Алгоритмические тесты и случайность относительно классов мер

Л. Биенвеню[*], П. Гач[†], К. Рохас[‡], М. Хойруп[§], А. Шень[¶]

Аннотация

В этой работе приводятся некоторые новые результаты об алгоритмической случайности по отношению к классам мер, а также подробно излагаются известные (но не опубликованные подробно) результаты об алгоритмических тестах случайности.

Мы начинаем с переформулировки определения случайности по Мартин-Лёфу в терминах тестов случайности (функций, измеряющих степень "неслучайности" последовательностей). Приводится формула, выражающая значение универсального теста в терминах префиксной сложности. Рассматриваются также варианты определения дефекта случайности для конечных слов, связанные с универсальным тестом.

Далее рассматривается (введённое ещё Мартин-Лёфом) понятие бернуллиевой последовательности (как последовательности, не противоречащей гипотезе о том, что все испытания независимы и имеют одинаковую вероятность успеха). Показано, что определение с помощью универсального теста эквивалентно первоначальному определению Мартин-Лёфа и что последовательность является бернуллиевой тогда и только тогда, когда она случайна по Мартин-Лёфу относительно бернуллиевой меры B_p при некотором p (с оракулом для p).

Затем этот же вопрос (о сравнении тестов относительно классов мер и тестов как функции двух аргументов — последовательности и меры) применяется к произвольным эффективно замкнутым классам мер в канторовском пространстве. Изучаются свойства ортогональных классов мер и указываются предположения, в которых два понятия случайности (равномерная и безоракульная) совпадают.

В заключение рассматриваются обобщения некоторых из указанных результатов на случай произвольных метрических пространств.

1 Введение

Эта работа может рассматриваться как продолжение [10] (которая в свою очередь является развитием давних идей Л. Левина) и [12].

[*]Laurent Bienvenu, LIAFA, CNRS & Université Paris Diderot, Paris 7, Case 7014, 75205 Paris Cedex 13, France, e-mail: `Laurent dot Bienvenu at liafa dot jussieu dot fr`

[†]Peter Gács, Department of Computer Science, Boston University, 111 Cummington st., Room 138, Boston, MA 02215, e-mail: `gacs at bu dot edu`

[‡]Cristobal Rojas, Department of Mathematics, University of Toronto, Bahen Centre, 40 St. George St., Toronto, Ontario, Canada, M5S 2E4, e-mail: `crojas at math dot utoronto dot ca`

[§]Mathieu Hoyrup, LORIA – B248, 615, rue du Jardin Botanique, BP 239, 54506 Vandœuvre-lès-Nancy, France, e-mail: `Mathieu dot Hoyrup at loria dot fr`

[¶]Alexander Shen, LIF, Universitè Aix – Marseille, CNRS, 39, rue Joliot-Curie, 13453 Marseille cedex 13, France, on leave from ИППИ РАН, Б. Каретный, 19, Москва. Supported by NAFIT ANR-08-EMER-008-01, RFBR 0901-00709-a grants. e-mail: `sasha dot shen at gmail dot com`.

Хорошо известны различные варианты определения понятия случайной последовательности нулей и единиц, соответствующие равномерному распределению (бросанию честной монеты). Большинство этих результатов естественно переносится на случай произвольного вычислимого распределения вероятностей на пространстве Ω бесконечных последовательностей нулей и единиц.

Наша цель — исследовать возможности определения случайности в более общей ситуации, когда распределение на Ω не является вычислимым (или когда рассматривается случайность в других пространствах, не только в Ω). Для этой цели мы рассматриваем тест случайности $\mathbf{t}(\omega, P)$ как функцию двух переменных, последовательности ω и меры P. Большие значения такого теста, интуитивно говоря, соответствуют ситуациям, когда гипотеза о том, что последовательность ω получилась в результате случайного выбора по мере P, неправдоподобна.

Кроме того, следуя [15], мы будем рассматривать тесты относительно классов мер, обладающих свойством типа компактности. Такой тест, $t_{\mathcal{C}}(\omega)$, измеряет, насколько неправдоподобным кажется появление последовательности ω в результате случайного процесса, распределение вероятностей которого (нам неизвестное) принадлежит классу $\mathcal{C}$. Мы покажем, что для класса бернуллиевых мер (независимые испытания с одинаковой вероятностью успеха) возникающее понятие случайности относительно этого класса совпадает с введённым Мартин-Лёфом в [19].

Для классов, меры в которых попарно ортогональны (в некотором эффективном смысле), мы получаем разложение теста случайности по данной мере на два: один проверяет случайность относительно класса мер, а второй проверяет (достаточно грубо) соответствие последовательности конкретной мере. Для случая бернуллиевых мер в качестве второго теста можно взять просто закон больших чисел и проверять, что предельная частота действительно равна декларируемой вероятности успеха. Аналогичное разбиение возможно и для других классов, соответствующих эргодическим стационарным процессам.

Определение случайности с помощью равномерных тестов $\mathbf{t}(\omega, P)$, вообще говоря, не обладает некоторыми интуитивно желательными свойствами (скажем, не монотонно по P в естественном смысле). Но для случая эффективно ортогональных классов оно равносильно другому, "слепому" определению случайности, в котором рассматриваются лишь тесты, вычисление которых не использует меру P как оракул.

Статья начинается с переформулировки определения случайности по Мартин-Лёфу (относительно вычислимых мер) в терминах тестов. Значения теста мы рассматриваем как количественную характеристику "неслучайности" последовательности, и считаем случайными последовательностями те, для которых тест конечен. Мы рассматриваем два вида тестов (ограниченные в среднем и ограниченные по вероятности) и показываем, что они близки друг к другу.

