# Применение методов теории информации для генерации случайных последовательностей. Полиномиальный метод Бабкина для экстракции истинно случайной последовательности 0 и 1 из не равновероятной бернуллиевской последовательности фотоотсчетов.

Начнем с того, что такое истинно случайная последовательность – это такая последовательность, все элементы которой равновероятны и независимы друг от друга. Получить такую последовательность – достаточно сложная задача, но выполнимая.

Существует ли, например, в классической физике понятие вероятности появления события? Фундаментально – нет, так как все законы классической физики есть эволюция системы из заданных начальных условий. Если повторять эксперимент с одинаковыми начальными условиями, результат будет получен с вероятностью 1.

Однако настоящая вероятность есть в квантовой механике, ведь если расписать эволюцию системы, при определенных переходах получим неоднозначый ответ на математическом уровне, базируемый на вероятности. То есть при одних и тех же условиях, на фундаментальном уровне будут получаться разные исходы, которые, что немаловажно, абсолютно независимы. А значит, теоретически имеем право достать из квантового мира истинную случайность. Остается понять, как это сделать.

Закроем глаза на техническую сторону и будем считать, что мы умеем определять в какое из двух возможных состояний редуцировался фотон. Вероятность выпадения каждого из состояний экспериментально можно сделать близким к ½, тогда можно было бы записывать 0 при попадании первый детектор и 1 при попадании на второй. В этом случае события почти равновероятны, но очень сильно ограничена скорость, так как после детектирования требуется время на релаксацию. Поэтому придумали способ лучше.

Хотим запускать фотоны друг за другом быстро. Тогда определять какой фотон куда пришел сложно, а также нужно много детекторов, чтобы определение шло параллельно. Привяжем систему к таймеру и с определенным временным интервалом будем записывать показания детекторов. Таким образом получается последовательность единиц (попадание хотя бы одной частицы в детекторы за текущий интервал) и нулей (не попадание ни одной). Теперь мы не можем гарантировать равновероятное выпадение 0 и 1, но можно утверждать, что события выпадения 1 и 0 независимы, а также имеют бернуллиевское распределение. Дальше стоит математическая задача составления равновероятной последовательности из независимой бернуллиевской.

Оказывается, что если разбить полученную нами последовательность на большие блоки длины n, то максимальное колличество случайных бит будет задаваться бинарной энтропией Шенона с коэффициентом $n$. То есть, для заданных $p$ и $q$ вероятностей выпадения 0 и 1 бернуллиевского распределения существует фундаментальная мера случайности, которую можно получить. Также расчитывается параметр n, при котором колличество возможных для получения случаных бит выходит на максимум.

Стоит отметить, что выпадение 1 в эксперименте получается гораздо реже, чем 0. Обозначим вероятности $p$ и $q$ соответсвенно, то есть $q > p$. Теперь разделим

последовательность на блоки длиной $n$. Затем образуем классы, каждый из которых содержит только такие блоки, у которых между собой равно колличество нулей и единиц. Таким образом, в первом классе будет только один блок со всеми 0, второй класс должен содержать одну единицу в каждом блоке, третий две и т.д. Пример для $n = 3$:

$$\begin{matrix} 0 & 0 & 0 \end{matrix} \qquad \begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix} \qquad \begin{matrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{matrix} \qquad \begin{matrix} 1 & 1 & 1 \end{matrix}$$

Колличество блоков в классе можно посчитать как $C_n^k$, где $k$ – колличество единиц в блоке этого класса. А теперь ход конем, так как выпадение 0 и 1 – события независимые, вероятность получить в последовательности конкретный блок $P = p^k q^{n-k}$ и она одинакова для всех последовательностей одного класса.

Имеем n классов, в каждом из которых $C_n^k$ равновероятных блоков. Тогда чтобы получить равновероятную последовательность, необходимо как-нибудь занумеровать каждый блок. Пусть в классе четное колличество блоков, тогда всегда можно разбить этот класс на подклассы таким образом, чтобы каждый подкласс состоял из $2^{n_i}$ блоков, где $i$ – индекс подкласса. Делается это для того, чтобы блоки каждого подкласса можно было занумеровать двоичной строкой, а состоять она будет из $n_i$ бит. Если в классе нечетное колличество блоков, то один просто выбрасывается, и работа продолжается с четным. Для примера с $n = 3$, первый и последний классы идут в утиль, во втором и третем остается по одному подклассу, а значит нумерация каждого блока делается с помощью одного бита.

Итак, выходит, что из каждого куска длины $n$ изначальной бернуллиевской последовательности мы получаем, по изначально приготовленной таблице классов, его собственный номер в двоичной системе, причем каждый бит этого номера истинно случаен.

Теперь вопрос к КПД этого алгоритма, то есть, как много случайных бит мы получаем из блока длины $n$ и можно ли больше? Оказывается, для каждого значения $p$ существует такое $n$, при котором достигается максимум получаемых истинно случайных бит. То есть больше случайности вытащить из изначальной последовательности не получится. Этот предел задается бинарной энтропийной функцией

$$h(p) = -pln(p) - (1-p)ln(1-p)$$

То есть это максимально возможная доля случайных бит полученная из изначальных бит. Чтобы приблизиться к этому пределу, необходимо брать $n$ как можно больше. Казалось бы, все отлично, но уже при $n = 100$, колличество воможных блоков равно $2^{100}$, а такую таблицу данных хранить невозможно. И здесь на помощь приходит метод Бабкина.

Итак, пусть у нас есть блок длиной $n$, в нем есть $k$ единиц. Мы можем вычислять его номер находу по формуле

$$N(i_1, i_2, \dots, i_k) = C_{i_1-1}^1 + C_{i_2-1}^2 + \cdots + C_{i_k-1}^k$$

Где $i_k$ – порядковый номер k-той единицы, начиная с 0. Однако вычисления N происходят в десятичной системе, поэтому, переводя значение N в бинарный вид, получаем тот самый набор истинно случайных единиц и нулей.

Для выполнения такой операции необходима таблица значений $C_{i_k-1}^{k}$ от номера $i_k$ k-той единицы, то есть таблица $n \times n$. Максимальное значение биномиального коэффициента может быть $2^n$, а значит записано $n$ битами. Следовательно, на такую таблицу потребуется $n^3$ бит памяти, что несравненно меньше таблицы $2^n$ номеров блоков, которая бы потребовалась без использования этого гениального метода.