P\_{1}P\_{2}P\_{3}...P\_{N}}+\frac{P\_{1}P\_{2}Q\_{3}n\_{3}}{P\_{1}P\_{2}P\_{3}...P\_{N}}+…+\frac{P\_{1}P\_{2}P\_{3}…Q\_{N}n\_{N}}{P\_{1}P\_{2}P\_{3}...P\_{N}}+\frac{P\_{1}P\_{2}P\_{3}...P\_{N}N}{P\_{1}P\_{2}P\_{3}...P\_{N}}$ .

Сократим слагаемые:

$x=\frac{Q\_{1}n\_{1}}{P\_{1}P\_{2}P\_{3}...P\_{N}}+\frac{Q\_{2}n\_{2}}{P\_{2}P\_{3}...P\_{N}}+\frac{Q\_{3}n\_{3}}{P\_{3}...P\_{N}}+...+\frac{Q\_{N}n\_{N}}{P\_{N}}+N$ .

Снова воспользуемся тем, что $Q\_{n}+P\_{n}=1$ и раскроем скобки:

х=$ \frac{n\_{1}-P\_{1}n\_{1}}{P\_{1}P\_{2}P\_{3}...P\_{N}}+\frac{n\_{2}-P\_{2}n\_{2}}{P\_{2}P\_{3}...P\_{N}}+\frac{n\_{3}-P\_{3}n\_{3}}{P\_{3}...P\_{N}}+...+\frac{n\_{N}-P\_{N}n\_{4}}{P\_{N}}+N.$

Снова сократим, где возможно:

х=$ \frac{n\_{1}}{P\_{1}P\_{2}P\_{3}...P\_{N}}+\frac{n\_{2}-n\_{1}}{P\_{2}P\_{3}...P\_{N}}+\frac{n\_{3}-n\_{2}}{P\_{3}...P\_{N}}+...+\frac{n\_{N}-n\_{N-1}}{P\_{N}}-n\_{N}+N.$

Это и есть искомая формула. Она удобна тем, что там используются только вероятности выпадения верных символов.

1. Вычисление потерь.

Из того, что было сказано выше, можно сформулировать, что потеря – это количество шагов, на которое, вследствие выпадения неверного символа, мы вернулись назад. Например, если данная нам последовательность –101011, и первые пять символов выпали верно, а шестой – нет, то получился ряд 101010, и потеря здесь составляет два – именно на столько символов мы «вернулись назад» в получении искомой последовательности (теперь, если убрать первые два символа из ряда, мы вернемся к ряду 1010). Таким образом, для вычисления потери для каждого из символов нужно сравнивать несколько последних символов из ряда, полученного при выпадении вместо нужного символа неверного, с началом данной последовательности (в данном примере из шести символов ряда четыре последних совпадают с началом последовательности, поэтому потеря была 6-4=2).

Для удобства вычисления потерь мы разработали специальный метод. Он состоит в том, что данную последовательность нужно записать в нижнюю строку таблицы (которая имеет номер 0), а в каждую следующую – сдвинув её на 1 символ вправо.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 5 |  |  |  |  |  | 1 |
| 4 |  |  |  |  | 1 | 0 |
| 3 |  |  |  | 1 | 0 | 1 |
| 2 |  |  | 1 | 0 | 1 | 0 |
| 1 |  | 1 | 0 | 1 | 0 | 1 |
| 0 | **1** | **0** | **1** | **0** | **1** | **1** |

Теперь нужно, двигаясь снизу вверх, брать строку и сравнивать со строкой номер 0, начиная с левого символа. Как только найдено несовпадение, нужно от символа на выбранной строке двигаться вниз до нулевой строки (также в клетке, где было найдено несоответствие можно поставить \*, это пригодиться нам для исследования в дальнейшем). У того символа, к которому мы попадем, будет потеря, равная номеру строки, от которой мы начинали движение. Это несовпадение двух строк означает, что если до данного символа получившийся ряд совпадает с последовательностью, а этот символ оказался неверным, то этот ряд совпадет с последовательностью, если из его начала убрать столько символов, сколько указывает номер строки. После этого нужно переходить к сравнению следующей строки, поскольку позже этих несовпадающих символов строки уже не будут совпадать как раз из-за них.

