

Моделирование распределений

В.В. Некруткин

кафедра статистического моделирования <http://statmod.ru>, матмех СПбГУ
Материал к специальному курсу, 2014 г.

Содержание

Предисловие	4
1 Введение. Общие положения	5
1.1 Источник случайности и моделирующие формулы	5
1.2 Язык описания алгоритмов моделирования	6
1.3 Точные и приближенные методы	8
1.4 Две характеристики алгоритмов моделирования	9
2 Табличные методы моделирования дискретных распределений	11
2.1 Таблицы дискретных распределений	11
2.2 Табличный метод обратных функций	11
2.2.1 Простейший табличный метод	12
2.2.2 Последовательный метод обратных функций	14
2.2.3 Метод дихотомии	17
2.2.4 Метод Чжень	19
2.3 О методе Уолкера	22
2.4 Еще о методе обратных функций	23
3 Моделирование дискретных распределений, зависящих от параметров	25
3.1 Моделирование биномиального распределения	25
3.2 Моделирование распределения Пуассона	26
3.3 Моделирование геометрического распределения	29
4 Общие методы моделирования	33
4.1 Метод обратных функций	33
4.1.1 Примеры	35
4.2 Метод отбора	36
4.2.1 Идея метода	36
4.2.2 Теория метода	37
4.2.3 Обсуждение. Варианты и примеры	41
4.3 Метод дискретной декомпозиции распределений	45
4.3.1 Общая схема и примеры	45
4.3.2 Модификации метода дискретной декомпозиции	47
4.3.3 Общий метод декомпозиции.	51
5 Приемы моделирования различных распределений	55
5.1 Полярные методы	55
5.1.1 Равномерное распределение на окружности \mathbb{S}^1	55
5.1.2 Нормальное распределение	56
5.1.3 Равномерное распределение на сфере \mathbb{S}^2	57
5.1.4 Распределение Коши	58
5.2 Моделирование многомерных гауссовских распределений	59
5.3 Разные распределения	61
5.3.1 Показательное распределение	61
5.3.2 Степенное распределение	62
5.3.3 Гамма-распределение	62
5.3.4 Бета-распределение	64

6	О моделировании с дискретным источником случайности	66
6.1	Корневые диаграммы и сложность моделирования	66
6.2	Оптимальные $D_M(\mathcal{P}, f)$ -диаграммы	69
6.3	Связь с традиционным моделированием	71
7	Приложение 1. Преобразования случайных векторов	73
7.1	Общее утверждение	73
7.2	Частные случаи, полезные для моделирования	74
7.2.1	Аффинные преобразования	74
7.2.2	Переход к полярным и сферическим координатам	75
7.2.3	Гамма-распределение и распределение Дирихле	76
7.2.4	Распределения порядковых статистик	78
8	Приложение 2. Многомерные гауссовские распределения	80
8.1	Общий случай	80
8.2	Невырожденный случай	81
9	Приложение 3. Упражнения	82
	Список литературы	99

Предисловие

Излагаемый здесь материал соответствует (несколько расширенной) первой части курса лекций «Статистическое моделирование», в течение ряда десятилетий читаемого на одноименной кафедре (и одноименной специализации) математико-механического факультета Санкт-Петербургского (Ленинградского) университета.

Эта первая часть посвящена принципам и основным техническим приемам генерирования¹ случайных величин и векторов, имеющих заданное распределение (для краткости часто употребляют несколько жаргонный термин «моделирование распределений»²). Поскольку материал носит учебный характер, в него включено большое число упражнений.

Дополнительную информацию по этой тематике можно найти в отечественных (например, [1] – [4]) и иностранных (см. [5] – [9]) учебниках и монографиях, посвященных как различным аспектам собственно моделирования, так и его применениям.

Особенно следует отметить книгу [10], которая фактически является подробным обзором и справочником по генерированию случайных величин. Там можно найти достаточно эффективные, глубокие и сложные алгоритмы, изложение которых невозможно в относительно кратком лекционном курсе.

В целом изложение рассчитано на учащихся, достаточно свободно владеющим стандартным аппаратом теории вероятностей, основанным на теории меры.

¹Мы будем употреблять термины «моделирование» и «генерирование» как синонимы.

²Этот термин и вынесен в заголовок.

1 Введение. Общие положения

1.1 Источник случайности и моделирующие формулы

При моделировании случайных явлений необходимо получать реализации случайных величин (и случайных векторов) с заданным распределением. Мы не будем сейчас задаваться вопросом, существует ли на самом деле в нашем мире случайность (пусть этим занимаются философы) и зачем ее моделировать, а сразу приступим к делу.

Ясно, что если мы хотим моделировать некоторое сложное случайное явление, то нам нужно иметь достаточно простой и универсальный источник случайности. Здесь и далее мы предполагаем, что источником случайности является (вообще говоря, бесконечная) последовательность $\alpha_1, \alpha_2, \dots$ независимых равномерно распределенных на отрезке $(0, 1]$ случайных величин. Этот выбор является общепринятым.³

Пусть теперь для простоты \mathcal{P} — распределение, заданное на борелевских подмножествах прямой. С чисто теоретической точки зрения промоделировать распределение⁴ \mathcal{P} означает найти такое $n \geq 1$ и такую измеримую функцию $f : (0, 1]^n \mapsto \mathbb{R}$, что случайная величина $\xi = f(\alpha_1, \dots, \alpha_n)$ имеет распределение \mathcal{P} (мы для краткости будем записывать этот факт как $\mathcal{L}(\xi) = \mathcal{P}$). Конечно, это «определение» автоматически переносится на случай случайной величины, принимающей значения в абстрактном измеримом пространстве (D, \mathcal{D}) . Удобно называть функцию f *моделирующей функцией*, а само равенство $\xi = f(\alpha_1, \dots, \alpha_n)$ — *моделирующей формулой*.⁵

Как уже говорилось, такое «определение» является чисто теоретическим. С точки зрения практики у него есть два дополнительных аспекта, относящихся к компьютерной реализации методов моделирования.

Согласно теории вероятностей, случайная величина ξ — это измеримая функция $\xi : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}, \mathcal{B}_R)$, где Ω — некоторое абстрактное множество (*пространство элементарных событий*), оснащенное σ -алгеброй \mathcal{F} (σ -алгеброй событий), а \mathcal{B}_R — борелевская σ -алгебра подмножеств множества \mathbb{R} . Что касается распределения $\mathcal{P} = \mathcal{P}_\xi$ случайной величины ξ , то оно является нормированной на единицу мерой, определенной на σ -алгебре \mathcal{B}_R равенством $\mathcal{P}(B) = \mathbb{P}(\xi \in B)$, $B \in \mathcal{B}_R$.

Конечно, за исключением того простого случая, когда σ -алгебра \mathcal{F} состоит из конечного числа событий (а случайная величина ξ дискретна), таких объектов в природе просто не существует.

Поэтому не удивительно, что работая на компьютере и обращаясь к генератору псевдослучайных чисел (или физическому генератору, это не важно), вы получаете не некую измеримую функцию α_1 , а просто какое-то число $a_1 \in (0, 1]$. А если обращаетесь к генератору n раз, то получаете (вместо n независимых равномерно распределенных на отрезке $(0, 1]$ случайных величин $\alpha_1, \dots, \alpha_n$) числа a_1, \dots, a_n . Иначе говоря, найдя такую функцию f , что случайная величина $\xi = f(\alpha_1, \dots, \alpha_n)$ имеет заданное распределение и, реализовав это преобразование на компьютере, вы получите на выходе число $x = f(a_1, \dots, a_n)$. И это тоже принято называть моделированием.

Как согласуются эти два взгляда на моделирование? Согласование на самом деле хорошо известно и традиционно для статистических исследований. Мы интерпретируем числа a_1, \dots, a_n как независимые реализации случайных величин $\alpha_1, \dots, \alpha_n$. Эти слова означают, что мы интерпретируем процесс работы генератора как (независимый от всего предыдущего) выбор точки ω_0 в некотором абстрактном пространстве элементарных событий Ω . Тем самым $a_i = \alpha_i(\omega_0)$ и соответствие двух языков установлено. Иначе говоря, изучая теоретические свойства какого-то метода моделирования, мы используем язык случайных величин и действуем, так сказать, до компьютерного эксперимента, а после этого эксперимента работаем с полученными реализациями этих

³Существуют и другие подходы. Так, в статье Д.Кнута и Э.Яо [11] (см. также [10, гл. 15]) в качестве источника случайности рассматривается последовательность независимых случайных величин, имеющих симметричное распределение Бернулли. См. также раздел 6, где кратко рассматриваются некоторые обобщения этого подхода.

⁴Точнее — «промоделировать случайную величину, имеющую распределение \mathcal{P} ».

⁵В дальнейшем мы увидим, что моделирующая функция может иметь даже бесконечное число аргументов, но для понимания сути дела сейчас нам эти тонкости не нужны.

случайных величин.⁶

Второй практический аспект моделирования связан с трудоемкостью вычислений. Одно и то же распределение \mathcal{P} может быть промоделировано с помощью различных⁷ моделирующих функций, и естественно ожидать, что на это потребуются разные машинные ресурсы. Более того, даже для фиксированной моделирующей формулы $\xi = f(\alpha_1, \dots, \alpha_n)$ могут существовать различные способы вычисления функции f , требующие разных затрат. А поскольку стандартной ситуацией является получение большой выборки (объема, скажем, тысяч или десятков тысяч) из распределения \mathcal{P} , то скорость вычисления функции f приобретает первостепенное значение.

Поэтому результатом решения задачи «промоделировать распределение \mathcal{P} » является не просто моделирующая функция, а подробно расписанный, обоснованный и проанализированный алгоритм моделирования.⁸

1.2 Язык описания алгоритмов моделирования

Для записи алгоритмов мы будем применять условный язык, не совпадающий ни с одним языком программирования. В принципе, этот язык достаточно прост и, возможно, не требует специальных пояснений. Тем не менее, учитывая специфику моделирования, мы приведем здесь его описание.

В наиболее полной записи каждый алгоритм состоит из следующих частей.

1. Название алгоритма (обычно это аббревиатура названия метода моделирования) и/или словесное описание моделируемого распределения (расшифровка аббревиатуры).
2. Словесное описание входных и выходных данных алгоритма.
3. Тело алгоритма.

Тело алгоритма разбивается на строки, иногда занумерованные натуральными числами. Это разбиение служит для удобства чтения алгоритма. Каждая строка должна быть достаточно краткой и обозримой, хотя, конечно, конкретное деление на строки — дело условное. В начале каждой строки, если это необходимо, может стоять краткий словесный комментарий, объясняющий смысл производимых операций.

Текст, соответствующий каждой строке, состоит из операторов, разделенных знаком «;». Порядок выполнения операторов естественный, если не оговорено противное. Перечислим основные операторы и соглашения, используемые в алгоритмических описаниях.

1. Оператор присвоения « $a \leftarrow 1$ »;
2. Стандартные логические операторы «and» и «or».
3. Условный оператор: «If A then B » или «If A then B else C ». Если A представляет собой составной логический оператор, то этот оператор записывается в скобках: «If (A_1 and A_2) then B ».

⁶Конечно, остается вопрос, насколько на самом деле числа a_1, \dots, a_n похожи на повторную независимую выборку из равномерного распределения на отрезке $(0, 1]$. Этот вопрос относится к проверке качества генераторов псевдослучайных чисел и пока нам не интересен.

⁷Естественно, функции $(0, 1]^n \mapsto \mathbb{R}$, совпадающие с точностью до n -мерной лебеговой меры ноль, мы не считаем различными, так что на самом деле речь идет об различных классах эквивалентности моделирующих функций. Поскольку на практике с этими тонкостями проблем не бывает, мы не будем дальше заострять на них внимание.

⁸Строго говоря, окончательным результатом должна являться компьютерная программа моделирования, но мы останавливаемся на алгоритмическом уровне, не вдаваясь в тонкости используемого языка программирования и т.д. Еще точнее, часто будут описываться не алгоритмы в строгом смысле этого слова, а лишь их алгоритмические схемы.

4. Оператор завершения STOP. Поскольку операции ввода и вывода в файл не предусматриваются, то приходится описывать числа и массивы, подаваемые на вход алгоритма или являющиеся результатом работы алгоритма. По этой же причине появляется необходимость в операторе, обозначающем завершение работы алгоритма.
 5. Циклы организованы при помощи операторов вида «While A do ...» (здесь A — логическое выражение), «Do ... while A » и «For $i = m$ to n step a do ...». В случае, когда $a = 1$, используется сокращенная запись «For $i = m$ to n do ...».
 6. В алгоритме разрешается проверка любых соотношений типа $a = b$, $a < b$, $a \leq b$, $a > b$, $a \neq b$ и т.д.
 7. В алгоритме разрешается использовать все стандартные арифметические действия (« $a + b$ », « $a - b$ », « $a * b$ », « a/b »), а также возведение в степень (« $a \wedge b$ »). Также можно использовать любые стандартные функции (\ln , \cos , \arctan , $\sqrt{}$, $\lfloor \cdot \rfloor$, $\lceil \cdot \rceil$ и пр.). Полный список таких функций мы устанавливать не будем.
 8. Моделирование случайной величины, равномерно распределенной на отрезке $(0, 1]$ реализуется в алгоритме специальным оператором $\text{Get}(\cdot)$. А именно, выражение $\text{Get}(\alpha)$ означает *получить реализацию α из равномерного распределения на $(0, 1]$* . Иначе говоря, $\text{Get}(\alpha)$ обозначает обращение к генератору⁹ и присвоение результата этого обращения переменной α .¹⁰ Подразумевается, что последовательные применения оператора $\text{Get}(\cdot)$ соответствуют получению независимых реализаций равномерного распределения. Для сокращения записи иногда будут применяться операторы вида « $\text{Get}(\alpha_1, \alpha_2)$ », что предполагается эквивалентным « $\text{Get}(\alpha_1); \text{Get}(\alpha_2)$ ».
- Буква α (с индексом или без) зарезервирована для обозначения равномерно распределенной на отрезке $(0, 1]$ случайной величины. Наличие разных индексов у нескольких случайных величин, обозначенных буквой α , подразумевает, что эти случайные величины независимы.
9. В алгоритм можно включать процедуры, предполагая, что они где-то уже описаны. Например, если в алгоритме используется реализация случайной величины η , имеющей некоторое распределение \mathcal{Q} и если для структуры алгоритма не важно, каким способом эта реализация получена, то мы будем писать $\eta \leftarrow \mathcal{Q}(\cdot)$, где точка в качестве аргумента подчеркивает, что мы имеем дело с моделированием, а не просто с присвоением.¹¹
- Другие процедуры могут быть описаны словесно.
10. Использование выражений вида $\eta \leftarrow \mathcal{Q}(\cdot)$ предполагает, что у нас уже есть обозначения для некоторого класса распределений. Приведем эти обозначения для наиболее употребительных распределений (точные определения распределений и их необходимые свойства вводятся и обсуждаются в тех разделах, где говорится о моделировании этих распределений). Обозначения других распределений будут вводиться по мере надобности.
 - $U(D)$ — равномерное распределение на измеримом подмножестве D евклидова пространства \mathbb{R}^d (предполагается, что $0 < \text{mes}_d(D) < \infty$, где mes_d — d -мерная мера Лебега). В частности, обозначение $U(a, b)$ используется для равномерного распределения на отрезке (a, b) .

⁹Под «генератором» здесь и далее понимается некий абстрактный механизм, поставляющий независимые реализации случайных величин, равномерно распределенных на $(0, 1]$. Этот термин перекидывает мостик к практическому моделированию, где используются так называемые *генераторы псевдослучайных чисел*.

¹⁰Обычно генераторы псевдослучайных чисел требуют инициализации. Здесь это обстоятельство игнорируется.

¹¹В случае, когда мы имеем дело с некоторым стандартным распределением, последнее замечание может быть проигнорировано. Например, допускается запись $\xi \leftarrow N(0, 1)$.

- $U_n(X)$ — равномерное распределение на конечном множестве X , имеющем мощность n (как обычно, природа множества X не имеет значения).
- $\text{Ber}(p)$ — распределение Бернулли с параметром $p \in (0, 1)$.
- $\text{Bin}(n, p)$ — биномиальное распределение с параметрами (n, p) , $n \geq 1$, $p \in (0, 1)$.
- $\text{Geom}(p)$ — геометрическое распределение с параметром $p \in (0, 1)$.

Здесь и далее под геометрическим распределением подразумевается распределение числа неудач (а не числа испытаний) в испытаниях Бернулли с вероятностью успеха p до первого успеха. Иначе говоря, наименьшее возможное значение случайной величины, имеющей распределение $\text{Geom}(p)$, равно нулю.

- $\text{NB}(k, p)$ — отрицательно-биномиальное распределение с параметрами (k, p) , $k > 0$, $p \in (0, 1)$.

Как и в предыдущем случае, отрицательно-биномиальное распределение сосредоточено на множества $\{0, 1, \dots\}$, причем каждое из этих чисел имеет положительную вероятность.

- $\Pi(\lambda)$ — распределение Пуассона с параметром $\lambda > 0$.
- $\text{Exp}(\lambda)$ — показательное (экспоненциальное) распределение с параметром λ .

Распределение предполагается параметризованным таким образом, что его математическое ожидание равно $1/\lambda$.

- $\text{Gamma}(k, \mu)$ — гамма-распределение с параметром формы $k > 0$ и параметром масштаба $\mu > 0$.

Отметим, что $\text{Gamma}(1, \mu) = \text{Exp}(\mu)$.

- $N(a, \sigma^2)$ — нормальное распределение с математическим ожиданием $a \in \mathbb{R}$ и дисперсией $\sigma^2 > 0$.
- $N_d(\bar{a}, \Sigma)$ — многомерное (d -мерное) гауссовское распределение со средним $\bar{a} \in \mathbb{R}^d$ и ковариационной матрицей Σ . Формально Σ является произвольной симметричной неотрицательно определенной $d \times d$ матрицей.

Приведем теперь простой пример записи алгоритма моделирования.

Алгоритм ВВ (Binomial Bernoulli)

Моделирование биномиального распределения $\text{Bin}(n, p)$ через испытания Бернулли

Входные данные: n, p .

Результат: ξ .

1. (Инициализация) $S \leftarrow 0$;
2. (Моделирование числа успехов) For $i = 1$ to n do (Get(α); If $\alpha < p$ then $S \leftarrow S + 1$);
3. (Завершение) $\xi \leftarrow S$; STOP.

В алгоритме 1.2 переменная i (счетчик цикла) обозначает номер моделируемого испытания Бернулли, а переменная S — (текущее) число успехов в этих испытаниях. Поскольку метод моделирования весьма прост, специального обоснования алгоритма не требуется.¹²

1.3 Точные и приближенные методы

Мы будем заниматься только теоретически точными алгоритмами моделирования и не будем рассматривать приближенные методы. Приближенные методы моделирования могут быть весьма привлекательными, но их применять можно только тогда, когда есть уверенность, что допускаемая

¹²Далее мы увидим, что это далеко не всегда так.

ошибка не скажется на конечном результате. Иначе говоря, нужно учитывать особенности той задачи, ради решения которой применяется моделирование.

Например, в ряде учебников в качестве хорошего приближенного метода моделирования стандартного нормального распределения $N(0, 1)$ предлагается такой прием. Моделируются случайные величины $\alpha_1, \dots, \alpha_n$, и в качестве случайной величины, приближенно имеющей распределение $N(0, 1)$, берется случайная величина $\eta^{(n)} = (\alpha_1 + \dots + \alpha_n - n/2)/\sqrt{n/12}$. При $n = 12$ эта формула выглядит особенно привлекательно: $\eta^{(12)} = (\alpha_1 + \dots + \alpha_{12}) - 6$.

Действительно, Центральная Предельная Теорема (теорема П. Леви) утверждает, что при $n \rightarrow \infty$ распределение случайной величины $\eta^{(n)}$ слабо сходится к $N(0, 1)$, а более внимательный анализ показывает, что плотность распределения случайной величины $\eta^{(12)}$ близка к плотности распределения $N(0, 1)$ в довольно широкой окрестности нуля. В то же время $\eta^{(12)} \in (-6, 6)$ с вероятностью 1, а носитель нормального распределения — вся прямая.

Отсюда становится ясно, что если в решаемой задаче существенным является лишь поведение плотности нормального распределения в окрестности нуля, то применение случайной величины $\eta^{(12)}$ вместо случайной величины $\xi \in N(0, 1)$ может быть оправдано. Если же на результат оказывают влияния очень редкие события типа $\{\omega : |\xi(\omega)| > 10\}$, то использовать $\eta^{(12)}$ нельзя.

Кроме этого общего соображения есть и другое, более профессиональное. Считается, что если есть некоторый приближенный метод моделирования распределения \mathcal{P} , то на его основе можно придумать теоретически точный метод, имеющий практически те же характеристики (например, быстродействие). Конечно, доказать такое утверждение во всей его полноте невозможно.

1.4 Две характеристики алгоритмов моделирования

Учитывая эти обстоятельства, мы будем рассматривать две основные характеристики алгоритмов моделирования: *трудоемкость* и *требуемую память*. Трудоемкость C (от complexity, сложность) описывает поведение (среднего) времени выполнения программы, написанной согласно рассматриваемому алгоритму. Конечно, здесь не может идти речь о каких-то точных числах, имеется в виду общий характер поведения этого времени в зависимости от параметров распределения (или в сравнении с другими методами). Важность такой характеристики как трудоемкость объясняется тем, что моделирование распределений обычно осуществляется много раз в больших циклах. В серьезных задачах скорость моделирования оказывается важным фактором даже на современных компьютерах.

Например, трудоемкость алгоритма **ВВ** пропорциональна числу обращений к генератору, которое равно n . В то же время эта трудоемкость мало зависит от параметра p (от значения p зависит только число присвоений $S \leftarrow S + 1$, поскольку это быстрая операция, ею можно пренебречь). Таким образом мы можем написать, что для алгоритма **ВВ** выполнено соотношение $C = C(n, p) \asymp n$ равномерно по p .¹³ Тем самым алгоритм будет обладать малым быстродействием при больших n .

Что касается требуемой памяти M (memory), то для алгоритма **ВВ** все очень просто: $M = O(1)$, так как практически вся память идет на хранение тела программы. Может показаться, что для современных компьютеров величина M не имеет большой роли. Однако при решении сложных задач, когда моделирование различных распределений является лишь одной из многих составных частей всей программы, необходимо следить и за этой характеристикой.

Табличные и нетабличные методы. Несколько условное деление методов моделирования на *табличные* и *нетабличные* не является совершенно четким, но может оказаться полезным. Дело в том, что алгоритмы моделирования могут быть использованы на практике, грубо говоря, в двух

¹³Для положительных последовательностей a_n и b_n пишут $a_n \asymp b_n$, если отношение a_n/b_n отделено от нуля и бесконечности. Иначе говоря, в этом случае предполагается, что $0 < c_1 \leq a_n/b_n \leq c_2 < \infty$ при достаточно больших n для некоторых постоянных c_1, c_2 .

ситуациях.¹⁴ Одна из них состоит в том, что один и тот же алгоритм без изменений работает в цикле. Иначе говоря, нам нужно промоделировать одно и тоже распределение много (на практике — тысячи и десятки тысяч) раз. В этом случае имеет смысл один раз провести некоторую предварительную работу с тем, чтобы до предела убыстрить собственно моделирующую часть программы, которая и будет работать в цикле. Поскольку эта предварительная работа часто (но не всегда) сводится к вычислению некоторых таблиц (массивов), то такие методы называются *табличными*.

Для табличных методов естественно разделять трудоемкость предварительной и моделирующей частей алгоритма, причем главную роль имеет трудоемкость моделирующей части, так как она выполняется много раз, а предварительная — только один.

Другая ситуация возникает, когда распределение нужно промоделировать один или небольшое число раз. Тогда нет необходимости отделять предварительную часть алгоритма, которая часто бывает весьма трудоемка, от моделирующей части. Подобного рода алгоритмы будем называть *нетабличными*.

Стандартный пример «нетабличной ситуации» возникает, когда нужно моделировать распределение, зависящее от параметров, причем параметры не являются фиксированными, а могут меняться в процессе счета. Тогда результат работы предварительной части алгоритма тоже будет зависеть от параметров (которые и подаются на вход алгоритма), и разделение алгоритма на предварительную и моделирующую части становится бессмысленным.

Алгоритм **ВВ**, конечно же, не является табличным — поскольку на вход алгоритма подаются лишь параметры n и p , они вполне могут измениться при следующем обращении к программе моделирования.

Ясно, что при сравнении алгоритмов, моделирующих одно и тоже распределение, нужно сравнивать «табличные» методы с «табличными», а «нетабличные» методы с «нетабличными». Сравнение табличных методов с нетабличными большого смысла не имеет.

¹⁴Бывают и промежуточные случаи.

2 Табличные методы моделирования дискретных распределений

2.1 Таблицы дискретных распределений

В этом разделе мы будем рассматривать табличные методы моделирования конечных дискретных распределений. А именно, будем считать, что существует такое конечное подмножество X множества¹⁵ \mathbb{R} , что $\mathcal{P}(X) = 1$. Конечно, в этом случае распределение \mathcal{P} определяется набором чисел $p_x = \mathcal{P}(\{x\})$ при $x \in X$. Удобно считать, что все числа p_x положительны. Тогда, обозначив $n = \text{card } X$ и перенумеровав в каком-то порядке элементы множества X , мы можем представить распределение \mathcal{P} в виде *таблицы распределения*

$$\mathcal{P}: \begin{pmatrix} x_1 & \dots & x_k & \dots & x_n \\ p_1 & \dots & p_k & \dots & p_n \end{pmatrix}, \quad (2.1.1)$$

где $X = \{x_1, \dots, x_n\}$, $p_n = \mathcal{P}(\{x_n\})$ и $\sum_{k=1}^n p_k = 1$.

Из этого описания следует, что распределение \mathcal{P} порождает много таблиц распределения, каждая из которых однозначно определяет \mathcal{P} . В некоторых случаях существует «естественная» таблица распределения. Например, если X представляет собой множество, состоящее из первых n членов натурального ряда, то естественно положить $x_i = i$ (хотя, конечно, можно взять и $x_i = n - i + 1$).

Все табличные методы моделирования конечных дискретных распределения основываются на некоторой таблице распределения и имеют на входе число n и два массива x_1, \dots, x_n и p_1, \dots, p_n .¹⁶ При этом для некоторых из этих методов результат не зависит от вида таблицы (2.1.1), а для других — зависит.

Среди последних наиболее популярен так называемый *метод обратных функций*.¹⁷ Его мы и будем сейчас рассматривать.

2.2 Табличный метод обратных функций

Пусть конечное дискретное распределение задано таблицей (2.1.1). Положим $s_i = p_1 + \dots + p_i$ при $0 \leq i \leq n$ (обычно эти числа называют *накопленными вероятностями*). Очевидно, $s_0 = 0$ и $s_n = 1$. *Метод обратных функций* задается моделирующей функцией $f: (0, 1] \mapsto X$ такой, что

$$f(x) = x_i, \quad \text{если } s_{i-1} < x < s_i; \quad i = 1, \dots, n. \quad (2.2.1)$$

Поскольку $\mathbb{P}(\alpha = s_i) = 0$, то значения функции f в точках s_i не имеют значения, поэтому в (2.2.1) они опущены. Обычно полагают $f(s_i) = x_{i-1}$ или $f(s_i) = x_i$ в зависимости от удобства записи алгоритма. В дальнейшем мы тоже будем так поступать, специально это не оговаривая.

Ясно, что вид этой функции существенно зависит от порядка, приписываемого элементам множества X . Соответствующая моделирующая формула может быть записана в виде

$$\xi = \sum_{i=1}^n x_i \mathbb{I}_{(s_{i-1}, s_i]}(\alpha), \quad (2.2.2)$$

где $\mathbb{I}_A(x)$ обозначает *индикатор* множества A , равный 1 или 0 в зависимости от того, принадлежит x множеству A или нет.

Формула (2.2.2) проясняет суть дела. Отрезок $(0, 1]$ делится на n отрезков (окон) вида $(s_{i-1}, s_i]$. Если α попадает в i -е из этих окон, то случайная величина ξ принимает значение x_i . Поскольку длина окна $(s_{i-1}, s_i]$ равна p_i , а все числа x_j различны, то $\mathbb{P}(\xi = x_i) = p_i$.

¹⁵Вообще говоря, природа множества X не имеет значения: его элементами могут быть вектора, картинки, слова и т.д. Нам удобнее везде считать, что $X \subset \mathbb{R}$. В дальнейшем множество X , вне зависимости от того, является ли оно конечным или счетным, мы будем называть *носителем* дискретного распределения \mathcal{P} .

¹⁶Конечно, нам достаточно знать числа p_1, \dots, p_{n-1} , но мы будем пренебрегать такой экономией.

¹⁷Происхождение этого названия будет объяснено позже.

Хотя моделирующая функция (2.2.1) (и, следовательно, моделирующая формула (2.2.2)) однозначно определяется таблицей распределения (2.1.1), существует много способов вычисления этой функции в произвольной точке x . На самом деле, конечно, речь идет о различных способах поиска интервала $(s_{i-1}, s_i]$, содержащего точку x , а «произвольность» x означает, что вместо числа x мы рассматриваем случайную величину α , равномерно распределенную на $(0, 1]$.

Прежде чем исследовать эти способы в общем виде, рассмотрим один частный случай.

2.2.1 Простейший табличный метод

Начнем с совсем простой ситуации. Пусть в таблице распределения (2.1.1) все p_i одинаковы (и, следовательно, равны $1/n$). Тогда \mathcal{P} является равномерным распределением на множестве X , то есть $\mathcal{P} = U_n(X)$.

Поскольку¹⁸ $\mathbb{P}(\lceil n\alpha \rceil = k) = 1/n$ при $k = 1, \dots, n$, то распределение \mathcal{P} очень просто моделируется. Запишем сразу соответствующий алгоритм.

Алгоритм $U_n(X)$ (Uniform on X)

Моделирование равномерного распределения на множестве $X = \{x_1, \dots, x_n\}$

Входные данные: n , массив (x_1, \dots, x_n) .

Результат: ξ .

1. (Выбор номера окна) $\text{Get}(\alpha); \tau \leftarrow \lceil n * \alpha \rceil$;

2. (Результат) $\xi \leftarrow x_\tau$; STOP.

Так как случайная величина τ равномерно распределена на множестве $\{1, \dots, n\}$, то специального обоснования этот алгоритм не требует. Обсудим теперь свойства этого алгоритма. Прежде всего алгоритм реализует метод обратных функций, так как $\xi = x_i$ тогда и только тогда, когда $\alpha \in (i/n, (i+1)/n]$.

Далее, алгоритм не является табличным (предварительная часть алгоритма отсутствует). Требуемая память M имеет вид $n + O(1)$, так как нужно держать в памяти числовой массив X .¹⁹

Что касается трудоемкости, то тут возникают некоторые условности, связанные со скоростью вычисления значений различных функций на компьютере. В данной ситуации нам нужно знать, сильно ли зависит скорость вычисления функции $\lceil \cdot \rceil$ от значений аргумента. Принято считать, что в широком диапазоне значений x время вычисления значения $\lceil x \rceil$ меняется мало.²⁰ Если принять это предположение, то трудоемкость алгоритма $U_n(X)$ окажется почти независимой от n . Таким образом, мы будем читать, что $C = O(1)$.

Развивая идею алгоритма $U_n(X)$, мы приходим к *простейшему табличному методу*²¹ моделирования распределения (2.1.1) при условии, что вероятности p_i имеют вид N_i/N , где $N = N_1 + \dots + N_n$ и N_i — целые положительные числа. Иначе говоря, мы рассматриваем таблицу распределения вида

$$\mathcal{P}: \begin{pmatrix} x_1 & \dots & x_k & \dots & x_n \\ N_1/N & \dots & N_k/N & \dots & N_n/N \end{pmatrix}. \quad (2.2.3)$$

Предполагается, что на вход алгоритма поступают целые N_i (а не вещественные $p_i = N_i/N$).

¹⁸Проверьте!

¹⁹Здесь используются два предположения. Во-первых, мы предположили, что $x_i \in \mathbb{R}$. Если, скажем, x_i — d -мерные вектора, то M будет иметь вид $dn + O(1)$. Во-вторых, если число x_i легко вычисляется по значению индекса i (запишем это как $x_i = h(i)$), то массива (x_1, \dots, x_n) вообще заводить не нужно, а в алгоритме вместо $\xi \leftarrow x_\tau$ появится $\xi \leftarrow h(\tau)$. (Ярким примером этой ситуации является случай $x_i = i$.) Тогда $M = O(1)$. В дальнейшем мы будем вспоминать об этих обстоятельствах только тогда, когда они будут действительно существенны.

²⁰И вычислительные эксперименты это подтверждают.

²¹Согласно терминологии [10] — *table look-up method*.

Идею (и обоснование) метода можно описать следующим образом. Сформируем массив («таблицу») T_1, \dots, T_N , таким образом, что

$$T_j = x_l \quad \text{при} \quad \sum_{i=1}^{l-1} N_i < j \leq \sum_{i=1}^l N_i, \quad l = 1, \dots, n. \quad (2.2.4)$$

Массив (2.2.4) устроен очень просто: первые N_1 его элементов равны x_1 , вторые N_2 равны x_2 и т.д. Обозначим $\tau = \lceil N\alpha \rceil$. Тогда, как легко видеть, случайная величина $\xi = T_\tau$ будет иметь распределение (2.2.3). Действительно,

$$\mathbb{P}(\xi = x_l) = \mathbb{P}(T_\tau = x_l) = \mathbb{P}\left(\sum_{i=1}^{l-1} N_i < \tau \leq \sum_{i=1}^l N_i\right) = N_l/N,$$

так как случайная величина τ равномерно распределена на множестве $\{1, \dots, N\}$.²² Конечно, описанный метод — это метод обратных функций.²³ Запишем соответствующий алгоритм, поместив в предварительную часть формирование массива T_j .

Алгоритм TS (Table Simplest)

Моделирование конечного дискретного распределения с $p_i = N_i/N$, ($i = 1, \dots, n$, $N = N_1 + \dots + N_n$)

Предварительные операции: формирование массива (T_1, \dots, T_N)

Входные данные: N , массивы (x_1, \dots, x_n) и (N_1, \dots, N_n) .

Результат: массив (T_1, \dots, T_N) .

1. (Инициализация) $i \leftarrow j \leftarrow 1$; $k \leftarrow N_1$;
2. (Запись чисел x_j в массив) While $k \leq N$ do
 - While $i \leq k$ do $(T_i \leftarrow x_j; i \leftarrow i + 1)$;
 - $j \leftarrow j + 1$; $k \leftarrow k + N_j$;
3. (Завершение) STOP.

Моделирование

Входные данные: N , массив (T_1, \dots, T_N) .

Результат: ξ .

1. (Выбор номера окна) Get(α); $\tau \leftarrow \lceil N * \alpha \rceil$;
2. (Результат) $\xi \leftarrow T_\tau$; STOP.

Моделирующая часть комментариев не требует — она точно такая же, как алгоритм $U_n(X)$. Обсудим предварительную часть алгоритма. Заполнение массива (T_1, \dots, T_N) происходит последовательно блоками. Блок номер j состоит из N_j одинаковых чисел x_j , всего блоков n . Переменная j алгоритма — счетчик числа блоков, переменная i — счетчик номера элемента в массиве (T_1, \dots, T_N) . Переменная k ответственна за переход от блока к блоку — как только номер i превышает текущее значение k , число x_j меняется на x_{j+1} . Останов организован по переменной k .

Алгоритм **TS** имеет ярко выраженный табличный характер. При больших N и n его предварительная часть может оказаться весьма трудоемкой, в то время как моделирующая (с учетом наших

²²На самом деле этот метод отличается от описанного выше метода моделирования дискретного равномерного распределения на конечном множестве $T = (T_1, \dots, T_N)$ только тем дополнительным допущением, что некоторые значения T_i могут совпадать.

²³А почему?

соглашения о скорости вычисления значений функции $\lceil \cdot \rceil$) обладает трудоемкостью $C = O(1)$. Поскольку подразумевается, что предварительная часть алгоритма обрабатывает только один раз, а моделирующая — много раз, то принято общую трудоемкость алгоритма определять по его моделирующей части.

Тем самым по быстройдействию метод является очень привлекательным. Подводный камень обнаружится, когда мы проанализируем память M , которая (опять-таки для моделирующей части алгоритма) равна $N + O(1)$. А число N — знаменатель в представлении $p_i = N_i/N$. Ясно, что этот знаменатель может быть очень большим.

Более того, у этого алгоритма наблюдается весьма неестественное свойство, которое (для наглядности) лучше проиллюстрировать на примере. Возьмем $n = 2$ и рассмотрим распределение

$$\mathcal{P}: \begin{pmatrix} x_1 & x_2 \\ p_1 & p_2 \end{pmatrix}$$

с $p_1 = 3/10$ и $p_2 = 7/10$. Для моделирования этого распределения методом **TS** требуется массив T длины $N = 10$. Это, конечно, совсем немного.

Изменим теперь эти вероятности: положим $p_1 = (3 \cdot 10^5 + 1)/10^6$ и $p_2 = (7 \cdot 10^5 - 1)/10^6$. Вероятности изменились на 10^{-6} , а длина N вспомогательного массива подскочила до 10^6 ! Получается, что алгоритм **TS** обладает большой неустойчивостью памяти: грубо говоря, чем меньше мы меняем вероятности, тем больше нам требуется память. Конечно, это странное свойство.

Сделаем вывод. Хотя алгоритм **TS** и является быстрым, область его применения сильно ограничена величиной знаменателя N в представлении $p_i = N_i/N$. А поскольку $n \leq N$ (все вероятности должны быть положительны), то и число значений моделируемого распределения не может быть слишком большим.

Возвратимся теперь к методу обратных функций для моделирования общего конечного дискретного распределения вида (2.1.1).

2.2.2 Последовательный метод обратных функций

Ввиду своей простоты и наглядности *последовательный метод обратных функций* занимает (не слишком ему соответствующее) первое место во многих учебниках по моделированию. Рассмотрим формулу (2.2.2) и положим

$$\tau = \sum_{i=1}^n i \mathbb{I}_{(s_{i-1}, s_i]}(\alpha). \quad (2.2.5)$$

Согласно (2.2.5) значение τ равно номеру i того окна $(s_{i-1}, s_i]$, куда попадает α . Легко видеть, что τ может быть представлено в виде

$$\tau = \min\{i : \alpha \leq s_i\}. \quad (2.2.6)$$

Действительно, если $s_{i-1} < \alpha \leq s_i$ (это эквивалентно $\tau = i$), то $\alpha > s_j$ при всех $j < i$, и $\alpha \leq s_j$ при всех $j > i$. Формула (2.2.6) наводит на мысль, что можно вычислять τ , последовательно сравнивая α с s_1, s_2 , и так далее до тех пор, пока первый раз не реализуется неравенство $\alpha \leq s_i$ при каком-то i . Полученное число i и есть значение τ .²⁴

Это идея и стоит в основе последовательного метода обратных функций. Запишем соответствующий алгоритм. Поскольку у нас распределение предполагается заданным своей таблицей, а для вычисления τ нам нужны не вероятности p_i , а накопленные вероятности s_i , то отнесем вычисление накопленных вероятностей в предварительную часть алгоритма.²⁵

²⁴Так как $\mathbb{P}(\alpha = s_j) = 0$, то, в принципе, не важно, проверять ли выполнения неравенств $\alpha \leq s_j$ или $\alpha < s_j$. Мы все равно получим распределение (2.1.1).

²⁵Конечно, вычислять s_i можно и в моделирующей части при помощи рекуррентного соотношения $s_i = s_{i-1} + p_i$, но в табличных методах считается грамотным выносить в предварительную часть все, что можно, — это хоть немного, но убыстрит моделирование.

Алгоритм TIS (Table Inverse Sequential)

Моделирование конечного дискретного распределения последовательным методом обратных функций

Предварительные операции: формирование массива накопленных вероятностей (s_1, \dots, s_n)

Входные данные: n , массив (p_1, \dots, p_n) .

Результат: массив (s_1, \dots, s_n) .

1. (Инициализация) $i \leftarrow 1$; $s_1 \leftarrow p_1$;
2. (Вычисление накопленных вероятностей) While $i \leq n$ do ($i \leftarrow i + 1$; $s_i \leftarrow s_{i-1} + p_i$);
3. (Завершение) STOP.

Моделирование

Входные данные: массивы (x_1, \dots, x_n) и (s_1, \dots, s_n) .

Результат: ξ .

1. (Инициализация) $\text{Get}(\alpha)$; $i \leftarrow 1$;
2. (Поиск интервала) While $\alpha > s_i$ do $i \leftarrow i + 1$;
3. (Завершение) $\xi \leftarrow x_i$; STOP.

Проанализируем свойства этого алгоритма.²⁶ Требуемая память M (согласно нашей договоренности) равна, конечно, $2n + O(1)$. Что касается трудоемкости, то она определяется числом шагов в цикле «Поиск интервала». Поскольку в каждом шаге цикла основной операцией является проверка неравенства $\alpha > s_i$, то можно сказать, что трудоемкость определяется числом проверок этих неравенств. Но число этих сравнения зависит от α и, следовательно, случайно. Действительно, если $\alpha \leq s_1$, то алгоритм заканчивает работу после одной проверки, а если $\alpha > s_{n-1}$, то требуется n проверок. Легко видеть,²⁷ что на самом деле число сравнений равно τ , где τ определено в (2.2.5).

Здесь мы видим существенное отличие стохастических алгоритмов от детерминированных: если считать, что и в том и в другом случае трудоемкость определяется числом каких-то основных операций, то в детерминированных алгоритмах (при фиксации всех параметров) это число четко определено, а стохастическом алгоритме может быть случайным.²⁸ В этом случае трудоемкость считают пропорциональной среднему числу таких операций.