Затем мы приводим формулу, которая выражает значение (ограниченного в среднем) теста через префиксную сложность (и даже два варианта такой формулы — с максимумом и с суммой). Эта формула является количественным уточнением критерия случайности Левина – Шнорра (в форме с префиксной сложностью, как в статье Чейтина). Далее мы обсуждаем некоторые варианты движения в обратную сторону: как от дефекта случайности для бесконечных последовательностей перейти к дефекту случайности конечных.

Далее мы определяем понятие теста бернуллиевости (частный случай случайности относительно класса мер, в данном случае — класса бернуллиевых мер). Мы показываем, что множество бернуллиевых последовательностей, для которых этот тест конечен, совпадает с объединением по всем $p \in [0, 1]$ множеств последовательностей, случайных в смысле Мартин-Лёфа относительно бернуллиевой меры B_p, при этом в определении случайности добавляется оракул для p. Для этого мы вводим понятие равномерного бернуллиева теста и устанавливаем

количественный вариант указанного результата: дефект бернуллиевости равен точной нижней грани (по p) дефектов случайности относительно каждой из мер B_p.

После этого мы вводим понятие равномерного теста (уже не ограничиваясь конкретным классом мер) и соответствующее ему понятие равномерной случайности (которое для случая вычислимых мер совпадает с определением Мартин-Лёфа).

Бернуллиевы меры обладают тем свойством, что видя случайную по одной из этих мер последовательность, можно восстановить значение p как предельную частоту единиц (закон больших чисел). Мы показываем, что для подобных классов мер различные определения случайности (равномерное и "слепое", когда тест не использует меру) равносильны. Это утверждение обобщается и на меры в произвольных конструктивных метрических пространствах.

Наконец, введём некоторые полезные обозначения.

Мы будем писать $f(x) \stackrel{\cdot}{<} g(x)$ для положительных функций f и g, если указанное неравенство выполняется с точностью до мультипликативной константы, то есть $f(x) \leqslant cg(x)$ для некоторого c и для всех x.

Запись $f(x) \stackrel{\cdot}{=} g(x)$ означает $f(x) \stackrel{\cdot}{<} g(x)$ и $g(x) \stackrel{\cdot}{<} f(x)$.

Обозначения $f(x) \stackrel{+}{<} g(x)$ и $f(x) \stackrel{+}{=} g(x)$ имеют аналогичный смысл (неравенство с точностью до аддитивной константы).

Через Λ мы обозначаем пустое слово (строку нулевой длины). Длина слова x обозначается $|x|$. Запись $x \sqsubseteq y$ или $y \sqsupseteq x$ означает, что слово x является началом (префиксом) слова y. Элементы бесконечной последовательности x (а также буквы слова x) обозначаются $x(1), x(2), \ldots$; её начало длины n обозначается $x(1:n)$.

Часто мы рассматриваем функции со значениям в множестве $\overline{\mathbb{R}}_+ = [0, +\infty]$ (неотрицательные функции, возможно, бесконечные в некоторых точках). Множества натуральных и действительных чисел обозначаются $\mathbb{N}$ и $\mathbb{Z}$; через $\mathbb{B}$ иногда обозначается множество $\{0, 1\}$.

Логарифмы, если не указано иное, берутся по основанию 2.

2 Случайность последовательностей относительно вычислимых мер

Мы начнём с определения случайности бесконечных двоичных последовательностей относительно вычислимых мер.

2.1 Перечислимые снизу функции на Ω

Определение 2.1 (Канторовское пространство)**.** *Множество* $\{0,1\}^{\mathbb{N}}$ *бесконечных двоичных последовательностей мы называем* двоичным канторовским пространством *и обозначаем* Ω. *Для каждого конечного двоичного слова* x *мы рассматриваем* интервал $x\Omega$, *состоящий из всех последовательностей, начинающихся на* x. *Интервалы являются* базисными открытыми множествами *стандартной топологии канторовского пространства;* открытыми *являются произвольные объединения интервалов.*

Понятие открытого множества, как и другие топологические понятия, имеет эффективный аналог.

Определение 2.2. Эффективно открытыми *множествами называют объединения перечислимых семейств интервалов.* Эффективно замкнутыми *называются дополнения эффективно открытых множеств.*

Далее можно определить эффективные G_δ *множества как счётные пересечения* $\cap_i U_i$ *последовательности эффективно открытых множеств* U_i; *при этом требуется, чтобы* U_i

было эффективно открыто равномерно по i (алгоритм получает на вход i и перечисляет интервалы, образующие U_i).

Будем называть функцию $t\colon \Omega \to [0,\infty]$ перечислимой снизу*, если*

(а) *для любого рационального r множество $U_r = \{\omega \mid r < t(\omega)\}$ открыто,*

(б) *и, более того, U_r эффективно открыто равномерно по r (существует алгоритм, который получает r и перечисляет интервалы, образующие U_r).*

Условие (а) означает, что функция t полунепрерывна снизу, так что перечислимость снизу является эффективизацией понятия полунепрерывности.

Понятие перечислимой снизу функции в дальнейшем играет важную роль. Его можно определить различными (эквивалентными) способами. Вот одна из таких переформулировок.

Определение 2.3. *Функция с рациональными значениями, определённая на Ω, называется* базисной*, если её значение на последовательности ω определяется конечным началом ω.*

Если это начало имеет длину N, то функция может принимать до 2^N значений, и её можно задать таблицей из 2^N строк, в которой для каждого варианта начала (двоичного слова длины N) указано рациональное значение функции. Поэтому базисные функции можно считать конструктивными объектами.

Следующее предложение легко следует из определений:

Предложение 2.4. *Перечислимые функции и только они являются поточечными пределами вычислимых возрастающих последовательностей базисных функций.*

Разность двух базисных функций тоже является базисной функцией, поэтому вместо пределов возрастающих функций можно говорить о суммах рядов, составленных из неотрицательных базисных функций.

Вот ещё один вариант определения перечислимых снизу функций на Ω.

