О методе Монте-Карло, статистических испытаниях, имитации случайных событий и Matlab - инструментарии

Как известно, случайное событие - это такое событие, которое может произойти, а может и не произойти заранее не предсказуемым с достоверностью образом при одних и тех же обстоятельствах и условиях. Причину случайности обычно связывают с влиянием большого числа факторов. Количество таких факторов столь велико, что их невозможно контролировать. Влияние каждого из них незначительно, но в совокупности их влияние на исход наблюдений за событием оказывается непредсказуемым.

Случайность события означает, что если обстоятельства появления (или не появления) события воспроизвести в реальности или мысленно (виртуально), то при каждом таком воспроизведении событие может появиться, а может и не появиться безо всякой закономерности в последовательности появлений.

Однако одни случайные события в такой череде наблюдений появляются чаще, а другие - реже. Числовой мерой частоты появления случайного события является вероятность его появления. Случайные события принято обозначать заглавными буквами латинского алфавита. Для обозначения вероятности используют заглавные или строчные латинские буквы p (от корня латинского слова probabilitas), q и некоторые другие. Вероятность любого случайного события лежит в пределах от нуля до единицы: чем ближе вероятность события к единице, тем чаще появляется это случайное событие в череде наблюдений.

Строгое определение вероятности случайного события основано на математическом понятии числовой меры, определяемой на множестве так называемых элементарных случайных событий. Однако числовой метод Монте-Карло, которым мы будем решать вероятностные задачи, основан на так называемом "статистическом" определении вероятности случайного события. А именно, рассматривается череда воспроизведений одних и тех же условий появления случайного события. В силу случайности каждый раз в этой череде событие может непредсказуемо появиться или не появиться. Отношение числа m появлений события к общему числу N воспроизведений условий $p^* = m / N$ принимают за **оценку вероятности события** и называют частотой появления события. При таком подходе к понятию вероятности кардинальное предположение - с увеличением числа повторных воспроизведений условий наблюдения события до бесконечности это отношение p^* приближается к точному значению вероятности сколь угодно близко («частотное» определение вероятности случайного события).

В череде воспроизводимых условий появления случайного события важнейшим условием является независимость каждого воспроизведения от всех других в череде. Тем самым каждый раз обеспечиваются идентичные, одинаковые условия для случайного события, никак не связанные с предыдущими и последующими исходами в череде повторения ситуации, при которой может произойти случайное событие.

Описанное "частотное", или "статистическое" определение вероятности имеет много недостатков и противоречий. Например, не определен предельный переход для частоты, который, как оказывается, требует использования понятия вероятности, которое само определяется через этот предел. Однако, эти теоретические математические и логические трудности и противоречия не мешают использовать приведенное определение для решения задач, что и будет осуществлено в данном пособии.

Суть **метода Монте-Карло**, иначе ещё называемом более длинно **методом статистических испытаний** или **методом статистического моделирования**, состоит в многократном воспроизведении одних и тех же условий появления случайного события и подсчёте частоты его появления. При больших N значение частоты с достаточной для практического применения точностью совпадает с "точным" значением вероятности, которое для простых задач (всех задач, которые рассматриваются в пособии) можно вычислить аналитически.

Поскольку для аналитического решения вероятностей задачи также строится некоторая математическая модель - вероятностная схема, для которой и отыскивается аналитическое решение, то, по сути, речь идёт о сравнении результатов моделирования одной и той же задачи, описанной ее условием, двумя методами: аналитически и численно методом Монте-Карло.

Поскольку в методе Монте-Карло речь идёт об огромном количестве (сотни тысяч и миллионы раз) повторения одних и тех же условий появления случайного события, то практически эта процедура осуществима лишь с использованием компьютера. Следовательно, реализация метода всегда связана с построением алгоритма и написанием программы для реализации этого алгоритма. Важнейшее преимущество метода Монте-Карло в том, что для многих вероятностных задач разработка алгоритма и программы оказывается проще, чем аналитическое решение задачи.

Рассмотрим простейший пример. Бросают игральную кость (идеально симметричный кубик из однородного материала с пронумерованными гранями). Какова вероятность, что при одном бросании кости выпадет пять или шесть очков?