В нашем случае

$$C \asymp \mathbb{E}\tau = \sum_{i=1}^n ip_i. \quad (2.2.7)$$

Видно, что трудоемкость C зависит не только от моделируемого распределения \mathcal{P} , но и от таблицы его распределения, то есть от способа упорядочивания вероятностей p_1, \dots, p_n . Например, если p_1 близко к 1, то с большой вероятностью алгоритм TIS закончит свою работу за 1 шаг, а если $p_n \approx 1$, то с вероятностью, близкой к 1, ему понадобится n шагов. Иначе говоря, при фиксированном n величина $\mathbb{E}\tau$, определяющая трудоемкость последовательного метода обратных функций удовлетворяет неравенству

$$1 < \mathbb{E}\tau < n \quad (2.2.8)$$

²⁶Поскольку $s_n = 1$, то в предварительной части эту величину можно было не вычислять, а в моделирующей части не проводить проверку очевидного равенства $\alpha < s_n$. Выигрыш, однако, мог бы получить совсем маленький, а, как окажется в дальнейшем, структура приведенной моделирующей части алгоритма TIS удобна и для моделирования дискретных распределений с бесконечным носителем.

²⁷Действительно?

²⁸А может и не быть, см. Алгоритм BB.

в зависимости от моделируемого распределения \mathcal{P} . При этом обе границы точны в том смысле, что $E\tau$ может быть как сколь угодно близкой к 1, так и сколько близкой к n .²⁹

Таким образом, при фиксированном большом n в зависимости от распределения \mathcal{P} и от выбранной таблицы этого распределения алгоритм **TIS** может быть как очень быстрым, так и очень медленным. Это, конечно, не очень хорошее свойство.

Если трудоемкость алгоритма зависит от параметров, имеет смысл изучать не только самый лучший и самый худший случаи, но и некоторую «среднюю» ситуацию. Поясним, как это можно делать для алгоритма **TIS**. Рассмотрим случайный вектор (π_1, \dots, π_n) такой, что $\pi_1 + \dots + \pi_n = 1$ и предположим, что распределение этого вектора не меняется после любой перестановки координат. Это означает, в частности, что все π_i имеют одинаковое распределение. Поэтому $E\pi_i = 1/n$. Если в алгоритме **TIS** взять вместо исходных вероятностей p_1, \dots, p_n случайные величины π_1, \dots, π_n , то правая часть (2.2.7) станет случайной величиной. А среднюю трудоемкость алгоритма естественно считать пропорциональной математическому ожиданию этой случайной величины. Оно, конечно, равно $(1 + 2 + \dots + n)/n = (n + 1)/2$. То есть в среднем трудоемкость алгоритма растет с ростом n линейно, хотя и в 2 раза медленнее, чем для самых «плохих» распределений.

В случае, когда распределение \mathcal{P} фиксировано, возникает естественный вопрос: как выбрать такую таблицу этого распределения, чтоб получившаяся трудоемкость алгоритма **TIS** была наименьшей? Ответ, в общем, является понятным: нужно упорядочить вероятности p_1, \dots, p_n по убыванию. Докажем этот простой факт в случае, когда носитель дискретного распределения \mathcal{P} не обязательно является конечным.

Лемма 1. Пусть p_1, \dots, p_n, \dots — положительные числа, в сумме дающие 1. Обозначим p_i^* результат упорядочивания чисел p_i по убыванию. Тогда

$$\sum_{i \geq 1} ip_i \geq \sum_{i \geq 1} ip_i^*. \quad (2.2.9)$$

Доказательство. Положим $q_i = \sum_{j \geq i} p_j$. Тогда, представляя ip_i как $p_i + \dots + p_i$ (i раз) и, группируя нужным образом члены полученного ряда, получим

$$\sum_{i \geq 1} ip_i = (p_1 + \dots + p_n + \dots) + (p_2 + \dots + p_n + \dots) + \dots (p_i + \dots + p_n + \dots) + \dots = \sum_{i \geq 1} q_i.$$

Поэтому достаточно доказать, что $q_i^* \stackrel{\text{def}}{=} \sum_{j \geq i} p_j^* \leq q_i$. Но $1 - q_i = p_1^* + \dots + p_{i-1}^* \geq p_1 + \dots + p_{i-1} = 1 - q_i$ (слева — сумма $j - 1$ наибольших из p_1, \dots, p_n, \dots чисел, справа — сумма первых из них). Поэтому утверждение доказано. \square

Вернемся к моделированию. Из утверждения Леммы 1 следует, что, предварительно упорядочив по убыванию вероятности $p_i = \mathcal{P}(\{x_i\})$, перейдя к соответствующей таблице распределения и применив после этого последовательный метод обратных функций, мы уменьшим исходную трудоемкость (2.2.7).

На этом основан прием, который приводит к *модифицированному последовательному методу обратных функций*. При этом предполагается, что входными данными для алгоритма все равно являются массивы (x_i, \dots, x_n) и (p_1, \dots, p_n) . Иначе говоря, мы считаем, что существует какая-то естественная упорядоченность чисел x_i , не связанная с величиной вероятностей p_i .³⁰

Не приводя полностью соответственного алгоритма, обозначим лишь его структуру.

²⁹Поскольку мы предположили, что все вероятности p_i положительны, то ни одна из границ точно достигаться не может.

³⁰Если, например, $x_i = i$, то такое предположение совершенно естественно.

Алгоритм TISM (Table Inverse Sequential Modified)

Моделирование конечного дискретного распределения модифицированным последовательным методом обратных функций

Предварительные операции

Входные данные: n , массивы (p_1, \dots, p_n) и (x_1, \dots, x_n) .

Результат: массивы (x_1^*, \dots, x_n^*) и (s_1^*, \dots, s_n^*) .

В предварительные операции входят:

1. Упорядочивание по убыванию чисел p_1, \dots, p_n вместе с одновременной перестройкой массива (x_1, \dots, x_n) .
Результаты — массивы (p_1^*, \dots, p_n^*) и (x_1^*, \dots, x_n^*) ;
2. Вычисление новых накопленных вероятностей $s_i^* = p_1^* + \dots + p_i^*$.
Результат — массив (s_1^*, \dots, s_n^*) .

Моделирование

Входные данные: массивы (x_1^*, \dots, x_n^*) и (s_1^*, \dots, s_n^*) .

Результат: ξ .

Эта часть алгоритма **TISM** получается из моделирующей части алгоритма **TIS** заменой s_i, x_i на s_i^*, x_i^* .

Посмотрим, что происходит с трудоемкостью метода **TISM**. Легко показать, что теперь при фиксированном n вместо (2.2.8) мы имеем неравенство $1 < \mathbb{E}\tau \leq (n+1)/2$, причем правая граница достигается на дискретном равномерном распределении. Докажем это.

Пусть $p_1 \geq p_2 \geq \dots \geq p_n > 0$ и $\sum_{i=1}^n p_i = 1$. Тогда $\sum_{i=1}^k p_i \geq k/n$. Действительно, если $\sum_{i=1}^k p_i < k/n$, то $p_k < 1/n$ (остальные вероятности в этой сумме не меньше p_k) и, следовательно, $\sum_{i=k+1}^n p_i \leq (n-k)p_k < (n-k)/n$. Поэтому сумма всех вероятностей p_i оказывается меньше 1. Таким образом, применение модифицированного метода обратных функций приводит к уменьшению трудоемкости моделирования самого «плохого» распределения примерно в 2 раза.

Замечание 1. Обсудим некоторые особенности алгоритма **TISM**.

1. Прежде всего, моделирующая функция алгоритма **TISM** отличается (вообще говоря) от моделирующей функции соответствующего алгоритма **TIS**, так как методы определяются не моделируемыми распределениями, а таблицами этих распределений. Таким образом, уменьшение трудоемкости при переходе от **TIS** к **TISM** достигается не более совершенным способом вычисления моделирующей функции, а изменением этой функции.
2. При $p_i = 1/n$ моделирующие функции алгоритмов совпадают. Более того, они совпадают с моделирующей функцией алгоритма $U_n(X)$. Тем самым мы получили первый пример, когда разные способы вычисления одной и той же моделирующей функции приводят к разительно отличающимся трудоемкостям.
3. Если $x_i = i$, то метод **TIS** при соответствующей переделке требует $n + O(1)$ ячеек памяти. Будучи модифицированным до **TISM** он потребует уже $2n + O(1)$ ячеек. В некоторых ситуациях это может оказаться существенным.

2.2.3 Метод дихотомии

Поскольку метод обратных функций сводится к нахождению номера τ окна, куда попадает случайная величина α , а число окон конечно, то естественно попробовать искать это окно *методом дихотомии*³¹.

³¹В англоязычной литературе — *bisection method*, поэтому иногда используется термин «метод бисекции».

Суть метода состоит в следующем. Сначала находится число n_1 , примерно равное $n/2$ (конечно, при нечетном n это можно делать разными способами). Затем проверяется условие $\alpha \leq s_{n_1}$ (напомним, что накопленные вероятности считаются уже вычисленными). Если это неравенство выполняется, то следующая проверка происходит с $n_2 \approx n_1/2$, в противном случае берется $n_2 \approx n_1 + (n - n_1)/2 = (n + n_1)/2$ и т.д. В результате каждой итерации мощность множества индексов I , которому принадлежит случайная величина τ , уменьшается примерно в 2 раза вне зависимости от значений вероятностей p_i . Приведем один из вариантов реализации этой идеи.

Алгоритм TID (Table Inverse Dichotomy)

Моделирование конечного дискретного распределения методом дихотомии

Предварительные операции: формирование массива накопленных вероятностей (s_1, \dots, s_n)

Входные данные: n , массив (p_1, \dots, p_n) .

Результат: массив (s_1, \dots, s_n) .

Алгоритм совпадает с предварительной частью алгоритма TIS.

Моделирование

Входные данные: n , массивы (x_1, \dots, x_n) и (s_1, \dots, s_n) .

Результат: ξ .

1. (Инициализация) $\text{Get}(\alpha)$; $i \leftarrow 1$; $j \leftarrow n$;
2. (Дихотомия) While $i < j$ do
 - $k \leftarrow \lfloor (i + j)/2 \rfloor$; If $\alpha \leq s_k$ then $j \leftarrow k$ else $i \leftarrow k + 1$;
3. (Завершение) $\xi \leftarrow x_i$; STOP.

Прокомментируем структуру алгоритма TID. Переменные i и j задают (текущее) множество $I = \{i, \dots, j\}$, которому принадлежит τ . Поэтому начальные значения есть $i = 1$ и $j = n$, а останов происходит при $i = j$. Разбиение множества I на 2 части проводится вычислением $k = \lfloor (i + j)/2 \rfloor$. В зависимости от того, в какую из этих частей попало τ (проверка этого осуществляется при помощи неравенства $\alpha \leq s_k$), идет пересчет левой или правой границы множества I .

Оценим сверху трудоемкость алгоритма TID, по прежнему соотнося ее с числом T проверок неравенств $\alpha \leq s_k$ или, что тоже самое, числом делений множества $\{1, \dots, n\}$ на две части. Если текущее множество I имеет вид $\{i, \dots, j\}$, то оно состоит из $l = j - i + 1$ элементов и разбивается выбором числа $k = \lfloor (i + j)/2 \rfloor$ на два подмножества, имеющих соответственно $k - i + 1$ и $j - k$ элементов. Легко видеть, что оба этих числа лежат между $(l - 1)/2$ и $(l + 1)/2$.

Таким образом, за m шагов алгоритма мощность множества, локализирующего положение случайной величины τ , не будет превосходить величины $\max(1, T_m^+)$, где

$$T_m^+ = n/2^m + 1/2 + \dots + 1/2^m < n/2^m + 1, \quad (2.2.10)$$

и будет не меньше, чем

$$T_m^- = n/2^m - 1/2 - \dots - 1/2^m > n/2^m - 1.$$

Из (2.2.10) следует, что любое число m , при котором $2^m > n$, является оценкой сверху трудоемкости алгоритма. Минимальное такое m равно $\lfloor \log_2(n) \rfloor + 1$. В то же время неравенство $T_m^- > 1$ удовлетворяется, как нетрудно видеть, при $m < \lceil \log_2(n) \rceil - 1$.

Таким образом, в алгоритме TID для величины T выполнено неравенство

$$\lfloor \log_2(n) \rfloor - 1 < T \leq \lfloor \log_2(n) \rfloor + 1 \quad (2.2.11)$$

и, следовательно, трудоемкость имеет вид $C \asymp \log_2 n$ при $n \rightarrow \infty$.

Тем самым по своим качествам алгоритм TID (при той же моделирующей функции и тех же затратах на память) разительно отличается от последовательного метода обратных функций. А именно,

1. трудоемкость C алгоритма **TID** практически не зависит от распределения \mathcal{P} при фиксированной мощности n носителя этого распределения;
2. трудоемкость C растет медленно (логарифмически) с ростом n , в то время как в алгоритме **TIS** рост средней трудоемкости — линейный.

2.2.4 Метод Чжень

Описанные выше варианты метода обратных функций являются вполне традиционными. В этом разделе мы опишем *метод Чжень*,³² гораздо менее известный, очень естественный, простой для реализации, и за счет увеличения требуемой памяти приводящий к неожиданно малой трудоемкости. Этот метод восходит к статье Н. С. Chen and Y. Asau [12].

Идея метода следующая. Рассмотрим таблицу распределения (2.1.1), вычислим накопленные вероятности s_i , выберем некоторое натуральное $m > 1$ и разобьем отрезок $(0, 1]$ на отрезки (*локализирующие окна*) вида $\Omega_j = ((j-1)/m, j/m]$, $j = 1, \dots, m$. Номер ν локализирующего окна, куда попадает случайная величина α , найти очень просто: $\nu = \lceil m\alpha \rceil$.

Накопленные вероятности s_i разбивают окно Ω_ν на несколько интервалов, и все, что нам осталось — это найти, в какой из этих интервалов попала случайная величина α . Таким образом, мы возвращаемся к исходной постановке задачи, но вместо отрезка $(0, 1]$ имеем дело со (случайным) отрезком длины $1/m$, который при больших m разбит, вообще говоря, на гораздо меньшее число отрезков, чем исходный.

Прежде чем оформлять эту идею в виде алгоритма, проиллюстрируем ее на простом примере. Возьмем $n = 5$ и рассмотрим таблицу распределения (2.1.1) с вероятностями $p_1 = 1/6$, $p_2 = 1/4$, $p_3 = 5/12$ и $p_4 = p_5 = 1/12$. Накопленные вероятности равны $s_1 = 1/6$, $s_2 = 5/12$, $s_3 = 5/6$ и $s_4 = 11/12$ (как всегда, $s_0 = 0$ и $s_5 = 1$). Соответствующая моделирующая функция несколько условно изображена на Рис. 1.

Теперь выберем $m = 4$ и разобьем отрезок $(0, 1]$ на «окна» вида $((i-1)/4, i/4]$, $i = 1, \dots, 4$. (границы этих окон отмечены жирными точками). Случайная величина $\nu = \lceil 4\alpha \rceil$ задает нам номер окна, куда попала α . Если $\nu = 1$, то моделируемая случайная величина ξ может принять значение x_1 (при $\alpha \leq s_1$) или x_2 (при $\alpha > s_1$). Аналогично, если $\nu = 2$ то либо $\xi = x_2$ при $\alpha \leq s_2$ либо $\xi = x_3$ при выполнении противоположного неравенства. А если $\nu = 3$, то однозначно $\xi = x_3$. Таким образом, вычислив значение ν , мы, вообще говоря, значительно сужаем область нашего поиска. Осталось понять, как можно организовать этот новый поиск.

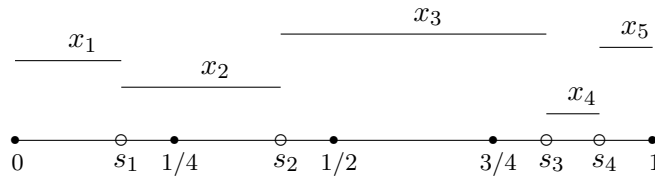


Рис. 1: Иллюстрация к методу Чжень

Вернемся к общему случаю и определим

$$r_j = \min\{i : s_i > (j-1)/m\} \text{ для } j = 1, \dots, m.$$

Кроме того, для удобства положим $r_{m+1} = n$. Смысл чисел r_j понятен: если промежуток $\Omega_j = ((j-1)/m, j/m]$ содержит какие-нибудь накопленные вероятности, то r_j — номер наименьшей из них, если же нет, то $r_j = r_{j+1}$.

³²В [10, Гл. III.2.4] употребляется термин *the method of guide tables*.

Иначе говоря, окно Ω_j содержит накопленные вероятности s_i с номерами $i = r_j, \dots, r_{j+1} - 1$, и знание чисел r_1, \dots, r_{m+1} дает нам полную информацию о распределении накопленных вероятностей s_i между различными окнами.

Так, в примере Рис. 1 мы видим, что $r_1 = 1$ (это всегда так по определению), $r_2 = 2$, $r_3 = r_4 = 3$ и $r_5 = 5$ (тоже по определению). Отсюда, например, мы видим, что окно $(0, 1/4]$ разбито накопленными вероятностями на $r_2 - r_1 + 1 = 2$ части, окно $(1/4, 1/2]$ — тоже на две части, в окне $(1/2, 3/4]$ накопленных вероятностей не содержится (так как $r_3 = r_4$), а последнее окно разбито на $r_5 - r_4 + 1 = 3$ части.

Зная номер локализирующего окна j и номер $i = r_j$ первой накопленной вероятности, превосходящей левую границу $(j-1)/m$ этого окна, остается определить, в какой из интервалов, разбивающих накопленными вероятностями окно Ω_j , попала случайная величина α . Алгоритмически проще всего это сделать последовательно, начиная с проверки неравенства $\alpha > s_i$ и затем (в случае надобности) увеличивая i на единицу. Так как $\alpha \leq j/n$, то число таких проверок не превзойдет $r_{j+1} - r_j + 1$.

Иначе говоря, введя обозначение $\nu = \lceil m\alpha \rceil$, мы будем использовать представление

$$\tau = \min_{i \geq r_\nu} \{i : \alpha \leq s_i\}$$

вместо более традиционного $\tau = \min_{i \geq 1} \{i : \alpha \leq s_i\}$.

Реализуем полученные соображения в виде алгоритма. Нахождение массива чисел r_j (как и вычисление накопленных вероятностей s_i), конечно, должны быть отнесены в подготовительные операции.

Алгоритм ChA (Cheng-Asau)

Моделирование конечного дискретного распределения последовательным методом Чжень

Предварительные операции: формирование массива накопленных вероятностей (s_1, \dots, s_n) и вспомогательного массива (r_1, \dots, r_m)

Входные данные: n, m , массив (p_1, \dots, p_n) .

Результат: массивы (s_1, \dots, s_n) и (r_1, \dots, r_m) .

1. (Инициализация) $i \leftarrow 1$; $s_1 \leftarrow p_1$;
2. (Подсчет s_i) For $k = 1$ to m do $(s_k \leftarrow s_{k-1} + p_k)$;
3. (Поиск окна и заполнение массива) For $j = 1$ to m do
 - (Первое s_i , большее $(j-1)/m$) While $ms_i \leq j-1$ do $i \leftarrow i+1$;
 - (Заполнение массива) $r_j \leftarrow i$;
3. (Завершение) STOP.

Моделирование

Входные данные: m , массивы (x_1, \dots, x_n) , (s_1, \dots, s_n) и (r_1, \dots, r_m) .

Результат: ξ .

1. (Инициализация) Get(α); $j \leftarrow \lceil m * \alpha \rceil$; $i \leftarrow r_j$;
2. (Поиск интервала) While $\alpha > s_i$ do $i \leftarrow i+1$;
3. (Завершение) $\xi \leftarrow x_i$; STOP.

Моделирующая часть не требует комментариев. В предварительной части первые два пункта посвящены вычислению накопленных вероятностей, а остальные — заполнению массива r_j .

Обратимся теперь к характеристикам моделирующей части алгоритма. Требуемая память, очевидно, равна $2n + m + O(1)$. Что касается трудоемкости метода (здесь, как и в методе **TIS**, она определяется средним числом T проверок неравенства $\alpha > s_i$), то ее можно оценить сверху следующим образом.

Вероятность выбрать локализирующее окно $((j-1)/m, j/m]$ равна $1/m$. Накопленные вероятности s_i делят это окно на $r_{j+1} - r_j + 1$ частей, что является оценкой сверху для числа проверок неравенств $\alpha \leq s_i$ последовательным перебором i начиная с $i_0 = r_j$. Поэтому

$$T \leq \frac{1}{m} \sum_{j=1}^m (r_{j+1} - r_j + 1) = 1 + \frac{1}{m} \sum_{j=1}^m (r_{j+1} - r_j) = 1 + (r_{m+1} - r_1)/m = 1 + (n - 1)/m. \quad (2.2.12)$$

Таким образом, метод Чжень, по-прежнему являющийся одним из вариантов метода обратных функций, демонстрирует удивительные свойства: путем увеличения требуемой памяти (это регулируется параметром m алгоритма) он позволяет величине T стать сколь угодно близкой к единице!

Например, если мы выберем $m = n$, то память увеличится примерно в 1.5 раза, в то время как T будет ограничена числом 2 равномерно не только по всем распределениям с одинаковой мощностью носителя (это мы наблюдали и в методе дихотомии), но и по n .

Кратко отметим возможные модификации метода Чжень.

1. Если в предварительную часть алгоритма включить дополнительно упорядочивание по убыванию вероятностей, соответствующих каждому локализирующему окну $((j-1)/m, j/m]$, то (см. комментарии к методу **TISM**) вместо (2.2.12) мы получим неравенство

$$T \leq \frac{1}{2m} \sum_{j=1}^m (r_{j+1} - r_j + 2) = 1 + (n - 1)/2m. \quad (2.2.13)$$

Тем самым неравенства $T < 2$ можно гарантировать уже при $m < n/2$, то есть при меньших затратах памяти, чем в основном алгоритме **ChA**.

2. Посмотрим теперь, к чему приводит упорядочивание вероятностей p_i глобально (а не в каждом локализирующем окне отдельно).

Тогда вероятности, соответствующие каждому локализирующему окну, тоже будут упорядочены по убыванию, кроме, может быть, самой первой по порядку вероятности.

Пусть теперь q_1, \dots, q_k — положительные вероятности, в сумме составляющие 1, причем $q_2 \geq q_3 \geq \dots \geq q_k$. Обозначим $t_i = q_i/(1 - q_1)$. Тогда

$$\begin{aligned} q_1 + 2q_2 + \dots + kq_k &= 1 + q_2 + 2q_3 + \dots + (k-1)q_k = \\ &= 1 + (1 - q_1)(t_2 + 2t_3 + \dots + (k-1)t_k) \leq 1 + (1 - q_1)k/2 < 1 + k/2. \end{aligned} \quad (2.2.14)$$

Отсюда

$$T \leq \frac{1}{2m} \sum_{j=1}^m (r_{j+1} - r_j + 3) = 1.5 + (n - 1)/2m. \quad (2.2.15)$$

Если нас интересует, например, ограничение $T < 3$, то неравенство (2.2.15) гарантирует его при $m > n/3$, в то время как неравенство (2.2.12) — лишь при $m > n/2$.

При очень больших m предварительное упорядочивание вероятностей p_i эффекта не дает. Например, если $m > 1/\min(p_1, \dots, p_n)$, то, очевидно, $r_{j+1} - r_j + 1 \in \{1, 2\}$, и упорядочивание теряет смысл.³³

³³Формально это отражено в неравенствах (2.2.14) и (2.2.15): если $k = 1$, то неравенство (2.2.14) становится очень неточным, слагаемое 1.5 в правой части (2.2.15) отражает именно этот эффект. А именно, если n фиксировано, а

3. Перейдем теперь к тому варианту алгоритма **ChA**, когда поиск внутри каждого локализирующего окна производится методом дихотомии. Согласно (2.2.11), поиск внутри окна $((j-1)/m, j/m]$ заканчивается не более чем за $1 + \log_2(r_{j+1} - r_j + 1)$ шагов. Поэтому

$$\begin{aligned} T &\leq \frac{1}{m} \sum_{j=1}^m (1 + \log_2(r_{j+1} - r_j + 1)) = 1 + \frac{1}{m} \sum_{j=1}^m (\log_2(r_{j+1} - r_j + 1)) = \\ &= 1 + \frac{1}{m} \log_2 \left(\prod_{j=1}^m (r_{j+1} - r_j + 1) \right). \end{aligned}$$

Так как максимум произведения $\prod_{j=1}^m x_j$ на множестве $x_j > 0, x_1 + \dots + x_m = k$ достигается при $x_j = k/m$ и равен $(k/m)^m$, то

$$T \leq 1 + \frac{1}{m} \log_2 \left(\frac{m+n}{m} \right)^m = 1 + \log_2(1 + n/m).$$

2.3 О методе Уолкера

Возьмем $m \geq 1/\min(p_1, \dots, p_n)$ в методе **ChA**. Тогда, очевидно, каждое окно Ω_j будет разбиваться накопленными вероятностями s_i не более чем на 2 интервала. Следовательно, дискретное конечное распределение (2.1.1), обладающее носителем X мощности n , может быть представлено в виде равновероятной смеси m распределений $\mathcal{P}^{(i)}$, сосредоточенных не более чем в 2-х точках, если только число $m = m(\mathcal{P})$ достаточно большое.

Идея метода Уолкера³⁴ [13] состоит том, что это утверждение остается верным и при $m = n$, если только распределения $\mathcal{P}^{(i)}$ выбраны соответствующим образом. Конечно, такой факт очевиден при $p_i = 1/n$. Он также ясен при $n = 2$: если $p_1 > p_2$, то распределение $\mathcal{P}^{(1)}$ сосредоточено в точке x_1 , а распределение $\mathcal{P}^{(2)}$ имеет вид

$$\mathcal{P}^{(2)} : \begin{pmatrix} x_1 & x_2 \\ 1 - 2p_2 & 2p_2 \end{pmatrix}$$

Докажем соответствующее утверждение в общем виде. Рассмотрим распределение \mathcal{P} , определенное в (2.1.1).

Предложение 1. Распределение \mathcal{P} может быть представлено в виде

$$\mathcal{P} = \sum_{i=1}^n n^{-1} \mathcal{P}^{(i)}, \quad (2.3.1)$$

где распределения $\mathcal{P}^{(i)}$ определяются таблицами

$$\mathcal{P}^{(i)} : \begin{pmatrix} a_i & b_i \\ \pi_i & 1 - \pi_i \end{pmatrix}$$

с $a_i, b_i \in X = \{x_1, \dots, x_n\}$ и $\pi_i \in [0, 1]$.

$m \rightarrow \infty$, то правая часть (2.2.15) стремится к 1.5, а не к 1. Поскольку неравенство (2.2.12) не зависит от порядка p_i в таблице распределения \mathcal{P} , то оценка трудоемкости алгоритма **ChA** при предварительном упорядочении вероятностей p_i по убыванию приобретает вид

$$T < \min(1 + (n-1)/m, 1.5 + (n-1)/2m).$$

³⁴В англоязычной литературе — *alias method*.

Доказательство. Доказательство проводится по индукции. При $n = 1, 2$ утверждение верно. Проведем индукционный переход $n - 1 \mapsto n$. Не уменьшая общности, будем считать, что вероятности p_1, \dots, p_n упорядочены по убыванию. Тогда $p_1 \geq 1/n$ и $p_n \leq 1/n$. Если $p_1 = p_n = 1/n$, то распределение \mathcal{P} — равномерное на множестве X , и все доказано. Иначе $p_1 > 1/n$ и $p_n < 1/n$. Тем самым мы можем представить \mathcal{P} в виде

$$\begin{aligned} \mathcal{P} &= \mathcal{P}_n = \frac{1}{n} \begin{pmatrix} x_n & x_1 \\ np_n & 1 - np_n \end{pmatrix} + \frac{n-1}{n} \begin{pmatrix} x_1 & x_2 & \dots & x_{n-1} \\ p_1^* & p_2^* & \dots & p_{n-1}^* \end{pmatrix} = \\ &= \frac{1}{n} \begin{pmatrix} x_n & x_1 \\ np_n & 1 - np_n \end{pmatrix} + \frac{n-1}{n} \mathcal{P}_{n-1}, \end{aligned} \quad (2.3.2)$$

где $p_1^* = (np_1 + np_n - 1)/(n-1)$ и $p_i^* = p_i n/(n-1)$ при $i \geq 2$, причем $\sum_{i=1}^{n-1} p_i^* = 1$ и $p_i > 0$ при $i = 1, \dots, n-1$.

Применяя индукционное предположение, получаем требуемое. \square

Имея представление (2.3.1), несложно представить себе моделирующую часть алгоритма Уолкера. Входными данными будут число n и 3 массива длины n : $A = (a_1, \dots, a_n)$, $B = (b_1, \dots, b_n)$, а также массив вероятностей $\Pi = (\pi_1, \dots, \pi_n)$. При этом, вообще говоря, массивы A и B нужны вне зависимости от того хранятся ли числа x_i в массиве или вычисляются по мере надобности.

Сама моделирующая часть выглядит следующим образом. Генерируя две равномерно распределенные на $(0, 1]$ случайные величины α_1, α_2 , полагаем $\tau = \lceil n\alpha_1 \rceil$. Если $\alpha_2 < \pi_\tau$, то полагаем $\xi = a_\tau$, а иначе $\xi = b_\tau$.³⁵

Предварительная часть (если не заботиться о ее трудоемкости) может быть организована по-разному. Например, можно следовать доказательству Предложения 1: сначала найти максимальную p_i и минимальную p_j из вероятностей p_1, \dots, p_n , а потом перейти от распределения $\mathcal{P} = \mathcal{P}_n$ к распределению \mathcal{P}_{n-1} , выкинув из таблицы (2.1.1) j -й столбец, пересчитав p_i как $p_i \leftarrow (np_i + np_j - 1)/(n-1)$ и преобразовав остальные вероятности p_k по правилу $p_k \leftarrow p_k n/(n-1)$. Одновременно осуществляются присвоения $a_1 = x_j$, $b_1 = x_i$ и $\pi_1 = np_j$.

Затем нужно сделать такую же операцию с распределением \mathcal{P}_{n-1} и так далее. При больших n это, конечно, очень трудоемкий метод.

Существуют гораздо более быстрые (и более трудные для понимания) способы формирования массивов A , B и Π , восходящие к [14] (см. также [10, Гл. III.4], [15] или [7]). Мы их, однако, касаться здесь не будем.³⁶

2.4 Еще о методе обратных функций

Рассмотрим распределение \mathcal{P} , задаваемое таблицей

$$\mathcal{P}: \begin{pmatrix} 1 & \dots & k & \dots & n \\ p_1 & \dots & p_k & \dots & p_n \end{pmatrix}. \quad (2.4.1)$$

По определению, метод обратных функций для моделирования этого распределения сводится к вычислению случайной величины $\tau = \min\{i : \alpha \leq s_i\}$, где $s_i = p_1 + \dots + p_i$. Тогда $\mathcal{L}(\tau) = \mathcal{P}$.

Ясно, что такая операция использует далеко не всю «случайность», содержащуюся в α . Например, если $n = 2$ и $p_1 = p_2 = 1/2$, то для получения τ достаточно знать лишь старший двоичный бит случайной величины α , остальные биты в моделировании не используются. Следующее утверждение дает теоретическую возможность исправить этот недостаток.

³⁵Поскольку $\lceil n\alpha_1 \rceil$ и $\{n\alpha_1\}$ независимы, а $\{n\alpha_1\} \in U(0, 1)$, то вместо α_2 можно использовать случайную величину $n\alpha_1 - \tau + 1$.

³⁶Не только потому, что это заняло бы много места. Просто мы условились пренебрегать трудоемкостью предварительной части алгоритма.

Предложение 2. Пусть $\alpha \in U(0, 1)$. Рассмотрим положительные вероятности p_1, \dots, p_i, \dots , в сумме дающие единицу, и положим $s_i = p_1 + \dots + p_i$ с $s_0 = 0$. Пусть $\tau = \min\{i : \alpha \leq s_i\}$. Тогда

1. случайная величина $\beta = (\alpha - s_{\tau-1})/p_\tau$ равномерно распределена на $(0, 1]$;
2. случайные величины β и τ независимы.

Доказательство. Прежде всего, $\mathbb{P}(\tau = i) = p_i$ при $i \geq 1$. Далее, при $x \in (0, 1]$

$$\begin{aligned} \mathbb{P}(\beta < x, \tau = i) &= \mathbb{P}((\alpha - s_{i-1})/p_i < x, s_{i-1} < \alpha \leq s_i) = \\ &= \mathbb{P}(\alpha < s_{i-1} + p_i x, s_{i-1} < \alpha \leq s_i) = \mathbb{P}(s_{i-1} < \alpha < s_{i-1} + p_i x) = p_i x, \end{aligned}$$

откуда немедленно следует требуемое. □

Замечание 2. 1. Если $p_i = 1/n$ при всех $i = 1, \dots, n$, то $(\alpha - s_{\tau-1})/p_\tau = \{n\alpha\}$ и мы приходим к уже использовавшемуся результату о том, что случайные величины $\lceil n\alpha \rceil$ и $\{n\alpha\}$ независимы, причем $\{n\alpha\} \in U(0, 1)$.

2. Результат Предложения 2 говорит о том, что во всех случаях, когда некоторое дискретное распределение моделируется методом обратных функций, мы можем сэкономить одно обращение к генератору, заменив его вычислением случайной величины β .³⁷

3. Можно заметить, что при доказательстве Предложения 2 нигде не используется конечность числа вероятностей p_i . Тем самым доказанный результат выходит за рамки табличных методов обратных функций.

³⁷Если среди вероятностей p_i встречаются очень маленькие числа, то на практике к такому приему нужно подходить с осторожностью.

3 Моделирование дискретных распределений, зависящих от параметров

В этом разделе мы рассматриваем только наиболее употребительные методы моделирования наиболее употребительных дискретных распределений.

3.1 Моделирование биномиального распределения

Биномиальное распределение $\text{Bin}(n, p)$ с параметрами $n \in \mathbb{N}$ и $p \in (0, 1)$ можно задать с помощью таблицы распределения

$$\mathcal{P} : \begin{pmatrix} 0 & 1 & \dots & k & \dots & n \\ p_0 & p_1 & \dots & p_k & \dots & p_n \end{pmatrix},$$

где $p_k = C_n^k p^k (1-p)^{n-k}$.

Один из методов моделирования распределения $\text{Bin}(n, p)$ уже рассматривался в разделе 1 (см. Алгоритм **ВВ** на стр. 8). Этот метод был основан на том, что распределение $\text{Bin}(n, p)$ — это распределение числа успехов в n испытаниях Бернулли с вероятностью успеха p .

Другой популярный метод является просто последовательным методом обратных функций, но в нетабличном варианте. А именно, накопленные вероятности s_k здесь вычисляются не заранее, а в том же цикле, где проверяются неравенства $\alpha \leq s_k$. Точнее, вероятности p_k рекуррентно пересчитываются одна через другую, а накопленные вероятности s_k последовательно вычисляются через s_{k-1} и p_k .

Поскольку для биномиального распределения $p_0 = (1-p)^n$ и

$$\frac{p_k}{p_{k-1}} = \frac{n-k+1}{k} \frac{p}{1-p}$$

при $k = 1, \dots, n$, то мы приходим к следующему алгоритму.

Алгоритм **BIS** (Binomial Inverse Sequential)

Моделирование распределения $\text{Bin}(n, p)$ последовательным методом обратных функций

Входные данные: n, p .

Результат: ξ .

1. (Инициализация) $c \leftarrow p/(1-p)$; $s \leftarrow r \leftarrow (1-p) \wedge n$; $k \leftarrow 0$; Get(α);
2. (Пересчет вероятностей и поиск окна) While $\alpha > s$ do
 - $k \leftarrow k+1$; $r \leftarrow r * c * (n-k+1)/k$; $s \leftarrow s+r$;
3. (Завершение) $\xi \leftarrow k$; STOP.

Дадим краткие комментарии к алгоритму. Переменная s обозначает текущую накопленную вероятность s_k , а переменная r — текущую вероятность p_k . Поскольку первая накопленная вероятность равна p_0 , то инициализация переменных s и r одинакова. Переменная c заведена для того, чтобы не делить p на $1-p$ в цикле.

Трудоёмкость метода вычисляется точно также как и в табличном методе обратных функций.³⁸ А именно,

$$C \asymp \sum_{k=0}^n (k+1)p_k = \mathbb{E}\xi + 1 = np + 1 \sim np$$

³⁸С учетом того, что нумерация вероятностей начинается с нуля.

при больших np . Таким образом, трудоемкость алгоритма **BIS** зависит от n и p , причем при фиксированном n наихудшим случаем является $p \approx 1$. Имеется простой способ улучшить этот наихудший случай.

Заметим, что если $\xi \in \text{Bin}(n, p)$, то $\eta = n - \xi \in \text{Bin}(n, 1 - p)$. Поэтому, если $p > 0.5$, то можно применить алгоритм **BIS** к моделированию числа неудач в n испытаниях Бернулли с вероятностью успеха p , а потом перейти к числу успехов. Конечно, это равносильно применению последовательного метода обратных функций «справа налево», а не «слева направо». Запишем соответствующий алгоритм.

Алгоритм **BISM** (Binomial Inverse Sequential Modified)

Моделирование $\text{Bin}(n, p)$ модифицированным последовательным методом обратных функций

Входные данные: n, p .

Результат: ξ .

1. (Инициализация)

- If $p \leq 0.5$ then $t \leftarrow p$ else $t \leftarrow 1 - p$;
- $c \leftarrow t/(1 - t)$; $s \leftarrow r \leftarrow (1 - t) \wedge n$; $k \leftarrow 0$; Get(α);

2. (Пересчет вероятностей и поиск окна) While $\alpha > s$ do

- $k \leftarrow k + 1$; $r \leftarrow r * c * (n - k + 1)/k$; $s \leftarrow s + r$;

3. (Завершение) If $p \leq 0.5$ then $\xi \leftarrow k$ else $\xi \leftarrow n - k$; STOP.

Комментария требует лишь введение переменной t , которая необходима при завершении алгоритма, чтобы (если нужно) перейти от моделирования числа неудач в испытаниях Бернулли к моделированию числа успехов. Трудоемкость алгоритма **BISM** имеет вид $C \asymp n(\min(p, 1 - p))$.

Таким образом, алгоритмы **BIS** и **BISM** становятся весьма трудоемкими при больших n .³⁹

Большие n приносят и еще одну неприятность в эти алгоритмы: операция возведения в большую степень $(1 - t)^n$ при реализации на компьютере может привести к исчезновению порядка. С этим, однако, бороться легко. Поскольку сумма независимых случайных величин $\xi_1 \in \text{Bin}(n_1; p)$ и $\xi_2 \in \text{Bin}(n_2; p)$ имеет распределение $\xi \in \text{Bin}(n_1 + n_2; p)$, то достаточно разложить большое n на сумму относительно небольших слагаемых n_i , промоделировать распределения $\text{Bin}(n_i; p)$ и сложить результаты. Общий вид трудоемкости при этом останется тем же.

3.2 Моделирование распределения Пуассона

Распределение Пуассона $\Pi(\lambda)$ с параметром $\lambda > 0$ описывается с помощью бесконечной таблицы распределения

$$\mathcal{P}: \begin{pmatrix} 0 & 1 & \dots & k & \dots \\ p_0 & p_1 & \dots & p_k & \dots \end{pmatrix},$$

где $p_k = \lambda^k e^{-\lambda}/k!$. Поскольку $p_k/p_{k-1} = \lambda/k$, а $p_0 = e^{-\lambda}$, то для распределения $\Pi(\lambda)$ легко реализуется последовательный нетабличный метод обратных функций.

Алгоритм **PIS** (Poisson Inverse Sequential)

Моделирование распределения $\Pi(\lambda)$ последовательным методом обратных функций

Входные данные: λ .

Результат: ξ .

³⁹Как и алгоритм **BB**. Некоторые методы моделирования биномиального распределения, трудоемкости которых остаются ограниченными с ростом n , можно найти в [10, гл. X.4]

1. (Инициализация) $s \leftarrow r \leftarrow \exp(-\lambda)$; $k \leftarrow 0$; Get(α);
2. (Пересчет вероятностей и поиск окна) While $\alpha > s$ do
 - $k \leftarrow k + 1$; $r \leftarrow r * \lambda/k$; $s \leftarrow s + r$;
3. (Завершение) $\xi \leftarrow k$; STOP.

Трудоёмкость снова легко считается:

$$C \asymp \sum_{k=0}^{\infty} (k+1)p_k = \mathbb{E}\xi + 1 = \lambda + 1.$$

Другой популярный (и на самом деле имеющий гораздо более глубокий смысл) алгоритм моделирования распределения $\Pi(\lambda)$ требует предварительного доказательства нескольких утверждений.

Лемма 2. Пусть η_1, \dots, η_n — независимые случайные величины, имеющие показательное распределение с параметром μ . Обозначим $S_n = \eta_1 + \dots + \eta_n$. Тогда случайная величина S_n имеет функцию распределения $G_{n,\mu}$ и плотность распределения $g_{n,\mu}$, задаваемые при $x > 0$ формулами

$$G_{n,\mu}(x) = 1 - \sum_{j=0}^{n-1} \frac{(\mu x)^j}{j!} e^{-\mu x}, \quad g_{n,\mu}(x) = \frac{\mu^n x^{n-1}}{(n-1)!} e^{-\mu x}. \quad (3.2.1)$$

При $x \leq 0$ обе функции равны нулю.⁴⁰

Доказательство. Прежде всего заметим, что непрерывная функция $G_{n,\mu}$ является функцией распределения: она монотонно возрастает, равна нулю на отрицательной полуоси и стремится к 1 при $x \rightarrow \infty$.

Непосредственным дифференцированием убеждаемся, что $G'_{n,\mu}(x) = g_{n,\mu}(x)$ при $x > 0$. Действительно, при $x > 0$

$$G'_{n,\mu}(x) = - \sum_{j=1}^{n-1} \frac{\mu^j x^{j-1}}{(j-1)!} e^{-\mu x} + \sum_{j=0}^{n-1} \frac{\mu^{j+1} x^j}{j!} e^{-\mu x} = \frac{\mu^n x^{n-1}}{(n-1)!} e^{-\mu x}.$$

Тем самым осталось доказать, что S_n имеет плотность распределения $g_{n,\mu}$. Это легко проверяется по индукции. При $n = 1$, очевидно, $g_{1,\mu}(x) = \mu e^{-\mu x}$, $x > 0$. Так что база индукции доказана.