Определение 2.5 (Порождающие функции)**.** *Будем говорить, что определённая на двоичных словах функция T со значениями в $[0,+\infty]$ является* перечислимой снизу*, если множество пар $\langle x, r\rangle$, где x — двоичное слово, а r — рациональное число, меньшее $T(x)$, перечислимо.*

Для каждой такой функции T определим функцию t на бесконечных последовательностях, положив

$$t(\omega) = \sup_{x \sqsubseteq \omega} T(x);$$

будем говорить, что функция T порождает *функцию t.*

Предложение 2.6. *При этом порождаются все перечислимые снизу функции на Ω и только они.*

Можно наложить дополнительные ограничения на порождающую функцию T, сохранив возможность порождать любую перечислимую снизу функцию на Ω. Например, можно требовать, чтобы функция T была монотонной (это значит, что $T(x) \leqslant T(y)$ при $x \sqsubseteq y$). В самом деле, от любой функции можно перейти к монотонной, положив $T'(x) = \max_{z \sqsubseteq x} T(z)$. Можно также потребовать, чтобы функция T принимала рациональные значения и была бы вычислимой (а не только перечислимой снизу). В самом деле, поскольку в определении участвует $\sup T(x)$ по всем x, являющимся началом ω, вместо увеличения $T(x)$ для некоторого x можно увеличить значения функции T для всех его продолжений достаточно большой длины, и эта задержка позволяет сделать функцию T вычислимой.

Следующее наблюдение использует компактность канторовского пространства. Среди всех функций T, порождающих данную функцию t, можно выбрать максимальную, положив

$$T(x) = \inf_{\omega \sqsupseteq x} t(\omega).$$

Предложение 2.7. *Определённая таким образом функция T перечислима снизу и порождает t. В её определении можно заменить* inf *на* min.

Доказательство. Очевидно, что порождаемая функция не превосходит t. С другой стороны, если $t(\omega) > r$, то в силу полунепрерывности снизу это верно в некоторой окрестности ω, и потому $T(x) \geqslant r$ для некоторого начала $x \sqsubseteq \omega$. Таким образом, порождаемая функция совпадает с t.

Остаётся убедиться, что T перечислима снизу. В самом деле, $r < \inf_{\omega \sqsupseteq x} t(\omega)$ тогда и только тогда, когда существует $r' > r$, для которого $r' < t(\omega)$ для всех $\omega \sqsupseteq x$. Последнее условие может быть переформулировано так: открытое множество тех последовательностей ω, для которых $t(\omega) > r'$, покрывает интервал $x\Omega$. Это открытое множество (по определению перечислимости снизу) есть объединение перечислимого семейства интервалов, и в силу компактности уже конечное число интервалов образует подпокрытие. Поскольку это в какой-то момент обнаруживается, указанное свойство перечислимо, и квантор существования по r' сохраняет перечислимость.

Полунепрерывность снизу также гарантирует, что минимум на компактном множестве достигается, так что inf можно заменить на min.

2.2 Тесты случайности

Мы предполагаем, что читатель знаком с основными понятиями теории меры (хотя бы для канторовского пространства Ω). Напомним, что мера (распределение вероятностей) P на Ω задаётся своими значениями на цилиндрах $x\Omega$. Эти значения задают неотрицательную действительную функцию на двоичных словах, которую мы обозначаем той же буквой, что и саму меру:

$$P(x) = P(x\Omega).$$

При этом

$$P(\Lambda) = 1, \quad P(x) = P(x0) + P(x1),$$

и любая неотрицательная функция на двоичных словах, обладающая этими двумя свойствами, соответствует мере на Ω.

Среди всех мер выделяются вычислимые.

Определение 2.8 (Вычислимые меры)**.** *Действительное число x называется* вычислимым, *если существует алгоритм, который по любому рациональному $\varepsilon > 0$ указывает рациональное приближение к x с абсолютной погрешностью не более ε.*

Вычислимые действительные числа можно определять также как пределы последовательностей $x_1, x_2, \ldots$, для которых $|x_n - x_{n+k}| \leqslant 2^{-n}$.

Функция, определённая на словах (или иных конструктивных объектах) и принимающая действительные значения, называется вычислимой, *если её значения вычислимы равномерно по входу, то есть существует алгоритм, который по входу и по $\varepsilon > 0$ указывает ε-приближение к значению функции на этом входе.*

Мера на Ω называется вычислимой, *если вычислима функция $P\colon \{0,1\}^* \to [0,1]$, соответствующая этой мере.*

Пусть фиксирована вычислимая (вероятностная) мера на Ω.

Определение 2.9 (Тест случайности по вычислимой мере)**.** *Перечислимая функция* $t\colon \Omega \to [0, +\infty]$ *называется* (ограниченным в среднем) тестом *относительно меры* P (P*-тестом*), *если её интеграл* (*математическое ожидание*) *по мере* P *не превосходит* 1*:*

$$\int t(\omega)\, dP(\omega) \leqslant 1.$$

Последовательность ω проходит *тест* t*, если* $t(\omega)$ *конечно* (*напомним, что мы рассматриваем перечислимые снизу функции, которые могут принимать бесконечные значения*).

Последовательность ω *называется* случайной *по мере* P*, если она проходит все* P*-тесты.*

Интуитивный смысл этого определения можно описать так: $t(\omega)$ отражает "количество закономерностей" в ω. Строя тест, мы можем объявить "закономерностью" любое (эффективно обнаруживаемое) свойство последовательности, надо только следить, чтобы их было не слишком много, иначе интеграл превысит границу.

Это определение эквивалентно (даёт тот же класс случайных последовательностей) классическому определению Мартин-Лёфа (см. ниже). Но сначала отметим, что среди тестов существует универсальный (максимальный):

Теорема 2.10. *Для любой вычислимой меры* P *существует универсальный* (*максимальный*) *тест* u*: это означает, что для любого* P*-теста* t *найдётся константа* c*, при которой*

$$t(\omega) \leqslant c \cdot u(\omega)$$

при всех $\omega \in \Omega$.