Аналитическое решение. При одном бросании кости может выпасть любое из шести: 1,2,...,6, - очков. В силу симметрии кости любой из шести исходов равно возможен. Следовательно, на каждый из них приходится одна шестая часть случаев, а на появление пяти или шести приходится треть случаев. Эта треть и есть искомая вероятность, т.е. 1/3.

Решение методом Монте-Карло. Берём игральную кость - кубик и бросаем ее много раз, записывая череду результатов. Затем подсчитываем число случаев, при которых появилась шестёрка или пятёрка, и вычисляем их долю от общего числа бросков. Например, если сделано 1000 повторных бросаний кости, а пятёрки или шестёрки появились в 342 случаях, то оценка вероятности - частота появлений пяти или шести очков - равна 0,342.

Для демонстрации применения метода Монте-Карло, а также аналитических возможностей программной математической системы Matlab в пособии рассмотрены решения задач из всемирно известного, выдержавшего многочисленные переиздания «Сборника задач по курсу Теории Вероятностей и Математической Статистики» В.Е. Гмурмана [1]. Из сборника выбраны задачи, которые отличаются оригинальностью постановки. Соседние с ними задачи, как правило, отличаются лишь числовыми исходными данными или непринципиальными вариациями по содержанию.

Часть задач решается методом Монте-Карло на удивление просто и наглядно, реализация метода требует лишь несколько строк программного кода. Другие задачи требуют определенной изощренности в программировании. Срабатывает закон рычага: избавляемся от логических усилий по аналитическому решению, приобретаем необходимость усилий в построении программного кода. Для студентов, выбравших в качестве направления своей будущей деятельности информационные технологии, второй путь – написание программного кода – по мнению, автора, более предпочтителен.

В качестве технического средства для реализации метода Монте-Карло выбран инструментарий Matlab. В основном, это скрипты, но используются также m-функции и инструментарий для аналитических преобразований и расчетов MUPAD. Поскольку изучение этих средств, конечно, выходит, за рамки данного пособия, которое нацелено на другие цели, то автор рассчитывает, что читатели либо владеют базовыми основам работы в Matlab, либо могут их освоить через справочную систему этого пакета и примеры, приведенные в этих справках, а также получить информацию из интернета.

Еще один полезный и рекомендуемый к применению подход – это запись алгоритма, приведенного в пособии, средствами знакомого читателю языка программирования и сравнение достоинств и недостатков разных сред разработки для решения рассматриваемых задач.

В учебном пособии рассмотрены некоторые задачи из книги [1[1]]. Показано использование метода статистических испытаний (метода Монте-Карло) для приближенного решения задач. Решение сопровождается текстами скриптов и программ средствами системы Matlab для реализации решения. Так называемые точные решения взяты из [1]. Кроме решения, найденного методом Монте-Карло, приведен расход времени на получение решение. Затраты времени зависят от используемого компьютера. Аналитические преобразования выполнены с помощью инструментария MUPAD системы Matlab. Следует учитывать, что приводимые решения дают лишь численную оценку точного результата, которая в некоторых задачах может отличаться от точного ответа. Номера задач совпадают с номерами задач цитируемого пособия.

Скрипты, в которых записан код решающей программы, сопровождается комментарии, которые поясняют главные этапы решения.

§ 1. Классическое и статистическое определение вероятности

<u>Задача 1.</u> Брошены две игральные кости. Найти вероятность того, что сумма очков на выпавших гранях – четная, причем на грани хотя бы одной из костей появится шестерка.

Точный ответ: 5/36 = 0,1389.

```
% Гмурман Задача 1
clear,clc % очистка рабочей области и командного окна
N=10000000; % количество проб
m=0; % количество удачных проб
% имитация N бросаний двух игральных костей
x1=randi([1,6],N,1); x2=randi([1,6],N,1);
for p=1:N % проверка появления события - "успеха"
   if (mod(x1(p,1)+x2(p,1),2)==0 && ((x1(p,1)==6) | (x2(p,1)==6)))
      m=m+1; % накопление числа успехов
   end
end
Pz=m/N % точечная оценка искомой вероятности
```

<u>Ответ</u>: Оценка Pz = 0,1390.

<u>Задача 3</u>. Задумано двузначное число. Найти вероятность того, что задуманным числом окажется случайно названное двузначное число.