Для индукционного перехода $n-1 \rightarrow n$ представим S_n в виде $S_{n-1} + \eta_n$ и применим формулу свертки для плотностей распределений. Получим, что при $x > 0$

$$\begin{aligned} g_{n,\mu}(x) &= \int_{-\infty}^{\infty} g_{n-1,\mu}(x-y) g_{1,\mu}(y) dy = \int_0^x \frac{\mu^{n-1} (x-y)^{n-2}}{(n-2)!} e^{-\mu(x-y)} \mu e^{-\mu y} dy = \\ &= \frac{\mu^n}{(n-2)!} e^{-\mu x} \int_0^x (x-y)^{n-2} dy = \frac{\mu^n}{(n-2)!} e^{-\mu x} \frac{x^{n-1}}{n-1}. \end{aligned}$$

Утверждение доказано. □

Предложение 3. Пусть $t > 0$. Рассмотрим независимые случайные величины $\{\eta_j\}_{j \geq 1}$, имеющие показательное распределение с параметром μ . Как и в Лемме 2, положим $S_n = \eta_1 + \dots + \eta_n$. Наконец, обозначим $\tau = \max\{n \geq 0 : S_n < t\}$, полагая $S_0 = 0$. Тогда $\mathcal{L}(\tau) = \Pi(\mu t)$.

Доказательство. Найдем вероятность того, что $\tau = n$. При $n = 0$

$$\mathbb{P}(\tau = 0) = \mathbb{P}(\eta_1 \geq t) = e^{-\mu t}.$$

Рассмотрим $n > 0$. Поскольку $\eta_i \geq 0$, то, согласно Лемме 2,

$$\begin{aligned} \mathbb{P}(\tau = n) &= \mathbb{P}(S_n < t, S_{n+1} \geq t) = \mathbb{P}(S_n < t) - \mathbb{P}(S_n < t, S_{n+1} < t) = \\ &= \mathbb{P}(S_n < t) - \mathbb{P}(S_{n+1} < t) = G_{n,\mu}(t) - G_{n+1,\mu}(t) = \frac{(\mu t)^n}{n!} e^{-\mu t}, \end{aligned}$$

что и требовалось доказать. □

⁴⁰На самом деле каждое из равенств (3.2.1) задает гамма-распределение $\text{Gamma}(n, \mu)$, см. Предложение 16 раздела 7.2.3 Приложения 1.

Следствие 1. Пусть $\alpha_1, \dots, \alpha_n, \dots$ — последовательность независимых равномерно распределенных на $(0, 1]$ случайных величин. Положим при $\lambda > 0$

$$\tau = \max \left\{ n \geq 0 : \prod_{i=1}^n \alpha_i > e^{-\lambda} \right\}. \quad (3.2.2)$$

Тогда $\tau \in \Pi(\lambda)$.

Доказательство. Так как⁴¹ $\ln(1/\alpha) \in \text{EXP}(1)$ при $\alpha \in U(0, 1)$ и

$$\tau = \max \left\{ n \geq 0 : \sum_{i=1}^n \ln \alpha_i > -\lambda \right\} = \max \left\{ n \geq 0 : \sum_{i=1}^n -\ln \alpha_i < \lambda \right\},$$

то ссылка на Предложение 3 завершает доказательство. \square

Пользуясь Следствием 1, легко записать соответствующий алгоритм.

Алгоритм РЕХР (Poisson Exponential)

Моделирование распределения $\Pi(\lambda)$ через случайные величины с показательным распределением

Входные данные: $\lambda > 0$.

Результат: ξ .

1. (*Инициализация*) $i \leftarrow 0$; $p \leftarrow \exp(-\lambda)$; $\text{Get}(\alpha)$; $q \leftarrow \alpha$;
2. (*Пересчет вероятностей и поиск окна*) While $q \geq p$ do
 - $\text{Get}(\alpha)$; $q \leftarrow q * \alpha$; $i \leftarrow i + 1$
3. (*Завершение*) $\xi \leftarrow i$; STOP.

Легко видеть, что при осуществлении этого алгоритма ровно $\xi + 1$ раз происходит обращение к генератору случайных чисел. Поэтому трудоемкость алгоритма имеет вид $C \asymp \lambda$, то есть он (как и алгоритм **PIS**) малопригоден при больших λ .⁴²

Кроме того, в инициализации обоих алгоритмов присутствует оператор $p \leftarrow \exp(-\lambda)$, который при больших λ может приводить к исчезновению порядка. Так как сумма независимых случайных величин $\xi_1 \in \Pi(\lambda_1)$ и $\xi_2 \in \Pi(\lambda_2)$ имеет распределение $\Pi(\lambda_1 + \lambda_2)$, то, разбив λ на несколько не очень больших слагаемых λ_i , используя один из алгоритмов **PIS** или **РЕХР** для каждого из λ_i сложив полученные результаты ξ_i , мы получим $\xi = \sum_i \xi_i \in \Pi(\lambda)$.

Как уже не раз говорилось, все методы моделирования случайных величин предназначены для работы в циклах. До рассмотрения Алгоритма **РЕХР** не возникало сомнения, что результаты работы каждого из таких циклов будут независимы — просто потому, что рассмотренные методы моделирования содержали детерминированное число обращений к генератору псевдослучайных чисел. Например, метод обратных функций использует ровно одно обращение к генератору, а алгоритм **ВВ** — n обращений.

С алгоритмом **РЕХР** ситуация другая. Представим себе, что мы 2 раза подряд обращаемся к работе этого алгоритма и получаем случайные величины ξ_1 и ξ_2 . При первом обращении нам для этого понадобились случайные величины $\alpha_1, \dots, \alpha_{\xi_1+1}$, а при втором — $\alpha_{\xi_1+2}, \dots, \alpha_{\xi_1+\xi_2+2}$. То есть случайная величина ξ_1 присутствует в записи не только первой группы используемых равномерно распределенных случайных величин (это естественно), но и второй тоже. Почему же тогда ξ_1 и ξ_2 независимы?⁴³

Имея в виду не только Алгоритм **РЕХР**, но и другие похожие случаи, мы докажем общее утверждение, устранившее эти сомнения.

⁴¹Т.к. $\mathbb{P}(-\ln \alpha > x) = \mathbb{P}(\alpha < e^{-x}) = e^{-x}$ при $x > 0$, см. также раздел 4.1.1.

⁴²Алгоритмы, адаптированные к большому λ , можно найти в [10, гл. X.3].

⁴³Более того, возникает сомнение в том, что случайная величина ξ_2 имеет то же распределение, что и ξ_1 .

Пусть $\beta_1, \dots, \beta_n, \dots$ — независимые одинаково распределенные случайные величины, принимающие значения в измеримом пространстве (H, \mathcal{H}) . Рассмотрим \mathcal{H}^k -измеримые подмножества D_k множеств H^k и положим $\tau_0 = 0$, а при $n \geq 1$

$$\tau_n = \tau_{n-1} + \begin{cases} \min \{ \ell \geq 1 : (\beta_{\tau_{n-1}+1}, \dots, \beta_{\tau_{n-1}+\ell}) \in D_\ell \} & \text{при } \tau_{n-1} < \infty, \\ +\infty & \text{при } \tau_{n-1} = \infty. \end{cases}$$

Предложение 4. Если $\mathbb{P}(\tau_1 < \infty) = 1$, то для любого $n > 1$

1. случайные величины $\tau_1, \tau_2 - \tau_1, \dots, \tau_n - \tau_{n-1}$ независимы и одинаково распределены;
2. случайные величины $\beta_{\tau_1}, \beta_{\tau_2}, \dots, \beta_{\tau_n}$ независимы и одинаково распределены.

Доказательство. Для простоты записи будем доказывать эти факты для $n = 2$. Переход на общий случай не представляет труда. Обозначим при $i \leq j$ события

$$A_{i,j} = \left(\bigcup_{\ell=i}^{j-1} \{ (\beta_i, \dots, \beta_\ell) \notin D_{\ell-i+1} \} \right) \cup \{ (\beta_i, \dots, \beta_j) \in D_{j-i+1} \},$$

которые измеримы относительно σ -алгебр $\sigma(\beta_i, \dots, \beta_j)$. Заметим, что события A_{i_1, j_1} и A_{i_2, j_2} независимы при $i_2 > j_1$, причем $\mathbb{P}(A_{i,j}) = \mathbb{P}(A_{i+h, j+h})$ для любого $h > 0$.

Введем ограниченные \mathcal{H} -измеримые функции $f_1, f_2 : H \mapsto \mathbb{R}$. Тогда при $k < m$

$$\begin{aligned} \mathbb{E}(f_1(\beta_{\tau_1})f_2(\beta_{\tau_2}), \tau_1 = k, \tau_2 = m) &= \mathbb{E}(f_1(\beta_k)f_2(\beta_m), \tau_1 = k, \tau_2 = m) = \\ &= \mathbb{E}(f_1(\beta_k)f_2(\beta_m), A_{1,k}, A_{k+1,m}) = \mathbb{E}(f_1(\beta_k), A_{1,k}) \mathbb{E}(f_2(\beta_m), A_{k+1,m}) = \\ &= \mathbb{E}(f_1(\beta_k), A_{1,k}) \mathbb{E}(f_2(\beta_{m-k}), A_{1,m-k}) = \mathbb{E}(f_1(\beta_k), \tau_1 = k) \mathbb{E}(f_2(\beta_{m-k}), \tau_1 = m - k). \end{aligned}$$

Положив в последнем равенстве $f_1 = f_2 \equiv 1$ и $\ell = m - k$, получим, что

$$\mathbb{P}(\tau_1 = k, \tau_2 - \tau_1 = \ell) = \mathbb{P}(\tau_1 = k) \mathbb{P}(\tau_1 = \ell).$$

Далее,

$$\begin{aligned} \mathbb{E}(f_1(\beta_{\tau_1})f_2(\beta_{\tau_2})) &= \sum_{m>k \geq 1} \mathbb{E}(f_1(\beta_{\tau_1})f_2(\beta_{\tau_2}), \tau_1 = k, \tau_2 = m) = \\ &= \sum_{k \geq 1} \sum_{m-k \geq 1} \mathbb{E}(f_1(\beta_k), \tau_1 = k) \mathbb{E}(f_2(\beta_{m-k}), \tau_1 = m - k) = \\ &= \sum_{k \geq 1} \mathbb{E}(f_1(\beta_k), \tau_1 = k) \sum_{\ell \geq 1} \mathbb{E}(f_2(\beta_\ell), \tau_1 = \ell) = \mathbb{E}f_1(\beta_{\tau_1}) \mathbb{E}f_2(\beta_{\tau_1}). \end{aligned}$$

Утверждение доказано. \square

Для того, чтобы применить Предложение 4 к алгоритму **РЕХР**, нужно положить $H = (0, 1]$, $D_k = \{(x_1, \dots, x_k) \in H^k : \prod_{i=1}^k x_i \leq e^{-\lambda}\}$ и $\beta_i = \alpha_i$. Тогда случайная величина τ , определенная в (3.2.2), будет связана с τ_1 равенством $\tau_1 = \tau + 1$. Первое утверждение Предложения 4 показывает, что Алгоритм **РЕХР** может быть использован в цикле.

3.3 Моделирование геометрического распределения

Геометрическое распределение $\text{Geom}(p)$ с параметром $p \in (0, 1)$ описывается с помощью бесконечной таблицы распределения

$$\mathcal{P} : \begin{pmatrix} 0 & 1 & \dots & k & \dots \\ p_0 & p_1 & \dots & p_k & \dots \end{pmatrix},$$

где $p_k = p(1-p)^k$.

Два способа моделирования геометрического распределения достаточно ясны и не требуют особого обоснования. Первый из них — моделирование испытаний Бернулли с вероятностью успеха p

до первого успеха с подсчетом числа неудач,⁴⁴ а второй — нетабличный последовательный метод обратных функций, основанный на пересчете $p_n = (1 - p)p_{n-1}$ с $p_0 = p$.

Запишем соответствующие алгоритмы.

Алгоритм GeomB (Geometric Bernoulli)

Моделирование геометрического распределения $\text{Geom}(p)$ через испытания Бернулли

Входные данные: p .

Результат: ξ .

1. (Инициализация) $i \leftarrow k \leftarrow 0$;
2. (Моделирование числа неудач) While $i = 0$ do (Get(α); If $\alpha < p$ then $i \leftarrow 1$ else $k \leftarrow k + 1$);
3. (Завершение) $\xi \leftarrow k$; STOP.

Поскольку в алгоритме нам нужно $\xi + 1$ раз моделировать испытания Бернулли, то

$$C \asymp \mathbb{E}\xi + 1 = \frac{1-p}{p} + 1 = \frac{1}{p}.$$

Запишем теперь алгоритм метода обратных функций.

Алгоритм GeomIS (Geometric Inverse Sequential)

Моделирование распределения $\text{Geom}(p)$ последовательным методом обратных функций

Входные данные: p .

Результат: ξ .

1. (Инициализация) $s \leftarrow r \leftarrow p$; $k \leftarrow 0$; $q \leftarrow 1 - p$; Get(α);
2. (Пересчет вероятностей и поиск окна) While $\alpha > s$ do
 - $k \leftarrow k + 1$; $r \leftarrow r * q$; $s \leftarrow s + r$;
3. (Завершение) $\xi \leftarrow k$; STOP.

Как обычно в последовательном методе обратных функций,

$$C \asymp \sum_{i \geq 0} (i+1)p_i = \mathbb{E}\xi + 1 = 1/p.$$

Таким образом, оба метода оказываются плохими при малых p . Существует, однако, такой способ реализации метода обратных функций, при котором трудоемкость по крайней мере формально не зависит от p .

Действительно, накопленная вероятность $s_{n+1} = p_0 + \dots + p_n$ для геометрического распределения имеет вид

$$s_{n+1} = \sum_{i=0}^n p(1-p)^i = 1 - (1-p)^{n+1}.$$

Поэтому событие $\{\xi = n\}$ приобретает вид

$$\begin{aligned} \{\xi = n\} &= \{s_n < \alpha \leq s_{n+1}\} = \{1 - (1-p)^n < \alpha \leq 1 - (1-p)^{n+1}\} = \\ &= \{(1-p)^{n+1} \leq 1 - \alpha < (1-p)^n\} = \{(n+1) \ln(1-p) \leq \ln(1-\alpha) < n \ln(1-p)\} = \\ &= \left\{ n < \frac{\ln(1-\alpha)}{\ln(1-p)} \leq n+1 \right\}, \end{aligned}$$

⁴⁴Первое утверждение Предложения 4 может быть использована для этого метода, если положить $H = (0, 1]$, $D_k = [p, 1]^{k-1} \times (0, p)$ и $\beta_i = \alpha_i$.

и тем самым

$$\xi = \lfloor \ln(1 - \alpha) / \ln(1 - p) \rfloor. \quad (3.3.1)$$

Эту же формулу можно получить и по-другому. Пусть η — случайная величина, имеющая показательное распределение с параметром λ и $\xi = \lfloor \eta \rfloor$. Тогда при $n \geq 0$

$$\mathbb{P}(\xi = n) = \mathbb{P}(n \leq \eta < n + 1) = e^{-n\lambda} - e^{-(n+1)\lambda} = (1 - e^{-\lambda})e^{-n\lambda}.$$

Поскольку случайная величина $-\ln(1 - \alpha)/\lambda$ имеет показательное распределение с параметром λ , то взяв $\lambda = -\ln(1 - p)$, приходим к формуле (3.3.1).

При моделировании естественно заменить $1 - \alpha$ в формуле (3.3.1) на α . В результате приходим к следующему алгоритму.

Алгоритм GeomDIM (Geometric Direct Inverse Modified)

Моделирование распределения $\text{Geom}(p)$ модифицированным прямым методом обратных функций

Входные данные: p .

Результат: ξ .

1. (Моделирование) Get(α); $\xi \leftarrow \lfloor \ln(\alpha) / \ln(1 - p) \rfloor$; STOP.

Таким образом, для моделирования геометрического распределения методом **GeomDIM** требуется одно обращение к генератору псевдослучайных чисел, два вычисления логарифма и одно — целой части числа. Если считать, что время вычисления значений функций $\lfloor x \rfloor$ и $\ln(z)$ ограничено при всех x и $z \in (0, 1)$ одной и той же постоянной, то трудоемкость алгоритма имеет вид $O(1)$ равномерно по p .

По сравнению с предыдущими алгоритмами это выглядит немного странно, так как интуитивно ясно, что геометрическое распределение с маленьким параметром p должно быть, в принципе, сложным для моделирования.⁴⁵ Разберемся, в чем здесь дело.

На примере геометрического распределения имеет смысл поговорить о недостатках того классического подхода к понятию трудоемкости, которого мы придерживаемся (и будем придерживаться в дальнейшем).

Коротко говоря, дело в том, что за источник случайности у нас принято считать независимые равномерно распределенные на $(0, 1]$ случайные величины, которые по определению имеют счетное число двоичных разрядов. В то же время большинство генераторов псевдослучайных чисел вырабатывают последовательность чисел вида $a_i = k_i/M$, где M — большое целое число,⁴⁶ а $k_i \in X = \{0, \dots, M - 1\}$.

Это означает, что более адекватным источником случайности является последовательность $\{\varepsilon_i\}_{i \geq 1}$ независимых случайных величин, равномерно распределенных на множестве X . При $M = 2$ элементы теории моделирования (в основном, дискретных) случайных величин с таким источником случайности были разработаны в статье Кнута и Яо [11], некоторые результаты для произвольного M можно найти в разделе 6.

В этом разделе, в частности, доказано следующее утверждение. Пусть мы используем последовательность $\{\varepsilon_i\}_{i \geq 1}$ для моделирования дискретного распределения \mathcal{P} . Определим трудоемкость соответствующего алгоритма как среднее число использованных случайных величин ε_i ,⁴⁷ необходимое для получения одной случайной величины, имеющей распределение \mathcal{P} .

Тогда, во-первых, существует оптимальный (то есть имеющий минимальную трудоемкость $C_M(\mathcal{P})$) алгоритм, и во-вторых, эта минимальная трудоемкость удовлетворяет неравенству

$$\frac{\mathcal{H}_2(\mathcal{P})}{\log_2 M} \leq C_M(\mathcal{P}) < \frac{\mathcal{H}_2(\mathcal{P})}{\log_2 M} + \frac{M}{M - 1}, \quad (3.3.2)$$

где двоичная энтропия $\mathcal{H}_2(\mathcal{P})$ распределения \mathcal{P} с вероятностями p_1, \dots, p_n, \dots определяется формулой

$$\mathcal{H}_2(\mathcal{P}) = - \sum_i p_i \log_2 p_i.$$

⁴⁵ Чем меньше p , тем геометрическое распределение больше «размазано» по счетному множеству $\{0, 1, \dots\}$.

⁴⁶ Как правило, $M = 2^d$, так что числа a_i имеют конечное число двоичных разрядов, в то время как α_i — бесконечное число. Наиболее часто используется $d = 32$.

⁴⁷ То есть среднее число обращений к (идеальному) генератору псевдослучайных чисел.

Для геометрического распределения с параметром p , стремящемся к нулю, получаем, что

$$\begin{aligned}\mathcal{H}_2(\mathcal{P}) &= - \sum_{i=0}^{\infty} p(1-p)^i (\log_2 p + i \log_2(1-p)) = \\ &= \log_2(1/p) - \frac{1-p}{p} \log_2(1-p) \sim \log_2(1/p) \rightarrow \infty.\end{aligned}$$

Следовательно, если мы хотим промоделировать геометрическое распределение методом обратных функций⁴⁸, то при фиксированном M и при достаточно малом p мы будем иметь ощутимые потери в точности моделирования. При моделировании распределения с большой энтропией следует использовать методы, в которых задействовано несколько обращений к генератору псевдослучайных чисел.⁴⁹ Как уже говорилось, более подробно об этих проблемах написано в разделе 6.

⁴⁸Или любым другим методом, использующим одно обращение к генератору.

⁴⁹Это относится и к дискретному равномерному распределению, двоичная энтропия которого равна $\log_2 n$.

4 Общие методы моделирования

Не существует общепринятого списка методов моделирования, которые могли быть названы «общими» или «универсальными». Тем не менее, мы выделим 3 метода, идеи которых в той или иной степени (и в той или иной комбинации) используются в большинстве практических алгоритмов.

4.1 Метод обратных функций

Пусть \mathcal{P} — распределение, заданное на борелевских подмножествах \mathbb{R} . Обозначим F функцию распределения распределения \mathcal{P} , то есть положим $F(x) = \mathcal{P}((-\infty, x])$. Метод обратных функций для моделирования распределения \mathcal{P} основан на следующем утверждении.

Лемма 3. Определим функцию $G : (0, 1) \mapsto \mathbb{R}$ равенством

$$G(y) = \inf\{x \in \mathbb{R} : F(x) > y\}. \quad (4.1.1)$$

Тогда 1. при $y \in (0, 1)$ и $t \in \mathbb{R}$ неравенства $G(y) < t$ и $y < F(t)$ эквивалентны;
2. функция G является монотонно неубывающей и непрерывной справа.

Доказательство. 1. Обозначим $A_y = \{x \in \mathbb{R} : F(x) > y\}$ и докажем, что $A_y = (z, +\infty)$, где $z = G(y)$. Так как F монотонно неубывает, то $A_y = [z, +\infty)$, где левая скобка " $<$ " означает, что точка z может принадлежать или не принадлежать множеству A_y . Кроме того, $F(z) \leq y$. Действительно, пусть $F(z) > y$. Поскольку функция F непрерывна слева, то из этого неравенства следует, что $F(z^*) > y$ при некотором $z^* < z$, что противоречит (4.1.1).

Теперь результат первого пункта леммы становится очевидным, так как при фиксированном y каждое из неравенств $G(y) < t$ и $y < F(t)$ означает, что $t \in A_y$.

2. Монотонность функции G следует из включения $\{x : F(x) > y_1\} \supset \{x : F(x) > y_2\}$, верного при $y_1 < y_2$. Пусть функция G не является непрерывной справа в точке $y \in (0, 1)$. Это значит, что существует такая последовательность $y_n \downarrow y$ и такое число t , что $G(y_n) > t > G(y)$ при $n > n_0$.

Согласно первому утверждению леммы эти неравенства эквивалентны неравенствам $F(t) > y$ и $F(t) \leq y_n$. Поскольку $y_n \downarrow y$, получаем противоречие. \square

Следствие 2. В условиях Леммы 3 случайная величина $G(\alpha)$ имеет распределение \mathcal{P} .

Доказательство. Это следует из цепочки равенств $\mathbb{P}(G(\alpha) < x) = \mathbb{P}(\alpha < F(x)) = F(x)$. \square

Метод моделирования одномерного распределения \mathcal{P} с функцией распределения F при помощи моделирующей функции (4.1.1) называется *методом обратных функций*.

Опишем три характерные ситуации поведения функции G .⁵⁰ Пусть $x \in \mathbb{R}$ и $y \in (0, 1)$ связаны соотношением $F(x) = y$.

1. Если в окрестности точки x функция распределения F обратима, то $G(y) = x$.
2. Пусть $(x_1, x_2]$ — максимальный (и не пустой) интервал постоянства функции F , содержащий точку x . Тогда $G(y) = x_2$.
3. Если $y_1 \stackrel{\text{def}}{=} F(x + 0) > y$, то $G(z) = x$ при любом $z \in [y, y_1)$.

Есть два стандартных варианта использования метода обратных функций.

⁵⁰Убедитесь, что эти описания верны.

1. Обозначим $x_* = \inf\{x : F(x) > 0\}$ и $x^* = \sup\{x : F(x) < 1\}$ и предположим, что $F(x_*) = 0$ и $F(x^*) = 1$ (если $x_* = -\infty$ или $x^* = +\infty$, то соответствующие равенства считаются автоматически выполненными). Предположим, что сужение F_* функции F на промежуток (x_*, x^*) обратимо. Тогда при $y \in (0, 1)$

$$G(y) = \inf\{x : F(x) > y\} = \inf\{x : F_*(x) > y\} = \inf\{x : x > F_*^{-1}(y)\} = F_*^{-1}(y). \quad (4.1.2)$$

Равенство (4.1.2), конечно, и дало название методу.

2. Рассмотрим дискретное распределение, задаваемое таблицей (2.1.1), и предположим, что $x_i < x_j$ при $i < j$. Обозначим $s_i = p_1 + \dots + p_i$. Тогда функция распределения, задаваемая распределением \mathcal{P} , принимает значения $F(x) = 0$ при $x \leq x_1$, $F(x) = s_i$ при $x_i < x \leq x_{i+1}$ ($i = 1, \dots, n-1$) и $F(x) = 1$ при $x > x_n$.

Поэтому

$$G(y) = \inf\{x : F(x) > y\} = \begin{cases} x_1 & \text{при } 0 < y < s_1 \\ x_i & \text{при } s_i \leq y < s_{i+1}, \quad i = 1, \dots, n-1. \end{cases}$$

Это в точности соответствует определению (2.2.1) и объясняет термин «метод обратных функций», использовавшийся в разделах 2.2 и 3.⁵¹

Легко доказать следующий факт, проясняющий метод обратных функций несколько с другой стороны.

Предложение 5. Пусть \mathcal{P} — некоторое распределение с функцией распределения F . Предположим, что $H : (0, 1) \mapsto \mathbb{R}$ — монотонно неубывающая измеримая функция такая, что $\mathcal{L}(H(\alpha)) = \mathcal{P}$. Тогда $H(y) = G(y)$ почти всюду по мере Лебега на $(0, 1)$. Если к тому же H является непрерывной справа, то $H(y) = G(y)$ при всех $y \in (0, 1)$.

Доказательство. Возьмем $t \in \mathbb{R}$ и рассмотрим множество $A_t = \{y : H(y) < t\}$. Ввиду монотонности функции H это множество имеет вид $(0, a_t)$, где угловая скобка " $)$ " означает, что точка a_t может как принадлежать, так и не принадлежать множеству A_t . Поскольку $\mathbb{P}(H(\alpha) < t) = F(t)$, то $a_t = F(t)$.

Отсюда и из Леммы 3 следует, что для любого $t \in \mathbb{R}$

$$\{y : H(y) < t \leq G(y)\} = (0, F(t)) \cap [F(t), 1) \subset \{F(t)\}.$$

Таким образом, $\mathbb{P}(H(\alpha) < t \leq G(\alpha)) = 0$ для любого $t \in \mathbb{R}$. Поскольку

$$\mathbb{P}(H(\alpha) < G(\alpha)) = \mathbb{P}\left(\bigcup_{r \in \mathbb{R}_0} \{H(\alpha) < r \leq G(\alpha)\}\right) \leq \sum_{r \in \mathbb{R}_0} \mathbb{P}(H(\alpha) < r \leq G(\alpha)),$$

где \mathbb{R}_0 — множество рациональных чисел, то $\mathbb{P}(H(\alpha) < G(\alpha)) = 0$. Аналогично доказывается, что $\mathbb{P}(H(\alpha) > G(\alpha)) = 0$, и поэтому функции G и H совпадают почти всюду.

Если две функции совпадают почти всюду, то отличаться они могут только в точках разрыва одной из этих функций. Поскольку в точках разрыва функции G и H непрерывны справа, то этот вариант тоже исключается. \square

Замечание 3. Конечно, вместо моделирующей формулы $\xi = G(\alpha)$ можно использовать формулу $\xi = G(1 - \alpha)$. Возникающая при этом моделирующая функция $G_1(y) = G(1 - y)$ является монотонно невозрастающей и для нее верен аналог Предложения 5. Учитывая это соображение, моделирование распределения \mathcal{P} с помощью функции (4.1.1) (или с помощью функции G_1) можно назвать *монотонным моделированием*. Термин «метод обратных функций», тем не менее, является более традиционным.

⁵¹Отметим, что в разделе 2.2 этот термин используется безотносительно к тому, упорядочены ли числа x_i по возрастанию, или нет.

4.1.1 Примеры

Перейдем к примерам. Эти примеры относятся к нескольким стандартным распределениям и соответствуют преобразованию (4.1.2). При записи плотностей и функций распределения условимся, что формула для плотности приводится только для тех аргументов, для которых плотность не равна нулю, а формула для функции распределения — для тех аргументов, где эта функция отлична от нуля и единицы.

1. *Показательное распределение* EXP(λ) с параметром $\gamma > 0$ имеет плотность распределения

$$p(x) = \lambda e^{-\lambda x}, \quad x > 0.$$

Здесь $x_* = 0$, $x^* = +\infty$, $F_*(x) = 1 - e^{-\lambda x}$ при $x > 0$ и $G(y) = -\ln(1 - y)/\lambda$ при $y \in (0, 1)$. Таким образом, моделирование показательного распределения методом обратных функций осуществляется как $\xi \leftarrow -\ln(1 - \alpha)/\lambda$. Поскольку $\mathcal{L}(1 - \alpha) = U(0, 1)$, то на практике используют моделирующую формулу

$$\xi \leftarrow -\ln(\alpha)/\lambda.$$

2. *Распределение Рэлея* можно задать плотностью

$$p(x) = x e^{-x^2/2}, \quad x > 0. \quad (4.1.3)$$

Получаем, что $x_* = 0$, $x^* = +\infty$, $F_*(x) = 1 - e^{-x^2/2}$ при $x > 0$ и $G(y) = \sqrt{-2 \ln(1 - y)}$. Снова переходя от α к $1 - \alpha$, получаем традиционную моделирующую формулу

$$\xi \leftarrow \text{sqrt}(-2 * \ln(\alpha)). \quad (4.1.4)$$

3. *Степенное распределение* с параметром $\gamma > 0$ имеет плотность

$$p(x) = \gamma x^{\gamma-1}, \quad x \in (0, 1).$$

В этом случае $x_* = 0$, $x^* = 1$, $F_*(x) = x^\gamma$, $G(y) = y^{1/\gamma}$ и метод обратных функций приводит к моделирующей формуле

$$\xi \leftarrow \alpha \wedge (1/\gamma).$$

4. *Распределение Парето* с параметром $\gamma > 0$ задается плотностью

$$p(x) = \gamma x^{-\gamma-1}, \quad x > 1.$$

Здесь $x_* = 1$, $x^* = +\infty$, $F_*(x) = 1 - x^{-\gamma}$ и $G(y) = (1 - y)^{-1/\gamma}$. Как и раньше, заменяя α на $1 - \alpha$, приходим к моделирующей формуле

$$\xi \leftarrow \alpha \wedge (-1/\gamma).$$

Отсюда и из предыдущего примера можно заметить, что если ξ имеет степенное распределение с параметром γ , то $1/\xi$ имеет распределение Парето с тем же параметром.

5. *Распределение Коши* имеет плотность распределения

$$p(x) = \frac{1}{\pi} \frac{1}{1 + x^2}, \quad x \in \mathbb{R}.$$

В этом случае $x_* = -\infty$, $x^* = +\infty$,

$$F_*(x) = F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan(x)$$

при $x \in \mathbb{R}$ и $G(y) = \tan(\pi(y - 0.5)) = -\cot(\pi y) = \cot(\pi(1 - y))$. Тем самым (заменяя α на $1 - \alpha$), получаем моделирующую формулу

$$\xi \leftarrow \cot(\pi * \alpha).$$

Отметим, что случайная величина $\tan(\pi\alpha)$ также имеет распределение Коши,⁵² но соответствующая моделирующая функция в этом случае не является монотонной.

Подводя некоторые итоги, можно сказать, что метод обратных функций для моделирования распределения с обратимой функцией F_* будет достаточно эффективным, если существует точная и быстрая процедура вычисления обратной к этой функции.

4.2 Метод отбора

Метод отбора (в другой терминологии — метод исключения или метод фон Неймана, в англоязычной литературе — rejection method) является весьма распространенным, универсальным и (при грамотной реализации) эффективным методом моделирования распределений. Принятое здесь изложение метода отбора отличается от традиционного несколько большей общностью.

4.2.1 Идея метода

Проиллюстрируем идею метода отбора на простейшем примере, восходящем к Дж. фон Нейману. Рассмотрим плотность распределения $p(x)$, $x \in \mathbb{R}$, и обозначим H подграфик этой плотности: $H = \{(x, y) : 0 \leq y < p(x)\}$. Ясно, что двумерная мера Лебега mes_2 множества H равна 1. Введем случайный вектор $\bar{\xi} = (\xi_1, \xi_2)$, равномерно распределенный в множестве H и найдем распределение случайной величины ξ_1 . Получим, что

$$\mathbb{P}(\xi_1 < z) = \int_{\{(x,y) \in H: x < z\}} dx dy = \int_{-\infty}^z dx \int_0^{p(x)} dy = \int_{-\infty}^z p(x) dx.$$

Таким образом, случайная величина ξ_1 имеет плотность распределения p .

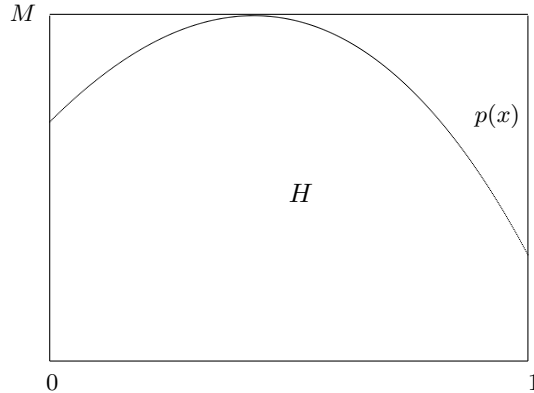


Рис. 2: Иллюстрация к идее метода отбора

⁵²Если случайная величина ξ имеет распределение Коши, то $\eta = 1/\xi$ тоже имеет распределение Коши.

Отсюда появляется идея моделирования распределения \mathcal{P} с плотностью распределения p : если мы сумеем получить реализацию случайного вектора, равномерно распределенного в подграфике плотности p , то первая координата этого вектора будет иметь нужное нам распределение.

Предположим теперь дополнительно, что плотность p сосредоточена на отрезке $[0, 1]$ и ограничена постоянной M (см. рис 2, где $M = \text{vraisup } p$). В этом случае имеется «очевидная»⁵³ процедура моделирования равномерного распределения в области H . А именно, моделируя равномерное распределение в прямоугольнике $[0, 1] \times [0, M]$, нужно проверить, лежит ли полученная точка под графиком плотности p или нет. Если лежит, то первая координата этой точки является результатом моделирования, если нет — процедуру нужно повторить.

Иначе говоря, получив 2 случайные величины α_1, α_2 , мы проверяем неравенство $p(\alpha_1) > M\alpha_2$ и повторяем эту процедуру до тех пор, пока это неравенство не выполнится.

Идея выглядит привлекательной, но возникают вопросы, что делать, если плотность p не является ограниченной и/или если она положительна на всей прямой (а почему обязательно на прямой?).

4.2.2 Теория метода

Мы начнем с необходимых нам сведений об абсолютной непрерывности мер и производных Радона-Никодима.

О производных Радона-Никодима. Пусть (D, \mathcal{D}) — некоторое измеримое пространство и ν, μ — меры, заданные на \mathcal{D} . Мера ν называется *абсолютно непрерывной* относительно меры μ , если для любого $A \in \mathcal{D}$ такого, что $\mu(A) = 0$, выполняется равенство $\nu(A) = 0$. Свойство абсолютной непрерывности записывается как $\nu \prec \mu$.

Имеет место следующий вариант теоремы Радона-Никодима (ее доказательство можно найти, например, в [16, гл. II.8.1]).

Теорема 1. Если μ — σ -конечная мера, то для того, чтобы конечная мера ν была абсолютно непрерывна относительно μ , необходимо и достаточно, чтобы существовала такая \mathcal{D} -измеримая функция r , что для любого $A \in \mathcal{D}$

$$\nu(A) = \int_A r d\mu. \quad (4.2.1)$$

При этом функция r единственна с точностью до множества μ -меры ноль.

Кроме того, для любой функции $f \in \mathbb{L}^1(d\nu)$

$$\int_D f d\nu = \int_D f r d\mu.$$

Функция r называется *производной Радона-Никодима* меры ν относительно меры μ и обозначается $r = d\nu/d\mu$. Иногда этот же факт записывается как $d\nu = r d\mu$.

Замечание 4. 1. Достаточность в Теореме 1 очевидна. Так же ясно, что замена в правой части равенства (4.2.1) функции r на другую функцию, совпадающую μ -почти всюду с r , не меняет левую часть. Суть Теоремы 1 состоит, конечно, в доказательстве необходимости.

2. На самом деле утверждение Теоремы 1 остается в силе, если ν является конечным зарядом. Именно в таком виде эта теорема используется в теории метода Монте-Карло.

Нам понадобится одно следствие из теоремы Радона-Никодима.

⁵³Кавычки появились из-за того, что эту процедуру все равно нужно обосновывать.

Следствие 3. Пусть (D, \mathcal{D}) — измеримое пространство. Рассмотрим распределения \mathcal{P} , \mathcal{Q} и σ -конечную меру μ , определенные на \mathcal{D} , и предположим, что $d\mathcal{P} = p d\mu$ и $d\mathcal{Q} = q d\mu$. Тогда имеют место следующие утверждения.

1. Распределение \mathcal{P} абсолютно непрерывно относительно \mathcal{Q} тогда и только тогда, когда

$$\mu\{x : p(x) \neq 0, q(x) = 0\} = 0. \quad (4.2.2)$$

2. При этом \mathcal{Q} -почти всюду

$$\frac{d\mathcal{P}}{d\mathcal{Q}}(x) = r(x) \stackrel{\text{def}}{=} \begin{cases} p(x)/q(x) & \text{при } q(x) \neq 0, \\ 0 & \text{при } q(x) = 0. \end{cases} \quad (4.2.3)$$

Доказательство. Не умаляя общности, будем считать, что функции p и q неотрицательны.

1. Предположим, что условие (4.2.2) выполнено. Тогда при любом $A \in \mathcal{D}$

$$\begin{aligned} \mathcal{P}(A) &= \int_A p(x) \mu(dx) = \int_{A \cap \{x: q(x) \neq 0\}} p(x) \mu(dx) + \int_{A \cap \{x: q(x) = 0\}} p(x) \mu(dx) = \\ &= \int_{A \cap \{x: q(x) \neq 0\}} \frac{p(x)}{q(x)} q(x) \mu(dx) + \int_{A \cap \{x: q(x) = 0, p(x) \neq 0\}} p(x) \mu(dx) = I_1 + I_2. \end{aligned}$$

Согласно (4.2.2), $I_2 = 0$. Поэтому

$$\begin{aligned} \mathcal{P}(A) &= I_1 = \int_{A \cap \{x: q(x) \neq 0\}} \frac{p(x)}{q(x)} \mathcal{Q}(dx) = \int_{A \cap \{x: q(x) \neq 0\}} r(x) \mathcal{Q}(dx) = \\ &= \int_{A \cap \{x: q(x) \neq 0\}} r(x) \mathcal{Q}(dx) + \int_{A \cap \{x: q(x) = 0\}} r(x) \mathcal{Q}(dx) = \int_A r(x) \mathcal{Q}(dx). \end{aligned}$$

Таким образом, если условие (4.2.2) выполнено, то $\mathcal{P} \prec \mathcal{Q}$ и $d\mathcal{P}/d\mathcal{Q} = r$.

2. Пусть теперь $\mathcal{P} \prec \mathcal{Q}$. Положим $A = \{x : p(x) \neq 0, q(x) = 0\}$. По определению, $\mathcal{Q}(A) = 0$, поэтому $\mathcal{P}(A) = 0$. То есть

$$0 = \mathcal{P}(A) = \int_{\{x: p(x) \neq 0, q(x) = 0\}} p(x) \mu(dx). \quad (4.2.4)$$

Подинтегральная функция в правой части (4.2.4) строго положительна, следовательно, множество интегрирования имеет μ -меру ноль. \square

Теорема о методе отбора. Алгоритмическая схема метода. Вернемся теперь к методу отбора. Пусть (D, \mathcal{D}) — некоторое измеримое пространство и \mathcal{P} , \mathcal{Q} — распределения, заданные на \mathcal{D} . Будем предполагать, что распределение \mathcal{Q} *мажорирует* распределение \mathcal{P} , то есть что существует производная $r = d\mathcal{P}/d\mathcal{Q}$ Радона-Никодима меры \mathcal{P} по мере \mathcal{Q} , обладающая свойством $\text{vraisup}_{\mathcal{Q}} r \leq M < \infty$.

Предложение 6. Пусть $\eta_1, \dots, \eta_n, \dots$ и $\alpha_1, \dots, \alpha_n, \dots$ — независимые в совокупности случайные величины, причем $\mathcal{L}(\eta_i) = \mathcal{Q}$, а $\alpha_i \in U(0, 1)$. Обозначим

$$\tau = \min(n : r(\eta_n) > M\alpha_n). \quad (4.2.5)$$

Тогда имеют место следующие утверждения.

1. Случайная величина $\tau - 1$ имеет геометрическое распределение с параметром $1/M$.
2. Случайная величина η_τ имеет распределение \mathcal{P} .
3. Случайные величины τ и ξ_τ независимы.