В частности, универсальность гарантирует, что всякая проходящая тест u последовательность проходит все другие тесты. Тем самым множество последовательностей, проходящих тест u, совпадает с множеством случайных последовательностей.

Доказательство. Будем перечислять все алгоритмы, задающие перечислимые снизу функции. (Такой алгоритм порождает возрастающую последовательность базисных функций.) Не все эти перечислимые функции будут тестами (интеграл может превысить единицу), но мы можем их фильтровать и не пропускать очередную функцию, пока не будет установлено, что её интеграл меньше (скажем) 2. (Напомним, что мера P вычислима, поэтому если интеграл меньше 2, то мы сможем в этом убедиться.) Такая фильтрация пропустит все тесты и ещё некоторые функции, которые превосходят тесты не более чем вдвое. Остаётся сложить все профильтрованные функции с коэффициентами, сумма которых не превосходит $1/2$, скажем, $1/2^{i+2}$.

В такой форме тесты случайности фигурировали в [8].

Убедимся, что это определение эквивалентно классическому определению Мартин-Лёфа:

Определение 2.11. *Пусть* P *— вычислимое распределение вероятностей на* Ω*. Последовательность открытых множеств* $U_1, U_2, \ldots$ *называется* тестом Мартин-Лёфа*, если множество* U_i *эффективно открыто* (*равномерно по* i) *и его мера не превосходит* 2^{-i}*.*

Последовательность ω проходит *этот тест, если она не принадлежит пересечению* $\cap_i U_i$*. Такие пересечения* (*а также все их подмножества*) *называют* эффективно нулевыми множествами.

Можно было бы рассматривать только эти пересечения (а не все их подмножества): такие множества было бы логично называть *эффективными нулевыми множествами.* Они являются эффективными G_δ-множествами (пересечениями последовательности равномерно эффективно открытых множеств).

Как мы уже говорили, это определение равносильно приведённому выше:

Теорема 2.12. *Последовательность проходит все тесты Мартин-Лёфа тогда и только тогда, когда она проходит все ограниченные в среднем тесты.*

Доказательство. Если t — ограниченный в среднем тест, то множество U_i тех ω, для которых $t(\omega) > 2^i$, эффективно открыто и имеет меру не более 2^{-i}, так что получается тест Мартин-Лёфа. Поэтому если ω проходит все тесты Мартин-Лёфа, то $t(\omega)$ конечно.

С другой стороны, из любого теста Мартип-Лёфа $\{U_i\}$ легко сделать ограниченный в среднем тест. Положим $t_i(\omega)$ равным 2^i внутри U_i и нулю вне U_i; функция t_i перечислима снизу и имеет среднее не больше 1, то есть представляет собой тест. Остаётся сложить, скажем, t_{2i} с весами 2^{-i}: если ω лежит в U_{2i}, то такая сумма будет не меньше 2^i.

В дальнейшем, говоря о случайности, мы будем иметь в виду случайность в смысле (любого из) этих определений определений, если не оговорено противное.

Ограниченные в среднем тесты не только отделяют случайные последовательности от неслучайных, но и классифицируют случайные последовательности: чем больше значение теста, тем ближе последовательность к неслучайным. Удобно перейти при этом к логарифмической шкале:

Определение 2.13. *Фиксируем некоторый универсальный (ограниченный в среднем) P-тест $\mathbf{t}_P(\omega)$. Через $\mathbf{d}_P(\omega)$ обозначим логарифм этого теста:*

$$\mathbf{t}_P(\omega) = 2^{\mathbf{d}(\omega)}.$$

Можно сказать, что $\mathbf{d}_P(\omega)$ измеряет в битах "дефект случайности" последовательности ω (количество закономерностей в ω).

При нашем определении дефект бывает отрицательным (и даже может быть равным $-\infty$): интеграл от теста не больше 1, и потому тест имеет значения, меньшие единицы. Можно изменить универсальный тест (взять его полусумму с постоянным тестом, везде равным единице) и добиться того, чтобы дефект был всегда не меньше -1. Можно также сделать дефект целочисленным (заменив каждое значение теста на максимальную степень двойки, меньшую его). Чтобы избавиться от отрицательных дефектов, можно разрешить тестам иметь любые конечные средние (не обязательно меньше единицы):

Предложение 2.14. *Функция $\mathbf{d}_P$ является максимальной (с точностью до константы) перечислимой снизу функцией на Ω, для которой P-среднее от $2^{\mathbf{d}_P(\cdot)}$ конечно.*

Замечание 2.15.

1. *Оригинальное определение Мартин-Лёфа также может быть использовано для измерения дефекта случайности. Именно, можно считать, что элементы множества U_i имеют дефект i или больше. Этот способ измерения дефекта, использованный в* [29]*, эквивалентен ограниченным по вероятности тестам (см. следующий раздел).*

2. *Мы определили функцию $\mathbf{d}_P(x)$ отдельно для каждой меры P (с точностью до константы). В дальнейшем мы определим (также с точностью до константы) функцию $\mathbf{d}(\omega, P)$ двух аргументов, которая будет для каждой вычислимой меры P совпадать с $\mathbf{d}_P$ (с той же точностью).*

2.3 Тесты, ограниченные по вероятности и в среднем

Приведённое выше определение ограниченного в среднем теста в каком-то смысле аналогично определению префиксной колмогоровской сложности; есть и другой вариант определения,

который больше похож на обычную сложность. (Определение префиксной и обычной сложности можно найти, например, в [25, 18]; мы используем эти понятия лишь как образец для аналогий, и потому не обсуждаем их подробно.)

Ослабим требования на тесты: вместо условия $\int t(\omega)\,dP(\omega) \leqslant 1$ будем требовать, чтобы для любого $c > 0$ множество последовательностей ω, для которых $t(\omega) > c$, имело бы меру не больше $1/c$. (Это условие следует из прежнего согласно неравенству Чебышёва.) Такие тесты будем называть *ограниченными по вероятности*.