---

[1] Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике: Учеб. пособие для студентов вузов/В.Е. Гмурман. — 9-е изд., стер. - М.: Высш. шк., 2004. – 404 с.: ил. **ISBN 5-06-004212-X**

Точный ответ: 1/90 = 0,0111.

```
% Гмурман Задача 3
clear, clc % очистка рабочей области и командного окна
N=10000000; % количество проб
% имитация N случайных целых двузначных чисел
x=randi([10,99],N,1);
z=randi([10,99]); % случайное задуманное число
% подсчет количества удачных проб
ma=sum(length(find(x==z))); Pa=ma/N
```

Ответ: Pa = 0,0111.

Задача 9. Найти вероятность того, что при бросании трех игральных костей шестерка выпадет на одной (безразлично какой) кости, если на гранях двух других костей выпадут числа очков, не совпадающие между собой (и не равные шести).

Точный ответ: $P = ½ = 0{,}5$.

```
% Гмурман Задача 9
clear, clc % очистка рабочей области и командного окна
N=10000000; % количество проб
m=0; % количество удачных проб
% имитация N бросаний трех игральных костей
x1=randi([1,6],N,1); x2=randi([1,6],N,1); x3=randi([1,6],N,1);
for p=1:N % проверка появления события - "успеха"
% проверка успеха и накопление числа успехов
  if x1(p,1)==6 && x2(p,1)~=6 && x3(p,1)~=6
    if x2(p,1)~=x3(p,1)  m=m+1;  end
  end
  if x2(p,1)==6 && x3(p,1)~=6 && x1(p,1)~=6
    if x3(p,1)~=x1(p,1)  m=m+1;  end
  end
  if x3(p,1)==6 && x1(p,1)~=6 && x2(p,1)~=6
    if x1(p,1)~=x2(p,1)  m=m+1;  end
  end
end
Pz=m/N % точечная оценка искомой вероятности
```

Ответ: Pz = 0,2780. Проверить правильность решения, приведенного в задачнике.

Ответ, приведенный в «Руководстве», ошибочный. Точное решение аналитически можно получить, например, следующими рассуждениями. Каждую кость бросают независимо от других. Воспользуемся формулой умножения вероятностей независимых событий. Вероятность появления шестерки на первой кости равна 1/6. Вероятность того, что на второй кости выпадет не шестерка, а любое другое число очков, равна 5/6. Вероятность появления любого из четырех очков, неравных очкам на первых двух костях, равна 4/6. Общая вероятность появления шестерки на 1-й кости и неравных шести и отличающихся очков на 2-й и 3-й костях равна $1/6 \cdot 5/6 \cdot 4/6$. Деля циклическую перестановку 1-й,2-й и 3-й костей и суммируя, найдем вероятность появления шестерки на одной из костей при других и неравных очках на двух других костях: $3 \cdot (1/6 \cdot 5/6 \cdot 4/6) = 5/18 = 0{,}2778$.

Задача 11. В ящике 10 одинаковых деталей, помеченных номерами 1, 2, …, 10. Наудачу извлекли шесть деталей. Найти вероятность того, что среди извлеченных деталей окажутся: а) деталь №1; б) детали №1 и №2.

Точные ответы: $Pa = 0{,}6$; $Pb = 1/3 = 0{,}3333$.

```
% Гмурман Задача 11
clear, clc % очистка рабочей области и командного окна
tic % запуск таймера
N=1000000; % количество проб
ma=0; mb=0;% количество удачных проб
for p=1:N % повторение проб
  % имитация извлечения 6 деталей из 10 случайным образом
  y=randsample(10,6);
  % проверка успеха вопросов а) и б) и накопление числа успехов
  if length(find(y==1))>0 ma=ma+1;
    if length(find(y==2))>0 mb=mb+1; end
  end
end
Pa=ma/N
Pb=mb/N
Tm = toc % продолжительность работы программы
```

Ответы: $Pa = 0{,}6000$. $Pb = 0{,}3331$. $Tm =$ 20,0956 секунд.

Задача 13. В конверте среди 100 фотокарточек находится одна разыскиваемая. Из конверта наудачу извлечены 10 карточек. Найти вероятность того, что среди них окажется нужная.

Точный ответ: $P = 0{,}1$.

```
% Гмурман Задача 13
clear, clc % очистка рабочей области и командного окна
tic % запуск таймера
N=1000000; m=0; % количества всех и удачных проб
for p=1:N % повторение проб
  % имитация извлечения 6 деталей из 10 случайным образом
```