 Например, если сравнивать вторую строку с нулевой: первые символы - 1/1, вторые – 0/0, третьи – 1/1, а четвертые – 0/1, поэтому идем от четвертого символа в строке №2 вниз и доходим до шестого символа нулевой строки, и именно у него будет потеря 2.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 5 |  |  |  |  |  | 1 |
| 4 |  |  |  |  | 1 | 0 |
| 3 |  |  |  | 1 | 0 | 1 |
| 2 |  |  | 1 | 0 | 1 | 0 |
| 1 |  | 1 | 0 | 1 | 0 | 1 |
| 0 | **1** | **0** | **1** | **0** | **1** | **1** |
| П. |  |  |  |  |  | **2** |

Если на один и тот же символ мы опускаемся из нескольких строк (в данном примере в шестой символ мы опускаемся из второй и четвертой строки), то нас интересует результат самой ранней строки, т.е. наименьшая потеря.

Однако после сравнения всех строк не всем символам будет назначена потеря. Происходит это из-за того, что данная таблица определяет потери, только если они не максимальные, т.е. прогресс не был потерян полностью. Поэтому во всех символах, где потеря не была назначена таблицей, она равна их порядковому номеру, если считать слева, также расставим звездочки для этих потерь и назовем диагональ, на которой они стоят (выходящая из левого символа нулевой строки вправо-вверх) *верхней диагональю*. Таблица со всеми обозначенными звездочками для 101011:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 5 |  |  |  |  | \* | 1 |
| 4 |  |  |  |  | 1 | 0\* |
| 3 |  |  | \* | 1\* | 0 | 1 |
| 2 |  |  | 1 | 0 | 1 | 0\* |
| 1 | \* | 1\* | 0 | 1 | 0 | 1 |
| 0 | **1** | **0** | **1** | **0** | **1** | **1** |
| П. | **1** | **1** | **3** | **3** | **5** | **2** |

Отсюда следует, что каждое отдельное числовое значение потери может появиться не больше двух раз — на символе, порядковый номер которого равен этой потере (на верхней диагонали) и где-то правее. Теперь имеет смысл попытаться определить, какие бывают комбинации потерь, а какие нет. Один из подходов к ответу на этот вопрос – изучить свойства расположения звездочек и их связь с числами в таблицах и составить алгоритм, который сможет по набору потерь (звездочек), стоящих в определенном порядке, составить соответствующую им последовательность.

Выявленные свойства:

* + - 1. Если звездочка стоит в клетке и **не** находится на верхней диагонали, символ в ней **не** совпадает с символом в нулевой строке под ней, т.к. звездочки расставлены только в местах, где таковые два символа не совпадают.
			2. На каждой из диагоналей (вправо-вверх) стоит один и тот же символ, т.к. при заполнении таблицы на каждой следующей строке символы сдвигались на один вправо.
			3. Если звездочка стоит на высоте k и находится больше, чем на один символ под верхней диагональю, то k+1-й символ совпадает с первым символом последовательности, т.к. такая звездочка означает, что если вместо того символа, над которым она стоит, выпадет неверный, то из получившегося ряда надо убрать первые k символов, и, начиная с k+1-го символа, этот ряд совпадает с началом последовательности.
			4. Если звёздочка стоит на строчку ниже первой диагонали, то символ под ней в нулевой строке не совпадает с началом последовательности, ведь в клетке с этой звездочкой стоит первый символ последовательности.

Таким образом, чтобы составить последовательность по данной расстановке звездочек нужно идти слева направо. Первый символ может быть или 1, или 0 (мы всегда будем для удобства брать 1), для второго символа есть два положения звездочки (см. рисунок ниже) – на первой диагонали (А), и тогда он совпадает с первым; и на первой строчке (Б), тогда в клетке с ней уже стоит первый символ, и мы можем определить второй. Для третьего символ существует уже три положения звездочки: на верхней диагонали (В), тогда он тоже совпадает с первым; на первой и второй строчках (Г,Д), но там уже стоят первый и второй символы, поэтому мы можем определить третий и т.д., поэтому для каждого данного набора звездочек можно составить последовательность.

|  |  |  |  |
| --- | --- | --- | --- |
| 3 |  |  | В |
| 2 |  | А | Г |
| 1 | \* | Б | Д |
| 0 | **1** |  |  |

Но не все комбинации звездочек «верные» - если сначала по звездочкам восстановить последовательность, а потом для нее снова составить набор звездочек (набор потерь), эти два набора могут отличаться.

|  |  |  |  |
| --- | --- | --- | --- |
| 3 |  |  |  |
| 2 |  | \* | **1** |
| 1 | \* | **1** | **1\*** |
| 0 | **1** | **1** | **0** |
| *Для получившейся последовательности 110 снова составим потери* |

|  |  |  |  |
| --- | --- | --- | --- |
| 3 |  |  |  |
| 2 |  | \* | \***1** |
| 1 | \* | **1** | **1** |
| 0 | **1** | **1** | **0** |
| *По набору звездочек 112 восстановим последовательность* |

Это означает, что данный в начале набор звездочек (в нашем примере 122) не соответствует никакой последовательности.