Доказательство. Будем доказывать все утверждения одновременно. Возьмем $A \in \mathcal{D}$. При $n \geq 1$ получим, что

$$\begin{aligned}\mathbb{P}(\eta_\tau \in A, \tau = n) &= \mathbb{P}(\eta_n \in A, r(\eta_1) \leq M\alpha_1, \dots, r(\eta_{n-1}) \leq M\alpha_{n-1}, r(\eta_n) > M\alpha_n) = \\ &= (\mathbb{P}(r(\eta) \leq M\alpha))^{n-1} \mathbb{P}(\eta \in A, r(\eta) > M\alpha),\end{aligned}$$

где η и α независимы, $\mathcal{L}(\eta) = \mathcal{Q}$ и $\alpha \in U(0, 1)$. Далее, так как $r(x) \leq M$ при \mathcal{Q} -п.в. x , то

$$\mathbb{P}(\eta \in A, r(\eta) > M\alpha) = \int_{x \in A, r(x) > M\alpha} \mathcal{Q}(dx) dt = \int_A \mathcal{Q}(dx) \int_0^{r(x)/M} dt = \frac{1}{M} \int_A r d\mathcal{Q} = \frac{1}{M} \mathcal{P}(A).$$

Поэтому $\mathbb{P}(r(\eta) \leq M\alpha) = 1 - 1/M$ и

$$\mathbb{P}(\eta_\tau \in A, \tau = n) = \frac{1}{M} \left(1 - \frac{1}{M}\right)^{n-1} \mathcal{P}(A).$$

Отсюда сразу видно, что $\tau - 1 \in \text{Geom}(1/M)$. Кроме того,

$$\mathbb{P}(\xi \in A) = \sum_{n=1}^{\infty} \mathbb{P}(\eta_\tau \in A, \tau = n) = \sum_{n=1}^{\infty} \frac{1}{M} \left(1 - \frac{1}{M}\right)^{n-1} \mathcal{P}(A) = \mathcal{P}(A).$$

Следовательно, остальные утверждения тоже доказаны. \square

Замечание 5. Из приведенного доказательства следует, что $\mathbb{P}(\eta \in A | r(\eta) > M\alpha) = \mathcal{P}(A)$ для любого $A \in \mathcal{A}$. Иногда именно это равенство считают обоснованием метода отбора.

Прежде чем разбирать алгоритмические аспекты метода отбора, обсудим условие $\text{vraisup}_{\mathcal{Q}} r < \infty$, используемое в Предложении 6. А именно, обозначим $M_0 = \text{vraisup}_{\mathcal{Q}} r$ и покажем, что число $M_0 - 1$ (если оно мало) можно считать мерой близости распределений \mathcal{P} и \mathcal{Q} по вариации.

Предложение 7. 1. $M_0 \geq 1$, причем $M_0 = 1$ тогда и только тогда, когда $\mathcal{P} = \mathcal{Q}$.
2. $\sup_{A \in \mathcal{D}} |\mathcal{P}(A) - \mathcal{Q}(A)| \leq 2(M_0 - 1)$.

Доказательство. 1. Действительно, если $M_0 < 1$, то $1 = \mathcal{P}(D) = \int_D r d\mathcal{Q} \leq M_0 < 1$, чего быть не может.

Пусть теперь $M_0 = 1$. Обозначив $A = \{x : r(x) < 1\}$, $B = \{x : r(x) = 1\}$ и $C = \{x : r(x) > 1\}$. По условию, $\mathcal{Q}(C) = 0$. Поэтому, если $\mathcal{Q}(A) > 0$, то

$$1 = \mathcal{P}(D) = \int_D r d\mathcal{Q} = \int_A r d\mathcal{Q} + \int_B r d\mathcal{Q} + \int_C r d\mathcal{Q} < \mathcal{Q}(A) + \mathcal{Q}(B) = 1,$$

и мы пришли к противоречию. Следовательно, $\mathcal{Q}(\{x : r(x) = 1\}) = 1$ и $\mathcal{P} = \mathcal{Q}$.

2. Так как

$$0 = \int_D (r - 1) d\mathcal{Q} = \int_D (r - 1)_+ d\mathcal{Q} - \int_D (r - 1)_- d\mathcal{Q},$$

то

$$\int_D |r - 1| d\mathcal{Q} = \int_D (r - 1)_+ d\mathcal{Q} + \int_D (r - 1)_- d\mathcal{Q} = 2 \int_D (r - 1)_+ d\mathcal{Q} = 2 \int_{r > 1} (r - 1) d\mathcal{Q} \leq 2(M_0 - 1)$$

и

$$\sup_A |\mathcal{P}(A) - \mathcal{Q}(A)| = \sup_A \left| \int_A (r - 1) d\mathcal{Q} \right| \leq \sup_A \int_A |r - 1| d\mathcal{Q} = \int_D |r - 1| d\mathcal{Q} \leq 2(M_0 - 1). \quad (4.2.6)$$

Тем самым, если число M_0 близко к 1, то распределения \mathcal{P} и \mathcal{Q} близки по вариации. \square

Запишем общую алгоритмическую схему метода отбора, вытекающую из Предложения 6.

Алгоритмическая схема метода отбора

Моделирование распределения \mathcal{P} методом отбора с мажорирующим распределением \mathcal{Q}

Входные данные: функция $r = d\mathcal{P}/d\mathcal{Q}$, мажоранта M .

Результат: ξ .

1. (*Отбор*) Do $(\eta \leftarrow \mathcal{Q}(\cdot); \text{Get}(\alpha))$ while $r(\eta) \leq M\alpha$;

2. (*Завершение*) $\xi \leftarrow \eta$; STOP.

Что касается трудоемкости метода, то она пропорциональна среднему числу моделирований случайного вектора (η, α) , то есть величине $\mathbb{E}\tau = M$.

Покажем возможность использования метода отбора в цикле, опираясь на второе утверждение Предложения 4. Возьмем $H = D \times (0, 1)$ и $\beta_i = (\eta_i, \alpha_i)$. Кроме того, положим $G = \{(x, t) \in H : r(x) > Mt\}$ и определим $D_k = (H \setminus G)^{k-1} \times G$. Тогда случайная величина τ_1 Предложения 4 совпадает с τ , определенной в (4.2.5), а $\beta_{\tau_1} = (\eta_\tau, \alpha_\tau)$. Следовательно, структура метода отбора описывается Предложением 4.

Посмотрим теперь на предварительный пример раздела 4.2.1 с точки зрения Предложения 6. В примере Рисунка 2 $D = [0, 1]$, $\mathcal{D} = \mathcal{B}_{[0,1]}$ — борелевская σ -алгебра подмножеств отрезка $[0, 1]$, а распределение \mathcal{P} определено своей плотностью p , которая ограничена постоянной M .

Введем меру $\mu = \text{mes}_1$ на отрезке $[0, 1]$ и возьмем $\mathcal{Q} = \mu$. Тогда $d\mathcal{P}/d\mu = p$, $d\mathcal{Q}/d\mu = 1$ и, согласно Следствию 3, существует производная Радона-Никодима $r = d\mathcal{P}/d\mathcal{Q} = p \leq M$.

Поэтому для моделирования распределения \mathcal{P} мы должны на каждом шаге общей алгоритмической схемы отбора моделировать случайные величины α_1, α_2 и проверять неравенство $p(\alpha_1) > M\alpha_2$. Как только это неравенство выполнится, результатом работы алгоритма окажется $\xi = \alpha_1$. Именно эта процедура словесно описана в разделе 4.2.1.

Наконец, приведем один иллюстративный пример.⁵⁴ Пусть задача состоит в моделировании «хвоста» нормального распределения с плотностью

$$p(x) = p_a(x) = C(a)e^{-x^2/2}, \quad x \geq a > 0, \quad (4.2.7)$$

где $C(a) = 1/\int_a^\infty e^{-x^2/2}dx$. Таким образом, у нас $D = [a, +\infty)$, а $\mathcal{P} = \mathcal{P}_a$ задается плотностью (4.2.7) относительно сужения меры Лебега на D . В качестве $\mathcal{Q} = \mathcal{Q}_a$ возьмем распределение с плотностью

$$q(x) = q_a(x) = xe^{(a^2-x^2)/2}, \quad x \geq a > 0. \quad (4.2.8)$$

Распределению \mathcal{Q} при $x \geq a > 0$ соответствует функция распределения $F(x) = F_a(x) = 1 - e^{(a^2-x^2)/2}$, поэтому \mathcal{Q} может быть промоделировано методом обратных функций, который (после замены α_1 на $1 - \alpha_1$) приводит к моделирующей формуле $\eta = \sqrt{a^2 - 2\ln(\alpha_1)}$.

Далее, производная Радона-Никодима r является ограниченной, так как

$$r(x) = p(x)/q(x) = x^{-1}C(a)e^{-a^2/2} \leq M = M(a) = a^{-1}C(a)e^{-a^2/2} \quad \text{при } x > a.$$

Неравенство $r(\eta) \leq M\alpha_2$, которое нужно проверять в методе отбора, очевидным образом переписывается как $\alpha_2\eta \geq a$, или, учитывая квадратный корень в моделирующей формуле для η , как $\alpha_2^2\eta^2 > a^2$.

Отсюда сразу же приходим к следующему алгоритму.

Алгоритм метода отбора для распределения (4.2.7)

Моделирование распределения (4.2.7) методом отбора с мажорирующим распределением (4.2.8)

Входные данные: $a > 0$.

Результат: ξ .

1. (*Инициализация*) $b \leftarrow a * a/2$;

2. (*Отбор*) Do $(\text{Get}(\alpha_1, \alpha_2); \beta \leftarrow b - \ln(\alpha_1); \gamma \leftarrow \alpha_2 * \alpha_2)$ while $\beta * \gamma \leq b$;

⁵⁴ см. [10, гл. IX.1.2]

3. (Завершение) $\xi \leftarrow \text{sqrt}(2 * \beta)$; STOP.

Таким образом, для получения одной реализации случайной величины ξ нужно в среднем: $2M(a)$ раз обратиться к генератору псевдослучайных чисел, $M(a)$ раз сосчитать логарифм и один раз извлечь квадратный корень.

Что касается трудоемкости $M(a)$ обсуждаемого метода отбора, то, как нетрудно видеть,

$$\begin{aligned} \int_a^\infty e^{-x^2/2} dx &= - \int_a^\infty x^{-1} d(e^{-x^2/2}) = - \left(x^{-1} e^{-x^2/2} \right) \Big|_a^\infty + \int_a^\infty e^{-x^2/2} dx^{-1} = \\ &= a^{-1} e^{-a^2/2} - \int_a^\infty x^{-2} e^{-x^2/2} dx = a^{-1} e^{-a^2/2} + \int_a^\infty x^{-3} d(e^{-x^2/2}) = \\ &= a^{-1} e^{-a^2/2} + \left(x^{-3} e^{-x^2/2} \right) \Big|_a^\infty - \int_a^\infty e^{-x^2/2} dx^{-3} = a^{-1} e^{-a^2/2} - a^{-3} e^{-a^2/2} + 3 \int_a^\infty x^{-4} e^{-x^2/2} dx. \end{aligned} \quad (4.2.9)$$

Поэтому

$$\int_a^\infty e^{-x^2/2} dx \sim e^{-a^2/2} / a$$

при $a \rightarrow \infty$ и $M(a) \rightarrow 1$ при тех же условиях. Более того, при $a > 1$ из (4.2.9) следует, что $1 < M(a) < a^2/(a^2 - 1)$, так что при больших a алгоритм оказывается достаточно эффективным.

4.2.3 Обсуждение. Варианты и примеры

1. Отбор с ограничениями. Обсудим, при каких условиях имеет смысл использовать метод отбора.

Поскольку распределение \mathcal{P} моделируется с использованием моделирования распределения \mathcal{Q} , то \mathcal{P} должно быть «сложным для моделирования», а \mathcal{Q} — «простым для моделирования».

Число M должно быть близким к 1. Согласно (4.2.6) это означает, что распределения \mathcal{P} и \mathcal{Q} должны быть близки по вариации.

Обычно реализация этих требований приводит к тому, что производная Радона-Никодима r оказывается сложна для вычислений. Есть несколько способов борьбы с этим явлением, основанных на том, что на самом деле нам нужно не вычислять функцию r в точке η , а проверять неравенство $r(\eta) > M\alpha$. Поэтому достаточно преобразовать это неравенство в более простое эквивалентное. Например, иногда помогает логарифмирование обеих частей неравенства.

Другой (и часто более эффективный) прием⁵⁵ заключается в том, чтобы найти такие простые в вычислительном смысле функции r_1 и r_2 , что $0 \leq r_1(x) \leq r(x) \leq r_2(x) \leq M$ при всех $x \in D$.

Тогда можно, например, сначала проверять условие $r_1(\eta) > M\alpha$ (в этом случае моделирование закончено), затем, если $r_1(\eta) \leq M\alpha$, проверять условие $r_2(\eta) \leq M\alpha$ (при его выполнении нужно моделировать новую пару (η, α)), и лишь в случае невыполнения обоих этих неравенств переходить к проверке неравенства $r(\eta) > M\alpha$. Если при этом окажется, что вероятность

$$\mathbb{P}(r_1(\eta) < M\alpha < r_2(\eta)) = M^{-1} \int_D (r_2(x) - r_1(x)) \mathcal{Q}(dx)$$

мала, то непосредственно проверять неравенство $r(\eta) > M\alpha$ придется редко.⁵⁶

Представим одну из возможных алгоритмических схем отбора с ограничениями.

Алгоритмическая схема метода отбора с ограничениями

Моделирование распределения \mathcal{P} методом отбора с ограничениями и мажорирующим распределением \mathcal{Q}

⁵⁵В англоязычной литературе этот прием называется «squeeze principle», см. [10, гл. III.3.6]. Поэтому функции r_1, r_2 иногда называются *сжимающими*.

⁵⁶Конечно, при реальном моделировании одна из функций r_1, r_2 может отсутствовать. Формально этому соответствует выбор $r_2 \equiv M$ или $r_1 \equiv 0$.

Входные данные: $r = d\mathcal{P}/d\mathcal{Q}$, мажоранта M , сжимающие функции r_1, r_2 .
Результат: ξ .

1. (Основной цикл отбора)
 - Do
 - Do (Отбор с вычислением функций r_1 и r_2)
 - Get(α); $\eta \leftarrow \mathcal{Q}(\cdot)$;
 - If $r_1(\eta) > M * \alpha$ then ($\xi \leftarrow \eta$; STOP);
 - while $r_2(\eta) \leq M * \alpha$;
 - while $r(\eta) \leq M * \alpha$;
2. $\xi \leftarrow \eta$; STOP.

Приведем пример, иллюстрирующий смысл отбора с ограничениями.
Рассмотрим распределение \mathcal{P} с плотностью

$$p(x) = 1 + 0.01 \sin(2\pi x), \quad x \in [0, 1]. \quad (4.2.10)$$

Если сравнить распределение \mathcal{P} с распределением $\mathcal{Q} = U(0, 1)$, то бросается в глаза, что, с одной стороны, \mathcal{P} и \mathcal{Q} близки, а с другой — распределение \mathcal{P} гораздо более сложно для моделирования, чем \mathcal{Q} . Поскольку

$$r(x) = \frac{d\mathcal{P}}{d\mathcal{Q}}(x) = p(x) \leq M = 1.01,$$

то неравенство $r(\eta) \leq M\alpha$ записывается как $100 + \sin(2\pi\alpha_1) \leq 101\alpha_2$, и общий алгоритм отбора приобретает следующий вид.

Алгоритм метода отбора для распределения (4.2.10)

Моделирование распределения (4.2.10) методом отбора из распределения $U(0, 1)$

Входные данные: $\vartheta = 8 \arctan(1)$.
Результат: ξ .

1. (Отбор) Do Get(α_1, α_2) while $100 + \sin(\vartheta * \alpha_1) \leq 101 * \alpha_2$;
2. (Завершение) $\xi \leftarrow \alpha_1$; STOP.

В этом алгоритме для получения одной реализации ξ нам нужно в среднем 2.02 раз обращаться к генератору псевдослучайных чисел и в среднем 1.01 раз считать функцию \sin . Если условиться, что вычисление функции \sin является трудоемкой операцией, то имеет смысл рассмотреть вариант отбора с ограничениями.

Заметим, что $\min_x p(x) = 0.99$, поэтому выберем в качестве функции r_1 эту постоянную. Что касается r_2 , то не будем ее для простоты вводить вовсе. Отметим, что неравенство $r_1(\eta) > M\alpha$ эквивалентно неравенству $\alpha_2 < 99/101$, поэтому алгоритм приобретает следующий вид.

Алгоритм метода отбора с ограничениями для распределения (4.2.10)

Моделирование распределения (4.2.10) методом отбора с ограничениями

Входные данные: $\vartheta = 8 \arctan(1)$, $a = 99/101$.
Результат: ξ .

1. (Основной цикл отбора)
 - Do (Отбор со сжимающей функцией r_1)
 - Get(α_1, α_2); If $\alpha_2 < a$ then ($\xi \leftarrow \alpha_1$; STOP);
 - while $100 + \sin(\vartheta * \alpha_1) \leq 101 * \alpha_2$;
2. $\xi \leftarrow \alpha_1$; STOP.

В этом варианте отбора функция \sin считается лишь с вероятностью $0.02/1.01 \approx 0.0198$ на каждом шаге цикла. Поэтому среднее число подсчетов синуса для получения одной реализации ξ равно 0.02, что существенно меньше, чем в предыдущем алгоритме.

2. Метод отбора для дискретных распределений. Конечно, общая алгоритмическая схема отбора, представленная в разделе 4.2.2, годится не только для абсолютно непрерывных распределений в евклидовых пространствах. Покажем, как она выглядит для дискретных распределений.

Пусть наша задача — промоделировать распределение \mathcal{P} , сосредоточенное на конечном или счетном множестве $D = \{x_1, \dots, x_n, \dots\}$ и определенное неотрицательными вероятностями $p_i = p(x_i) = \mathcal{P}(\{x_i\})$.

Введем на подмножествах D считающую меру μ равенством $\mu(\{x_i\}) = 1$. Тогда $\mathcal{P} \prec \mu$ и соответствующая производная $p(x_i)$ Радона-Никодима равна $d\mathcal{P}(\{x_i\})/d\mu(\{x_i\}) = p_i$. Действительно, для любого $A \subset D$

$$\mathcal{P}(A) = \sum_{i: x_i \in A} p_i = \sum_{i: x_i \in A} p(x_i)\mu(\{x_i\}) = \int_A p d\mu.$$

Аналогичным образом любое другое распределение \mathcal{Q} , сосредоточенное на D , абсолютно непрерывно относительно считающей меры μ с производной Радона-Никодима $q(x_i) = q_i$, где $q_i = \mathcal{Q}(\{x_i\})$. Поэтому, согласно Следствию 3, условие существования производной Радона-Никодима переписывается в виде⁵⁷ «если $p_i \neq 0$, то $q_i \neq 0$ », а сама производная равна

$$r_i = r(x_i) = \begin{cases} \frac{p_i}{q_i} & \text{при } q_i \neq 0; \\ 0 & \text{при } q_i = 0. \end{cases}$$

Поэтому, если отношение p_i/q_i ограничено сверху постоянной M , то общий алгоритм метода отбора сохраняет свой вид с заменой неравенств $r(\eta) \leq M\alpha$ на $p_\eta/q_\eta \leq M\alpha$.

3. Моделирование сужений распределений. В примере раздела 4.2.1 обсуждалась следующая процедура моделирования равномерного распределения в подграфике H плотности p (см. рис. 2): нужно моделировать равномерное распределение в прямоугольнике $[0, 1] \times [0, M]$ до тех пор, пока полученный случайный вектор η не попадет в H . Этот вектор и будет иметь нужное нам распределение.

Конечно, такая интуитивно очевидная процедура нуждается в обосновании. Кроме того, она структурно похожа на метод отбора, но формально отличается от него, так как на каждом шаге метода отбора моделируется не только случайная величина η , но и $\alpha \in U(0, 1)$. В обсуждаемом же примере α отсутствует. В этом нужно разобраться.

Начнем с определения сужения и расширения распределений. Пусть на измеримом пространстве (D, \mathcal{D}) задано распределение \mathcal{Q} и пусть G — измеримое подмножество D такое, что $\mathcal{Q}(G) > 0$. Каждое измеримое множество $A \in \mathcal{D}$ порождает подмножество $A \cap G$ множества G , а семейство $\mathcal{D}_G = \{A \cap G, A \in \mathcal{D}\}$ является σ -алгеброй.

Таким образом мы приходим к измеримому пространству (G, \mathcal{D}_G) . Распределение $\mathcal{P} = \mathcal{Q}|_G$, определенное на (G, \mathcal{D}_G) равенством

$$\mathcal{P}(B) = \mathcal{Q}|_G(B) = \frac{\mathcal{Q}(B)}{\mathcal{Q}(G)},$$

называется *сужением распределения \mathcal{Q} на G* . Это определение имеет очевидную интерпретацию: если случайная величина ξ имеет распределение \mathcal{Q} , то $\mathcal{Q}|_G$ — это условное распределение ξ при условии $\xi \in G$.

Пусть теперь на измеримом пространстве (G, \mathcal{D}_G) задано распределение \mathcal{P} . Тогда его можно *расширить* до распределения $\mathcal{P}^D(A)$, заданного на (D, \mathcal{D}) с помощью формулы

$$\mathcal{P}^D(A) = \mathcal{P}(A \cap G), \quad A \in \mathcal{D}.$$

⁵⁷Здесь используется тот факт, что единственное множество, имеющее считающую меру ноль, является пустым.

Конечно, распределения \mathcal{P} и \mathcal{P}^D различны (они определены на разных пространствах). Но, поскольку $\mathcal{P}^D(D \setminus G) = 0$, а $\mathcal{P}(A) = \mathcal{P}^D(A)$ при любом измеримом $A \subset G$, то с точки зрения моделирования этими различиями можно пренебречь.⁵⁸

Вернемся к моделированию. Рассмотрим измеримое пространство (D, \mathcal{D}) и его измеримое подмножество G . Будем считать, что на (D, \mathcal{D}) определено распределение \mathcal{Q} такое, что $\mathcal{Q}(G) > 0$. Наша цель — промоделировать сужение \mathcal{P} распределения \mathcal{Q} на множество G .

Будем отождествлять расширение \mathcal{P} на D с \mathcal{P} и найдем производную Радона-Никодима $r = d\mathcal{P}/d\mathcal{Q}$. Нетрудно видеть, что

$$r(x) = \begin{cases} \frac{1}{\mathcal{Q}(G)} & \text{при } x \in G, \\ 0 & \text{иначе.} \end{cases} \quad (4.2.11)$$

Действительно, при $A \in \mathcal{D}$

$$\mathcal{P}(A) = \mathcal{P}(A \cap G) = \frac{\mathcal{Q}(A \cap G)}{\mathcal{Q}(G)} = \frac{1}{\mathcal{Q}(G)} \int_A \mathbb{I}_G d\mathcal{Q} = \int_A r d\mathcal{Q},$$

где \mathbb{I}_A — индикатор множества A . Поскольку $r \leq M = 1/\mathcal{Q}(G)$, то мы можем моделировать распределение \mathcal{P} методом отбора с помощью распределения \mathcal{Q} .⁵⁹

Пусть $\eta \in \mathcal{Q}$. Тогда с учетом (4.2.11) основное неравенство метода отбора $r(\eta) > M\alpha$ приобретает вид

$$r(\eta) > M\alpha \iff \begin{cases} 1 > \alpha & \text{при } \eta \in G, \\ 0 > \alpha & \text{при } \eta \notin G. \end{cases}$$

Следовательно, событие $r(\eta) > M\alpha$ эквивалентно (с точностью до события $\alpha = 1$, которое имеет нулевую вероятность) событию $\eta \in G$, и для реализации метода отбора моделировать случайную величину α просто не нужно.

Запишем соответствующую алгоритмическую схему.

Алгоритмическая схема метода отбора для сужения распределения \mathcal{Q} на множество G

Результат: ξ .

1. (Отбор) До $\eta \leftarrow \mathcal{Q}(\cdot)$ while $\eta \notin G$;
2. (Завершение) $\xi \leftarrow \eta$; STOP.

Трудоёмкость метода равна $M = 1/\mathcal{Q}(G)$. Следовательно, чем множество G массивнее, тем отбор эффективнее. Время исполнения алгоритма зависит и от способа проверки принадлежности точки η множеству G . В сложных случаях, как и для общего алгоритма, можно использовать сжимающие функции.

Если взять $G_1 \subset G \subset G_2$ и положить $M \mathbb{I}_{G_1}(x) = r_1(x) \leq r(x) \leq r_2(x) = M \mathbb{I}_{G_2}(x)$, то при $i = 1, 2$ проверка неравенств $r_i(\eta) > M\alpha$ сведется к проверки принадлежности точки η множествам G_i .

Если множества G_i являются «простыми» и достаточно тесно аппроксимируют G , то отбор с такими ограничениями может оказаться достаточно эффективным. Приведем соответствующую алгоритмическую схему.

Алгоритмическая схема метода отбора с ограничениями для сужения распределений

⁵⁸С точки зрения моделирования: если $\alpha \in U(0, 1)$, то не все ли равно, где определено $\mathcal{L}(\alpha)$ — на борелевских подмножествах \mathbb{R} или на борелевских подмножествах $(0, 1]$?

⁵⁹Точнее, мы моделируем расширение на D сужения распределения \mathcal{Q} на множество G .

Результат: ξ .

1. (Основной цикл отбора)

- Do

- ◊ Do (Отбор с проверкой попадания в G_1 и G_2)

- $\eta \leftarrow \mathcal{Q}(\cdot);$

- If $\eta \in G_1$ then ($\xi \leftarrow \eta$; STOP);

- ◊ while $\eta \notin G_2$;

- while $\eta \notin G$;

2. $\xi \leftarrow \eta$; STOP.

4.3 Метод дискретной декомпозиции распределений

4.3.1 Общая схема и примеры

Рассмотрим измеримое пространство (D, \mathcal{D}) и конечное или счетное число распределений \mathcal{P}_i , заданных на этом пространстве. Кроме того, рассмотрим положительные вероятности q_i , в сумме дающие единицу. Тогда распределение

$$\mathcal{P} = \sum_{i \geq 1} q_i \mathcal{P}_i \quad (4.3.1)$$

называется *дискретной композицией* (по-другому — *дискретной смесью*) распределений \mathcal{P}_i с весами q_i , а само представление (4.3.1) — *дискретной декомпозицией* распределения \mathcal{P} . Конечно, если $(D, \mathcal{D}) = (\mathbb{R}^d, \mathcal{B}_d)$, то (4.3.1) эквивалентно аналогичной декомпозиции функции распределения F распределения \mathcal{P} :

$$F(x) = \sum_{i \geq 1} q_i F_i(x)$$

для любого $x \in \mathbb{R}^d$, где F_i — функция распределения распределения \mathcal{P}_i . Если же все распределения \mathcal{P}_i абсолютно непрерывны относительно одной и той же меры μ с производными Радона-Никодима p_i , то \mathcal{P} тоже абсолютно непрерывна относительно этой меры с производной Радона-Никодима $p = d\mathcal{P}/d\mu$ вида

$$p(x) = \sum_{i \geq 1} q_i p_i(x). \quad (4.3.2)$$

Выясним, как можно моделировать распределение \mathcal{P} , исходя из представления (4.3.1).

Предложение 8. Рассмотрим случайные величины ξ_n такие, что $\mathcal{L}(\xi_n) = \mathcal{P}_i$. Кроме того, пусть τ — независимая от $\{\xi_i, i \geq 1\}$ случайная величина с распределением $\mathbb{P}(\tau = i) = q_i$. Тогда случайная величина $\xi = \xi_\tau$ имеет распределение (4.3.1).

Доказательство. Пусть $A \in \mathcal{D}$. Тогда

$$\mathbb{P}(\xi \in A) = \sum_{i \geq 1} \mathbb{P}(\xi_\tau \in A, \tau = i) = \sum_i \mathbb{P}(\xi_i \in A, \tau = i) = \sum_{i \geq 1} \mathbb{P}(\xi_i \in A) \mathbb{P}(\tau = i) = \sum_{i \geq 1} q_i \mathcal{P}_i(A).$$

Доказательство закончено. □

Запишем теперь общую алгоритмическую схему, соответствующую Предложению 8. Обозначим \mathcal{Q} дискретное распределение, для которого $\mathcal{Q}(\{i\}) = q_i$.

Алгоритмическая схема метода дискретной декомпозиции

Результат: ξ .

1. (*Выбор компоненты смеси*) $i \leftarrow \mathcal{Q}(\cdot)$;
2. (*Моделирование компоненты*) $\xi \leftarrow \mathcal{P}_i(\cdot)$; STOP.

Одной из основных особенностей метода декомпозиции является принципиальная неоднозначность представления распределения \mathcal{P} в виде правой части равенства (4.3.1). С практической же точки зрения необходимо, чтобы распределения \mathcal{P}_i и \mathcal{Q} были удобными для моделирования.

Приведем несколько простых примеров использования метода дискретной декомпозиции.

Пример 1. Рассмотрим одномерное абсолютно непрерывное распределение \mathcal{P} с плотностью

$$p(x) = \begin{cases} 1/3 & \text{при } 0 \leq x \leq 1; \\ 2/3 & \text{при } 1 < x \leq 2; \\ 0 & \text{иначе.} \end{cases} \quad (4.3.3)$$

Конечно, это распределение можно моделировать разными способами (например, методом обратных функций), но нас интересует именно метод декомпозиции.

Легко представить плотность (4.3.3) в виде (4.3.2) (понятно, что здесь мера μ — это одномерная мера Лебега):

$$p(x) = \frac{1}{3} \mathbb{I}_{(0,1)}(x) + \frac{2}{3} \mathbb{I}_{(1,2)}(x) = \frac{1}{3} p_1(x) + \frac{2}{3} p_2(x). \quad (4.3.4)$$

Отсюда сразу же получается следующий алгоритм моделирования распределения (4.3.3).

Алгоритм дискретной декомпозиции (4.3.4) для моделирования распределения (4.3.3)

Результат: ξ .

1. (*Инициализация*) $\text{Get}(\alpha_1, \alpha_2)$;
2. (*Моделирование смеси*) If $3 * \alpha_1 < 1$ then $\xi \leftarrow \alpha_2$ else $\xi \leftarrow \alpha_2 + 1$;
3. (*Завершение*) STOP,

Однако разложение (4.3.4) не является единственно пригодным для моделирования. Например, можно использовать разложение

$$p(x) = \frac{2}{3} \mathbb{I}_{(0,2)}(x)/2 + \frac{1}{3} \mathbb{I}_{(1,2)}(x) = \frac{2}{3} p_1(x) + \frac{1}{3} p_2(x). \quad (4.3.5)$$

Тогда мы придем к следующему алгоритму.

Алгоритм дискретной декомпозиции (4.3.5) для моделирования распределения (4.3.3)

Результат: ξ .

1. (*Инициализация*) $\text{Get}(\alpha_1, \alpha_2)$;
2. (*Моделирование смеси*) If $3 * \alpha_1 < 2$ then $\xi \leftarrow 2 * \alpha_2$ else $\xi \leftarrow \alpha_2 + 1$;
3. (*Завершение*) STOP.

Отметим, что в разложении (4.3.4) носители плотностей p_1 и p_2 оказались дизъюнкты, в то время как в разложении (4.3.5) — нет.

Пример 2. Рассмотрим одномерное распределение \mathcal{P} с плотностью

$$p(x) = e^{-x-1} (1 + e^{e^{-x}}), \quad x > 0. \quad (4.3.6)$$

В отличие от предыдущего примера здесь наличие удобного для моделирования разложения вида (4.3.2) не бросается в глаза. Тем не менее,

$$\begin{aligned} p(x) &= e^{-x-1} + e^{-x-1} \sum_{k=0}^{\infty} \frac{1}{k!} (e^{-x})^k = e^{-x-1} + \sum_{k=0}^{\infty} \frac{e^{-1}}{k!} e^{-(k+1)x} = \\ &= e^{-1} e^{-x} + \sum_{k=0}^{\infty} \frac{e^{-1}}{(k+1)!} (k+1) e^{-(k+1)x} = e^{-1} e^{-x} + \sum_{m=1}^{\infty} \frac{e^{-1}}{m!} m e^{-mx} = \sum_{m=0}^{\infty} \frac{e^{-1}}{m!} p_m(x), \end{aligned}$$

где p_0 — плотность показательного распределения с параметром 1, а p_m при $m \geq 1$ — плотности показательных распределений с параметром m . Отсюда приходим к следующей алгоритмической схеме моделирования распределения (4.3.6) без уточнения способов моделирования показательного распределения и распределения Пуассона.

Алгоритм дискретной декомпозиции для моделирования распределения (4.3.6)

Результат: ξ .

1. (*Инициализация*) $\eta \leftarrow \text{Exp}(1); i \leftarrow \Pi(1);$
2. (*Моделирование смеси*) If $i = 0$ then $\xi \leftarrow \eta$ else $\xi \leftarrow \eta/i;$
3. (*Завершение*) STOP.

Как видно, удачное представление в виде смеси приводит к достаточно компактному способу моделирования не слишком простого распределения.

4.3.2 Модификации метода дискретной декомпозиции

На практике для моделирования сложных распределений редко используют один из изложенных общих методов в чистом виде. Как правило, эти методы применяются в виде различных комбинаций. Многочисленные примеры таких комбинаций можно найти в [13]. Здесь мы приведем две модификации метода дискретной декомпозиции, использующие, кроме того, идеи метода обратных функций и метода отбора.

1. Дискретная декомпозиция с обращением.

Вообще говоря, метод дискретной декомпозиции по определению требует как минимум двух обращений к генератору псевдослучайных чисел — для моделирования дискретного распределения \mathcal{Q} и для моделирования одного из распределений \mathcal{P}_i . Предложение 2 (см. раздел 2.4) показывает, что в случае, когда распределение \mathcal{Q} моделируется методом обратных функций, можно уменьшить количество обращений к генератору.

Точнее, там доказано, что если распределение \mathcal{Q} моделируется как $\tau = \min\{i : \alpha_1 \leq s_i\}$, где $s_i = q_1 + \dots + q_i$ — накопленные вероятности, то $\beta \stackrel{\text{def}}{=} (\alpha_1 - s_{\tau-1})/q_{\tau} \in U(0, 1)$ и случайные величины β, τ независимы. Тем самым мы можем использовать β для моделирования одного из распределений \mathcal{P}_i .

Этот прием выглядит наиболее эффектно, когда распределения \mathcal{P}_i (и, следовательно, распределение \mathcal{P}) определены на борелевских подмножествах \mathbb{R} .

Действительно, если для распределений \mathcal{P}_i применяются моделирующие формулы $G_i(\alpha_2)$, где $G_i : (0, 1) \mapsto \mathbb{R}$ — некоторые функции, то благодаря Предложению 2 в этих формулах вместо α_2 мы можем использовать β . Таким образом, в этом случае смесь (4.3.1) будет моделироваться с помощью одного обращения к генератору.

Конечно, наиболее стандартный выбор функций G_i связан с общим методом обратных функций раздела 4.1. Разберем этот случай подробнее, сформулировав соответствующий результат в виде отдельного утверждения.

Рассмотрим разложение (4.3.1) и обозначим F_i функции распределения одномерных распределений \mathcal{P}_i . Кроме того, положим (как в методе обратных функций) $G_i(y) = \inf\{x \in \mathbb{R} : F_i(x) > y\}$. Для дискретного распределения \mathcal{Q} определим $s_i = q_1 + \dots + q_i$ с $s_0 = 0$.

Предложение 9. Пусть $\alpha \in U(0, 1)$ и $\tau = \min\{i : \alpha \leq s_i\}$. Тогда случайная величина

$$\xi = G_\tau((\alpha - s_{\tau-1})/q_\tau) \quad (4.3.7)$$

имеет распределение \mathcal{P} .

Доказательство. Пусть $\alpha_1 \in U(0, 1)$ и α, α_1 независимы. Тогда, согласно Предложению 8, случайная величина $\xi_1 = G_\tau(\alpha_1)$ имеет распределение \mathcal{P} . По Предложению 2 это распределение не изменится, если α_1 заменить на $(\alpha - s_{\tau-1})/q_\tau$. \square

Замечание 6. Если носители распределений \mathcal{P}_i попарно дизъюнкты, то, как нетрудно видеть, моделирующая формула (4.3.7) совпадает с моделирующей формулой метода обратных функций для распределения \mathcal{P} .⁶⁰

Приведем несколько примеров.

Пример 1. Рассмотрим распределение (4.3.3). В разделе 4.3.1 приведены две дискретных декомпозиции этого распределения. Согласно Замечанию 6, разложение (4.3.4) для нас не представляет интереса, так его компоненты имеют дизъюнкты носители. Поэтому остановимся на разложении (4.3.5), которое в терминах функций распределений имеет вид

$$F(x) = \frac{2}{3} F_1(x) + \frac{1}{3} F_2(x),$$

где

$$F_1(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ x/2 & \text{при } 0 < x \leq 2, \\ 1 & \text{при } x > 2 \end{cases} \quad \text{и} \quad F_2(x) = \begin{cases} 0 & \text{при } x \leq 1, \\ x - 1 & \text{при } 1 < x \leq 2, \\ 1 & \text{при } x > 2. \end{cases}$$

Следовательно, в обозначениях Предложения 9 случайная величина τ принимает значения 1 и 2 с вероятностями $2/3$ и $1/3$, $G_1(y) = 2y$ и $G_2(y) = y + 1$. Поэтому мы получаем следующий алгоритм моделирования распределения (4.3.3).

Декомпозиция (4.3.5) с обращением для моделирования распределения (4.3.3)

Результат: ξ .

1. (*Инициализация*) $\text{Get}(\alpha); \beta \leftarrow 3 * \alpha;$
2. (*Моделирование смеси*) If $\beta < 2$ then $\xi \leftarrow \beta$ else $\xi \leftarrow \beta - 1$; STOP.

Заметим, что моделирующая функция здесь не является монотонной.

⁶⁰Покажите это.

Пример 2. Рассмотрим распределение с плотностью

$$p(x) = \frac{1}{n} \sum_{i=1}^n i^{-1} e^{-x/i}, \quad x > 0, \quad (4.3.8)$$

то есть равновероятную смесь показательных распределений EXP(1/i), $i = 1, \dots, n$.

Обычный метод дискретной декомпозиции раздела 4.3.1 приводит к следующему алгоритму.

Стандартный метод декомпозиции для моделирования распределения (4.3.8)

Результат: ξ .

1. (*Выбор компоненты смеси*) Get(α_1); $\tau \leftarrow \lceil n * \alpha_1 \rceil$;
2. (*Моделирование компоненты*) Get(α_2); $\xi \leftarrow -\tau * \ln(\alpha_2)$; STOP.

Использование варианта декомпозиции с обращением позволяет сэкономить (на практике — при относительно небольшом n) одно обращение к генератору псевдослучайных чисел.

Декомпозиция с обращением для моделирования распределения (4.3.8)

Результат: ξ .

1. (*Выбор компоненты смеси*) Get(α); $\tau \leftarrow \lceil n * \alpha \rceil$;
2. (*Моделирование компоненты*) $\xi \leftarrow -\tau * \ln(\tau - n * \alpha)$; STOP.

Обратимся теперь к другой модификации метода декомпозиции, связанной с методом отбора.

2. Метод декомпозиции-дополнения.

Рассмотрим разложение

$$\mathcal{P} = q \mathcal{P}_1 + (1 - q) \mathcal{P}_2, \quad (4.3.9)$$

где $0 < q < 1$ и $\mathcal{P}_1, \mathcal{P}_2$ — распределения в измеримом пространстве (D, \mathcal{D}) . Будем считать, что распределение \mathcal{P}_1 моделируется методом отбора с помощью распределения \mathcal{Q} , причем $r = d\mathcal{P}_1/\mathcal{Q}(x) \leq 1/q$ и используемая в отборе мажоранта M равна $1/q$. Тогда алгоритмическая схема моделирования декомпозиции (4.3.9) может выглядеть следующим образом.

Декомпозиция \mathcal{P} с отбором для \mathcal{P}_1

Входные данные: функция $r = d\mathcal{P}_1/d\mathcal{Q}$, число $q \in (0, 1)$.

Результат: ξ .

1. (*Выбор компоненты и моделирование \mathcal{P}_1*)
Get(β); If $\beta < q$ then
 - Do ($\eta \leftarrow \mathcal{Q}(\cdot)$; Get(α)) while $q * r(\eta) \leq \alpha$;
 - $\xi \leftarrow \eta$; STOP.
2. (*Моделирование \mathcal{P}_2 и завершение*)
 $\xi \leftarrow \mathcal{P}_2(\cdot)$; STOP.

Если через t_r обозначить среднее время проверки неравенства $q r(\eta) \leq \alpha$ и ввести обозначения t_U, t_Q, t_{P_2} для среднего времени моделирования распределений $U(0, 1)$, \mathcal{Q} и \mathcal{P}_2 соответственно, то среднее время моделирования случайной величины ξ с помощью описанного выше алгоритма будет иметь вид

$$t_U + q(t_Q + t_U + t_r)/q + (1 - q)t_{P_2} = t_Q + 2t_U + t_r + (1 - q)t_{P_2}. \quad (4.3.10)$$

В то же время, как следует из Предложения 6 раздела 4.2.2, $\mathbb{P}(qr(\eta) < \alpha) = q$. Это наводит на мысль, что в случае, когда выполнено неравенство $qr(\eta) < \alpha$, нужно (без специального розыгрыша компоненты декомпозиции) положить $\xi = \eta$, а при выполнении противоположного неравенства — моделировать распределение \mathcal{P}_2 . Докажем соответствующее утверждение.

Предложение 10. Рассмотрим разложение (4.3.9). Пусть η, ξ_2 и α — независимые случайные величины, имеющие распределения \mathcal{Q} , \mathcal{P}_2 и $U(0, 1)$. Предположим, что $r(x) \stackrel{\text{def}}{=} d\mathcal{P}_1/d\mathcal{Q}(x) \leq 1/q$ при \mathcal{Q} -почти всех x . Положим

$$\xi = \begin{cases} \eta & \text{при } qr(\eta) < \alpha, \\ \xi_2 & \text{иначе.} \end{cases} \quad (4.3.11)$$

Тогда $\mathcal{L}(\xi) = \mathcal{P}$.

Доказательство. При $A \in \mathcal{D}$ действуя так же, как в Предложении 6, получим, что

$$\begin{aligned} \mathbb{P}(\xi \in A) &= \mathbb{P}(\eta \in A, qr(\eta) < \alpha) + \mathbb{P}(\xi_2 \in A, qr(\eta) \geq \alpha) = \\ &= q\mathcal{P}_1(A) + \mathbb{P}(\xi_2 \in A)\mathbb{P}(qr(\eta) \geq \alpha) = q\mathcal{P}_1(A) + (1 - q)\mathcal{P}_2(A) = \mathcal{P}(A). \end{aligned}$$

Утверждение доказано. \square

Из Предложения 10 вытекает естественная алгоритмическая схема моделирования распределения (4.3.9), которую мы опускаем за очевидностью (конечно, проверку неравенства $qr(\eta) < \alpha$ можно проводить по-разному, в том числе и с помощью введения сжимающих функций).