В логарифмической шкале это определение может быть переформулировано так: P-мера множества последовательностей, имеющих дефект больше n, не превосходит 2^{-n}. Если ограничиться целыми значениями n, то мы приходим к классическому определению Мартин-Лёфа (см. замечание 2.15).

Легко видеть, что и при этом определении среди всех тестов существует максимальный (с точностью до умножения на константу). Ему соответствует максимальная (с точностью до аддитивной константы) функция дефекта. В самом деле, будем перечислять все тесты (и почти-тесты, где условие на меру вдвое ослаблено) и соответствующие им функции дефекта d_i. Затем возьмём их максимум (с аддитивными добавками, соответствующими весам):

$$\mathbf{d}(\omega) = \max_i [d_i(\omega) - i] - c.$$

Этот максимум может быть меньше d_i только на $i + c$; с другой стороны, множество тех ω, для которых $\mathbf{d}(\omega) > k$, представляет собой объединение множеств $\{\omega \mid d_i(\omega) > k + i + c\}$. Меры этих множеств не превосходят $O(2^{-k-i-c})$, и при подходящем c их объединение имеет меру не больше 2^{-k}, как и требуется.

Возникает естественный вопрос: как связаны значения универсального ограниченного в среднем теста $\mathbf{t}^{\mathrm{aver}}$ и универсального ограниченного по вероятности теста $\mathbf{t}^{\mathrm{prob}}$? Как мы видели, они оба бесконечны на одних и тех же последовательностях; более того, и конечные значения их близки:

Предложение 2.16.

$$\mathbf{d}^{\mathrm{aver}}(\omega) \stackrel{+}{<} \mathbf{d}^{\mathrm{prob}}(\omega) \stackrel{+}{<} \mathbf{d}^{\mathrm{aver}}(\omega) + 2\log \mathbf{d}^{\mathrm{aver}}(\omega)$$

Доказательство. Как мы видели, ограниченные в среднем тесты автоматически ограничены по вероятности, откуда следует первое неравенство. Чтобы доказать второе неравенство, рассмотрим произвольный ограниченный по вероятности тест и его логарифм $d(\omega)$. Покажем, что $d - 2\log d$ ограничен в среднем (в логарифмической шкале). В самом деле, событие "$d(\omega)$ находится между $i-1$ и i" имеет вероятность не более $1/2^{i-1}$, интеграл от $2^{d-2\log d}$ по этому множеству не превосходит $2^{-i+1}2^{i-2\log i} = O(1/i^2)$, и потому интеграл по всему пространству конечен.

Остаётся заметить, что неравенство $a \stackrel{+}{<} b + 2\log b$ следует из $b \stackrel{+}{>} a - 2\log a$. В самом деле, из $b \geqslant a - 2\log a$ следует $b \geqslant a/2$ (при достаточно больших a) и потому $\log a \leqslant \log b + 1$, так что $a \stackrel{+}{<} b + 2\log a \stackrel{+}{<} b + 2\log b$.

В общем случае вопрос о том, как связана ограниченность в среднем и ограниченность по вероятности, разбирается в статье [23]. Там показано (и это несложно), что если $u\colon [1, +\infty] \to [0, +\infty]$ — монотонная непрерывная функция, для которой $\int_1^\infty u(t)/t^2\,dt \leqslant 1$, то $u(t(\omega))$ является ограниченным в среднем тестом для любого ограниченного по вероятности теста t, и что это условие на u нельзя улучшить. (Наша оценка получается при $u(x) \sim x/\log^2 x$.)

Замечание 2.17. *Последнее предложение напоминает соотношение между простой и префиксной сложностями (как и с точки зрения соотношения между определениями — в одном ограничивается интеграл, в другом количество объектов, — так и по результатам). Важно иметь в виду, что сейчас разница между двумя величинами ограничена логарифмом* дефекта*, и потому мала, если последовательность близка к случайной, в то время как для разница между префиксной и обычной сложностями оценивается через логарифм самих этих величин (который велик для случайных объектов).*

Вопрос. Интересно было бы понять, отличаются ли два вида тестов лишь некоторым сдвигом шкалы или более существенным образом. Подтверждением такого более существенного различия могло бы служить семейство последовательностей ω_i и ω'_i, для которых

$$\mathbf{d}^{\mathrm{aver}}(\omega_i) - \mathbf{d}^{\mathrm{aver}}(\omega'_i) \to +\infty$$

при $i \to \infty$, но

$$\mathbf{d}^{\mathrm{prob}}(\omega_i) - \mathbf{d}^{\mathrm{prob}}(\omega'_i) \to -\infty.$$

Авторы не знают, существует ли такое семейство.

2.4 Формула для ограниченного в среднем дефекта

Эта формула использует понятие априорной вероятности (или префиксную сложность); напомним соответствующие определения (подробнее см. в [25, 18]).

Определение 2.18. *Множество двоичных слов называется* беспрефиксным*, если ни один из его элементов не является началом другого. Вычислимая частичная функция T из множества двоичных слов в себя называется* самоограниченным декомпрессором*, если её область определения является беспрефиксным множеством. Мы определяем сложность $KP_D(x)$ слова x относительно декомпрессора D как минимальную длину слова p, для которого $D(p) = x$. Среди всех самоограниченных декомпрессоров существует* оптимальный*, для которого функция KP_D минимальна с точностью до аддитивной константы. Эта минимальная функция (для некоторого фиксированного оптимального декомпрессора) называется* префиксной сложностью *и обозначается $KP(x)$.*

Величина $\mathbf{m}(x) = 2^{-KP(x)}$ называется дискретной априорной вероятностью *слова x.*

Название “априорная вероятность” связано с тем, что эта функция является максимальной (с точностью до постоянного множителя) в некотором классе вероятностных распределений. Мы приведём соответствующие определения и формулировки без доказательства.