Для того, чтобы можно было определить, есть ли для данного набора соответствующая ему последовательность мы разработали специальный алгоритм, основанный на свойствах III и IV.

Эти свойства говорят о том, что для k-го символа информацию о нем мы можем получить не только от звездочки, находящейся непосредственно над ним (назовем ее первой), но и от звездочки, стоящей в строчке с номером k-1(но больше, чем на 1 символ ниже первой диагонали), если она есть(назовем ее второй). Вторая звездочка говорит о том, что этот символ совпадает с первым, значит, если звездочка над ним будет говорить об обратном (т.е. символ в ней тоже совпадает с первым), то мы придем к противоречию.

Для того, чтобы понять, есть ли вторая звездочка, нужно от данного символа провести прямую вверх до пересечения с первой диагональю, а потом направо. Эта прямая пройдет по строке k, значит искомая звездочка на строку ниже. Если она нашлась, нужно от первой звездочки идти по диагонали влево-вниз до нулевой строки. Во всех клетках, по которым мы прошли, стоит один и тот же символ. Теперь нужно подняться вверх до клетки со звездочкой, в ней стоит символ, отличный от предыдущих. Так нужно повторять до тех пор, пока мы не дойдем до звездочки на верхней диагонали, причем каждый раз, когда мы будем попадать в нулевую строку, символ будет меняться. Когда мы дойдем до звездочки на первой диагонали, символ в нулевой строке под ней будет совпадать с первым, а все символы с нулевых строках, через которые мы прошли, чередуются. Если мы попадали в нулевую строку четное число раз, символ, с которого мы начинали (стоявший в клетке со звездочкой), не совпадает с первым, т.е. противоречия нет. Если мы попадали в нулевую строку нечетное число раз, то символ, с которого мы начинали, совпадает с первым, и мы приходим к противоречию (см. рисунок).

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 8 |  |  |  |  |  |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |  |  |  | **\*** |  |
| 6 |  |  |  |  |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  | \* |  |  |  |  |  |
| 2 |  | \* |  |  |  |  |  | **\***(1) |  |  |  |
| 1 |  |  | \* |  |  |  |  |  |  |  |  |
| 0 |  |  (1) |  (0) |  |  |  (1) |  | (0)/(1) |  |  |  |

Однако этот способ работает не для всех наборов звездочек. В том числе он не работает для набора 122, указанного в примере ранее. Там просто нет такого символа, у которого было бы сразу две звездочки. Тут остается только перебрать возможные варианты:

Для последовательности 11 существует три возможных положения третьей звездочки: А, на верхней диагонали; Б, на второй строчке; В, на первой строчке. Однако третьим символом может быть либо 1, либо 0, поэтому только два положения звездочки верны. В положении А звездочка находится, если третий символ совпадает с символом и в первой, и во второй строке. Т.к. символы в первой и второй строке совпадают, это возможно. В положении Б звездочка находится, если третий символ совпадает с символом в первой строке, но не во второй. Это не возможно, т.к. эти два символа совпадают. Наконец, в положении В звездочка находится, когда третий символ не совпадает с символом в первой строке (т.е. со вторым символом последовательности), такое возможно. Значит, случаи А и В возможны, а как раз случай Б нет.

|  |  |  |  |
| --- | --- | --- | --- |
| 3 |  |  | А |
| 2 |  | \* | 1 |
| 1 | \* | **1** | 1 |
| 0 | **1** | **1** |  |
| *Третий символ совпадает и с первым, и со вторым.* |

|  |  |  |  |
| --- | --- | --- | --- |
| 3 |  |  |  |
| 2 |  | \* | 1(Б) |
| 1 | \* | **1** | 1 |
| 0 | **1** | **1** |  |
| *Третий символ не совпадает с первым, но совпадает со вторым - противоречие* |

|  |  |  |  |
| --- | --- | --- | --- |
| 3 |  |  |  |
| 2 |  | \* | 1 |
| 1 | \* | **1** | 1(В) |
| 0 | **1** | **1** |  |
| *Третий символ не совпадает со вторым.* |

Итак:

* + 1. Выведена формула, по которой можно рассчитать среднее время выпадения данной последовательности.
		2. Составлен алгоритм, который может определить «потери» на каждом шаге, необходимые для подстановки в формулу.
		3. Частично составлен алгоритм, позволяющий определить, соответствует ли набору потерь какая-либо последовательность.