Важно подчеркнуть логическую структуру обсуждаемого метода: он следует из разложения распределения \mathcal{P} на две компоненты, одна из которых может быть удобно промоделирована методом отбора, согласованным с весами разложения. После этого цикл метода отбора заменяется на его первый шаг, причем в случае невыполнения условий этого шага происходит «дополнение» — моделируется вторая компонента смеси.

В англоязычной литературе этот метод носит название «the acceptance-complement method», см. [10, гл. II.5], мы здесь назовем его *методом декомпозиции-дополнения*.

Очевидно, что время моделирования случайной величины ξ согласно формуле (4.3.11) равно $t_\eta + t_U + t_r + (1 - q)t_{\mathcal{P}_2}$, что на t_U меньше, чем в (4.3.10).

Пример 1. Проиллюстрируем эти рассуждения на примере плотности

$$p(x) = 1 + \theta \sin(2\pi x), \quad \text{где } x \in [0, 1] \text{ и } 0 < \theta \leq 0.5. \quad (4.3.12)$$

Эта плотность при $\theta = 0.01$ уже появлялась в связи с обсуждением отбора с ограничениями, см. (4.2.10).

Взяв некоторое $0 < q < 1$, представим $p(x)$ в виде

$$p(x) = qp_1(x) + 1 - q, \quad \text{где } p_1(x) = 1 + \frac{\theta}{q} \sin(2\pi x). \quad (4.3.13)$$

Ясно, что при $q \geq \theta$ функция p_1 неотрицательна и, следовательно, является плотностью некоторого распределения. Тем самым (4.3.13) является частным случаем (4.3.9).

Чтобы применить к распределению (4.3.13) Предложение 10, выберем в качестве \mathcal{Q} равномерное распределение $U(0, 1)$. Тогда $r(x) = p_1(x)$ и $\max r(x) = 1 + \theta/q$. Для того, чтобы условия Предложения 10 были выполнены, осталось обеспечить неравенство $r(x) \leq 1/q$, которое выполняется, очевидно, при $q \leq 1 - \theta$.

Итак, выбрав $\theta \leq q \leq 1 - \theta$, приходим к следующему алгоритму.

Алгоритм декомпозиции-дополнения для распределения (4.3.12)

Входные данные: $\theta \in (0, 0.5]$, $q \in [\theta, 1 - \theta]$, $\vartheta = 8 \arctan(1)$.

Результат: ξ .

1. (Выбор первой компоненты) $\text{Get}(\alpha_1, \alpha_2)$; If $q + \theta * \sin(\vartheta * \alpha_1) \leq \alpha_2$ then $(\xi \leftarrow \alpha_1; \text{STOP})$;
2. (Вторая компонента) $\text{Get}(\alpha_3)$; $\xi \leftarrow \alpha_3$; STOP .

Ясно, что для реализации алгоритма понадобится в среднем $3 - q$ обращения к генератору псевдослучайных чисел и одна проверка неравенства $q + \theta \sin(\vartheta \alpha_1) \leq \alpha_2$. Если считать, что трудоемкость этой проверки не зависит от q , то следует выбирать q максимально возможным, то есть полагать $q = 1 - \theta$.

Как показано в [10, разд. II.5.4] метод декомпозиции-дополнения удобно использовать для моделирования распределений, близких к равномерному. Пример с распределением (4.3.12) относится именно к такому случаю.

4.3.3 Общий метод декомпозиции.

Формула (4.3.1) относится лишь к частному (дискретному) случаю утверждения, являющегося основой общего метода декомпозиции для моделирования распределений. Это утверждение можно сформулировать следующим образом.

Предложение 11. Пусть $(\Omega, \mathcal{F}, \mathbb{P})$ — вероятностное пространство. Рассмотрим отображение $\eta : (\Omega, \mathcal{F}) \mapsto (G, \mathcal{G})$ и функцию $f : (\Omega \times G, \mathcal{F} \times \mathcal{G}) \mapsto (\mathbb{R}, \mathcal{B}_R)$. Если $\mathbb{E}|f(\omega, \eta)| < \infty$ и для любого $t \in G$ случайная величина $f(\cdot, t)$ и отображение η независимы, то

$$\mathbb{E}f(\omega, T) = \int_G \mathbb{E}f(\omega, t) \mathcal{P}_\eta(dt). \quad (4.3.14)$$

Доказательство. Поскольку случайные величины $f(\cdot, t)$ и η независимы, то, не умаляя общности, можно считать, что они определены на разных вероятностных пространствах.⁶¹ Точнее, рассмотрим вероятностные пространства $(\Omega_1, \mathcal{F}_1, \mathcal{P}_1)$ и $(\Omega_2, \mathcal{F}_2, \mathcal{P}_2)$, являющиеся копиями $(\Omega, \mathcal{F}, \mathcal{P})$. Положим $\Omega^* = \Omega_1 \times \Omega_2$, $\mathcal{F}^* = \mathcal{F}_1 \times \mathcal{F}_2$ и $\mathcal{P}^* = \mathcal{P}_1 \otimes \mathcal{P}_2$.

Далее, для $\omega^* = (\omega_1, \omega_2) \in \Omega^*$ определим $f^*(\omega^*, t) = f(\omega_1, t)$ и $\eta^*(\omega^*) = \eta(\omega_2)$. Ясно, что $\mathcal{L}(f^*(\cdot, t)) = \mathcal{L}(f(\cdot, t))$, $\mathcal{L}(\eta^*) = \mathcal{L}(\eta)$, причем $f^*(\omega^*, t)$ и η^* независимы. Поэтому

$$\begin{aligned} \mathbb{E}f(\omega, \eta) &= \mathbb{E}^* f^*(\omega^*, \eta^*) = \int_{\Omega_1 \times \Omega_2} f(\omega_1, \eta(\omega_2)) \mathbb{P}_1(d\omega_1) \mathbb{P}_2(d\omega_2) = \\ &= \int_G \mathcal{P}_\eta(dt) \int_{\Omega_1} f(\omega_1, t) \mathbb{P}_1(d\omega_1) = \int_G \mathcal{P}_\eta(dt) \mathbb{E}f(\omega, t). \end{aligned}$$

Утверждение доказано. \square

К собственно моделированию имеет отношение следующее следствие из Предложения 11.

Следствие 4. Пусть отображения $\xi : (\Omega, \mathcal{F}) \mapsto (D, \mathcal{D})$ и $\eta : (\Omega, \mathcal{F}) \mapsto (G, \mathcal{G})$ заданы на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$. Предположим кроме того, что существует такое отображение $\beta_\cdot(\cdot) : (G \times \Omega, \mathcal{G} \times \mathcal{F}) \mapsto (D, \mathcal{D})$, что

$$\xi(\omega) = \beta_{\eta(\omega)}(\omega), \quad (4.3.15)$$

где при любом $t \in G$ случайная величина $\beta_t(\cdot)$ не зависит от η .

Тогда для любых $A \in \mathcal{D}$ и $B \in \mathcal{G}$

$$\mathbb{P}(\xi \in A, \eta \in B) = \int_B \mathbb{P}(\beta_t \in A) \mathcal{P}_\eta(dt). \quad (4.3.16)$$

В частности,

$$\mathbb{P}(\xi \in A) = \int_G \mathbb{P}(\beta_t \in A) \mathcal{P}_\eta(dt), \quad (4.3.17)$$

Доказательство. Для доказательства (4.3.16) достаточно применить Предложение 11 к функции $f(\omega, t) = \mathbb{I}_A(\beta_t(\omega)) \mathbb{I}_B(t)$. \square

⁶¹С практической точки зрения это означает, что при моделировании случайных величин $f(\cdot, t)$ и η используются наборы α_i с непересекающимися номерами.

Общая алгоритмическая схема моделирования случайной величины ξ , вытекающая из представления (4.3.17) Следствия 4, основана на задании функции β , удовлетворяющей (4.3.15), и выглядит следующим образом.⁶²

Алгоритмическая схема общего метода декомпозиции

Входные данные: функция β .

Результат: ξ .

1. (Выбор компоненты смеси η) $t \leftarrow \mathcal{P}_\eta(\cdot)$;
2. (Моделирование ξ) $\xi \leftarrow \beta_t$; STOP.

Замечание 7. Отметим частные случаи общего утверждения Следствия 4. Если η — дискретная случайная величина, то общая декомпозиция сводится к дискретной декомпозиции раздела 4.3.1. Рассмотрим абсолютно непрерывный случай.

Пусть совместное распределение ξ, η имеет плотность распределение (то есть производную Радона-Никодима) $p_{\xi\eta}(x, t)$ относительно меры $\mu(dx) \otimes \nu(dt)$ в $(D \times G, \mathcal{D} \times \mathcal{G})$. Тогда, как обычно, случайные величины ξ и η обладают плотностями

$$p_\xi(x) = \int_G p_{\xi\eta}(x, t) \nu(dt) \quad \text{и} \quad p_\eta(t) = \int_D p_{\xi\eta}(x, t) \mu(dx)$$

относительно мер μ и ν соответственно.

Обозначим $D_0 = \{t : p_\eta(t) \neq 0\}$ и положим

$$p_{\xi|\eta}(x|t) = \begin{cases} \frac{p_{\xi\eta}(x, t)}{p_\eta(t)} & \text{если } t \in D_0, \\ p_\xi(x) & \text{если } t \notin D_0. \end{cases} \quad (4.3.18)$$

Тогда, учитывая, что $\mathcal{P}_\eta(D_0) = 1$, получим цепочку равенств

$$\begin{aligned} \int_B \left(\int_A p_{\xi|\eta}(x|t) \mu(dx) \right) \mathcal{P}_\eta(dt) &= \int_{B \cap D_0} \left(\int_A \frac{p_{\xi\eta}(x, t) \mu(dx)}{p_\eta(t)} \right) \mathcal{P}_\eta(dt) + \int_{B \cap D_0^c} \mathbb{P}(\xi \in A) \mathcal{P}_\eta(dt) = \\ &= \int_{B \cap D_0} \left(\int_A p_{\xi\eta}(x, t) \mu(dx) \right) \nu(dt) = \mathbb{P}(\xi \in A, \eta \in B \cap D_0) = \mathbb{P}(\xi \in A, \eta \in B). \end{aligned}$$

Это означает, (см. (4.3.16)), что для получения реализации случайной величины β_t нужно моделировать распределение с плотностью (4.3.18) относительно меры $\mu(dx)$.⁶³

Понятно, что применение общего метода декомпозиции для моделирования требует немалого искусства,⁶⁴ поэтому этот метод применяется не так часто. Приведем несколько примеров его использования.

Пример 3. (Негативно-биномиальное распределение.)

Рассмотрим негативно-биномиальное распределение $\text{NB}(k, p)$ с параметрами $k > 0, p \in (0, 1)$. Это дискретное распределение задается таблицей

$$\mathcal{P} : \begin{pmatrix} 0 & 1 & \dots & n & \dots \\ p_0 & p_1 & \dots & p_n & \dots \end{pmatrix}, \quad \text{где } p_n = \frac{\Gamma(n+k)}{\Gamma(k)n!} p^k (1-p)^n.$$

⁶²Если наша цель — промоделировать вектор (ξ, η) , то вместо (4.3.17) необходимо пользоваться (4.3.16). Алгоритм при этом практически не меняется.

⁶³Плотность (4.3.18) называется *условной плотностью ξ при условии $\eta = t$* . В общем случае распределение случайной величины β_t — это регулярный вариант условного распределения ξ при условии $\eta = t$. Нам, однако, терминология и техника условных распределений здесь не понадобятся.

⁶⁴Нужно уметь находить как подходящее разложение (4.3.17) распределения \mathcal{P} , так и функцию β .

Конечно, такое распределение можно моделировать нетабличным методом обратных функций. Кроме того, при целом k случайная величина, равная числу неудач до k -го успеха в испытаниях Бернулли с вероятностью успеха p , имеет как раз распределение $NB(k, p)$. Можно, однако, использовать для моделирования распределения $NB(k, p)$ и метод декомпозиции.

Действительно, пусть случайная величина η имеет гамма-распределение $\text{Gamma}(k, \mu)$ с параметрами (k, μ) ,⁶⁵ а семейство случайных величин $\beta_\lambda(\cdot)$ обладает следующими свойствами: во-первых, $\beta_\lambda(\cdot)$ не зависит от η , во-вторых, $\beta_\lambda(\cdot) \in \Pi(\lambda)$ и, наконец, функция $\beta_\lambda(\omega)$ измерима по паре (λ, ω) .⁶⁶ Тогда, согласно Следствию 4,

$$\begin{aligned} \mathbb{P}(\beta_\eta = n) &= \int_0^\infty \mathbb{P}(\beta_\lambda = n) g_{k, \mu}(\lambda) d\lambda = \int_0^\infty \frac{\lambda^n}{n!} e^{-\lambda} \frac{\mu^k}{\Gamma(k)} \lambda^{k-1} e^{-\mu \lambda} d\lambda = \frac{\mu^k}{\Gamma(k) n!} \int_0^\infty \lambda^{k+n-1} e^{-(\mu+1)\lambda} d\lambda = \\ &= \frac{\mu^k}{\Gamma(k) n!} \frac{\Gamma(n+k)}{(\mu+1)^{n+k}} = \frac{\Gamma(n+k)}{\Gamma(k) n!} \left(\frac{\mu}{\mu+1} \right)^k \left(\frac{1}{\mu+1} \right)^n. \end{aligned}$$

Тем самым мы получили распределение $NB(k, p)$ с $p = \mu/(1+\mu)$. Поскольку параметр μ является параметром масштаба гамма-распределения, отсюда сразу же следует алгоритмическая схема моделирования распределения $NB(k, p)$. Так как здесь не уточняется метод моделирования распределения Пуассона, то вместо оператора $\xi \leftarrow \beta_\lambda$ в алгоритме записан оператор $\xi \leftarrow \Pi(\lambda)$.

Алгоритмическая схема моделирования распределения $NB(k, p)$ методом декомпозиции

Входные данные: k, p .

Результат: ξ .

1. (Инициализация) $\mu \leftarrow p/(1-p)$;
2. (Выбор компоненты смеси) $s \leftarrow \text{Gamma}(k, 1)$; $\lambda \leftarrow s/\mu$;
3. (Моделирование компоненты) $\xi \leftarrow \Pi(\lambda)$; STOP.

Конечно, трудоемкость этого метода будет зависеть от выбранных способов моделирования гамма-распределения и распределения Пуассона.

Пример 4. (Двумерное нормальное распределение.)

Плотность невырожденного двумерного нормального распределения, как известно,⁶⁷ имеет вид

$$p_{\xi_1 \xi_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left(\frac{(x_1-a_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1-a_1)(x_2-a_2)}{\sigma_1\sigma_2} + \frac{(x_2-a_2)^2}{\sigma_2^2} \right)}, \quad (4.3.19)$$

где a_1, a_2 — математические ожидания координат случайного вектора $(\xi_1, \xi_2)^T$, имеющего распределение (4.3.19), σ_1^2, σ_2^2 — соответствующие (положительные) дисперсии, а $\rho \in (-1, 1)$ — коэффициент корреляции между ξ_1 и ξ_2 .

Обсудим использование Следствия 4 для моделирования распределения (4.3.19), считая, что мы уже умеем моделировать стандартное нормальное распределение $N(0, 1)$. Тогда, если $\varepsilon \in N(0, 1)$, то $\sigma\varepsilon + a \in N(a, \sigma^2)$.

Прежде всего, положим $a_1 = a_2 = 0$. Нетрудно видеть,⁶⁸ что $\xi_1 \in N(0, \sigma_1^2)$. Кроме того, условная плотность $p_{\xi_2|\xi_1}(x_2|x_1) = p_{\xi_1\xi_2}(x_1, x_2)/p_{\xi_1}(x_1)$ является⁶⁹ плотностью нормального распределения $N(\theta, \tau^2)$ с $\theta = \theta(x_1) = \rho\sigma_2x_1/\sigma_1$ и $\tau = (1-\rho^2)\sigma_2^2$.

⁶⁵То есть имеет плотность $g_{k, \mu}(\lambda) = \mu^k \lambda^{k-1} e^{-\mu \lambda} / \Gamma(k)$, сосредоточенную на $(0, \infty)$. О моделировании гамма-распределения см. раздел 5.3.3.

⁶⁶Последнее свойство всегда выполняется при моделировании. Например, оба метода моделирования распределения Пуассона (см. раздел 3.2) именно такие.

⁶⁷см. Лемму 9 Приложения 2. Как из (8.2.1) получить (4.3.19)?

⁶⁸Проверьте это.

⁶⁹Проверьте!

Поэтому (см. Замечание 7) в нашем случае $\beta_t = \rho\sigma_2t/\sigma_1 + \sigma_2\sqrt{1-\rho^2}\varepsilon_2$, где $\varepsilon_2 \in N(0, 1)$ и не зависит от ξ_1 . Соответствующий вариант общей алгоритмической схемы декомпозиции может быть записан следующим образом.⁷⁰

Алгоритмическая схема моделирования двумерного нормального распределения

Входные данные: $a_1, a_2, \sigma_1, \sigma_2, \rho, \delta = \sqrt{1-\rho^2}$.

Результат: ξ_1, ξ_2 .

1. (Моделирование $N(0, \sigma_1^2)$) $\varepsilon_1 \leftarrow N(0, 1)$; $\xi_1 \leftarrow \sigma_1 * \varepsilon_1$;
2. (Моделирование условного распределения) $\varepsilon_2 \leftarrow N(0, 1)$; $\xi_2 \leftarrow \sigma_2 * (\rho * \varepsilon_1 + \delta * \varepsilon_2)$;
3. (Восстановление средних) $\xi_1 \leftarrow \xi_1 + a_1$; $\xi_2 \leftarrow \xi_2 + a_2$; STOP.

Замечание 8. Приведенный выше алгоритм имеет и другое обоснование. Действительно, поскольку речь идет о двумерном гауссовском распределении, то для моделирования ξ_2 можно использовать формулу (5.2.2) Леммы 4 раздела 5.2, положив $k = 1$. Это легко проверить,⁷¹ если заметить, что (одномерный) вектор $\bar{\beta}_2$ равен $\rho\sigma_1\sigma_2$, а (1×1) -матрица Σ_1 имеет вид (σ_1^2) .

Пример 5. (Равномерное распределение в подграфике плотности.)

Рассмотрим двумерную область D , являющуюся подграфиком плотности p_{ξ_1} распределения \mathcal{P}_1 , и поставим задачу моделирования распределения $U(D)$ с плотностью $p_{\xi_1, \xi_2}(x_1, x_2) = \mathbb{I}_D(x_1, x_2)$.⁷²

Если область D является неограниченной, то такая задача может оказаться непростой. Покажем, как ее можно решать методом декомпозиции в предположении, что существует достаточно простой алгоритм моделирования распределения \mathcal{P}_1 .

Заметим, что

$$p_{\xi_2 | \xi_1}(x_2 | x_1) = p_{\xi_1 \xi_2}(x_1, x_2) / p_{\xi_1}(x_1) = \frac{\mathbb{I}_D(x_1, x_2)}{p_{\xi_1}(x_1)} = \frac{1}{p_{\xi_1}(x_1)} \mathbb{I}_{(0, p_{\xi_1}(x_1))}(x_2),$$

то есть эта условная плотность является равномерной на $(0, p_{\xi_1}(x_1))$. Поэтому, если мы уже получили реализацию t случайной величины ξ_1 с плотностью p_{ξ_1} , то, согласно Замечанию 7, ξ_2 должна быть равномерно распределена на интервале $(0, s)$ с $s = p_{\xi_1}(t)$, что не представляет сложности для моделирования.

Проиллюстрируем эту идею в случае, когда p_{ξ_1} является плотностью распределения Парето с параметром $\gamma > 0$,⁷³ то есть когда $p_{\xi_1}(x_1) = \gamma x_1^{\gamma-1} \mathbb{I}_{(0,1)}(x_1)$. Согласно разделу 4.1.1, ξ_1 моделируется как $\xi_1 = \alpha_1^{1/\gamma}$. Тогда можно положить

$$\xi_2 = \alpha_2 \gamma \xi_1^{\gamma-1} = \gamma \alpha_2 \alpha_1^{(\gamma-1)/\gamma} = \gamma \alpha_1 \alpha_2 / \xi_1.$$

Запишем соответствующий алгоритм.

Алгоритмическая схема моделирования равномерного распределения в подграфике плотности Парето

Входные данные: $\gamma > 0$.

Результат: ξ_1, ξ_2 .

1. (Моделирование абсциссы) $\text{Get}(\alpha_1)$; $\xi_1 \leftarrow \alpha_1 \wedge (1/\gamma)$;
2. (Моделирование условного распределения) $\text{Get}(\alpha_2)$; $\xi_2 \leftarrow \gamma * \alpha_1 * \alpha_2 / \xi_1$; STOP.

⁷⁰Отметим, что в алгоритме сначала моделируется двумерное нормальное распределение с нулевыми средними, а потом делается очевидный переход к средним a_1, a_2 .

⁷¹Так проверьте!

⁷²Если $(\eta_1, \eta_2) \in U(D)$, то η_1 имеет плотность p_{ξ_1} .

⁷³Если $\gamma < 1$, то подграфик плотности Парето является неограниченным.

5 Приемы моделирования различных распределений

В этом разделе мы кратко остановимся на методах моделирования некоторых популярных распределений, рассматривая, как обычно, лишь простые для реализации алгоритмы.

5.1 Полярные методы

Группа полярных методов предназначена для моделирования распределений, так или иначе связанных с двумерными распределениями, инвариантными относительно вращений.

5.1.1 Равномерное распределение на окружности \mathbb{S}^1

По определению, случайный вектор $\bar{\xi} = (\xi_1, \xi_2)^T$ равномерно распределен на единичной окружности \mathbb{S}^1 с центром в нуле, если $\xi_1^2 + \xi_2^2 = 1$ с вероятностью 1 и если полярный угол φ вектора $\bar{\xi}$ равномерно распределен на $[0, 2\pi)$.

Из этого определения сразу же следует моделирующая формула для равномерного распределения на \mathbb{S}^1 :

$$\xi_1 = \cos(2\pi\alpha), \quad \xi_2 = \sin(2\pi\alpha). \quad (5.1.1)$$

Вычисление двух тригонометрических функций, однако, может оказаться трудоемкой операцией.⁷⁴ Стандартной альтернативой формуле (5.1.1) является использование метода отбора для моделирования равномерного распределения в единичном круге $\mathbb{B}_1(\mathbf{0}) = \{(x, y) : x^2 + y^2 < 1\}$ с центром в нуле с последующей нормировкой результата. Формальное обоснование этой достаточно очевидной процедуры следует из материала раздела 7.2.2.

Запишем соответствующий алгоритм.

Моделирование равномерного распределения на окружности \mathbb{S}^1 полярным методом

Результат: (ξ_1, ξ_2) .

1. (Отбор в круг)

- Do

Get(α_1, α_2); $\beta_1 \leftarrow 2 * \alpha_1 - 1$; $\beta_2 \leftarrow 2 * \alpha_2 - 1$; $d \leftarrow \beta_1 * \beta_1 + \beta_2 * \beta_2$;

- While $d > 1$;

2. (Нормировка) $t \leftarrow \sqrt{d}$; $\xi_1 \leftarrow \beta_1/t$; $\xi_2 \leftarrow \beta_2/t$; STOP.

Нетрудно видеть, что здесь трудоемкость отбора равна $M = 4/\pi \approx 1.27$. Поэтому для моделирования равномерного распределения на окружности требуется (в среднем) $8/\pi \approx 2.55$ обращений к генератору псевдослучайных чисел и одно извлечение квадратного корня.

Замечание 9. Заметим, что при использовании обсуждаемого метода теряется информация о полярном радиусе r случайного вектора, равномерно распределенного в $\mathbb{B}_1(\mathbf{0})$. Согласно разделу

⁷⁴Конечно, прямое вычисление одной из этих функций (например, \cos) можно заменить на ее выражение через вторую с присвоением соответствующего знака:

$$\cos(2\pi\alpha) = \begin{cases} \sqrt{1 - \sin^2(2\pi\alpha)} & \text{при } |\alpha - 0.5| > 0.25, \\ -\sqrt{1 - \sin^2(2\pi\alpha)} & \text{иначе.} \end{cases}$$

Это несколько убыстряет процедуру, но не решает полностью всю проблему.

7.2.1 Приложения 1, этот полярный радиус имеет плотность распределения $p_r(t) = 2t\mathbf{1}_{(0,1)}$ или, что тоже самое, функцию распределения

$$F_r(t) = \begin{cases} 0 & \text{при } t \leq 0, \\ t^2 & \text{при } 0 < t \leq 1, \\ 1 & \text{при } t > 1. \end{cases}$$

Поэтому $\mathbb{P}(r^2 < t) = \mathbb{P}(r < \sqrt{t}) = t$ при $t \in [0, 1]$ и, следовательно, случайная величина $d = r^2$ (полученная в алгоритме после отбора по условию $d \leq 1$) равномерно распределена на $(0, 1]$. Кроме того, эта случайная величина не зависит от моделируемой точки на окружности.

Следовательно, если полярный алгоритм моделирования равномерного распределения на \mathbb{S}^1 используется в цикле, то можно сэкономить одно обращение к генератору псевдослучайных чисел, используя для этой цели квадрат полярного радиуса d , полученный при предыдущем моделировании. При такой организации нам потребуется в среднем $8/\pi - 1 \approx 1.55$ обращений к генератору псевдослучайных чисел.⁷⁵

5.1.2 Нормальное распределение

Быстрому и качественному моделированию нормального распределения $N(a, \sigma^2)$ посвящено множество публикаций (см., например обзор в [10, гл. IX]; некоторые методы подробно описаны в [19, раздел 3.4.1]). Однако самые эффективные из разобранных там алгоритмов являются непростыми для реализации, поэтому мы остановимся лишь на наиболее популярных методах.

Конечно, нам достаточно уметь моделировать стандартное нормальное распределение $N(0, 1)$. Если в наличии имеется быстрая и точная процедура вычисления обратной функции Ψ к функции распределения

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

стандартного нормального распределения, то, вполне возможно, нужно использовать именно этот способ моделирования. Альтернативой является полярный метод, где моделируются две независимые случайные величины ξ_1, ξ_2 , каждая из которых имеет распределение $N(0, 1)$.

Действительно, поскольку совместная плотность распределения ξ_1, ξ_2 имеет вид

$$p(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}, \quad x, y \in \mathbb{R},$$

то, как следует из раздела 7.2.1 Приложения 1, полярный радиус r и полярный угол φ случайного вектора (ξ_1, ξ_2) независимы, причем полярный угол равномерно распределен на $[0, 2\pi)$, а полярный радиус имеет распределение Рэлея (4.1.3). Отсюда, применяя моделирующую формулу (4.1.4), сразу же приходим к представлению

$$\xi_1 = \sqrt{-2 \ln(\alpha_1)} \cos(2\pi\alpha_2), \quad \xi_2 = \sqrt{-2 \ln(\alpha_1)} \sin(2\pi\alpha_2), \quad (5.1.2)$$

где $\alpha_1, \alpha_2 \in U(0, 1)$.⁷⁶

Конечно, формулы (5.1.2) можно непосредственно использовать для моделирования (сразу двух независимых) реализаций случайных величин с нормальным распределением. Заменяя, как и в разделе 5.1.1, вычисление синуса и косинуса моделированием вектора, равномерно распределенного на \mathbb{S}^1 , приходим к следующему полярному алгоритму моделирования нормального распределения.

⁷⁵Конечно, возможно и другое использование случайной величины d , не связанное с моделированием равномерного распределения на \mathbb{S}^1 .

⁷⁶Иногда формулу (5.1.2) называют *преобразованием Бокса-Мюллера*.

Моделирование нормального распределения $N(0, 1)$ полярным методом

Результат: (ξ_1, ξ_2) .

1. (Отбор в круг)

- Do

$\text{Get}(\alpha_1, \alpha_2); \beta_1 \leftarrow 2 * \alpha_1 - 1; \beta_2 \leftarrow 2 * \alpha_2 - 1; d \leftarrow \beta_1 * \beta_1 + \beta_2 * \beta_2;$

- While $d > 1$;

3. (Моделирование) $\text{Get}(\alpha), t \leftarrow \text{sqrt}(-\ln(2 * \alpha)/d); \xi_1 \leftarrow \beta_1 * t; \xi_2 \leftarrow \beta_2 * t$; STOP.

Ясно, что этот алгоритм требует в среднем $8/\pi + 1$ обращений к генератору, а также по одному вычислению логарифма и квадратного корня. Поскольку в результате получается две независимые реализации нормального распределения, то на каждую из них приходится $4/\pi + 0.5 \approx 1.77$ обращений к генератору, а также по 0.5 вычислений логарифма и квадратного корня.

Для уменьшения числа обращений к генератору псевдослучайных чисел можно использовать идею Замечания 9. А именно, для моделирования полярного радиуса вместо α можно взять величину d , которая не зависит от полярного угла и равномерно распределена на $(0, 1]$. Тогда мы придем к модифицированному полярному методу.

Моделирование нормального распределения $N(0, 1)$ модифицированным полярным методом

Результат: (ξ_1, ξ_2) .

1. (Отбор в круг)

- Do

$\text{Get}(\alpha_1, \alpha_2); \beta_1 \leftarrow 2 * \alpha_1 - 1; \beta_2 \leftarrow 2 * \alpha_2 - 1; d \leftarrow \beta_1 * \beta_1 + \beta_2 * \beta_2;$

- While $d > 1$;

3. (Моделирование) $t \leftarrow \text{sqrt}(-\ln(2 * d)/d); \xi_1 \leftarrow \beta_1 * t; \xi_2 \leftarrow \beta_2 * t$; STOP.

Здесь на одну реализацию нормального распределения приходится $4/\pi \approx 1.27$ обращений к генератору.

5.1.3 Равномерное распределение на сфере \mathbb{S}^2

Результаты, изложенные в Приложении 1 (см. раздел 7.2.2), позволяют выдвинуть несколько идей по поводу моделирования равномерного распределения на сфере $\{(x, y, z) : x^2 + y^2 + z^2 = 1\}$.⁷⁷

1. Если у нас есть быстрый метод моделирования случайного вектора $\bar{\eta} \in \mathbb{R}^3$ с плотностью распределения вида $f(\sqrt{x^2 + y^2 + z^2})$, то нормировка этого вектора $\bar{\xi} = \bar{\eta}/\|\bar{\eta}\|$ даст нам искомый результат. У нас уже есть два естественных варианта реализации такой идеи.

- Можно в качестве распределения вектора $\bar{\eta}$ взять равномерное распределение в шаре $\{(x, y, z) : x^2 + y^2 + z^2 \leq 1\}$ и моделировать это распределение отбором из куба $[-1, 1]^3$. Иначе говоря, можно попробовать перенести на трехмерное пространство метод, описанный в разделе 5.1.1. Отбор из трехмерного куба во вписанный шар, однако, имеет среднее число циклов, равное $6/\pi \approx 1.91$, так что метод оказывается слишком трудоемким.⁷⁸

⁷⁷Интерес к равномерному распределению на \mathbb{S}^2 связан с необходимостью в ряде физических задач моделировать случайное направление в \mathbb{R}^3 .

⁷⁸В больших размерностях ситуация еще хуже, так как отношение объема шара к объему описанного вокруг него куба быстро стремится к нулю с увеличением размерности.

- Можно также взять $\bar{\eta} = (\eta_1, \eta_2, \eta_3)^T$, где $\eta_i \in N(0, 1)$ и независимы. Если у нас есть быстрый метод моделирования нормального распределения, этот метод глядится вполне приемлемым.⁷⁹

2. Как отмечено в разделе 7.2.2 Приложения 1, координаты случайного вектора, равномерно распределенного на сфере \mathbb{S}^2 , могут быть представлены в виде

$$\xi_1 = \cos(\varphi) \cos(\theta), \quad \xi_2 = \sin(\varphi) \cos(\theta), \quad \xi_3 = \sin(\theta), \quad (5.1.3)$$

где (см. формулу (7.2.8)) случайные величины φ и θ независимы, $\varphi \in U(0, 2\pi)$ и θ имеет плотность распределения $p_\theta(x) = 0.5 \cos(x) \mathbf{1}_{(-\pi/2, \pi/2)}(x)$.

Сама по себе моделирующая формула (5.1.3) вряд ли пригодна для практики. Можно, однако, заметить, то условное распределение вектора $(\xi_1, \xi_2)^T$ при условии $\xi_3 = t$ является равномерным на окружности радиуса $\sqrt{1 - t^2}$ с центром в нуле. Поскольку $\xi_3 \in U(-1, 1)$ (об этом также см. раздел 7.2.2), то появляется реальная альтернатива моделирующей формуле (5.1.3).

Обсудим алгоритм, реализующий эти рассуждения. Начинается алгоритм с моделирования равномерного распределения на окружности полярным методом раздела 5.1.1 (это нужно для генерации первых двух координат вектора $\bar{\xi}$), а получающаяся при этом случайная величина d , равномерно распределенная на $(0, 1]$, будет использоваться для моделирования ξ_3 .

Иначе говоря, после применения полярного метода раздела 5.1.1, мы получим случайный вектор $\bar{\eta} = (\eta_1, \eta_2)^T$, равномерно распределенный на \mathbb{S}^1 , и случайную величину $d \in U(0, 1)$, независимую с $\bar{\eta}$. Конечно, мы полагаем $\xi_3 = 2d - 1$. Что касается вектора $(\xi_1, \xi_2)^T$, то он будет получаться умножением $\bar{\eta}$ на величину $\sqrt{1 - \xi_3^2} = \sqrt{1 - (2d - 1)^2} = 2\sqrt{d(1 - d)}$.

Запишем соответствующий алгоритм.

Моделирование равномерного распределения на сфере \mathbb{S}^2 полярным методом

Результат: (ξ_1, ξ_2, ξ_3) .

1. (Отбор в круг)

- Do

Get(α_1, α_2); $\beta_1 \leftarrow 2 * \alpha_1 - 1$; $\beta_2 \leftarrow 2 * \alpha_2 - 1$; $d \leftarrow \beta_1 * \beta_1 + \beta_2 * \beta_2$;

- While $d > 1$;

3. (Нормировка и результат) $t \leftarrow 2 * \text{sqrt}(1 - d)$; $\xi_1 \leftarrow \beta_1 * t$; $\xi_2 \leftarrow \beta_2 * t$; $\xi_3 \leftarrow 2 * d - 1$; STOP.

Отметим, что множитель $2\sqrt{1 - d}$ в третьем пункте алгоритма получается из произведения нормировки $1/\sqrt{d}$ в полярном методе для \mathbb{S}^2 и множителя $2\sqrt{d(1 - d)}$ для ξ_1, ξ_2 .

Как и в полярном методе для моделирования равномерного распределения на окружности, здесь требуется (в среднем) $8/\pi \approx 2.55$ обращений к генератору псевдослучайных чисел и одно извлечение квадратного корня.

5.1.4 Распределение Коши

Обратные функции. В разделе 4.1.1 показано, что (модифицированный) метод обратных функций для моделирования распределения Коши сводится к использованию моделирующей формулы $\xi = \cot(\pi\alpha) = \cos(\pi\alpha)/\sin(\pi\alpha)$. Если функция котангенс считается достаточно быстро, то метод обратных функций оказывается вполне конкурентоспособным.

⁷⁹Более того, он, вообще говоря, годится для сферы любой не слишком большой размерности $d > 1$, поскольку трудоемкость этого метода растет линейно с ростом d .

Полярный метод. Так как вектор $(\sin(\pi\alpha), \cos(\pi\alpha))^T$ равномерно распределен на половине окружности $\{(x, y) : y > 0, x^2 + y^2 = 1\}$, то не представляет труда описать полярный метод моделирования распределения Коши: нужно методом отбора из прямоугольника $[-1, 1] \times [0, 1]$ получить вектор $\bar{\eta}$, имеющий равномерное распределение в полукруге $\{(x, y) : y > 0, x^2 + y^2 < 1\}$ и взять отношение его ординаты к абсциссе.⁸⁰

Моделирование распределения Коши полярным методом

Результат: ξ .

1. (Отбор в полукруге)

- Do
Get(α_1, α_2); $\beta \leftarrow 2 * \alpha_1 - 1$; $d \leftarrow \beta * \beta + \alpha_2 * \alpha_2$;
- While $d > 1$;

3. (Результат) $\xi \leftarrow \alpha_2 / \beta$; STOP.

В среднем метод требует $8/\pi$ обращений к генератору. Кроме того, «бесплатно» моделируется случайная величина $d \in U(0, 1)$, независимая от ξ .

Моделирование через нормально распределенные случайные величины. Хорошо известно, что отношение двух независимых $N(0, 1)$ -распределенных случайных величин имеет распределение Коши.⁸¹ Этот факт можно использовать при моделировании, если есть быстрый метод генерирования нормального распределения.⁸²

5.2 Моделирование многомерных гауссовских распределений

При моделировании многомерных гауссовских распределение (определение и свойства см. в Приложении 2) источником случайности естественно считать последовательность $\{\varepsilon_n, n \geq 1\}$ независимых нормально распределенных случайных величин со нулевым средним и единичной дисперсией. Как обычно в таких случаях, метод моделирования $\varepsilon_n \in N(0, 1)$ мы уточнять не будем.

Итак, рассмотрим два наиболее популярных метода моделирования распределения $N_d(\bar{a}, \Sigma)$.

Метод разложения ковариационной матрицы. Пусть $\mu_1 \geq \dots \geq \mu_d \geq 0$ — собственные числа матрицы Σ , упорядоченные по убыванию с учетом кратности. Как и в Лемме 8 Приложения 2, обозначим U_1, \dots, U_d соответствующую ортонормированную систему собственных векторов и положим $U = [U_1 : \dots : U_d]$. Кроме того, обозначим Λ диагональную матрицу с элементами $\lambda_{jj} = \mu_j$. Наконец, пусть $\bar{\mathcal{E}}$ — d -мерный вектор с компонентами $\varepsilon_1, \dots, \varepsilon_d$.

Как следует из той же Леммы 8, в этих условиях $U\Lambda^{1/3}\bar{\mathcal{E}} \in N_d(\mathbf{0}, \Sigma)$. Следовательно,

$$\bar{\xi} \stackrel{\text{def}}{=} \bar{a} + U\Lambda^{1/2}\bar{\mathcal{E}} \in N_d(\bar{a}, \Sigma). \quad (5.2.1)$$

Формула (5.2.1) по существу является моделирующей формулой для распределения $N_d(\bar{a}, \Sigma)$. Соответствующий метод (*метод разложения ковариационной матрицы*) может считаться универсальным, если только есть уверенность в достаточной точности нахождения собственных чисел и собственных векторов матрицы Σ .⁸³

⁸⁰ Нормировать вектор $\bar{\eta}$ на его длину не нужно, так как нас интересует только котангенс соответствующего угла.

⁸¹ Проверьте!

⁸² Конечно, полярный метод раздела 5.1.2 для этих целей не годится.

⁸³ В некоторых случаях, когда матрица Σ имеет простую структуру, оказывается возможным находить в явном виде линейное преобразование, переводящее $\bar{\mathcal{E}}$ в $\bar{\xi}$.

Замечание 10. Обозначим $k = \max(j : \mu_j > 0)$ и предположим, что $k < d$. Кроме того, пусть $U^{(k)} = [U_1 : \dots : U_k]$, Λ_k — диагональная $k \times k$ матрица с $\lambda_{jj} = \mu_j > 0$, и $\mathcal{E}_k = (\varepsilon_1, \dots, \varepsilon_k)^T$. Тогда, как нетрудно видеть,⁸⁴ (5.2.1) переписывается как $\bar{\xi} = \bar{a} + U^{(k)} \Lambda_k^{1/2} \bar{\mathcal{E}}_k$. Таким образом, если матрица Σ вырождена и имеет ранг k , то для моделирования распределения $N_d(\bar{a}, \Sigma)$ нужно k (а не d) раз генерировать стандартное нормальное распределение.

Невырожденное распределение. Итерационный метод. Итерационный метод моделирования многомерного невырожденного гауссовского распределения основан на следующем простом утверждении.⁸⁵

Рассмотрим симметричную положительно определенную (и, следовательно, невырожденную) матрицу Σ размерности $d \times d$. При $1 \leq j \leq d$ обозначим Σ_j подматрицу матрицы Σ , находящуюся на пересечении первых j столбцов с первыми j строками Σ . Последний столбец матрицы Σ_j обозначим $(\bar{\beta}_j^T, \sigma_j^2)^T$, где σ_j^2 — j -й диагональный элемент матрицы $\Sigma = \Sigma_d$, а $\bar{\beta}_j \in \mathbb{R}^{j-1}$.

Кроме того, для произвольного вектора $\bar{a} = (a_1, \dots, a_d)^T \in \mathbb{R}^d$ обозначим $\bar{a}_j = (a_1, \dots, a_j)^T$.

Лемма 4. 1. Для любого $1 \leq k < d$ выполнено неравенство $\Delta_{k+1}^2 \stackrel{\text{def}}{=} \sigma_{k+1}^2 - \bar{\beta}_{k+1}^T \Sigma_k^{-1} \bar{\beta}_{k+1} > 0$.
2. Пусть $\bar{\xi}_k \in N_k(\bar{a}_k, \Sigma_k)$ и $\varepsilon_{k+1} \in N(0, 1)$, причем ε_{k+1} и $\bar{\xi}_k$ независимы. Обозначим

$$\xi_{k+1} = a_{k+1} + \bar{\beta}_{k+1}^T \Sigma_k^{-1} (\bar{\xi}_k - \bar{a}_k) + \Delta_{k+1} \varepsilon_{k+1} \quad (5.2.2)$$

и положим $\bar{\xi}_{k+1} = (\bar{\xi}_k^T, \xi_{k+1})^T$. Тогда $\bar{\xi}_{k+1} \in N_k(\bar{a}_{k+1}, \Sigma_{k+1})$.