Определение 2.19. *Функция $f\colon \{0,1\}^* \to [0,\infty)$ называется* дискретной полумерой*, если $\sum_x f(x) \leqslant 1$.*

Перечислимые снизу дискретные полумеры можно описать как выходные распределения вероятностных алгоритмов, использующих датчик случайных чисел и выдающих на выход некоторое слово (если алгоритм останавливается; с некоторой вероятностью он может и не остановиться).

Предложение 2.20. *Функция $\mathbf{m}(x)$ является перечислимой снизу дискретной полумерой, максимальной в этом классе с точностью до константы: для любой перечислимой снизу дискретной полумеры f найдётся такая константа c, что $c \cdot \mathbf{m}(x) \geqslant f(x)$ при всех x.*

Теперь можно указать явную формулу для универсального ограниченного в среднем теста случайности:

Предложение 2.21. *Для данной вычислимой меры P универсальный ограниченный в среднем тест $\mathbf{t}_P$ задаётся формулой:*

$$\mathbf{t}_P(\omega) \doteq \sum_{x \sqsubseteq \omega} \frac{\mathbf{m}(x)}{P(x)}.$$

(*Если $P(x) = 0$, соответствующая дробь считается бесконечной.*)

Доказательство. Перечислимая снизу функция на бесконечных последовательностях определяется как предел возрастающей последовательности базисных функций. Можно представлять себе это возрастание так: в каждый момент каждое слово x имеет некоторый неотрицательный рациональный "вес" $w(x)$, а значение функции на последовательности равно сумме весов всех её начал. Постепенно веса (изначально равные нулю) увеличиваются; в каждый момент лишь конечное число весов не равны нулю.

В терминах весов условие ограниченности в среднем записывается как

$$\sum_x P(x)w(x) \leqslant 1,$$

поэтому при умножении весов на $P(x)$ оно в точности соответствует определению дискретной полумеры. Заметим, что вычислимость меры P гарантирует, что перечислимость снизу сохраняется в обе стороны (при умножении и делении на P).

Более формально, функция

$$\sum_{x \sqsubseteq \omega} \frac{\mathbf{m}(x)}{P(x)}$$

является перечислимым снизу ограниченным в среднем тестом; её интеграл в точности равен $\sum_x \mathbf{m}(x)$. С другой стороны, любой перечислимый снизу тест может быть представлен в терминах увеличения весов, и предельные значения этих весов, умноженные на P, образуют перечислимую снизу полумеру.

(Заметим, что второе преобразование не однозначно: веса можно перераспределять между двоичным словом и его продолжениями без изменения функции на бесконечных последовательностях.)

В этом рассуждении нам было важно, что P (во второй части рассуждения) и $1/P$ (в первой) перечислимы снизу.

Оказывается, что в этом предложении можно заменить сумму на точную верхнюю грань:

Теорема 2.22.

$$\mathbf{t}_P(\omega) \doteq \sup_{x \sqsubseteq \omega} \frac{\mathbf{m}(x)}{P(x)} \doteq \sum_{x \sqsubseteq \omega} \frac{\mathbf{m}(x)}{P(x)} \tag{1}$$

Первое из равенств можно переписать в логарифмической шкале:

$$\mathbf{d}_P(\omega) \stackrel{+}{=} \sup_{x \sqsubseteq \omega} [-\log P(x) - KP(x)]$$

Доказательство. Поскольку верхняя грань не превосходит суммы, требует доказательства только неравенство в одну сторону: надо объяснить, почему верхняя грань не сильно меньше суммы.

Для данного теста t рассмотрим функции t_i (для всех $i \in \mathbb{Z}$), определённые так: $t_i(\omega) = 2^i$, если $t(\omega) > 2^i$, и равно нулю в противном случае. Все они перечислимы снизу, и их сумма отличается от t не более чем вдвое (в ту или другую сторону). Кроме того, для любой точки ω величина $\sum_i t_i(\omega)$ превосходит $\sup_i t_i(\omega)$ не более чем вдвое.

Преобразуем каждое t_i в сумму весов, как описано выше. При этом, поскольку t_i имеет только два значения (нулевое и ненулевое), можно считать, что вершины ненулевого веса образуют беспрефиксное множество (на каждой ветви есть максимум одна такая вершина).

Складывая веса вдоль каждой ветви ω, мы получим $\sum_i t_i(\omega)$, то есть универсальный тест $t(\omega)$ (с точностью до константы). Если же вместо суммы весов брать максимум, то мы получим нечто меньшее, но уменьшение будет не более чем в два раза, поскольку мы складываем различные степени двойки. (При этом важно, что для каждого t_i в отдельности переход от суммы к максимуму ничего не меняет, поскольку вдоль каждой ветви только один член ненулевой.)

Это рассуждение остаётся в силе, если разрешить ненулевые веса не для всех вершин, а только для слов определённых длин. Пусть у нас имеется некоторая вычислимая возрастающая последовательность длин $n_1 < n_2 < n_3 < \ldots$.

Теорема 2.23.

$$\mathbf{d}_P(\omega) \stackrel{+}{=} \sup_k [-\log P(\omega(1:n_k)) - KP(\omega(1:n_k))].$$

(Здесь $\omega(1:n)$ означает начало последовательности ω длины n.)

Доказательство. Переходя от перечислимых функций на последовательностях к суммам весов, можно выбирать веса только разрешённых длин.

Эта теорема позволяет дать естественную характеризацию дефекта случайности двумерных массивов (которые с точки зрения топологии и меры ничем не отличаются от одномерных, и потому определение ограниченного в среднем дефекта случайности на них очевидно переносится). А именно, достаточно сравнивать вероятность и сложность, скажем, для квадратов с центром в начале координат. (В самом деле, можно расположить клетки плоскости в последовательность таким образом, чтобы эти квадраты соответствовали началам последовательности, и сослаться на предыдущую теорему.)

Историческое отступление

Формула для дефекта случайности является количественным уточнением следующего критерия:

Теорема 2.24 (Критерий случайности в терминах префиксной сложности)**.** *Последовательность ω случайна по вычислимой мере P тогда и только тогда, когда разность $-\log P(x) - KP(x)$ ограничена сверху для всех её начал.*

Этот критерий был впервые сформулирован в [4] со ссылкой на Шнорра; доказательство (для произвольной меры) было опубликовано впервые в [8]. Ещё до этого Шнорр и Левин (независимо в [22] и [15]) сформулировали близкий критерий случайности, использующий несколько другой вид сложности ("монотонную сложность"). Приведём её определение и соответствующую формулировку критерия. (В цитированной работе Шнорра используется несколько другой вид сложности, но позже Шнорр также использовал вариант сложности, введённый Левиным.)

Определение 2.25 (Монотонная сложность)**.** *Будем называть два слова* совместными*, если одно из них является началом другого. Рассмотрим перечислимое множество A пар слов $\langle x, y\rangle$, обладающее таким свойством: если $\langle p, q\rangle \in A$, $\langle p', q'\rangle \in A$ и p совместно с p', то q*

совместно с q'. Такое множество ("монотонный декомпрессор") задаёт отображение множества конечных и бесконечных последовательностей в себя, определяемое такой формулой (и обозначаемое той же буквой A):

$$A(p) = \sum\{x \mid (\exists p' \sqsubseteq p)\langle p', x\rangle \in A\}.$$

Здесь p — конечная или бесконечная последовательность, p' и x — двоичные слова, а sup *понимается как наименьшее общее продолжение, которое может быть конечным или бесконечным. (Условие на A гарантирует, что общее продолжение существует.)*

Далее мы определяем монотонную сложность $KM_A(x)$ слова x относительно A как минимальную длину слова p, для которого $A(p) \sqsupseteq x$. Среди всех монотонных декомпрессоров существует оптимальный, сложность относительно него минимальна (с точностью до константы). Фиксируем оптимальный декомпрессор V и положим $KM(x) = KM_V(x)$.

Замечание 2.26. *Частным случаем монотонных декомпрессоров являются отображения, задаваемые* машинами Тьюринга с оракулом. *Представим себе машину M со односторонней входной лентой (только для чтения), на которой написана конечная или бесконечная последовательность p. Машина также имеет рабочую ленту, а также одностороннюю выходную ленту (только для записи). В процессе работы на этой ленте появляется конечная или бесконечная последовательность $M(p)$ (работа может закончиться, если машина придёт в заключительное состояние или выйдет за границу входного слова, или продолжаться бесконечно, если не будет ни того, ни другого). Легко убедиться, что отображение $p \mapsto M(p)$ будет монотонным декомпрессором (однако не все монотонные декомпрессоры соответствуют таким машинам, так что получается несколько более узкий класс отображений — что, впрочем, само по себе не гарантирует, что получится существенно другая функция сложности).*

Монотонные декомпрессоры (или машины с оракулом описанного вида) могут быть использованы для определения другого вида априорной вероятности: *априорной вероятности на дереве* (которую можно также назвать *непрерывной априорной вероятностью*).

Определение 2.27. *Подадим на вход монотонного декомпрессора A последовательность независимых случайных битов и посмотрим на выходное распределение на конечных и бесконечных последовательностях. Обозначим через $M_A(x)$ вероятность того, что выходная последовательность будет иметь начало x.*

Легко видеть, что функция M_A принимает неотрицательные значения, перечислима снизу, $M_A(\Lambda) = 1$ (здесь Λ — пустое слово) и что $M_A(x) \geqslant M_A(x0) + M_A(x1)$.

Функции, обладающие указанными свойствами, называются перечислимыми снизу полумерами на дереве (*или* непрерывными полумерами).

Используя ту же конструкцию, что и для оптимального декомпрессора, можно доказать такое утверждение [29]:

Предложение 2.28. (а) *Всякая перечислимая снизу полумера на дереве является выходным распределением M_A для некоторого монотонного декомпрессора A.*

(б) *Среди всех перечислимых снизу полумер на дереве существует максимальная (с точностью до умножения на константу).*

Определение 2.29. *Фиксируем некоторую максимальную перечислимую снизу полумеру на дереве и назовём её* априорной вероятностью на дереве, *или* непрерывной априорной вероятностью. *Обозначение:* $\mathbf{a}(x)$.

Соотношение между априорной вероятностью на дереве и монотонной сложностью отчасти напоминает соотношение между дискретной априорной вероятностью и префиксной сложностью. Однако в этом случае $2^{-KM(x)}$, хотя и является перечислимой снизу полумерой на дереве, не является максимальной [9]. Другими словами, $KA(x) = -\log \mathbf{a}(x)$ не превосходит монотонной сложности, но может быть меньше её (и разница не ограничена).

Теперь можно сформулировать упомянутый критерий случайности; его доказательство, технически не сложное, можно найти в [18, 6, 25].

Предложение 2.30. *Для вычислимой меры P на Ω и последовательности $\omega \in \Omega$ следующие свойства равносильны:*

(i) *ω случайна по мере P;*
(ii) $\limsup_{x \sqsubset \omega}[-\log P(x) - KM(x)] < \infty$;
(iii) $\liminf_{x \sqsubset \omega}[-\log P(x) - KM(x)] < \infty$;
(iv) $\limsup_{x \sqsubset \omega}[-\log P(x) - KA(x)] < \infty$;
(v) $\liminf_{x \sqsubset \omega}[-\log P(x) - KA(x)] < \infty$;

Критерий случайности с префиксной сложностью имеет два отличия: в нём разность (ограниченная сверху для случайных последовательностей) не всегда ограничена снизу (в отличие от последнего критерия); кроме того, в нём $\limsup$ нельзя заменить на $\liminf$.