Доказательство. 1. По условию, $X^T \Sigma_{k+1} X > 0$ для любого $X \neq \mathbf{0}$. Возьмем $X = (Y^T, -1)^T$ где $Y = \Sigma_k^{-1} \bar{\beta}_{k+1}$. Так как

$$\Sigma_{k+1} = \begin{pmatrix} \Sigma_k & \bar{\beta}_{k+1} \\ \bar{\beta}_{k+1}^T & \sigma_{k+1}^2 \end{pmatrix},$$

то

$$\begin{aligned} 0 < X^T \Sigma_{k+1} X &= (Y^T, -1) \begin{pmatrix} \Sigma_k & \bar{\beta}_{k+1} \\ \bar{\beta}_{k+1}^T & \sigma_{k+1}^2 \end{pmatrix} \begin{pmatrix} Y \\ -1 \end{pmatrix} = (Y^T, -1) \begin{pmatrix} \Sigma_k Y - \bar{\beta}_{k+1} \\ \bar{\beta}_{k+1}^T Y - \sigma_{k+1}^2 \end{pmatrix} = \\ &= Y^T \Sigma_k Y - 2 \bar{\beta}_{k+1}^T Y + \sigma_{k+1}^2 = \bar{\beta}_{k+1}^T \Sigma_k^{-1} \Sigma_k \Sigma_k^{-1} \bar{\beta}_{k+1} - 2 \bar{\beta}_{k+1}^T \Sigma_k^{-1} \bar{\beta}_{k+1} + \sigma_{k+1}^2 = \\ &= \sigma_{k+1}^2 - \bar{\beta}_{k+1}^T \Sigma_k^{-1} \bar{\beta}_{k+1}. \end{aligned}$$

2. Конечно, достаточно рассмотреть случай $\bar{a} = \mathbf{0}$. Прежде всего, случайный вектор $\bar{\eta}_{k+1} = (\bar{\xi}_k^T, \varepsilon_{k+1})^T$ является гауссовским (это проверяется непосредственно по определению, см. Приложение 2). Следовательно, вектор $\bar{\xi}_{k+1} = (\bar{\xi}_k^T, \xi_{k+1})^T$ тоже гауссовский как результат линейного преобразования вектора $\bar{\eta}_{k+1}$.⁸⁶

Поэтому нам осталось проверить, что ковариации случайной величины ξ_{k+1} с координатами случайного вектора $\bar{\xi}_{k+1}$ ровно такие, как нужно. Действительно, так как $\mathbb{E} \bar{\xi}_k \bar{\xi}_k^T = \Sigma_k$, то

$$\mathbb{E} \xi_{k+1} \bar{\xi}_k^T = \mathbb{E} (\bar{\beta}_{k+1}^T \Sigma_k^{-1} \bar{\xi}_k + \Delta_{k+1} \varepsilon_{k+1}) \bar{\xi}_k^T = \bar{\beta}_{k+1}^T \Sigma_k^{-1} \mathbb{E} \bar{\xi}_k \bar{\xi}_k^T = \bar{\beta}_{k+1}^T \Sigma_k^{-1} \Sigma_k = \bar{\beta}_{k+1}^T.$$

⁸⁴Проверьте!

⁸⁵Существует различные способы доказательства Леммы 4 (например, можно явно выписать линейное отображение $(\bar{\xi}_k^T, \varepsilon_{k+1})^T \mapsto \bar{\xi}_{k+1}$ и затем пересчитать соответствующие плотности). Наиболее естественный из этих способов использует тот факт, что условное математическое ожидание ξ_{k+1} при условии $\bar{\xi}_k$ совпадает с линейной регрессией ξ_{k+1} на $\bar{\xi}_k$ (ср. Пример 4 раздела 4.3.3 про моделирование двумерного нормального распределения). Приведенное ниже доказательство выбрано из соображений замкнутости и компактности изложения.

⁸⁶Это доказано в Лемме 8 Приложения 2.

Кроме того,

$$\begin{aligned}\mathbb{E}\xi_{k+1}^2 &= \mathbb{E}(\bar{\beta}_{k+1}^T \Sigma_k^{-1} \bar{\xi}_k \bar{\xi}_k^T \Sigma_k^{-1} \bar{\beta}_{k+1}) + \Delta_{k+1}^2 = \\ &= \bar{\beta}_{k+1}^T \Sigma_k^{-1} \mathbb{E}(\bar{\xi}_k \bar{\xi}_k^T) \Sigma_k^{-1} \bar{\beta}_{k+1} + \sigma_{k+1}^2 - \bar{\beta}_{k+1}^T \Sigma_k^{-1} \bar{\beta}_{k+1} = \sigma_{k+1}^2.\end{aligned}$$

Утверждение доказано. \square

Замечание 11. Для фиксированного k второе утверждение леммы остается верным, если отказаться от предположения, что матрица Σ_{k+1} невырожденная (достаточно ее неотрицательной определенности). Если $\text{rank } \Sigma_{k+1} < k+1$, то $\Delta_{k+1}^2 = 0$.⁸⁷

Формулы (5.2.2) при $1 \leq k < d$ фактически являются моделирующими для распределения $N_d(\bar{a}, \Sigma)$. К ним нужно только присоединить инициализацию: $\xi_1 = a_1 + \sigma_1 \varepsilon_1$.

В том виде, в котором он записан, алгоритм, основанный на Лемме 4, вообще говоря, очень трудоемкий.⁸⁸ Тем не менее он может быть напрямую использован при относительно небольших d (см. Замечание 8 раздела 4.3.3 для $d = 2$) при помощи следующего приема, позволяющего проводить вычисление векторов $\beta_{k+1}^T \Sigma_k^{-1}$ в явном виде.

Пусть в условиях Леммы 4 средние a_j равны нулю при любом j . Кроме того, обозначим r_{ij} элементы матрицы Σ . Тогда равенства (5.2.2) можно переписать в виде

$$\xi_{k+1} = a_{k+1,1}\varepsilon_1 + \dots + a_{k+1,k}\varepsilon_k + a_{k+1,k+1}\varepsilon_{k+1}, \quad k = 0, \dots, d-1, \quad (5.2.3)$$

где $a_{i,j}$ — некоторые коэффициенты. Если, как и в (5.2.1), ввести d -мерные вектора-столбцы $\bar{\varepsilon}$ и $\bar{\xi}$ с компонентами $\varepsilon_1, \dots, \varepsilon_d$ и ξ_1, \dots, ξ_d соответственно, то равенства (5.2.3) переписутся в матричной форме $\bar{\xi} = A\bar{\varepsilon}$, где A — нижняя треугольная матрица с элементами a_{ij} (то есть $a_{ij} = 0$ при $i < j$).

Тогда, очевидно, $\Sigma = \mathbb{E}\bar{\xi}\bar{\xi}^T = \mathbb{E}A\bar{\varepsilon}\bar{\varepsilon}^T A^T = AA^T$, и мы приходим к классическому разложению Холецкого положительно определенной матрицы. Рекуррентные формулы, выражающие элементы a_{ij} через r_{ij} хорошо известны: $a_{11} = \sqrt{r_{11}}$, $a_{1j} = r_{1j}/a_{11}$ при $1 < j \leq d$,

$$a_{ij} = \frac{r_{ij} - \sum_{\ell=1}^{j-1} r_{i\ell} r_{\ell j}}{a_{jj}} \quad \text{при } i > j, \quad a_{jj} = \sqrt{r_{jj} - \sum_{k=1}^{j-1} a_{jk}^2}.$$

5.3 Разные распределения

В этом разделе мы затронем методы моделирования нескольких популярных распределений. Как всегда, будут рассматриваться лишь простые для реализации алгоритмы.

5.3.1 Показательное распределение

Стандартный метод моделирования распределения EXP(1) — это (модифицированный) метод обратных функций, приводящий к моделирующей формуле $\xi = -\ln(\alpha)$ (см. раздел 4.1.1).

Если положить $d = 1$, $k = 1$, $k_1 = k_2 = 1$ и $\mu = 1$ во втором утверждении Предложения 16 (см. раздел 7.2.3 Приложения 1), то можно получить метод моделирования, альтернативный методу обратных функций. Действительно, в этом случае оказывается, что случайные величины $\xi_1 = \eta\alpha$ и $\xi_2 = \eta(1 - \alpha)$ показательны распределены с параметром 1 и независимы, если только независимы $\eta \in \text{Gamma}(2, 1)$ и $\alpha \in U(0, 1)$.⁸⁹ Поскольку случайная величина η может быть промоделирована как $-\ln(\alpha_1\alpha_1)$,⁹⁰ то мы приходим к следующему алгоритму.

⁸⁷Убедитесь в этом.

⁸⁸Алгоритм получается рекуррентным и при каждом k нужно обращаться матрицу Σ_k .

⁸⁹Таким образом, η «расщепляется» на 2 показательные независимые случайные величины.

⁹⁰См. Лемму 2 раздела 3.2.

Моделирование показательного распределения расщеплением $\text{Gamma}(2, 1)$

Результат: (ξ_1, ξ_2) .

1. (*Моделирование гамма-распределения*) $\text{Get}(\alpha_1, \alpha_2); \eta \leftarrow -\ln(\alpha_1 * \alpha_2);$
2. (*Результат*) $\text{Get}(\alpha); \xi_1 \leftarrow \eta * \alpha; \xi_2 \leftarrow \eta - \xi_1; \text{STOP}.$

Ясно что приведенный алгоритм требует 3 обращения к генератору псевдослучайных чисел и одно вычисление логарифма, но при этом выдает 2 независимых реализации показательного распределенной случайной величины. Таким образом, на одну реализацию приходится 1.5 обращений к генератору и 0.5 вычислений логарифма. Следовательно, описанный метод расщепления $\text{Gamma}(2, 1)$ будет иметь меньшую трудоемкость, чем метод обратных функций, если обращение к генератору стоит дешевле, чем вычисление логарифма. Конечно, использование другого метода моделирования распределения $\text{Gamma}(2, 1)$ может изменить упорядоченность трудоемкостей этих методов.

Замечание 12. Легко представить себе обобщение приведенного метода, связанное с расщеплением случайной величины $\text{Gamma}(n, 1)$ в сумму n независимых показательных случайных величин. Для этого нужно использовать второе утверждение Предложения 16 с $d = n - 1$, $k = n$, $k_1 = \dots = k_{d+1} = 1$ и $\mu = 1$. Кроме того, Замечание 17 раздела 7.2.4 Приложения 1 позволяет моделировать соответствующее распределение Дирихле как набор приращений n порядковых статистик, полученных из случайных величин $\alpha_1, \dots, \alpha_n$.

Если случайная величина $\eta \in \text{Gamma}(n, 1)$ моделируется как $-\ln(\alpha_1 \dots \alpha_n)$,⁹¹ то для получения одной реализации распределения $\text{EXP}(1)$ требуется $2 - 1/n$ обращений к генератору и одно вычисление логарифма. Кроме того, нужно дополнительное время (и, возможно, дополнительная память) на упорядочивание (и вычисление приращений) $(n - 1)$ -ой реализации случайной величины, имеющей распределение $U(0, 1)$. Ясно, что при больших n этот метод может быть более трудоемким, чем метод обратных функций. Возможно, ситуация изменится, если гамма-распределение будет моделироваться более экономичным образом.

Ряд других методов моделирования показательного распределения⁹² приведен в [10, гл. IX.2].

5.3.2 Степенное распределение

Метод обратных функций для моделирования степенного распределения с плотностью $p_\gamma(x) = \gamma x^{\gamma-1}$ при $x \in (0, 1)$ и $\gamma > 0$ уже обсуждался в разделе 4.1.1. Этот метод порождает моделирующую формулу $\xi = \alpha^{1/\gamma}$ и поэтому его трудоемкость мало зависит от параметра γ . В то же время она может оказаться достаточно большой для практики, так как возведение числа в вещественную степень производится с помощью вычисления логарифма и экспоненты.

При целых $\gamma = n \geq 1$ существует естественная альтернатива методу обратных функций. Она основана на том, что максимум из n независимых одинаково распределенных случайных величин, имеющих функцию распределения F , распределен согласно функции распределения F^n . Отсюда сразу же видно, что случайная величина $\xi = \max(\alpha_1, \dots, \alpha_n)$ имеет степенное распределение с параметром n .⁹³ При небольших n этот способ может оказаться эффективнее метода обратных функций.

В некоторых случаях имеет смысл использовать оператор извлечения квадратного корня, так как он реализуется значительно быстрее, чем оператор возведения в вещественную степень.

5.3.3 Гамма-распределение

Плотность гамма-распределения $\text{Gamma}(k, \mu)$ с параметрами $k, \mu > 0$ задается формулой

$$p_{k,\mu}(x) = \mu^k x^{k-1} e^{-\mu x} / \Gamma(k), \quad x > 0.$$

⁹¹Опять-таки Лемма 2 раздела 3.2.

⁹²Некоторые из них более эффективны, чем описанные выше.

⁹³Заметим, что этот простой факт является частным случаем Следствия 6 (раздел 7.2.4 Приложения 1).

Параметр μ является *параметром масштаба*: если $\eta \in \text{Gamma}(k, 1)$, то $\xi = \eta/\mu \in \text{Gamma}(k, \mu)$.⁹⁴ Поэтому нам достаточно уметь моделировать гамма-распределение с $\mu = 1$. Параметр k (так называемый *параметр формы*) гораздо более важен. Многочисленные методы генерирования гамма-распределения, адаптированные к различным (большим и маленьким) значениям k , можно найти в [10, гл. IX.3]. Здесь мы ограничимся лишь несколькими простыми замечаниями.

Как уже отмечалось (см. Лемму 2 раздела 3.2), в частном случае целого $k > 1$ распределение $\text{Gamma}(k, 1)$ может быть промоделировано очень просто с помощью моделирующей формулы $\xi = -\ln(\alpha_1 \dots \alpha_k)$. При относительно небольших k это выглядит вполне приемлемо.⁹⁵

Переход от целого к вещественному $k > 1$, в принципе, может быть осуществлен с помощью следующего простого факта:⁹⁶ если $\xi_1 \in \text{Gamma}(k_1, \mu)$, $\xi_2 \in \text{Gamma}(k_2, \mu)$ и ξ_1, ξ_2 независимы, то $\xi_1 + \xi_2 \in \text{Gamma}(k_1 + k_2, \mu)$. Поэтому, используя разложение $k = [k] + \{k\}$ и считая, что у нас есть приемлемый метод моделирования гамма-распределения с целым параметром формы, мы приходим к задаче моделирования распределения $\text{Gamma}(k, 1)$ с $k < 1$.

Если $k = 1/2$, то ситуация достаточно проста, так как случайная величина $\xi^2/2$ имеет распределение $\text{Gamma}(1/2, 1)$ при $\xi \in N(0, 1)$.⁹⁷ Приведем один из легко реализуемых методов, пригодных для любого $k < 1$. Заметим, что при $x > 0$

$$\begin{aligned} p_k(x) &= p_{k,1}(x) = \frac{1}{\Gamma(k)} x^{k-1} e^{-x} \leq \frac{1}{\Gamma(k)} x^{k-1} \mathbf{1}_{(0,1)}(x) + \frac{1}{\Gamma(k)} e^{-x} \mathbf{1}_{[1,\infty)}(x) = \\ &= \frac{1}{k\Gamma(k)} kx^{k-1} \mathbf{1}_{(0,1)}(x) + \frac{1}{e\Gamma(k)} e^{-(x-1)} \mathbf{1}_{[1,\infty)}(x) = M(k)q_k(x), \end{aligned} \quad (5.3.1)$$

где

$$M(k) = \frac{1}{k\Gamma(k)} + \frac{1}{e\Gamma(k)}$$

и

$$q_k(x) = \rho_k kx^{k-1} \mathbf{1}_{(0,1)}(x) + (1 - \rho_k) e^{-(x-1)} \mathbf{1}_{[1,\infty)}(x) \quad (5.3.2)$$

с $\rho_k = 1/(k\Gamma(k)M(k)) = e/(e + k)$. При этом равенство в (5.3.1) достигается как при $x = 1$ так и при $x \uparrow 1$. Ясно, что функция q_k является плотностью некоторого распределения.

Отсюда сразу видна идея моделирования: можно использовать метод отбора с мажорирующим распределением \mathcal{Q}_k , имеющим плотность q_k , а само распределение \mathcal{Q}_k моделируется как смесь (с весом ρ_k) степенного распределения с параметром k и показательного распределения, сдвинутого на 1 вправо (с весом $1 - \rho_k$). Заметим также, что, так как $k\Gamma(k) \rightarrow 1$ при $k \rightarrow 0$, то функция $M(k)$ ограничена на промежутке $(0, 1]$.

Поскольку при положительных x

$$\begin{aligned} r_k(x) \stackrel{\text{def}}{=} \frac{p_k(x)}{q_k(x)} &= \begin{cases} \frac{x^{k-1}e^{-x}/\Gamma(k)}{\rho_k kx^{k-1}} & \text{при } 0 < x < 1, \\ \frac{x^{k-1}e^{-x}/\Gamma(k)}{(1 - \rho_k)e^{-(x-1)}} & \text{при } x \geq 1 \end{cases} = \begin{cases} \frac{1}{k\Gamma(k)\rho_k} e^{-x} & \text{при } 0 < x < 1, \\ \frac{1}{e\Gamma(k)(1 - \rho_k)} x^{k-1} & \text{при } x \geq 1 \end{cases} = \\ &= M(k) \begin{cases} e^{-x} & \text{при } 0 < x < 1, \\ x^{k-1} & \text{при } x \geq 1 \end{cases} \leq M(k), \end{aligned}$$

⁹⁴Проверьте!

⁹⁵Еще раз подчеркнем, что при больших k (не обязательно целых) нужно использовать более продвинутые методы, трудоемкость которых не стремится к бесконечности при $k \rightarrow \infty$.

⁹⁶Проверьте!

⁹⁷Проверьте! Кстати, если у нас есть быстрый метод моделирования нормального распределения, то нам достаточно научиться моделировать гамма-распределение с параметром $k < 0.5$.

то при использовании метода отбора нужно с вероятностью ρ_k проверять неравенство $e^{-\eta_1} > \alpha$, а с вероятностью $1 - \rho_k$ — неравенство $\eta_2^{k-1} > \alpha$, где η_1 имеет степенное распределение с параметром k , а $\eta_2 - 1$ — распределение EXP(1). Тем самым мы приходим к следующему алгоритму.⁹⁸

Моделирование распределения Gamma($k, 1$) при $k < 1$ методом отбора с мажорирующим распределением (5.3.2).

Входные данные: $k \in (0, 1)$.

Результат: ξ .

1. (Пересчет параметров) $\rho \leftarrow e/(e+k)$; $a \leftarrow 1/k$; $d \leftarrow \rho/(1-\rho)$;
2. (Отбор)
 - Do
 - ◊ Get(α_1, α_2); $\beta \leftarrow \alpha_1/\rho$;
 - ◊ If $\beta < 1$ then $\eta \leftarrow \beta \wedge a$ else $\eta \leftarrow -\ln(d * (\beta - 1))$;
 - While ($\eta < 1$ and $\exp(-\eta) < \alpha_2$) or ($\eta \geq 1$ and $\eta \wedge (k-1) < \alpha_2$);
3. (Результат) $\xi \leftarrow \eta$; STOP.

Отметим, что в приведенном алгоритме использован результат Предложения 2 раздела 2.4, позволяющий при моделировании смеси распределений сэкономить одно обращение к генератору. Кроме этого, сделаны некоторые простые преобразования.⁹⁹

Функция $M(k)$ принимает максимальное значение ≈ 1.39 при $k \approx 0.8$. В то же время $M(k) \rightarrow 1$ при $k \rightarrow 0$, $M(1/2) \approx 1.34$ и $M(1) = 1 + 1/e \approx 1.37$. Таким образом, метод оказывается наименее трудоемким при k , близких к нулю.¹⁰⁰

5.3.4 Бета-распределение

Бета-распределение Beta(k_1, k_2) с параметрами $k_1, k_2 > 0$ определяется плотностью

$$p_{k_1, k_2}(x) = \frac{\Gamma(k_1 + k_2)}{\Gamma(k_1)\Gamma(k_2)} x^{k_1-1}(1-x)^{k_2-1},$$

сосредоточенной на множестве $(0, 1)$. Как отмечено в разделе 7.2.3 Приложения 1, бета-распределение является частным случаем распределения Дирихле.

Рассмотрим сначала некоторые частные случаи параметров k_1, k_2 . Ясно, что распределение Beta(1, 1) является равномерным на $(0, 1)$. При $k_2 = 1$ мы получаем степенное распределение с параметром $m = k_1$, моделирование которого обсуждалось в разделе 5.3.2.¹⁰¹

Кроме того, если $\xi \in \text{Beta}(k_1, k_2)$, то $1 - \xi \in \text{Beta}(k_2, k_1)$. Отсюда сразу же следует метод моделирования распределения Beta(1, k_2).

Если числа k_1, k_2 — целые, то из Следствия 6 (раздел 7.2.4 Приложения 1) можно вывести следующий способ моделирования распределения Beta(k_1, k_2). Согласно утверждению этого следствия, порядковая статистика $\alpha_{[k]}^{(n)}$, построенная по выборке $\alpha_1, \dots, \alpha_n$, имеет бета-распределение с параметрами $k, n - k + 1$. Поэтому, если положить $n = k_1 + k_2 - 1$, то случайная величина $\alpha_{[k_1]}^{(n)}$ будет иметь нужное нам распределение. При небольших (целых) k_1, k_2 этот способ может оказаться приемлемым.

⁹⁸Алгоритм Аренса и Дитера, см. [20], а также [10, стр. 425].

⁹⁹Эти преобразования можно продолжить. Например, вместо проверки неравенства $\exp(-\eta) < \alpha_2$ можно проверять эквивалентное неравенство $\eta > -\ln(\alpha_2)$. Кроме того, можно применить отбор с ограничениями и т.д.

¹⁰⁰Например, $M(0.1) \approx 1.09$.

¹⁰¹А если $k_2 \gtrsim 1$, то степенное распределение с параметром k_1 можно использовать в качестве мажорирующего для метода отбора.

Обсудим теперь общий случай. Существует большое число алгоритмов, достаточно эффективных либо для произвольных параметров k_1, k_2 , либо для ситуации, когда на эти параметры наложены ограничения.¹⁰² Достаточно полный обзор таких методов можно найти в [10, гл. IX.4]. Мы же здесь ограничимся одним простым для реализации методом, который является вполне конкурентоспособным, если только уже имеется хороший способ моделирования гамма-распределения с произвольными параметрами.

А именно, если положить $d = 1$ в Предложении 16 (раздел 7.2.3 Приложения 1), то окажется, что случайная величина $\eta_1/(\eta_1 + \eta_2)$ имеет бета-распределение с параметрами k_1, k_2 при условии, что случайные величины η_1, η_2 независимы и имеют гамма-распределения $\text{Gamma}(k_1, 1)$ и $\text{Gamma}(k_2, 1)$ соответственно. Этот прием представляется универсальным, хотя и не всегда менее трудоемким, чем более изощренные методы.¹⁰³

¹⁰²Например, есть алгоритмы, специально приспособленные для случаев $k_1 = k_2$, $\min(k_1, k_2) < 1$, $\max(k_1, k_2) > 1$ и т.д.

¹⁰³Все зависит от того, насколько быстро мы умеем моделировать гамма-распределения с нужными параметрами формы.

6 О моделировании с дискретным источником случайности

Материал этого раздела выходит за рамки традиционного моделирования, описанного во Введении, и, вообще говоря, может быть опущен при чтении.

Говоря в предыдущих разделах о моделировании различных распределений, мы под трудоемкостью алгоритмов понимали (среднее) число основных операций, необходимых для получения одной реализации случайной величины или случайного вектора. При этом предполагалось, что источником случайности является последовательность $\alpha_1, \dots, \alpha_n, \dots$ независимых случайных величин, имеющих равномерное распределение на интервале $(0, 1]$.¹⁰⁴

На самом деле, однако, генераторы псевдослучайных чисел порождают, как правило, последовательность чисел a_1, \dots, a_n, \dots таких, что для некоторого $M > 1$ выполняются равенства $a_i = x_i/M$, где $x_i \in \{0, \dots, M-1\}$.¹⁰⁵ Тем самым предположение о том, что случайные величины α_i принимают вещественные (а не рациональные) значения оказывается неверным и может служить источником ошибок.

С этой точки зрения более естественным является предположение, что источником случайности является последовательность независимых случайных величин $\varepsilon_1, \dots, \varepsilon_n, \dots$, каждая из которых равномерно распределена на множестве $\{0, \dots, M-1\}$.

Оказывается, что если ограничиваться моделированием дискретных распределений,¹⁰⁶ то такой подход дает возможность описывать оптимальные (в смысле среднего числа обращений к генератору) методы и алгоритмы моделирования. Более того, эти результаты могут быть полезны и для традиционного моделирования.

Начнем с некоторых общих положений.¹⁰⁷

6.1 Корневые диаграммы и сложность моделирования

Рассмотрим задачу моделирования дискретного распределения \mathcal{P} , сосредоточенного на множестве $X = \{x_1, \dots, x_N\}$ и заданного таблицей распределения

$$\mathcal{P}: \begin{pmatrix} x_1 & \dots & x_k & \dots & x_N \\ p_1 & \dots & p_k & \dots & p_N \end{pmatrix}, \quad (6.1.1)$$

предполагая, что источником случайности является последовательность независимых случайных величин $\{\varepsilon_i\}_{i \geq 1}$, каждая из которых имеет распределение

$$\mathcal{Q}: \begin{pmatrix} 0 & \dots & k & \dots & M-1 \\ q_0 & \dots & q_k & \dots & q_{M-1} \end{pmatrix} \quad (6.1.2)$$

с некоторыми вероятностями q_i . Для простоты записи числа N и M считаются конечными, хотя некоторые из дальнейших рассуждений верны и для случая, когда N и/или M бесконечны.

Традиционно считается, что такая постановка задачи восходит к [21], где обсуждается случай $M = N = 2$, $p_1 = p_2 = 1/2$ и $q_2 = q \in (0, 1)$, то есть рассматривается моделирование симметричных испытаний Бернулли с помощью несимметричных.¹⁰⁸

Нас будет интересовать конструирование и исследование алгоритмов моделирования в достаточной мере произвольных распределений (6.1.1) в случае, когда распределение (6.1.2) является простым и универсальным. Здесь можно выделить статью [11], где обсуждается оптимальное моделирование распределения \mathcal{P} с помощью случайных битов, то есть в том случае, когда распределение \mathcal{Q} представляет собой симметричное распределение Бернулли. Соответствующие алгоритмы опубликованы в [22] и [23].

¹⁰⁴ Кроме того, предполагалось, что все математические операции над числами производятся абсолютно точно. Мы сейчас не обращаем внимание на последнее обстоятельство.

¹⁰⁵ Иногда a_i имеют вид $(bx_i + c)/M$. Для нашего анализа это не является принципиальным.

¹⁰⁶ Для перехода к теоретически непрерывным распределениям нужно учитывать представление чисел с плавающей точкой. Мы здесь не будем этим заниматься.

¹⁰⁷ Изложение ведется на основе статьи [26].

¹⁰⁸ Решение этой задачи очень простое: если $\{\varepsilon_i, i \geq 1\}$ — последовательность независимых случайных величин с распределением $\text{Ber}(q)$, то нужно положить $\tau = \min\{i : \varepsilon_{2i-1} \neq \varepsilon_{2i}\}$ и взять $\xi = \varepsilon_{2\tau}$. Обоснуйте эту процедуру. Что можно сказать о ее трудоемкости?

Как правило, в общем случае моделирование распределения (6.1.1) с помощью независимых случайных величин ε_i , имеющих распределение (6.1.2), производится методом корневых диаграмм. Корневая диаграмма описывается набором D цепочек (i_1, \dots, i_r) с $i_\ell \in \{0, \dots, M-1\}$, удовлетворяющих условию префикса: если $(i_1, \dots, i_r) \in D$, то $(i_1, \dots, i_r, j_1, \dots, j_m) \notin D$ при любых $m \geq 1$ и $j_1, \dots, j_m \in \{0, \dots, M-1\}$.

Кроме того, предполагается, что

$$\sum_{(i_1, \dots, i_r) \in D} q_{i_1} \dots q_{i_r} = 1. \quad (6.1.3)$$

В этом случае мы будем говорить, что корневая диаграмма D *согласована с распределением* \mathcal{Q} и будем обозначать ее $D = D(\mathcal{Q})$.

Наконец, пусть корневая диаграмма $D(\mathcal{Q})$ связана с распределением \mathcal{P} при помощи отображения $f : D \mapsto X$ такого, что для любого $x_\ell \in X$

$$\sum_{f^{-1}(\{x_\ell\})} q_{i_1} \dots q_{i_r} = p_\ell \quad (6.1.4)$$

(конечно, (6.1.3) следует из (6.1.4)). Тогда употребляется термин *корневая диаграмма* $D(\mathcal{P}, \mathcal{Q}, f)$, *порождающая распределение* \mathcal{P} .

Положим теперь

$$\tau = \min\{n : (\varepsilon_1, \dots, \varepsilon_n) \in D\}. \quad (6.1.5)$$

Тогда, очевидно, (6.1.3) означает, что τ конечно почти всюду, равенство (6.1.4) переписывается в виде $\mathbb{P}(f(\varepsilon_1, \dots, \varepsilon_\tau) = x_\ell) = p_\ell$, и общая схема моделирования распределения \mathcal{P} выглядит следующим образом: ищется наименьшее n такое, что $(\varepsilon_1, \dots, \varepsilon_n) \in D$ и полагается $\xi = f(\varepsilon_1, \dots, \varepsilon_n)$.

Пример. Рассмотрим простейший случай $M = N = 2$, $q_0 = q_1 = 1/2$, $x_1 = 0$, $x_2 = 1$, $p_1 = 1/3$ и $p_2 = 2/3$. Тогда в качестве корневой диаграммы можно выбрать множество D , состоящее из элементов $(1), (01), (001), \dots$. Очевидно, равенство (6.1.3) при этом будет выполняться. Отображение $f : D \mapsto \{0, 1\}$ определим следующим образом: если цепочка $(0 \dots 01)$, входящая в D , имеет четное число нулей, то значение f на этой цепочке равно 1, иначе это значение равно нулю. Поскольку

$$1/3 = \sum_{j=1}^{\infty} 2^{-2j}, \quad \text{а} \quad 2/3 = \sum_{j=1}^{\infty} 2^{-2j+1},$$

то (6.1.4) тоже выполняется. Моделирующая процедура здесь очень простая: проводятся симметричные испытания Бернулли до первого успеха, если он наступает на нечетном испытании, то моделируемой случайной величине присваивается значение 1, если на четном — то 0.¹⁰⁹

Вернемся к общему случаю и будем интерпретировать случайные величины ε_i как результаты обращения к некоторому генератору случайных чисел, распределенных согласно (6.1.2).

Естественной характеристикой трудоемкости описанного выше метода моделирования является величина $\mathbb{E}\tau$ — среднее число случайных величин ε_j , необходимое для получения одной реализации случайной величины ξ . Иначе говоря, при моделировании распределения \mathcal{P} согласно корневой диаграмме $D(\mathcal{P}, \mathcal{Q}, f)$, величина $\mathcal{E}\tau$ — это просто среднее число обращений к генератору.

Точное выражение для $\mathbb{E}\tau$ хорошо известно в теории информации (например, [25, гл. 1 §4]). Для того, чтобы доказать соответствующий результат вероятностными методами, остановимся сначала на так называемом *тождестве Вальда*.

Лемма 5. Пусть $\xi_1, \dots, \xi_n, \dots$ — последовательность независимых случайных величин, имеющих математическое ожидание a и таких, что $\mathbb{E}|\xi_i| \leq b < \infty$ для всех i . Положим $S_n = \xi_1 + \dots + \xi_n$. Рассмотрим случайную величину τ с конечным математическим ожиданием, принимающую целые положительные значения и такую, что для любого $i \geq 1$ случайная величина ξ_i и событие $\tau \geq i$ независимы. Тогда у случайной величины S_τ существует математическое ожидание, причем $\mathbb{E}S_\tau = a\mathbb{E}\tau$.

¹⁰⁹Этот и другие подобные примеры подробно разобраны в [11].

Доказательство. Заметим, что

$$\begin{aligned} b\mathbb{E}\tau &= \sum_{i \geq 1} b\mathbb{P}(\tau \geq i) \geq \sum_{i \geq 1} \mathbb{E}|\xi_i| \mathbb{P}(\tau \geq i) = \sum_{i \geq 1} \mathbb{E}(|\xi_i|, \tau \geq i) = \sum_{i \geq 1} \sum_{n \geq i} \mathbb{E}(|\xi_i|, \tau = n) = \\ &= \sum_{n \geq 1} \sum_{i=1}^n \mathbb{E}(|\xi_i|, \tau = n) = \sum_{n \geq 1} \mathbb{E}(|\xi_1| + \dots + |\xi_n|, \tau = n) \geq \sum_{n \geq 1} \mathbb{E}(|S_n|, \tau = n) = \mathbb{E}|S_\tau|. \end{aligned} \quad (6.1.6)$$

Следовательно, $\mathbb{E}|S_\tau| < \infty$. Теперь, повторяя все преобразования (6.1.6) с заменой b на a , $|\xi_i|$ на ξ_i , неравенств на равенства, $|S_n|$ на S_n , и пользуясь тем что (как уже доказано), можно менять порядок суммирования в двойной сумме $\sum_{i \geq 1} \sum_{n \geq i}$, мы получим требуемый результат. \square

Замечание 13. 1. Если случайные величины ξ_i одинаково распределены (или, например, если они неотрицательны), то условие $\mathbb{E}|\xi_i| \leq b < \infty$ выполнено автоматически.
2. Если случайные величины ξ_i неотрицательны, а $\mathbb{E}\tau = \infty$, то равенство $\mathbb{E}S_\tau = a\mathbb{E}\tau$ все равно имеет место в том смысле, что $\mathbb{E}S_\tau = \infty$. Это снова следует из (6.1.6).

Для того, чтобы получить выражение для $\mathbb{E}\tau$, где τ определено в (6.1.5), нам понадобится понятие энтропии дискретного распределения.¹¹⁰

Определение 6.1. Пусть $K > 1$. Рассмотрим дискретное распределение \mathcal{R} с вероятностями $r_i \geq 0$, где i пробегает конечное или счетное число значений и $\sum_i r_i = 1$. Энтропией $H_K(\mathcal{R})$ распределения \mathcal{R} по основанию K называется число

$$H_K(\mathcal{R}) = - \sum_i r_i \log_K r_i, \quad (6.1.7)$$

через \log_K обозначен логарифм по основанию K и предполагается, что $x \log_K(x) = 0$ при $x = 0$.

Заметим, что носитель распределения \mathcal{R} в определении не участвует. Конечно, энтропии одного и того же распределения по различным основаниям отличаются только множителями. Кроме того, энтропия распределения с конечным носителем конечна, а если носитель распределения \mathcal{R} бесконечен, то энтропия $H_K(\mathcal{R})$ тоже может оказаться бесконечной.

Нам понадобятся следующие свойства энтропии.

Лемма 6. 1. Для распределения (6.1.2) с конечным M имеет место неравенство $H_K(\mathcal{Q}) \leq \log_K(M)$. Тем самым при фиксированном M максимальная энтропия достигается на распределении \mathcal{Q} с $q_i = 1/M$.
2. Задав разбиение множества индексов I вероятностей r_i на непересекающиеся подмножества I_j и положив $s_j = \sum_{i \in I_j} r_i$, мы получим новое распределение \mathcal{S} . Соотношение между энтропиями распределений \mathcal{R} и \mathcal{S} задается равенствами $H_K(\mathcal{R}) \geq H_K(\mathcal{S})$.

Доказательство. 1. Рассмотрим случайную величину ξ , принимающую значения $1/q_i$ с вероятностями q_i . Очевидно, $\mathbb{E}\xi = M$ и $\mathbb{E} \log_K(\xi) = H_K(\mathcal{Q})$. Так как функция \log_K является выпуклой вверх, то по неравенству Йенсена $\mathbb{E} \log_K(\xi) \leq \log_K(\mathbb{E}\xi)$.

2. Утверждение следует из цепочки преобразований

$$\begin{aligned} \sum_i r_i \log_K r_i &= \sum_j \sum_{i \in I_j} r_i (\log_K(r_i/s_j) + \log_K(s_j)) = \sum_j \sum_{i \in I_j} r_i \log_K(s_j) + \sum_j \sum_{i \in I_j} r_i \log_K(r_i/s_j) \leq \\ &\leq \sum_j \sum_{i \in I_j} r_i \log_K(s_j) = \sum_j s_j \log_K(s_j). \end{aligned}$$

Доказательство окончено. \square

Вернемся к математическому ожиданию случайной величины (6.1.5).

¹¹⁰Понятие энтропии широко используется в теории информации (см., например, [24] или [25]). Здесь мы ограничиваемся лишь теми свойствами энтропии, которые нам необходимы.

Предложение 12. Рассмотрим распределение \mathcal{Q}_D , имеющее носитель D и такое, что

$$\mathcal{Q}_D((i_1 \dots i_n)) = q_{i_1} \dots q_{i_n}.$$

Тогда для любого $K > 1$

$$H_K(\mathcal{Q}_D) = \mathbb{E} \tau H_K(\mathcal{Q}). \quad (6.1.8)$$

Доказательство. Обозначим $\eta_i = -\log_K(q_{\varepsilon_i})$. Конечно, случайные величины η_i положительны, независимы и одинаково распределены. Кроме того, заметим, что

$$\{\tau \geq i\} = \{\varepsilon_1 \notin D, (\varepsilon_1, \varepsilon_2) \notin D, \dots, (\varepsilon_1, \dots, \varepsilon_{i-1}) \notin D\}.$$

Поэтому событие $\{\tau \geq i\}$ и случайная величина η_i независимы. Далее,

$$\mathbb{E} \eta_i = -\sum_i q_i \log_K(q_i) = H_K(\mathcal{Q}) < \infty,$$

так как $M < \infty$. Следовательно, мы можем применить утверждение Леммы 5 (а также Замечание 13 в случае $\mathbb{E} \tau = \infty$) и получить равенство $\mathbb{E} S_\tau = \mathbb{E} \tau H_K(\mathcal{Q})$, где $S_n = \eta_1 + \dots + \eta_n$.

Осталось доказать, что $\mathbb{E} S_\tau = H_K(\mathcal{Q}_D)$. Действительно, если обозначить D_n подмножество D , состоящее из цепочек (i_1, \dots, i_n) длины n , то окажется, что

$$H_K(\mathcal{Q}_D) = -\sum_{n \geq 1} \sum_{(i_1, \dots, i_n) \in D_n} q_{i_1} \dots q_{i_n} \log_K(q_{i_1} \dots q_{i_n}). \quad (6.1.9)$$

С другой стороны, так как ввиду условия префикса $\{\tau = n\} = \{(\varepsilon_1, \dots, \varepsilon_n) \in D_n\}$, то

$$\mathbb{E} S_\tau = \sum_{n \geq 1} \mathbb{E}(S_n, \tau = n) = -\sum_{n \geq 1} \mathbb{E}(\log_K(\varepsilon_1 \dots \varepsilon_n), (\varepsilon_1, \dots, \varepsilon_n) \in D_n),$$

что в точности совпадает с правой частью (6.1.9). \square

Следствие 5. В условиях Предложения 12 имеет место неравенство¹¹¹ $\mathbb{E} \tau \geq H_K(\mathcal{P})/\log_K(M)$.

Доказательство. Согласно Лемме 6, $H_K(\mathcal{Q}) \leq \log_K M$, а $H_K(\mathcal{Q}_D) \geq H_K(\mathcal{P})$. Отсюда и из (6.1.8) сразу же следует требуемое. \square

6.2 Оптимальные $D_M(\mathcal{P}, f)$ -диаграммы

Итак, пусть у нас имеется бесконечная последовательность независимых одинаково распределенных случайных величин ε_i , принимающих с равной вероятностью значения $0, 1, \dots, M-1$ при некотором $M > 1$. В дальнейшем члены этой последовательности будем называть *случайными M -битами*.

Рассмотрим некоторую корневую диаграмму $D_M(\mathcal{P}, f)$, порождающую распределение \mathcal{P} , и обозначим

$$t_\ell(m) = \text{card} \{(i_1, \dots, i_m) : f((i_1, \dots, i_m)) = x_\ell\}.$$

Тогда условие (6.1.4) превратится в равенство $\sum_{m \geq 1} t_\ell(m)/M^m = p_\ell$.

Введем при $x \in (0, 1)$ функции

$$\nu^{(M)}(x) = \sum_{m \geq 0} \{M^m x\}/M^m$$

(здесь $\{z\}$ обозначает дробную часть числа z) и $e_m(x) = \lfloor M^m x \rfloor \pmod{M}$, так что

$$x = \sum_{m \geq 1} e_m(x) M^{-m}.$$

Пусть, наконец, $\xi = f(\varepsilon_1, \dots, \varepsilon_\tau)$.

¹¹¹Это неравенство носит в теории информации название *обратного неравенства Шеннона для кодирования канала без помех*.

Предложение 13. 1. Для любой корневой $D_M(\mathcal{P}, f)$ -диаграммы и любого $m \geq 0$ имеет место неравенство

$$\mathbb{P}(\tau > m) \geq \sum_{\ell} \{M^m p_{\ell}\} / M^m. \quad (6.2.1)$$

2. Равенства в (6.2.1) достигаются тогда и только тогда, когда $t_{\ell}(m) = e_m(p_{\ell})$ для всех m и ℓ , при этом

$$\mathbb{E}\tau = \sum_{\ell} \nu(p_{\ell}). \quad (6.2.2)$$

Доказательство. 1. Очевидно, $\mathbb{P}(\tau \leq m, \xi = x_{\ell}) = \sum_{1 \leq k \leq m} t_{\ell}(k) / M^k \leq p_{\ell}$, что эквивалентно неравенству

$$\sum_{1 \leq k \leq m} M^{m-k} t_{\ell}(k) \leq \lfloor M^m p_{\ell} \rfloor,$$

так как сумма в его левой части есть целое число. Отсюда сразу же получаем, что

$$\begin{aligned} \mathbb{P}(\tau > m) &= \sum_{\ell} \mathbb{P}(\tau > m, \xi = x_{\ell}) = \sum_{\ell} (p_{\ell} - \mathbb{P}(\tau \leq m, \xi = x_{\ell})) \geq \\ &\geq \sum_{\ell} (p_{\ell} - \lfloor M^m p_{\ell} \rfloor / M^m) = \sum_{\ell} \{M^m p_{\ell}\} / M^m. \end{aligned}$$

2. Если в (6.2.1) достигаются равенства, то

$$t_{\ell}(m) + M \sum_{1 \leq k < m} M^{m-k-1} t_{\ell}(k) = \lfloor M^m p_{\ell} \rfloor$$

для всех ℓ, m . Следовательно, $t_{\ell}(m) = \lfloor M^m p_{\ell} \rfloor \pmod{M} = e_m(p_{\ell})$.

Обратное утверждение очевидно. Представление (6.2.2) следует из определения функции $\nu(x)$ и равенства $\mathbb{E}\tau = \sum_{m \geq 0} \mathbb{P}(\tau > m)$. \square

Замечание 14. Результат Предложения 13 показывает, что выбор $t_{\ell}(m) = e_m(p_{\ell})$ обеспечивает не только минимальное среднее число обращений к генератору, но одновременно и минимальные значения вероятностей $\mathbb{P}(\tau > m)$ при всех $m \geq 1$. В частности, при этом достигается максимум вероятности $\mathbb{P}(\tau = 1)$, который равен

$$\max_{D_M(\mathcal{P}, f)} \mathbb{P}(\tau = 1) = \sum_{\ell} \lfloor M p_{\ell} \rfloor / M. \quad (6.2.3)$$

Предложение 14. Минимальное значение $\mathbb{E}\tau$ удовлетворяет неравенствам

$$\max(1, H_M(\mathcal{P})) \leq \min_{D_M(\mathcal{P}, f)} \mathbb{E}\tau < H_M(\mathcal{P}) + M/(M-1). \quad (6.2.4)$$

Доказательство. Левое неравенство уже обсуждалось в разделе 6.1. Для доказательства правого неравенства заметим, что

$$\nu(x) = \sum_{k \geq 1} k e_k(x) M^{-k}$$

при $x \in (0, 1)$, и определим функцию $H_M(x)$ равенством $H_M(x) = -x \log_M(x)$. Ввиду (6.2.2) достаточно показать, что $\nu(x) < H_M(x) + xM/(M-1)$.

Для $x \in (0, 1)$ найдем такие целые $m > 0$ и $1 \leq j < M$, что $\lfloor xM^m \rfloor = j$. Тогда

$$\begin{aligned} H_M(x) + xM/(M-1) - \nu(x) &= \sum_{k \geq 1} e_k(x) M^{-k} \left(-\log_M(x) + M/(M-1) - k \right) \geq \\ &\geq jM^{-m} \left(M/(M-1) - \log_M(j+1) \right) + \sum_{k > m} e_k(x) M^{-k} \left(m - k + M/(M-1) - \log_M(j+1) \right). \end{aligned} \quad (6.2.5)$$

Так как $M > j \geq 1$, то $0 < M/(M-1) - \log_M(j+1) < 1$, и поэтому первое слагаемое в правой части (6.2.5) положительно, а остальные — отрицательны. Следовательно,

$$\begin{aligned} H_M(x) + xM/(M-1) - \nu(x) &> jM^{-m} \left(M/(M-1) - \log_M(j+1) \right) + \\ &+ \sum_{k>m} (M-1)M^{-k} \left(M/(M-1) - \log_M(j+1) \right) - \sum_{k>m} (M-1)M^{-k}(k-m) = \\ &= M^{-m} \left((j+1) \left(M/(M-1) - \log_M(j+1) \right) - M/(M-1) \right). \end{aligned}$$

Для окончания доказательства осталось заметить, что $y(z/(z-1) - \log_z y) \geq z/(z-1)$ при $z \in (1, +\infty)$ и $y \in (1, z]$. \square

Замечание 15. Из (6.2.4) следует что $\mathbb{E}\tau \rightarrow 1$ при $M \rightarrow \infty$ для любого фиксированного распределения \mathcal{P} с конечной энтропией.

Таким образом, для любого $M > 1$ и любого дискретного распределения \mathcal{P} существует оптимальная корневая диаграмма $D_M(\mathcal{P}, f)$, доставляющая минимум среднему числу обращений к генератору при моделировании распределения \mathcal{P} с помощью M -битов.

Следующей задачей является построение алгоритмических схем для оптимального M -битового моделирования дискретных распределений. Такие алгоритмы можно найти в [26], однако они, вообще говоря, слишком сложны, не имеют практической ценности и поэтому здесь не приводятся.

Вместо этого мы обсудим применение полученных результатов к традиционному моделированию.

6.3 Связь с традиционным моделированием

Многие традиционные методы моделирования дискретных случайных величин используют преобразование вида $\xi = f(\alpha)$, где α — случайная величина, равномерно распределенная на $[0, 1]$, а f — специальным образом подобранная функция. Например, для моделирования равномерного распределения на множестве $\{1, \dots, N\}$ используют моделирующую формулу $\xi = \lceil N\alpha \rceil$, а для геометрического распределения с параметром $p \in (0, 1)$ — преобразование $\xi = \lfloor \ln(\alpha)/\ln(1-p) \rfloor$.

В общем случае этим свойством обладают все методы обратных функций, которые являются теоретически точными, если, в частности, α бесконечное число двоичных разрядов. Как уже говорилось, на практике это не так, и моделирование с помощью метода обратных функций, вообще говоря, приводит к погрешностям.

Попробуем описать ситуации, когда такие погрешности могут оказаться существенными. Предполагая, что источником случайности является последовательность случайных M -битов, мы можем для этого разными способами использовать результаты раздела 6.2.

Во-первых, как следует из неравенств (6.2.4), дискретное распределение \mathcal{P} с большой (намного большей единицы) M -энтропией $H_M(\mathcal{P})$ заведомо требует большого числа обращений к генератору даже при оптимальном моделировании. Соответствующий пример с геометрическим распределением уже разобран в разделе 3.3. Эти соображения, однако, не работают, если энтропия $H_M(\mathcal{P})$ лишь немного превышает 1.

Можно действовать и по-другому. Действительно, формула (6.2.3) позволяет сосчитать вероятность того, что при оптимального моделировании распределения (6.1.1) потребуется ровно одно обращение к генератору. Если эта вероятность окажется существенно меньше 1, то любой метод моделирования распределения (6.1.1), использующий ровно одно обращение к генератору, будет неточным.

Приведем примеры этого подхода для равномерного на множестве $\{1, \dots, N\}$ и геометрического распределений. В первом случае

$$\pi_{\max}(1) \stackrel{\text{def}}{=} \max \mathbb{P}(\tau = 1) = \Lambda \lfloor \Lambda^{-1} \rfloor$$

с $\Lambda = N/M$, что равно нулю при $N > M$. При $N \leq M$ ситуация другая. Например, при типичном выборе $M = 2^{32}$ для $N = 2^k \leq 2^{32}$ требуется ровно одно обращение к генератору, в то время как при $N = 2^k + j$ с $1 \leq j < 2^k$

$$\pi_{\max}(1) = 1 - 2^{-32+k} + j2^{-k} - j2^{-32},$$

что дает $\pi_{\max}(1) \approx 1 - 2^{-32+k}$ при маленьких j и больших k . Тем самым случай N , близких к M , требует внимательности и осторожности.

Таблица 1: Вероятности одного обращения к генератору. Геометрическое распределение.

d	k						
16	9	10	11	12	13	14	15
32	25	26	27	28	29	30	31
64	57	58	59	60	61	62	63
$\pi_{\max}(1)$	0.97	0.95	0.92	0.86	0.75	0.59	0.35

Что касается геометрического распределения, то удобное для теоретического анализа представление правой части (6.2.3), вообще говоря, отсутствует. В Таблице 1 приведены результаты компьютерных вычислений вероятности $\pi_{\max}(1)$ для $M = 2^d$ и различных p вида 2^{-k} .

Как и следовало ожидать, при k , близких к d , алгоритмы, использующие ровно одно обращение к генератору псевдослучайных чисел, будут весьма неточными.

Приближенное равенство чисел $\pi_{\max}(1)$ в Таблице 1 при разных d, k с одинаковой разностью $d - k$ объясняется тем, что при маленьких p число тех значений j , при которых $\lfloor Mp(1 - p)^j \rfloor = n > 0$, примерно равно $\ln(1 + 1/n)/p$.

Замечание 16. Если для какого-то распределения \mathcal{P} оказалось, что число $\pi_{\max}(1)$ близко к единице, это еще не означает, что любой теоретически точный (в смысле Введения) алгоритм моделирования распределения \mathcal{P} не приведет к серьезным ошибкам при использовании ϵ вместо α в моделирующей формуле одной переменной. Это означает лишь, что такие алгоритмы существуют.

Впрочем, многодесятилетняя мировая практика показывает, что использование хорошо опробованных методов для моделирования «разумных» распределений делает такие опасения напрасными.

7 Приложение 1. Преобразования случайных векторов

Преобразование случайных величин и векторов — естественный (и часто достаточно элементарный) способ нахождения моделирующих функций заданного распределения. В этом приложении приводится доказательство одного общего факта, относящегося к преобразованию случайных векторов, имеющих абсолютно непрерывные распределения, а также обсуждаются частные случаи, часто используемые при моделировании.

7.1 Общее утверждение

Предложение 15. Пусть $\bar{\xi} \in \mathbb{R}^d$ — d -мерный случайный вектор, обладающий распределением \mathcal{P}_ξ с плотностью распределения p_ξ , причем $\mathcal{P}_\xi(D) = 1$ для некоторого измеримого $D \subset \mathbb{R}^d$. Рассмотрим измеримое отображение $\phi : D \mapsto \mathbb{R}^d$ и предположим, что при $i = 1, \dots, n$ существуют открытые попарно непересекающиеся подмножества $D_i \subset \mathbb{R}^d$, удовлетворяющие следующим условиям.

1. Множества D и $\bigcup_{i=1}^n D_i$ совпадают \mathcal{P}_ξ -почти всюду.
2. Обозначим $\phi_i = \phi|_{D_i}$, $\phi(D_i) = G_i$ и предположим, что при любом i отображение $\phi_i : D_i \mapsto G_i$
 - является биекцией с $\psi_i = \phi_i^{-1}$;
 - непрерывно дифференцируемо с $\det \psi'_i \neq 0$ в G_i .

Тогда случайный вектор $\bar{\eta} = \phi(\bar{\xi})$ обладает плотностью $p_\eta(Y)$ и

$$p_\eta(Y) = \sum_i p_\xi(\psi_i(Y)) |\det \psi'_i(Y)| \mathbb{I}_{G_i}(Y). \quad (7.1.1)$$

Доказательство.

$$\begin{aligned} \mathbb{P}(\bar{\eta} \in A) &= \mathbb{P}(\phi(\bar{\xi}) \in A) = \sum_i \mathbb{P}(\phi(\bar{\xi}) \in A, \bar{\xi} \in D_i) = \sum_i \mathbb{P}(\phi_i(\bar{\xi}) \in A \cap G_i, \bar{\xi} \in D_i) = \\ &= \sum_i \mathbb{P}(\phi_i(\bar{\xi}) \in A \cap G_i) = \sum_i \int_{\{x: \phi_i(x) \in A \cap G_i\}} p_\xi(X) dX = \\ &= \sum_i \int_{A \cap G_i} p_\xi(\psi_i(Y)) |\det \psi'_i(Y)| dY = \int_A \sum_i p_\xi(\psi_i(Y)) |\det \psi'_i(Y)| \mathbb{I}_{G_i}(Y) dY, \end{aligned}$$

что и требовалось доказать. □

7.2 Частные случаи, полезные для моделирования

Перейдем теперь к примерам применения Предложения 15.

7.2.1 Аффинные преобразования

Общая формула. В условиях Предложения 15 возьмем $n = 1$ и при $X \in \mathbb{R}^d$ положим $\phi(X) = AX + b$, где A — невырожденная $d \times d$ матрица и $b \in \mathbb{R}^d$. Тогда $\psi(Y) = A^{-1}(Y - b)$ и (7.1.1) превращается в

$$p_\eta(Y) = |\det A|^{-1} p_\xi(A^{-1}(Y - b)). \quad (7.2.1)$$

Например, если случайный вектор $\bar{\xi} = (\xi_1, \dots, \xi_d)^T$ имеет плотность распределения $p_\xi(x_1, \dots, x_d)$ и нас интересует совместное распределение случайных величин $\eta_1 = \xi_1$, $\eta_j = \xi_j - \xi_{j-1}$, $j = 2, \dots, d$, то соответствующее отображение ϕ является линейным невырожденным с определителем, равным 1,¹¹² а обратное отображение ψ определяется равенствами $x_i = y_1 + \dots + y_i$, $i = 1, \dots, d$. Поэтому

$$p_\eta(y_1, \dots, y_d) = p_\xi(y_1, y_1 + y_2, \dots, y_1 + \dots + y_d). \quad (7.2.2)$$

Одномерный случай. В случае $d = 1$ аффинное отображение ψ определяется равенством $\psi(x) = ax + b$, условие невырожденности выглядит как $a \neq 0$ и формула (7.2.1) приобретает хорошо известный вид

$$p_\eta(y) = \frac{1}{|a|} p_\xi((y - b)/a). \quad (7.2.3)$$

С точки зрения моделирования это означает, что генерирование одномерного распределения с плотностью (7.2.3) не представляет труда, если только мы умеем получать реализации случайной величины ξ с плотностью распределения p_ξ : в этом случае достаточно положить $\eta = a\xi + b$.

Таких примеров много. Например, если нам нужно промоделировать распределение $\text{EXP}(\mu)$, то достаточно уметь моделировать случайную величину $\eta \in \text{EXP}(1)$, а потом сделать преобразование $\xi = \eta/\mu$. Другой стандартный пример: для моделирования любого распределения из параметрического семейства $\{N(a, \sigma^2), a \in \mathbb{R}, \sigma > 0\}$ достаточно промоделировать $\eta \in N(0, 1)$, а затем сделать преобразование $\xi = \sigma\eta + a$ (или $\xi = -\sigma\eta + a$).

Равномерное распределение в областях. Пусть случайный вектор $\bar{\xi}$ равномерно распределено в d -мерной области D (конечно, предполагается, что $0 < \text{mes}_d(D) < \infty$). Иначе говоря, пусть $p_\xi(X) = c \mathbb{I}_D(X)$, где $c = 1/\text{mes}_d(D)$.

Обозначим $G = AD + b$ — образ D при невырожденном преобразовании $\theta(X) = AX + b$. Тогда, так как $A^{-1}(Y - b) \in D$ тогда и только тогда, когда $Y \in G$, то (7.2.1) приобретает вид

$$p_\eta(Y) = c |\det A|^{-1} \mathbb{I}_G(Y).$$

Таким образом, случайный вектор $\bar{\eta}$ равномерно распределен в образе G множества D .

Ортонормированные линейные преобразования. Пусть $\bar{\xi} \in \mathbb{R}^d$ — случайный вектор с плотностью распределения p_ξ , которая имеет вид

$$p_\xi(X) = f(\|X\|) = f(\sqrt{X^T X}), \quad X = (x_1, \dots, x_d)^T.$$

¹¹²Проверьте!

Если $A : \mathbb{R}^d \mapsto \mathbb{R}^d$ — ортонормированная матрица и $\bar{\eta} = A\bar{\xi}$, то $|\det A| = 1$, $A^T = A^{-1}$, и по формуле (7.2.1)

$$p_{\eta}(Y) = f\left(\sqrt{Y^T A A^T Y}\right) = f(\|Y\|).$$

Значит, $\mathcal{L}(\bar{\xi}) = \mathcal{L}(A\bar{\xi})$.

Обозначим теперь $r = \|\bar{\xi}\|$ и $\bar{v} = \bar{\xi}/r$.¹¹³ Тогда $\|\bar{v}\| = 1$, $\bar{\xi} = r\bar{v}$ и (поскольку ортонормированное отображение сохраняет длины векторов) $A\bar{\xi} = rA\bar{v}$. Поэтому¹¹⁴ $\mathcal{L}(\bar{v}) = \mathcal{L}(A\bar{v})$.

Следовательно, распределение случайного вектора \bar{v} , лежащего на единичной d -мерной сфере с центром в нуле, инвариантно относительно любых ортонормированных преобразований этого вектора. Значит, $\mathcal{L}(\bar{v})$ является равномерным распределением на этой сфере.

7.2.2 Переход к полярным и сферическим координатам

Полярные координаты. Каждая точка $X = (x, y)^T \in \mathbb{R}^2 \setminus \{\mathbf{0}\}$ может быть однозначно представлена в виде $X = \|X\| \bar{e}$, где $\|\bar{e}\| = 1$. Полагая $s = \|X\|$ и $\bar{e} = (\cos t, \sin t)^T$, где $t \in [0, 2\pi)$, получаем биекцию $\phi : (x, y)^T \mapsto (s, t)^T$, действующую из $\mathbb{R}^2 \setminus \{\mathbf{0}\}$ в $(0, \infty) \times [0, 2\pi)$. Конечно, переменные (s, t) являются полярными координатами вектора X , а обратное отображение $\psi : (0, \infty) \times [0, 2\pi) \mapsto \mathbb{R}^2 \setminus \{\mathbf{0}\}$ имеет вид $x = s \cos t$, $y = s \sin t$ с якобианом $\det \psi'(s, t) = s$.

Если теперь рассмотреть случайный вектор $\bar{\xi} \in \mathbb{R}^2$ с плотностью распределения $p_{\xi}(x, y)$ и обозначить r, φ (случайные) полярные координаты этого вектора, то, так как в этом случае $n = 1$, мы получим из (7.1.1), что

$$p_{r,\varphi}(s, t) = s p_{\xi}(s \cos t, s \sin t) \mathbf{1}_{(0, \infty) \times [0, 2\pi)}(s, t). \quad (7.2.4)$$

Выражение (7.2.4) выглядит особенно просто, если существует такая функция $f : (0, \infty) \mapsto (0, \infty)$, что $p_{\xi}(x, y) = f(\sqrt{x^2 + y^2})$. В этом случае, очевидно,

$$p_{r,\varphi}(s, t) = s f(s) \mathbb{I}_{(0, \infty) \times [0, 2\pi)}(s, t) = 2\pi s f(s) \mathbb{I}_{(0, \infty)}(s) \frac{1}{2\pi} \mathbb{I}_{[0, 2\pi)}(t). \quad (7.2.5)$$

Это значит, что случайные величины r и φ независимы, $\varphi \in U(0, 2\pi)$, а r имеет плотность распределения $p_r(s) = 2\pi s f(s)$, $s > 0$.

Примеры.

1. Стандартным примером является равномерное распределение в единичном круге с центром в нуле, где $f(s) = \mathbb{I}_{(0, 1)}(s)/\pi$ и полярный радиус имеет плотность распределения $p_r(s) = 2s \mathbb{I}_{(0, 1)}(s)$.
2. Другой стандартный пример — совместное распределение двух независимых $N(0, 1)$ -распределенных случайных величин, где

$$p(x, y) = \frac{1}{2\pi} e^{-(x^2 + y^2)/2},$$

$f(s) = e^{-s^2/2}/2\pi$, и плотность распределения полярного радиуса — это плотность распределения Рэлея, то есть $p_r(s) = s e^{-s^2/2}$ при $s > 0$.

Понятно, как можно использовать подобного рода факты для моделирования: поскольку полярный радиус и полярный угол независимы, то их можно моделировать отдельно, а потом переходить обратно в евклидову систему координат, причем сложности могут возникнуть только при моделировании распределения полярного радиуса.

¹¹³Поскольку $\mathbb{P}(r = 0) = 0$, то эта операция корректна.

¹¹⁴Почему?

Сферические координаты и равномерное распределение на сфере. Аналогично полярным координатам на плоскости, каждый ненулевой вектор $X = (x, y, z)^T \in \mathbb{R}^3$ может быть однозначно представлен в виде $X = \|X\| \bar{e}$, где

$$\bar{e} = (\cos(t) \cos(u), \sin(t) \cos(u), \sin(u))^T, \quad t \in [0, 2\pi), \quad u \in [-\pi/2, \pi/2].$$

Это, конечно, соответствует переходу от евклидовой системы координат (x, y, z) к сферической системе (s, t, u) со сферическим радиусом $s = \|X\|$, долготой s и широтой u . Хорошо известно, что якобиан обратного отображения равен $s^2 \cos(u)$.

Поэтому, если случайный вектор $\bar{\xi} = (\xi_1, \xi_2, \xi_3)^T$ имеет плотность распределения $p_{\xi}(x, y, z)$, то сферические координаты r, φ, θ этого вектора имеют совместную плотность

$$p_{r,\varphi,\theta}(s, t, u) = p_{\xi}(s \cos(t) \cos(u), s \sin(t) \cos(u), s \sin(u)) s^2 \cos(u), \quad (7.2.6)$$

сосредоточенную в области $(0, \infty) \times [0, 2\pi) \times (-\pi/2, \pi/2)$. В случае, когда

$$p_{\xi}(x, y, z) = f(\sqrt{x^2 + y^2 + z^2}), \quad (7.2.7)$$

равенство (7.2.6) приобретает вид

$$p_{r,\varphi,\theta}(s, t, u) = 4\pi s^2 f(s^2) \mathbb{I}_{(0,\infty)}(s) \frac{1}{2\pi} \mathbb{I}_{(0,2\pi)}(t) \frac{\cos(u)}{2} \mathbb{I}_{(-\pi/2,\pi/2)}(u).$$

Таким образом, случайные величины r, φ и θ оказываются независимыми, причем долгота φ равномерно распределена на $(0, 2\pi)$, плотность $p_r(s)$ распределения r равна $4\pi s^2 f(s^2)$, а плотность $p_{\theta}(u)$ распределения широты θ сосредоточена на $(-\pi/2, \pi/2)$ и равна на этом интервале $0.5 \cos(u)$.

Например, если $\bar{\xi} = (\xi_1, \xi_2, \xi_3)^T$ — случайный вектор с независимыми $N(0, 1)$ -распределенными координатами, то его плотность распределения равна $(2\pi)^{-3/2} e^{-(x^2+y^2+z^2)/2}$, и длина $r = \|\bar{\xi}\|$ этого вектора будет иметь плотность распределения $\sqrt{2/\pi} s^2 e^{-s^2/2}$.

Как уже обсуждалось в разделе 7.2.1, если случайный вектор $\bar{\xi}$ имеет распределение (7.2.7), то вектор $\bar{v} = \bar{\xi}/\|\bar{\xi}\|$ равномерно распределен на поверхности сферы $\{(x, y, z) : x^2 + y^2 + z^2 = 1\}$. С другой стороны, координаты v_1, v_2, v_3 этого вектора выражаются через случайные величины φ и θ как

$$v_1 = \cos(\varphi) \cos(\theta), \quad v_2 = \sin(\varphi) \cos(\theta), \quad v_3 = \sin(\theta). \quad (7.2.8)$$

Поскольку θ имеет плотность $p_{\theta}(u) = \cos(u)/2$, $|u| \leq \pi/2$, а функция $\sin(u)$ обратима на этом интервале, то, согласно Предложению 15,

$$p_{v_3}(z) = 0.5 \cos(\arcsin(z)) |\arcsin'(z)| = 0.5 \cos(\arcsin(z)) \frac{1}{\cos(\arcsin(z))} = 0.5$$

при $|z| < 1$. Таким образом, $v_3 \in U(-1, 1)$. Конечно, такое же распределение имеет и скалярное произведение $v_t = \bar{v}^T \bar{t}$ для любого детерминированного вектора \bar{t} единичной длины.¹¹⁵

7.2.3 Гамма-распределение и распределение Дирихле

По определению, случайная величина ξ имеет гамма-распределение $\text{Gamma}(k, \mu)$ с параметрами $k > 0$ (параметр формы) и $\mu > 0$ (параметр масштаба), если у нее существует плотность распределения, сосредоточенная на $(0, \infty)$ и задаваемая формулой

$$p_{\xi}(x) = \frac{\mu^k}{\Gamma(k)} x^{k-1} e^{-\mu x}, \quad x > 0.$$

¹¹⁵Этот факт можно также вывести из известной теоремы школьной стереометрии о том, что площадь поверхности сферы, отсекаемая двумя параллельными плоскостями, пропорциональна расстоянию между этими плоскостями.

Случайный вектор $(\xi_1, \dots, \xi_d)^T$ имеет распределение Дирихле $D_d(k_1, \dots, k_{d+1})$ с положительными параметрами k_i , если его плотность распределения имеет вид

$$p_d(x_1, \dots, x_d) = \frac{\Gamma(k_1 + \dots + k_{d+1})}{\Gamma(k_1) \dots \Gamma(k_{d+1})} x_1^{k_1-1} \dots x_d^{k_d-1} (1 - x_1 - \dots - x_d)^{k_{d+1}-1}, \quad x_i > 0, \quad \sum_{i=1}^d x_i < 1.$$

Если $k_1 = \dots = k_{d+1} = 1$, то распределение $D_d(k_1, \dots, k_{d+1})$ равномерно в области $x_i > 0$, $\sum_{i=1}^d x_i < 1$. При $d = 1$ распределение Дирихле называется бета-распределением и обозначается $\text{Beta}(k_1, k_2)$.

Приведем одно утверждение, связывающее гамма-распределение и распределение Дирихле.

Предложение 16. 1. Пусть $\eta_1, \dots, \eta_d, \eta_{d+1}$ — независимые случайные величины, причем $\eta_i \in \text{Gamma}(k_i, \mu)$. Обозначим $\xi_{d+1} = \eta_1 + \dots + \eta_{d+1}$ и $\xi_i = \eta_i / \xi_{d+1}$ при $1 \leq i \leq d$. Тогда вектор $\bar{\xi} = (\xi_1, \dots, \xi_d)^T$ и случайная величина ξ_{d+1} независимы, причем $\bar{\xi} \in D_d(k_1, \dots, k_{d+1})$ и $\eta_{d+1} \in \text{Gamma}(k, \mu)$, где $k = k_1 + \dots + k_{d+1}$.

2. Пусть $\eta \in \text{Gamma}(k, \mu)$. Кроме того, рассмотрим случайный вектор

$$\bar{\pi} = (\pi_1, \dots, \pi_d)^T \in D_d(k_1, \dots, k_{d+1})$$

и положим $\pi_{d+1} = 1 - \pi_1 - \dots - \pi_d$. Если $\eta, \bar{\pi}$ независимы и $k = k_1 + \dots + k_{d+1}$, то при $i = 1, \dots, d+1$ случайные величины $\xi_i = \pi_i \eta$ независимы и $\xi_i \in \text{Gamma}(k_i, \mu)$.

Доказательство. 1. Применим Предложение 15 с $n = 1$, $D = (0, \infty)^{d+1}$ и

$$\phi(x_1, \dots, x_{d+1}) = \begin{cases} \frac{x_i}{x_1 + \dots + x_{d+1}} & \text{при } i = 1, \dots, d, \\ \frac{x_{d+1}}{x_1 + \dots + x_{d+1}} & \text{при } i = d+1. \end{cases}$$

Тогда обратное отображение ψ запишется как

$$\psi(y_1, \dots, y_{d+1}) = (\psi_1, \dots, \psi_{d+1})^T = \begin{cases} y_i y_{d+1} & \text{при } i = 1, \dots, d, \\ y_{d+1} (1 - y_1 - \dots - y_d) & \text{при } i = d+1, \end{cases} \quad (7.2.9)$$

где $y_i > 0$ и $y_1 + \dots + y_d < 1$. Ясно, что

$$\partial \psi_i / \partial y_j = \begin{cases} y_{d+1} & \text{при } i = j = 1, \dots, d, \\ 0 & \text{при } i, j \in \{1, \dots, d\}, i \neq j, \\ y_i & \text{при } i = 1, \dots, d, j = d+1, \\ 1 - (y_1 + \dots + y_d) & \text{при } i = j = d+1, \\ -y_{d+1} & \text{при } i = d+1, j < d. \end{cases}$$

Это означает,¹¹⁶ что $|\det \psi'| = y_{d+1}^d$ и, обозначив $k = k_1 + \dots + k_{d+1}$, получим, что совместная плотность распределения случайных величин ξ_1, \dots, ξ_{d+1} равна

$$\begin{aligned} p_\xi(y_1, \dots, y_{d+1}) &= \frac{\mu^k}{\prod_{j=1}^{d+1} \Gamma(k_j)} \prod_{j=1}^d (y_j y_{d+1})^{k_j-1} \left(1 - \sum_{j=1}^d y_j\right)^{k_{d+1}-1} y_{d+1}^{k_{d+1}-1} e^{-\mu y_{d+1}} y_{d+1}^d = \\ &= \frac{\Gamma(k)}{\prod_{j=1}^{d+1} \Gamma(k_j)} \prod_{j=1}^d y_j^{k_j-1} \left(1 - \sum_{j=1}^d y_j\right)^{k_{d+1}-1} \frac{\mu^k}{\Gamma(k)} y_{d+1}^{k-1} e^{-\mu y_{d+1}}, \end{aligned} \quad (7.2.10)$$

где $y_i > 0$ и $y_1 + \dots + y_d < 1$. Первое утверждение доказано.

2. Совместное распределение случайных величин $\pi_1, \dots, \pi_d, \eta$ имеет плотность распределения (7.2.10). Взаимно-однозначное преобразование, переводящее эти случайные величины в случайные величины ξ_1, \dots, ξ_{d+1} , имеет вид (7.2.9). Таким образом, второе утверждение Предложения 16 является обратным к первому, и, тем самым, становится очевидным.¹¹⁷ \square

¹¹⁶Проверьте!

¹¹⁷Действительно?

7.2.4 Распределения порядковых статистик

Пусть $\alpha_1, \dots, \alpha_n$ — независимые случайные величины, равномерно распределенные на $(0, 1)$. При каждом элементарном событии ω упорядочим числа $\alpha_i(\omega)$ по возрастанию и получим новые случайные величины $0 \leq \alpha_{[1]}^{(n)} \leq \dots \leq \alpha_{[i]}^{(n)} \leq \dots \leq \alpha_{[n]}^{(n)} \leq 1$. Эти случайные величины принято называть порядковыми статистиками (выборки из распределения $U(0, 1)$). Конечно, $\alpha_{[1]}^{(n)} = \min(\alpha_1, \dots, \alpha_n)$ и $\alpha_{[n]}^{(n)} = \max(\alpha_1, \dots, \alpha_n)$.

Предложение 15 позволяет легко получить следующий результат.

Предложение 17. Случайный вектор $\bar{\xi} = (\alpha_{[1]}^{(n)}, \dots, \alpha_{[n]}^{(n)})^T$ равномерно распределен в области $G = \{(x_1, \dots, x_n) : 0 < x_1 < \dots < x_i < \dots < x_n < 1\}$.

Доказательство. Обозначим Ψ множество перестановок чисел $1, \dots, n$. Зафиксируем $\psi \in \Psi$ с $\psi(1, \dots, n) = (i_1, \dots, i_n)$. При $X = (x_1, \dots, x_n)^T$ будем обозначать ψX вектор $(x_{i_1}, \dots, x_{i_n})^T$.

Обозначим $D_\psi = \psi G$. Тогда, очевидно, множества D_{ψ_1} и D_{ψ_2} дизъюнкты при $\psi_1 \neq \psi_2$. Кроме того,

$$\text{mes}_n\left((0, 1)^n \setminus \bigcup_{\psi \in \Psi} D_\psi\right) = 0.$$

Ясно, что при фиксированном $\psi \in \Psi$ отображение $\Theta_\psi : G \mapsto D_\psi$, поточечно определенное как $X \mapsto \psi X$, является линейным и невырожденным. При этом матрица отображения Θ_ψ ортонормированна и, следовательно, имеет определитель, по модулю равный 1.

В этих обозначениях отображение ϕ , преобразующее $\bar{\xi} = (\alpha_1, \dots, \alpha_n)^T$ в $\bar{\eta} = (\alpha_{[1]}^{(n)}, \dots, \alpha_{[n]}^{(n)})^T$ может быть (с точностью до множества меры ноль) записано как

$$\phi(x_1, \dots, x_n) = \Theta_\psi^{-1}(x_1, \dots, x_n) \quad \text{при } (x_1, \dots, x_n) \in D_\psi,$$

причем $\phi D_\psi = G$ для любого $\psi \in \Psi$. Тем самым (7.1.1) приобретает вид

$$p_\eta(Y) = \sum_{\psi \in \Psi} p_\xi(\psi Y) \mathbb{I}_G(Y) = n! \mathbb{I}_G(Y), \quad (7.2.11)$$

что и требовалось доказать. \square

Следствие 6. Пусть $1 \leq s \leq n$ и $1 \leq k_1 < \dots < k_s \leq n$. Тогда в условиях Предложения 17 последовательные разности порядковых статистик

$$\alpha_{[k_1]}^{(n)}, \alpha_{[k_2]}^{(n)} - \alpha_{[k_1]}^{(n)}, \dots, \alpha_{[k_s]}^{(n)} - \alpha_{[k_{s-1}]}^{(n)} \quad (7.2.12)$$

имеют совместное распределение $D_s(\ell_1, \dots, \ell_{s+1})$, где $\ell_1 = k_1$, $\ell_{s+1} = n - k_s + 1$ и $\ell_j = k_j - k_{j-1}$ при $1 < j \leq s$.

Доказательство. Достаточно доказать, что случайные величины $\alpha_{[k_1]}^{(n)}, \dots, \alpha_{[k_s]}^{(n)}$ имеют совместную плотность

$$p_{k_1 \dots k_s}(x_1, \dots, x_s) = \frac{\Gamma(n+1)}{\prod_{j=1}^{s+1} \Gamma(\ell_j)} \prod_{j=1}^{s+1} (x_j - x_{j-1})^{\ell_j-1}, \quad (7.2.13)$$

где $0 < x_1 < \dots < x_s < 1$, а переменные $x_0 \equiv 0$ и $x_{s+1} \equiv 1$ введены только для удобства записи. Это немедленно следует из равенства (7.2.2) и определения распределения Дирихле.

Для простоты будем доказывать равенство (7.2.13) при $s = 2$, $k_1 > 1$ и $k_s < n$. Согласно (7.2.11), плотность p_{k_1, k_2} совместного распределения случайных величин $\alpha_{[k_1]}, \alpha_{[k_2]}$ имеет вид

$$p_{k_1, k_2}(x_1, x_2) = n! \int_{S_{k_1-1}(0, x_1)} dz_1 \dots dz_{k_1-1} \int_{S_{k_2-k_1-1}(x_1, x_2)} dz_{k_1+1} \dots dz_{k_2-1} \int_{S_{n-k_1}(x_1, 1)} dz_{k_2+1} \dots dz_n,$$

где $0 < x_1 < x_2 < 1$, $0 < s < t$ и $S_j(s, t) = \{(z_1, \dots, z_j) : s < z_1 < \dots < z_j < t\}$. Поскольку¹¹⁸

$$\int_{S_j(s, t)} dz_1 \dots dz_j = (t - s)^j / j! ,$$

то

$$p_{k_1, k_2}(x_1, x_2) = \frac{n!}{(k_1 - 1)!(k_2 - k_1 - 1)!(n - k_2)!} x_1^{k_1-1} (x_2 - x_1)^{k_2-k_1-1} (1 - x_2)^{n-k_2} ,$$

что соответствует (7.2.13). Общий случай исследуется аналогично. \square

Замечание 17. В частности, из Следствия 6 следует, что совместное распределение приращений порядковых статистик $\alpha_{[1]}^{(n)}, \alpha_{[2]}^{(n)} - \alpha_{[1]}^{(n)}, \dots, \alpha_{[n]}^{(n)} - \alpha_{[n-1]}^{(n)}$ имеет распределение Дирихле $D_n(1, \dots, 1)$, которое, в свою очередь, является равномерным распределением в области $x_i > 0$, $\sum_{i=1}^d x_i < 1$.

¹¹⁸Почему?

8 Приложение 2. Многомерные гауссовские распределения

Определение 8.1. Говорят, что случайный вектор $\bar{\xi} = (\xi_1, \dots, \xi_d)^T \in \mathbb{R}^d$ имеет (d -мерное) гауссовское распределение, если для любого $\bar{t} \in \mathbb{R}^d$ случайная величина $(\bar{\xi}, \bar{t})$ имеет нормальное распределение¹¹⁹ с некоторыми параметрами a_t и σ_t^2 .

Нужные нам свойства многомерных гауссовских распределений объединены в три леммы.

8.1 Общий случай

Лемма 7. (Характеристическая функция гауссовского вектора)

Пусть $\bar{\xi} = (\xi_1, \dots, \xi_d)^T \in \mathbb{R}^d$ имеет гауссовское распределение. Тогда характеристическая функция $\varphi_{\bar{\xi}}(\bar{t})$ случайного вектора $\bar{\xi}$ имеет вид

$$\varphi_{\bar{\xi}}(\bar{t}) = \exp(i(\bar{a}, \bar{t})) \exp(-(\Sigma \bar{t}, \bar{t})/2), \quad (8.1.1)$$

где $\bar{a} = \mathbb{E}\bar{\xi}$ и Σ — ковариационная матрица вектора $\bar{\xi}$.

Доказательство. Прежде всего заметим, что компоненты гауссовского вектора обладают всеми моментами. Действительно, взяв j -ый орт в качестве \bar{t} , мы получим, что ξ_j имеет нормальное распределение. Следовательно, мы можем говорить о среднем \bar{a} и ковариационной матрице Σ вектора $\bar{\xi}$.

Сосчитаем теперь среднее a_t и дисперсию σ_t^2 скалярного произведения $(\bar{\xi}, \bar{t}) = \bar{t}^T \bar{\xi}$. Ясно, что $a_t = (\bar{a}, \bar{t})$. Для вычисления дисперсии достаточно рассмотреть случай $\bar{a} = \mathbf{0}$. Тогда

$$\sigma_t^2 = \mathbb{E}(\bar{\xi}, \bar{t})^2 = \mathbb{E}(\bar{t}^T \bar{\xi} \bar{\xi}^T \bar{t}) = \bar{t}^T \mathbb{E}(\bar{\xi} \bar{\xi}^T) \bar{t} = \bar{t}^T \Sigma \bar{t} = (\Sigma \bar{t}, \bar{t}).$$

Следовательно, характеристическая функция $\psi_t(\tau)$ скалярного произведения $(\bar{\xi}, \bar{t})$ равна

$$\psi_t(\tau) = \mathbb{E} \exp(i(\bar{\xi}, \bar{t})\tau) = \exp(i(\bar{a}, \bar{t})\tau) \exp(-(\Sigma \bar{t}, \bar{t})\tau^2/2) \quad (8.1.2)$$

Полагая $\tau = 1$ в (8.1.2), приходим к нужному нам результату. \square

Таким образом, распределение d -мерного гауссовского случайного вектора $\bar{\xi}$ полностью определяется его средним \bar{a} и ковариационной матрицей Σ . Поэтому мы будем использовать запись $\bar{\xi} \in N_d(\bar{a}, \Sigma)$. Отметим, что невырожденность ковариационной матрицы при этом не предполагается.

Следствие 7. Для того, чтобы компоненты гауссовского вектора $\bar{\xi} \in N_d(\bar{a}, \Sigma)$ были независимы, необходимо и достаточно, чтобы они были попарно некоррелированы.¹²⁰

Доказательство. Случайный вектор $\bar{\xi} = (\xi_1, \dots, \xi_d)^T$ имеет независимые компоненты тогда и только тогда, когда его характеристическая функция $\varphi_{\bar{\xi}}(t_1, \dots, t_d)$ распадается в произведение некоторых функций вида $\psi_j(t_j)$. Поскольку попарная некоррелированность компонент этого вектора эквивалентна диагональности его ковариационной матрицы, то утверждение следует из формулы (8.1.1). \square

Лемма 8. (Сохранение гауссовости при аффинных преобразованиях. Существование.)

1. Если $\bar{\xi} \in \mathbb{R}^d$, $\bar{\xi} \in N_c(\bar{a}, \Sigma)$, $\bar{b} \in \mathbb{R}^k$ и матрица A имеет d столбцов и k строк, то $\bar{\eta} \stackrel{\text{def}}{=} A\bar{\xi} + \bar{b} \in N_k(\bar{b} + A\bar{a}, A\Sigma A^T)$
2. Каковы бы ни были вектор $\bar{a} \in \mathbb{R}^d$ и $d \times d$ неотрицательно определенная матрица Σ , существует такое вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$ и такая случайная величина $\bar{\xi} : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}^d, \mathcal{B}_d)$, что $\bar{\xi} \in N_d(\bar{a}, \Sigma)$.

¹¹⁹Напомним, что по нашему соглашению постоянная a имеет нормальное распределение $N(a, 0)$.

¹²⁰Здесь нам удобно считать, что случайные величины ξ и η некоррелированы, если $\text{Cov}(\xi, \eta) = 0$.

Доказательство. 1. Гауссовость случайного вектора $\bar{\eta}$ следует из цепочки равенств

$$(\bar{\eta}, \bar{s}) = (A\bar{\xi} + \bar{b}, \bar{s}) = (\bar{b}, \bar{s}) + (\bar{\xi}, A^T \bar{s})$$

и Определения 8.1. Вычисляя среднее и ковариационную матрицу вектора $\bar{\eta}$, получаем окончательный результат.

2. Прежде всего, нам достаточно рассматривать случай $\bar{a} = \mathbf{0}$. Пусть $\mu_1 \geq \dots \mu_d \geq 0$ — собственные числа матрицы Σ с учетом их кратности. Обозначим U_1, \dots, U_d соответствующую ортонормированную систему собственных векторов и положим $U = [U_1 : \dots : U_d]$. Ясно, что матрица U является ортонормированной, причем $\Sigma = U\Lambda U^T$, где матрица Λ диагональна с элементами $\lambda_{jj} = \mu_j$.

Рассмотрим теперь случайный вектор $\bar{\eta} = (\eta_1, \dots, \eta_d)^T$ такой, что $\eta_j \in N(0, \mu_j)$ и потребуем, чтобы эти компоненты были независимы. Такие случайные величины существуют.¹²¹ Более того, так как линейная комбинация независимых нормальных случайных величин имеет нормальное распределение, то вектор $\bar{\eta}$ имеет гауссовское распределение. Точнее, $\bar{\eta} \in N_d(\mathbf{0}, \Lambda)$.

Теперь, положив $\bar{\xi} = U\bar{\eta}$, получим требуемое из первого утверждения леммы. \square

8.2 Невырожденный случай

Лемма 9. Пусть $\bar{\xi} \in N_d(\bar{a}, \Sigma)$. Если матрица Σ невырождена, то распределение вектора $\bar{\xi}$ абсолютно непрерывно с плотностью

$$p_{\xi}(X) = \frac{\det \Sigma^{-1/2}}{(2\pi)^{d/2}} e^{-(\Sigma^{-1}(X - \bar{a}), X - \bar{a})/2}, \quad X \in \mathbb{R}^d. \quad (8.2.1)$$

Доказательство. Нам снова достаточно рассматривать случай $\bar{a} = \mathbf{0}$. Действуя так же, как в Лемме 8, рассмотрим собственные числа μ_j матрицы Σ и соответствующие матрицы U и Λ . Так как матрица Σ невырождена, то $\mu_j > 0$ при всех j .

Введем гауссовский случайный вектор $\bar{\eta} = (\eta_1, \dots, \eta_d)^T$ с независимыми компонентами и такой, что $\eta_j \in N(0, \mu_j)$. Нетрудно видеть, что плотность распределения этого вектора имеет вид

$$p_{\eta}(Y) = \frac{\det \Lambda^{-1/2}}{(2\pi)^{d/2}} e^{-(\Lambda^{-1}Y, Y)/2}, \quad Y \in \mathbb{R}^d.$$

Положим теперь $\bar{\xi} = U\bar{\eta}$. Согласно Лемме 8, $\bar{\xi} \in N_d(\mathbf{0}, \Sigma)$. С другой стороны, так как $U^{-1} = U^T$, $\det \Sigma = \det \Lambda$ и $\Sigma^{-1} = U\Lambda^{-1}U^T$, то из (7.2.1) мы получаем, что

$$\begin{aligned} p_{\xi}(X) &= |\det U|^{-1} p_{\eta}(U^{-1}X) = \frac{\det \Lambda^{-1/2}}{(2\pi)^{d/2}} e^{-(\Lambda^{-1}U^{-1}Y, U^{-1}Y)/2} = \\ &= \frac{\det \Sigma^{-1/2}}{(2\pi)^{d/2}} e^{-(U\Lambda^{-1}U^TY, Y)/2} = \frac{\det \Sigma^{-1/2}}{(2\pi)^{d/2}} e^{-(\Sigma^{-1}Y, Y)/2}. \end{aligned}$$

Утверждение доказано. \square

¹²¹ Действительно, приведенное описание полностью определяет распределение случайного вектора $\bar{\eta}$ как меры на борелевской σ -алгебре в \mathbb{R}^d . А для любой вероятностной тройки $(\mathbb{R}^d, \mathcal{B}_d, \mathcal{P})$ существует такое вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$ и такой случайный вектор $\bar{\xi}: (\Omega, \mathcal{F}) \mapsto (\mathbb{R}^d, \mathcal{B}_d)$, что $\mathcal{L}(\bar{\xi}) = \mathcal{P}$. Для доказательства последнего факта достаточно взять $\Omega = \mathbb{R}^d$, $\mathcal{F} = \mathcal{B}_d$, $\mathbb{P} = \mathcal{P}$ и $\bar{\xi}(\omega) = \omega$ для любого $\omega \in \Omega$.

9 Приложение 3. Упражнения

Равномерные распределения в различных множествах

1. Промоделировать равномерное распределение в объединении интервалов $(-2, -1)$ и $(1, 3)$.
2. Промоделировать равномерное распределение в квадрате с вершинами в точках $(0, 0)$, $(1, 1)$, $(0, 2)$ и $(1, -1)$.
3. Промоделировать равномерное распределение в треугольнике с вершинами в точках $(0, 0)$, $(1, 0)$ и $(0, 1)$ без использования метода отбора.
4. Промоделировать равномерное распределение в произвольном треугольнике без использования метода отбора.
5. Промоделировать равномерное распределение в четырехугольнике с вершинами в точках $(0, 0)$, $(1, 3)$, $(3, 1)$ и $(4, 5)$, не используя метода отбора.
6. Промоделировать равномерное распределение в четырехугольнике с вершинами в точках $(0, 0)$, $(2, 0)$, $(2, 1)$ и $(1, 3)$, не используя метода отбора.
7. Промоделировать равномерное распределение в объединении прямоугольника $(-1, 1) \times (0, 1)$ и треугольника с вершинами в точках $(2, 0)$, $(3, 0)$ и $(3, 2)$ без использования метода отбора.
8. Промоделировать равномерное распределение в объединении треугольников с вершинами в точках $(0, 0)$, $(1, 0)$ и $(1, 1)$ и в точках $(0, 0)$, $(-1, 0)$ и $(-1, -1)$ без использования метода отбора.
9. Промоделировать равномерное распределение в выпуклом пятиугольнике с вершинами в точках $(0, 0)$, $(0, 2)$, $(2, 2)$, $(3, 0)$ и $(2, -2)$, не используя метода отбора.
10. Промоделировать равномерное распределение в правильном шестиугольнике с центром в нуле и со стороной, равной 1.
11. Промоделировать тремя способами равномерное распределение в объединении квадратов с вершинами в точках $((0, 0), (1, 1), (2, 0), (1, -1))$ и $((0, 0), (-1, 1), (-2, 0), (-1, -1))$ соответственно.
12. Промоделировать равномерное распределение в разности квадратов $[0, 1]^2$ и $[\varepsilon, 1 - \varepsilon]^2$. Число $\varepsilon \in (0, 0.5)$ — параметр.
13. Промоделировать равномерное распределение в эллипсе

$$\left\{ (x, y) : 0 \leq \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1 \right\}$$

двумя способами.

14. Промоделировать двумя способами равномерное распределение в объединении квадрата $[0, 1]^2$ и круга $\{(x, y) : x^2 + y^2 \leq 1\}$.
15. Промоделировать равномерное распределение в области

$$\{(x, y) : R_1^2 \leq x^2 + y^2 \leq R_2^2\}$$

двумя способами, один из которых имеет трудоемкость, равномерно ограниченную по параметрам $0 \leq R_1 < R_2$.

16. Промоделировать равномерное распределение в объединении кругов радиуса 2 и 3 с центрами в точках $(-2, 0)$ и $(1, 2)$ соответственно.

17. Промоделировать равномерное распределение в объединении эллипсов

$$\{(x, y) : 0 \leq x^2 + 2y^2 \leq 1\} \quad \text{и} \quad \{(x, y) : 0 \leq 3x^2 + y^2 \leq 1\}.$$

18. Промоделировать равномерное распределение в пересечении 2-х кругов $\{(x, y) : x^2 + y^2 \leq 1\}$ и $\{(x, y) : (x + \varepsilon)^2 + y^2 \leq 1\}$, где ε — маленькое число.

19. Промоделировать равномерное распределение в подграфике плотности показательного распределения EXP(1).

20. Промоделировать равномерное распределение в области

$$\{(x, y, z) : R_1^2 \leq x^2 + y^2 + z^2 \leq R_2^2\}$$

двумя способами, один из которых имеет трудоемкость, равномерно ограниченную по параметрам $0 \leq R_1 < R_2$.

21. Промоделировать равномерное распределение в цилиндре

$$\{(x, y, z) : 0 \leq x^2 + y^2 \leq 1, 0 \leq z \leq 1\}.$$

22. Промоделировать двумя способами равномерное распределение в прямом круговом конусе с высотой 1 и основанием $\{(x, y) : x^2 + y^2 \leq 1\}$.

23. Промоделировать равномерное распределение в правильном тетраэдре, не используя метод отбора.

24. Промоделировать равномерное распределение в единичном трехмерном шаре с центром в нуле, не используя трехмерный вариант метода отбора.

25. Промоделировать равномерное распределение в трехмерном подграфике совместной плотности двух независимых $N(0, 1)$ -распределенных случайных величин.

26. Промоделировать равномерное распределение на границе квадрата $[-1, 1]^2$ двумя способами.

27. Промоделировать равномерное распределение на объединении границ квадратов $[-2, 2]^2$ и $[-4, 4]^2$.

28. Промоделировать равномерное распределение на границе треугольника с вершинами в точках $(0, 0)$, $(1, 0)$ и $(0, 1)$ двумя способами.

29. Промоделировать равномерное распределение на границе правильного шестиугольника с центром в нуле и со стороной, равной 1.

30. Промоделировать равномерное распределение на объединении двух окружностей с центрами в точках $(0, 0)$ и $(-1, 1)$ и радиусами 1 и 2 соответственно.

31. Промоделировать равномерное распределение на объединении сфер

$$\{(x, y, z) : 0 \leq x^2 + y^2 + z^2 = 1\} \quad \text{и} \quad \{(x, y, z) : 0 \leq x^2 + y^2 + z^2 = 4\}.$$

32. Промоделировать равномерное распределение на поверхности цилиндра

$$\{(x, y, z) : 0 \leq x^2 + y^2 \leq 1, 0 \leq z \leq 1\}.$$

33. Промоделировать равномерное распределение на границе правильного тетраэдра.
34. Промоделировать равномерное распределение на поверхности прямого усеченного конуса с основаниями $\{(x, y, z) : x^2 + y^2 \leq 2, z = 0\}$ и $\{(x, y, z) : x^2 + y^2 \leq 1, z = 1\}$.

Одномерные абсолютно непрерывные распределения

1. Промоделировать методом отбора распределение с плотностью $p(x) = 3x^2$ на отрезке $[0, 1]$, используя $q_1(x) = 2x$ и $q_2(x) = 1$. Сравнить методы.
2. Промоделировать двумя способами распределение с плотностью $p(x) = 0.5\lambda e^{-\lambda|x-a|}$, где $a \in \mathbb{R}$ и $\lambda > 0$.
3. Промоделировать распределение с плотностью $p(x) = c \left(e^{-x} + e^x + xe^{-x^2} \right)$, $x \in [0, 1]$, не используя метод отбора.
4. Промоделировать распределение с плотностью $p(x) = c(x - x^2)$, $x \in [0, 1]$, двумя способами, один из которых не использует метод отбора.
5. Промоделировать двумя способами распределение с плотностью

$$p(x) = c \begin{cases} e^{-x} & \text{при } 0 < x \leq L, \\ 0 & \text{иначе.} \end{cases}$$

Сравнить трудоемкости моделирования в зависимости от L .

6. Промоделировать двумя способами распределение с плотностью

$$p(x) = c \begin{cases} e^{-|x|} & \text{при } |x| > a, \\ 1 & \text{иначе.} \end{cases}$$

Сравнить трудоемкости моделирования в зависимости от a .

7. Промоделировать двумя способами распределение с плотностью

$$p(x) = c(e^{-x}\mathbf{1}_{-1,+\infty}(x) + e^x\mathbf{1}_{-\infty,1}(x)).$$

8. Промоделировать распределение с плотностью $p(x) = 5(1 + (x - 1)^4/12)$, $x \in [0, 2]$.
9. Промоделировать методом отбора (3 способа) распределение с плотностью

$$p(x) = c \begin{cases} 1 & \text{при } -1 \leq x \leq 0, \\ x & \text{при } 0 < x \leq 1, \\ e^{-x} & \text{при } x > 1, \\ 0 & \text{иначе.} \end{cases}$$

10. Промоделировать двумя способами распределение с плотностью

$$p(x) = c \begin{cases} e^{-x} + e^{-x^2} & \text{при } x \geq 0, \\ -x + x^2 & \text{при } -1 < x < 0, \\ e^{2x} & \text{при } x < -1. \end{cases}$$

11. Провести моделирование распределения с плотностью

$$p(x) = c \begin{cases} e^{-(x-1)^2} & \text{при } x > 1, \\ x^2 & \text{при } |x| \leq 1, \\ e^x & \text{при } x < -1. \end{cases}$$

12. Промоделировать тремя способами распределение с плотностью

$$p(x) = c(x - \varepsilon x(1 - x)),$$

сосредоточенной на отрезке $[0, 1]$. Параметр ε мал и положителен.

13. Провести моделирование распределения с плотностью

$$p(x) = c \begin{cases} e^{-2|x|} & \text{при } |x| \leq 1, \\ e^{-|x|} & \text{иначе} \end{cases}$$

тремя способами.

14. Провести моделирование распределения с плотностью

$$p(x) = c \begin{cases} x^2 + (2\pi)^{-1/2} e^{-x^2/2} & \text{при } |x| \leq 2, \\ (2\pi)^{-1/2} e^{-x^2/2} & \text{иначе.} \end{cases}$$

15. Придумать несколько способов моделирования распределения с плотностью

$$p(x) = c \begin{cases} |x|^{-1} e^{-2x-x^2} & \text{при } |x| \geq 1, \\ 0 & \text{иначе.} \end{cases}$$

16. Придумать несколько способов моделирования распределения с плотностью

$$p(x) = c(1 + |x|)^{-3}.$$

17. Придумать несколько способов моделирования распределения с плотностью

$$p(x) = c e^{-x^2 - |x|^3}.$$

18. Промоделировать распределение с плотностью $p(x) = c e^{x^2 - |x|^3}$.

19. Придумать несколько способов моделирования распределения с плотностью

$$p(x) = c |x|^3 e^{-x^4} (1 + |x|^\alpha),$$

при положительном значении параметра α .

20. Промоделировать распределение с плотностью

$$p(x) = c \begin{cases} e^{-x^2} & \text{при } A \leq |x| \leq B, \\ 0 & \text{иначе} \end{cases}$$

тремя способами.

21. Промоделировать распределение с плотностью

$$p(x) = c \begin{cases} e^{-x^2} & \text{при } 0 \leq x \leq A, \\ 0 & \text{иначе} \end{cases}$$

при маленьком и большом A .

22. Промоделировать распределение с плотностью

$$p(x) = c \begin{cases} e^{-x^2} & \text{при } 1 \leq x \leq \infty, \\ 0 & \text{иначе} \end{cases}$$

и исследовать его трудоемкость.

Указание. Если Φ — функция распределения $N(0, 1)$, то $\Phi(1) \approx 0.8423 \dots$

23. Промоделировать распределение с плотностью

$$p(x) = c \begin{cases} e^{-x^2} & \text{при } 0 \leq x \leq \infty, \\ 0 & \text{иначе} \end{cases}$$

не моделируя нормальное распределение.

24. Промоделировать распределение с плотностью $p(x) = c(1 + |x|)^{-3}$.

25. Провести двумя способами моделирование распределения с плотностью

$$p(x) = c \begin{cases} 1 - x^4 & \text{при } -1 \leq x \leq 0, \\ x^2 & \text{при } 0 < x \leq 1, \\ 0 & \text{иначе.} \end{cases}$$

26. Промоделировать тремя способами распределение с плотностью

$$p(x) = c \begin{cases} x^2 & \text{при } 0 \leq x \leq 1, \\ c & \text{при } 1 < x \leq 2, \\ 0 & \text{иначе.} \end{cases}$$

27. Промоделировать тремя способами распределение с плотностью

$$p(x) = c \begin{cases} 1 & \text{при } 0 \leq x \leq 1, \\ x^2 & \text{при } 1 < x \leq 2, \\ 0 & \text{иначе.} \end{cases}$$

28. Промоделировать четырьмя способами распределение с плотностью

$$p(x) = c \begin{cases} x^2 & \text{при } 1 \leq x \leq 1, \\ 1 & \text{при } 1 < x \leq 5/3, \\ 0 & \text{иначе.} \end{cases}$$

29. Промоделировать тремя способами распределение с плотностью $p(x) = c(x^2 + x^3 + x^4)$, $x \in [0, 1]$.

30. Промоделировать распределение с плотностью $p(x) = ce^{-ax}(1 + e^{-bx})$, $x > 0$. Параметры a, b положительны.

31. Промоделировать распределение с плотностью $p(x) = c(x^{-2} + x^2e^{-x^3})$, $x > 1$.

32. Промоделировать распределение с плотностью

$$p(x) = c \begin{cases} x^2 & \text{при } 0 \leq x \leq 1, \\ x + 1 & \text{при } 1 < x \leq 2, \\ 0 & \text{иначе.} \end{cases}$$

33. Промоделировать распределение с плотностью

$$p(x) = c \begin{cases} x^2 & \text{при } 0 \leq |x| \leq 1, \\ |x| - 1 & \text{при } 1 < |x| \leq 2, \\ 0 & \text{иначе.} \end{cases}$$

34. Промоделировать распределение с плотностью

$$p(x) = c \begin{cases} e^{-(1-x)^2} & \text{при } x \geq 1, \\ 0 & \text{иначе.} \end{cases}$$

35. Промоделировать двумя способами распределение с плотностью

$$p(x) = c \begin{cases} e^{-|x|} & \text{при } x \geq 1, \\ 0 & \text{иначе.} \end{cases}$$

36. Промоделировать двумя способами распределение с плотностью

$$p(x) = c \begin{cases} x^{-2} & \text{при } |x| \geq 1, \\ |x| & \text{иначе.} \end{cases}$$

37. Промоделировать тремя способами распределение с плотностью

$$p(x) = c \begin{cases} e^x & \text{при } |x| \leq 1, \\ |x| & \text{при } 1 < |x| \leq 2, \\ 0 & \text{иначе.} \end{cases}$$

38. Промоделировать тремя способами распределение с плотностью

$$p(x) = \begin{cases} x/2 & \text{при } 0 \leq x \leq 1, \\ c(3-x) & \text{при } 1 < x \leq 2, \\ 0 & \text{иначе.} \end{cases}$$

39. Промоделировать тремя способами распределение с плотностью

$$p(x) = c \begin{cases} |x-3|/2 & \text{при } |x-3| \leq 2, \\ 1 & \text{при } 5 < x \leq 6, \\ 0 & \text{иначе.} \end{cases}$$

40. Промоделировать тремя способами распределение с плотностью

$$p(x) = c \begin{cases} |x-1| & \text{при } |x-1| \leq 1, \\ |x+1| & \text{при } |x+1| \leq 1, \\ 0 & \text{иначе.} \end{cases}$$

41. Промоделировать тремя способами распределение с плотностью

$$p(x) = c \begin{cases} 2-x & \text{при } 0 < x \leq 2, \\ x+1 & \text{при } -1 < x \leq 0, \\ 0 & \text{иначе.} \end{cases}$$

42. Промоделировать тремя способами распределение с плотностью

$$p(x) = c \begin{cases} 1 & \text{при } -1 < x \leq 0, \\ 2 - |x - 1| & \text{при } 0 < x \leq 2, \\ 0 & \text{иначе.} \end{cases}$$

43. Промоделировать тремя способами распределение с плотностью

$$p(x) = c \begin{cases} 1 - |x| & \text{при } 0 < |x| \leq 1, \\ 1 - ||x| - 3| & \text{при } 2 < |x| \leq 4, \\ 0 & \text{иначе.} \end{cases}$$

44. Промоделировать тремя способами распределение с плотностью

$$p(x) = c \begin{cases} 1 & \text{при } -2 < x \leq -1, \\ 2 & \text{при } 1 < x \leq 2, \\ |x| & \text{при } |x| \leq 1, \\ 0 & \text{иначе.} \end{cases}$$

45. Промоделировать тремя способами распределение с плотностью

$$p(x) = c \begin{cases} 1 & \text{при } 0 < x \leq 1, \\ 2 - x & \text{при } 1 < x \leq 2, \\ 1 - 2|x + 1/2| & \text{при } -1 < x \leq 0, \\ 0 & \text{иначе.} \end{cases}$$

46. Промоделировать двумя способами распределение с плотностью

$$p(x) = c \begin{cases} |x|^3 + e^{-|x|} & \text{при } |x| \leq 1, \\ e^{-2x} & \text{при } x > 1. \\ e^{3x} & \text{иначе.} \end{cases}$$

47. Промоделировать двумя способами распределение с плотностью

$$p(x) = c \begin{cases} e^x & \text{при } x \leq -2, \\ 1 & \text{при } -2 < x < 1. \\ e^{-(x-1)^2} & \text{иначе.} \end{cases}$$

48. Промоделировать двумя способами распределение с плотностью

$$p(x) = c \begin{cases} x^2 & \text{при } |x| \leq 2, \\ e^{-|x|} & \text{при } |x| > 2. \end{cases}$$

49. Промоделировать двумя способами распределение с плотностью

$$p(x) = c \begin{cases} 1/\sqrt{x} & \text{при } 0 < x \leq 1/2, \\ 1/\sqrt{1-x} & \text{при } 1/2 < x < 1. \end{cases}$$

50. Промоделировать двумя способами распределение с плотностью

$$p(x) = \frac{1}{\pi} \begin{cases} 1/\sqrt{x(1-x)} & \text{при } 0 < x \leq 1, \\ 0 & \text{иначе.} \end{cases}$$

51. Промоделировать распределение с плотностью

$$p(x) = \frac{5}{12} (1 + (x - 1)^4), \quad x \in [0, 2].$$

52. Промоделировать распределение с плотностью

$$p(x) = c (0.5 + (2x + 1)^3), \quad x \in [1, 2].$$

53. Промоделировать распределение с плотностью $p(x) = cx^2 e^{-x^2/2}$ двумя способами, один из которых не является методом отбора.

54. Промоделировать распределение с плотностью $p(x) = cx^3 e^{-x^2/2}$, $x > 0$, двумя способами, один из которых не является методом отбора.

55. Промоделировать распределение с плотностью $p(x) = cx^{5/3} e^{-x}$, $x > 0$.

56. Промоделировать случайную величину с плотностью распределения

$$p(x) = \frac{e^{-\lambda(1-x)}}{\lambda x + 1}, \quad x \in (0, 1).$$

Параметр λ положителен.

57. Промоделировать распределение с плотностью $p(x) = cx/(1 + x^4)$, $x > 0$.

58. Промоделировать распределение с плотностью $p(x) = c \cosh(x)$, $0 < x < 1$.

59. Промоделировать распределение с плотностью $p(x) = cx \cosh(x)$, $0 < x < 1$.

60. Промоделировать распределение с плотностью $p(x) = cx \sinh(R - x)$, $0 < x < R$.

Моделирование двумерных и многомерных распределений

1. Промоделировать двумя способами распределение с плотностью

$$p(x, y) = c \begin{cases} e^{-x-y} & \text{при } x^2 + y^2 \leq 1, \\ 0 & \text{иначе.} \end{cases}$$

2. Промоделировать распределение с плотностью

$$p(x, y) = c \begin{cases} e^{-x^2-y^2} & \text{при } x > 0, \quad y > 0, \\ 0 & \text{иначе} \end{cases}$$

3. Промоделировать двумя способами распределение с плотностью

$$p(x, y) = \begin{cases} c e^{-x^2-y^2} & \text{при } x + y > 0, \\ 0 & \text{иначе.} \end{cases}$$

4. Промоделировать распределение с плотностью

$$p(x, y) = c \begin{cases} e^{-x^2-y^2} & \text{при } -1 \leq x + y \leq 1, \quad |x| < 3, \quad |y| < 3, \\ 0 & \text{иначе} \end{cases}$$

5. Промоделировать распределение с плотностью

$$p(x, y) = c \begin{cases} e^{-x^2-y^2} & \text{при } |x+y| \leq 1, |x-y| \leq 1, \\ 0 & \text{иначе} \end{cases}$$

двумя способами.

6. Промоделировать двумя способами распределение с плотностью

$$p(x, y) = c \begin{cases} e^{-x^2-y^2} & \text{при } x^2 + y^2 \leq 1, \\ 0 & \text{иначе.} \end{cases}$$

7. Промоделировать двумя способами распределение с плотностью

$$p(x, y) = c \begin{cases} e^{-x^2-y^2} & \text{при } |x| < A, |y| \leq B, \\ 0 & \text{иначе.} \end{cases}$$

8. Промоделировать двумя способами распределение с плотностью

$$p(x, y) = \begin{cases} c e^{-x^2-y^2} & \text{при } x^2 + y^2 > 1, \\ 0 & \text{иначе.} \end{cases}$$

9. Промоделировать двумя способами распределение с плотностью

$$p(x, y) = \begin{cases} c e^{-x^2-y^2} & \text{при } 1 \leq x^2 + y^2 \leq 4, \\ 0 & \text{иначе.} \end{cases}$$

10. Промоделировать двумя способами распределение с плотностью

$$p(x, y) = \frac{c}{\sqrt{x^2 + y^2}} e^{-x^2-y^2}.$$

11. Промоделировать двумя способами распределение с плотностью

$$p(x, y) = c(x^2 + y^2) e^{-x^2-y^2}.$$

12. Промоделировать двумя способами распределение с плотностью

$$p(x, y) = c \begin{cases} x^2 + y^2 & \text{при } 0 \leq |x|, |y| \leq 1, \\ e^{-x^2-y^2} & \text{при } x^2 + y^2 > 4, \\ 0 & \text{иначе.} \end{cases}$$

13. Промоделировать двумя способами распределение с плотностью

$$p(x, y) = c \begin{cases} 1 & \text{при } 0 < x^2 + y^2 < 1, \\ e^{-x^2-y^2} & \text{при } x^2 + y^2 > 4, \\ 0 & \text{иначе.} \end{cases}$$

14. Промоделировать двумя способами распределение с плотностью

$$p(x, y) = c(x^2 |y| + |x| y^2) e^{-x^2-y^2}.$$

15. Промоделировать распределение с плотностью

$$p(x, y) = \begin{cases} c|x||y| & \text{при } |x| < 1, |y| < 1, \\ 0 & \text{иначе.} \end{cases}$$

16. Промоделировать распределение с плотностью

$$p(x, y) = \begin{cases} c|x||y| & \text{при } 0 \leq x^2 + y^2 \leq 1, \\ 0 & \text{иначе.} \end{cases}$$

тремя способами. Сравнить трудоемкости.

17. Промоделировать распределение с плотностью

$$p(x, y) = \begin{cases} \frac{c}{\sqrt{x^2 + y^2}} & \text{при } x^2 + y^2 \leq 1, \\ 0 & \text{иначе.} \end{cases}$$

18. Промоделировать распределение с плотностью

$$p(x, y) = c(x^2|y| + |x|y^2)e^{-(x^2+y^2)}.$$

19. Промоделировать распределение с плотностью

$$p(x, y) = \begin{cases} c(e^{-x-y} + e^{x+y}) & \text{при } 0 \leq |x|, |y| \leq 1, \\ 0 & \text{иначе.} \end{cases}$$

двумя способами.

20. Промоделировать трехмерное распределение с плотностью

$$p(x, y, z) = c \begin{cases} \frac{1}{1 + \|\mathbf{x}\|^2} \frac{1}{\|\mathbf{x}\|} & \text{при } \|\mathbf{x}\|^2 \leq A^2, \\ 0 & \text{иначе,} \end{cases}$$

где $\|\mathbf{x}\| = \sqrt{x^2 + y^2 + z^2}$.

21. Промоделировать трехмерное распределение с плотностью

$$p(x, y, z) = c \begin{cases} \frac{1}{1 + \|\mathbf{x}\|^2} & \text{при } \|\mathbf{x}\|^2 \leq A^2, x, y, z > 0, \\ 0 & \text{иначе,} \end{cases}$$

где $\|\mathbf{x}\| = \sqrt{x^2 + y^2 + z^2}$.

22. Промоделировать распределение с плотностью

$$p(x, y, z) = c(|x| + |y| + |z|)e^{-(|x|+|y|+|z|)}.$$

23. Промоделировать двумя способами распределение с плотностью

$$p(x, y) = \begin{cases} c|x|e^{-|y|} & \text{при } |x| + |y| \leq 1, \\ 0 & \text{иначе.} \end{cases}$$

24. Промоделировать двумя способами распределение с плотностью

$$p(x, y) = \begin{cases} cxy e^{-x-y} & \text{при } x^2 + y^2 \leq 1, \\ 0 & \text{иначе.} \end{cases}$$

25. Провести двумя способами моделирование распределения с плотностью

$$p(x, y) = \begin{cases} \frac{c}{1+x^2+y^2} & \text{при } x^2 + y^2 \leq 1, \\ 0 & \text{иначе.} \end{cases}$$

26. Промоделировать случайный вектор с плотностью распределения

$$p(x, y) = \begin{cases} c(x^2 + y^2) & \text{при } x^2 + y^2 \leq 1, \\ 0 & \text{иначе.} \end{cases}$$

27. Промоделировать случайный вектор с плотностью распределения

$$p(x, y) = \begin{cases} c(x^2 + y^2) & \text{при } |x + y| \leq 1, \\ 0 & \text{иначе.} \end{cases}$$

28. Промоделировать случайный вектор с плотностью распределения

$$p(x, y) = \begin{cases} c(x^2 + y^2) & \text{при } |x| + |y| \leq 1, \\ 0 & \text{иначе.} \end{cases}$$

29. Промоделировать случайный вектор с плотностью распределения

$$p(x, y, z) = c\sqrt{x^2 + y^2 + z^2} e^{-(x^2+y^2+z^2)}.$$

30. Промоделировать случайный вектор с плотностью распределения

$$p(x, y, z) = c\sqrt{x^2 + y^2 + z^2} e^{-(x^2+y^2+z^2)^2}.$$

31. Промоделировать четырехмерный случайный вектор с плотностью

$$p(x, y, z, t) = c \left(e^{-(x^2+y^2+z^2+t^2)} + p_1(x, y, z, t) \right),$$

где

$$p_1(x, y, z, t) = \begin{cases} |x||y||z||t| & \text{при } |x|, |y|, |z|, |t| < 1, \\ 0 & \text{иначе.} \end{cases}$$

32. Промоделировать случайный вектор с плотностью распределения

$$p(x, y, z) = \frac{c}{1 + (x^2 + y^2 + z^2)^2} \frac{1}{\sqrt{x^2 + y^2 + z^2}}.$$

33. Промоделировать двумя способами случайный вектор с плотностью распределения

$$p(x, y) = c \begin{cases} |\ln(xy)| & 0 \leq x, y < 1, \\ 0 & \text{иначе.} \end{cases}$$

34. Промоделировать двумя способами случайный вектор с плотностью распределения

$$p(x, y, z) = c \begin{cases} x^2 + y^2 + z^2 & 0 \leq x^2 + y^2 + z^2 < 1, \\ 0 & \text{иначе.} \end{cases}$$

35. Промоделировать двумя способами случайный вектор с плотностью распределения

$$p(x, y, z) = c \begin{cases} xyz & 0 \leq x < y < z < 1, \\ 0 & \text{иначе.} \end{cases}$$

36. Промоделировать двумя способами случайный вектор с плотностью распределения

$$p(x, y, z) = c \begin{cases} x + y + z & 0 \leq x, y, z < 1, \\ 0 & \text{иначе.} \end{cases}$$

Гауссовские распределения

1. Промоделировать двумерный гауссовский вектор со средним $(0, 0)^T$ и ковариационной матрицей

$$\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

2. Промоделировать трехмерный гауссовский вектор с нулевым средним и ковариационной матрицей

$$\Sigma = \begin{pmatrix} 1 & 0.5 & 1 \\ 0.5 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

3. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (2, 1, 0)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 5 & 3 & 7 \\ 3 & 2 & 4 \\ 7 & 4 & 10 \end{pmatrix}.$$

4. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (0, 1, 1)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 2 & 3 & 4 \\ 3 & 5 & 7 \\ 4 & 7 & 10 \end{pmatrix}.$$

5. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (0, -1, 1)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 6 & 5 \\ 1 & 5 & 6 \end{pmatrix}.$$

6. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (1, 1, 0)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 1 & 3 & 2 \\ 2 & 3 & 3 \\ 2 & 3 & 3 \end{pmatrix}.$$

7. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (1, 0, -1)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 5 \end{pmatrix}.$$

8. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (3, 0, 1)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 5 & 3 & 4 \\ 3 & 2 & 2 \\ 4 & 3 & 5 \end{pmatrix}.$$

9. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (0, 0, 1)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}.$$

10. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (0, 1, 2)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 2 & 0 & 3 \\ 0 & 2 & 1 \\ 3 & 1 & 5 \end{pmatrix}.$$

11. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (1, 0, 1)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 1 & 2 \\ 2 & 2 & 4 \end{pmatrix}.$$

12. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (1, 0, -1)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 3 & 1 & 2 \\ 1 & 3 & 2 \\ 2 & 2 & 2 \end{pmatrix}.$$

13. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (1, 2, -1)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 3 & 3 \\ 2 & 3 & 3 \end{pmatrix}.$$

14. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (0, 1, -1, 0)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 2 & 1 & 1 & -1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & -2 \\ -1 & 1 & -2 & 5 \end{pmatrix}.$$

15. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (2, 1, -1, 1)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 2 & -1 & 1 & -1 \\ -1 & 5 & 1 & 1 \\ 1 & 1 & 2 & -2 \\ 1 & 1 & -2 & 10 \end{pmatrix}.$$

16. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (-1, 1, -1, 2)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 5 & -4 \\ 0 & 1 & -4 & 5 \end{pmatrix}.$$

17. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (0, 1, -1, 3)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 2 \\ 1 & 0 & 2 & 0 \\ 1 & 2 & 0 & 3 \end{pmatrix}.$$

18. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (0, 1, 2, 2)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 3 & -1 & 4 \\ 1 & -1 & 5 & -1 \\ 1 & 4 & -1 & 6 \end{pmatrix}.$$

19. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (0, 1, -1, -1)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 3 & 2 & 2 & 1 \\ 2 & 2 & 1 & 0 \\ 2 & 1 & 2 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}.$$

20. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (0, 1, 1, -3)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 2 \\ 1 & 0 & 2 & 0 \\ 1 & 2 & 0 & 3 \end{pmatrix}.$$

21. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (0, 1, 1, -1)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 3 & -1 & 2 \\ 1 & -1 & 5 & 0 \\ 1 & 2 & 0 & 2 \end{pmatrix}.$$

22. Промоделировать гауссовский вектор со средним \bar{a} и ковариационной матрицей Σ , если

$$\bar{a} = (3, 1, -1, -1)^T \quad \text{и} \quad \Sigma = \begin{pmatrix} 3 & 2 & 1 & 1 \\ 2 & 6 & 2 & -1 \\ 1 & 2 & 1 & 0 \\ 1 & -1 & 0 & 1 \end{pmatrix}.$$

23. Решить задачи Примера в разделе 5.2.

Различные распределения

1. Промоделировать случайную величину ξ с распределением

$$\mathbb{P}(\xi = k) = c \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{при } 0 \leq k \leq N.$$

двумя способами. Число $c = c(N)$ считается известным.

2. Промоделировать случайную величину ξ с распределением

$$\mathbb{P}(\xi = k) = c_n C_n^k 2^{-n}, \quad 0 < k < n$$

двумя способами.

3. Промоделировать случайную величину ξ с распределением

$$\mathbb{P}(\xi = k) = c_n C_n^k 2^{-n}, \quad 0 < k < \lfloor n/2 \rfloor.$$

двумя способами.

4. Промоделировать распределение с плотностью

$$p(x) = \frac{1}{2\sqrt{2\pi}} \int_{-1}^1 e^{-(x-y)^2/2} dy.$$

5. Промоделировать двумя способами распределение, имеющее функцию распределения

$$F_p(x) = \frac{p}{1 - (1-p)x} \quad \text{при } x \in [0, 1].$$

Параметр p принимает значения из отрезка $(0, 1)$.

6. Промоделировать двумя способами распределение с плотностью

$$p(x) = c \begin{cases} \sum_{n \geq 0} n x^n 2^{-n-1} + 1/(2-x) & \text{при } x \in [0, 1], \\ 0 & \text{иначе.} \end{cases}$$

7. Промоделировать двумя способами распределение с плотностью

$$p(x) = c \begin{cases} e^{-1} \sum_{n \geq 1} \frac{x^n}{(n-1)!} + e^{x-1} & \text{при } x \in [0, 1], \\ 0 & \text{иначе.} \end{cases}$$

8. Пусть мы умеем моделировать распределение с плотностью $p(x)$. Как промоделировать распределение с плотностью $q(x) = 2xp(x^2)$?

9. Предлагается моделировать распределение \mathcal{P} двумя способами. Первый из них является методом отбора распределения \mathcal{Q}_1 . Во втором сначала методом отбора с помощью распределения \mathcal{Q}_1 моделируется распределение \mathcal{Q} , а потом моделируется \mathcal{P} отбором из \mathcal{Q} . Выяснить, разные это методы или нет. Все условия для методов отбора считаются выполненными.

Проверить результат на примере абсолютно непрерывных распределений \mathcal{P} , \mathcal{Q}_1 и \mathcal{Q} с плотностями $p(x) = 3x^2$, $q(x) = 2x$, $q_1(x) = 1$, где $x \in [0, 1]$.

10. Промоделировать распределение с плотностью

$$p(x, y) = \begin{cases} \frac{c}{x^2 + y^2} & \text{при } x^2 + y^2 \leq 1, \\ 0 & \text{иначе.} \end{cases}$$

11. k маленьких различных шариков раскладываются по n различным большим ящикам (все шарики помещаются в один ящик). Промоделировать двумя способами совместное распределение числа шаров, попавших в ящики с номерами 1, 2 и 3.

12. Промоделировать двумя способами случайную величину ξ с распределением $\mathbb{P}(\xi = n) = np^2(1-p)^{n-1}$, $n \geq 1$, $0 < p < 1$.

13. Промоделировать случайную величину ξ с распределением

$$\mathbb{P}(\xi = n) = \frac{3 + 2^{n+1}}{2^{2n+1}}, \quad n \geq 1.$$

14. Промоделировать биномиальное распределение $\text{Bin}(n, 0.5)$, если известно, что n является параметром распределения, лежащим в промежутке от n_0 до $n_0 + 10$, где n_0 — большое число. Какой метод следует выбрать, чтобы моделирование было достаточно быстрым, а требуемая память — не слишком большой? Приведите несколько вариантов решения.

Список литературы

- [1] И.М. Соболев (1973), Численные методы Монте-Карло, М. Наука.
- [2] С.М. Ермаков (1975), Метод Монте-Карло и смежные вопросы. 2-е изд., М., Наука.
- [3] С.М. Ермаков, Г.А. Михайлов (1982), Статистическое моделирование, М. Наука.
- [4] С.М. Ермаков (2009). Метод Монте-Карло в вычислительной математике. Вводный курс. Невский диалект, СПб.
- [5] R.Y.Rubinstein, B. Melamed (1997). Modern Simulation and Modelling, Wiley, NY.
- [6] Handbook of Simulation. Principles, Methodology, Advances, Applications, and Practice (1998), J. Banks (editor), Wiley, NY.
- [7] G.S. Fishman (1999), Monte Carlo. Concepts, algorithms and applications, Springer-Verlag, New York-Berlin-Amsterdam.
- [8] J.E. Gentle (2003), Random Number Generation and Monte Carlo Methods (Statistics and Computing), 2nd ed., Springer.
- [9] R.Y.Rubinstein, D.P. Kroese (2008), Simulation and the Monte Carlo Method, 2nd ed. Wiley Series in Probability and Statistics. Wiley-Interscience.
- [10] L. Devroye (1986), Non-uniform random variate generation, Springer-Verlag, New York.
- [11] Д. Кнут, Э. Яо. (1983), Сложность моделирования неравномерных распределений. Кибернетический сборник, новая серия, выпуск 19, М. Мир, стр. 97 – 158.
- [12] H.C. Chen and Y. Asau (1974), On generating random variates from an empirical distribution, AIIE Transactions, V.6, pp. 163 – 166.
- [13] A.J. Walker (1977), An efficient method for generating discrete random variables with general distributions, ACM Transactions on Mathematical Software, V.3, pp. 253 – 256.
- [14] R.A. Kronmal, A.V. Peterson (1979), On the alias method for generating random variables from a discrete distribution, The American Statistician, V.33, pp. 214 – 218.
- [15] M.D. Vose (1991), A linear algorithm for generating random numbers with a given distribution, IEEE Transactions on Software Engineering, V.17, 9, pp. 972 – 974.
- [16] М. Лозев (1962), Теория вероятностей, М., ИЛ.
- [17] П. Биллингсли (1977), Сходимость вероятностных мер, М., Наука.
- [18] А.Н. Ширяев (2004), Вероятность-1, М., Изд-во МЦНМО.
- [19] Д. Кнут (2007), Искусство программирования, т.2: Получисленные алгоритмы, М., Вильямс.
- [20] J.H. Ahrens and U. Dieter (1974), Computer methods for sampling from gamma, beta, Poisson and binomial distributions, Computing, v.12, pp. 223 - 246.
- [21] J. von Neumann (1951), Various techniques used in connection with random digits, Applied Math Series, V.12, p. 36–38.
- [22] Б.Б. Походзей (1983), Преобразование случайных битов в случайные величины с конечными дискретными распределениями, Вестник Ленингр. ун-та, Сер. 1, No 13, стр. 31–36.

- [23] Б.Б. Походзей (1985), Преобразование случайных битов в случайные величины с произвольными дискретными распределениями, Вестник Ленингр. ун-та, 1985. Сер. 1, No 1, стр. 39–43.
- [24] А.Файнштейн (1960), Теория информации, М., ИИЛ.
- [25] И. Чисар, Я. Кёрнер (1985), Теория информации. Теоремы кодирования для дискретных систем без памяти. Мир, Москва, 395 с.
- [26] Н.А. Воробьева, А.И. Коробейников, В.В. Некруткин (2012), Об оптимальном моделировании дискретных распределений, Вестник СПбГУ, Сер. 1, Вып. 3, С. 14–23.