В последнем можно убедиться на таком примере. Заметим, что к всякому слову x можно дописать некоторые биты, получив слово y с $KP(y) \geqslant |y|$ (где $|y|$ — длина слова y). В самом деле, если бы это было не так, то для продолжений слова x мы имели бы $\mathbf{m}(y) \geqslant 2^{-|y|}$ и сумма $\sum_y \mathbf{m}(y)$ была бы бесконечной. Построим последовательность, по очереди дописывая длинные участки из нулей, чтобы сделать сложность существенно меньше длины, а потом биты, которые вновь доводят сложность до длины (как мы только что видели, это всегда возможно). Такая последовательность не будет случайной по равномерной мере (поскольку $\limsup$ разности бесконечен), но имеет бесконечно много начал, у которых сложность не меньше длины, так что $\liminf$ конечен.

Формула для (ограниченного в среднем) дефекта случайности имеет любопытное следствие. Рассмотрим равномерную меру на последовательностях (соответствующую независимым бросаниям честной монеты). Эта мера инвариантна относительно перестановок, и отсюда легко следует, что вычислимые перестановки членов последовательности сохраняют случайность. Более того, они сохраняют и дефект случайности (с точностью до константы). Отсюда получаем такое следствие:

Предложение 2.31. *Максимальная разность $|x| - KP(x)$ для начал случайной последовательности ω изменяется при вычислимой перестановке членов последовательности не более чем на константу (зависящую от перестановки, но не от последовательности).*

Некоторые более общие результаты такого типа можно найти в [15, 16, 10].

Другое следствие известно под названием “леммы Миллера – Ю” (Miller–Yu ample access lemma):

Следствие 2.32. *Последовательность ω случайна относительно вычислимой меры P тогда и только тогда, когда*

$$\sum_{x \sqsubset \omega} 2^{-\log P(x) - KP(x)} < \infty.$$

Отсюда, кстати, можно получить другое доказательство уже упомянутого факта:

Следствие 2.33. *Для всякого слова x найдётся его продолжение y, у которого $KP(y) > |y|$.*

Доказательство. В самом деле, x является началом некоторой случайной последовательности, и у неё по лемме Миллера–Ю есть сколь угодно длинные начала, сложность которых больше длины.

2.5 Игровая интерпретация

Формула для дефекта случайности может быть интерпретирована в игровых терминах. Рассмотрим игру Алисы и Боба с неполной информацией. Алиса выбирает бесконечную последовательность нулей и единиц. Боб выбирает (не видя последовательности Алисы) слово x. Они встречаются и одновременно открывают свои ходы. После этого, если x является началом ω, то Алиса платит Бобу $2^{|x|}$ рублей. (Эта версия игры соответствует равномерной мере, то есть независимым бросаниям честной монеты; в общем случае Алиса платит Бобу $1/P(x)$.)

Как обычно для игр с неполной информацией, будем рассматривать *чистые* стратегии (возможности игроков, согласно правилам игры), и *смешанные стратегии* (распределения вероятностей на чистых стратегиях). Легко видеть, что *цена* этой игры (в смысле смешанных стратегий, как это обычно понимается для игр с неполной информацией) равна 1. В самом деле, Боб может указать пустое слово и получить 1 в любом случае. С другой стороны, если Алиса честно получает свою последовательность бросанием монеты, то математическое ожидание её проигрыша равно 1, как бы ни пошёл Боб.

Оказывается, что Боб может построить вероятностную стратегию, которая принесёт ему успех, если Алиса поленится бросать монеты и принесёт неслучайную последовательность. Рассмотрим вероятностный алгоритм D, который даёт на выходе двоичные слова (а может и ничего не дать с положительной вероятностью). Такой алгоритм является смешанной стратегией для Боба (если на выходе не появляется никакого слова, то Боб пропускает игру и ничего не получает).

Теперь можно заметить следующее:

(i) Для любой вероятностной стратегии Боба математическое ожидание её выигрыша (как функция от последовательности Алисы) является ограниченным в среднем тестом. (Отсюда уже следует, что это математическое ожидание будет конечным, если последовательность Алисы случайна в смысле Мартин-Лёфа.)

(ii) Если $m(x)$ — вероятность получить x на выходе алгоритма D, то математическое ожидание выигрыша Боба на ω равно

$$\sum_{x \sqsubseteq \omega} \frac{m(x)}{P(x)}.$$

(iii) Поэтому, если взять алгоритм, порождающий на выходе дискретную априорную вероятность $\mathbf{m}(x)$, то математическое ожидание выигрыша Боба будет универсальным тестом (по доказанной формуле для универсального теста).

Таким образом, использование априорной вероятности как смешанной стратегии позволяет Бобу (в среднем) наказать Алису бесконечным штрафом за любую неслучайность в её последовательности.

Можно рассматривать немного более общую игру и разрешить Бобу указывать (в качестве чистых стратегий) не одно слово x, а некоторую базисную функцию f с неотрицательными значениями. При этом его выигрыш (для последовательности ω, принесённой Алисой), равен $f(\omega)/\int f(\omega)\, dP(\omega)$. (Знаменатель делает средний выигрыш равным единице.) Ходу x в старой игре при этом соответствует базисная функция, равная $2^{|x|}$ на продолжениях x и нулю в остальных местах.

Такое обобщение не даёт по существу ничего нового: мы и так разрешаем смешанные стратегии, а базисную функцию можно представить смесью нескольких ходов. (Получив функцию

f, Боб может сделать ещё один вероятностный шаг и выбрать один из интервалов, на которых f постоянна, с соответствующей вероятностью.) Таким образом мы приходим к другой формуле для универсального теста:

$$\mathbf{t}_P(\omega) \doteq \sum_f \frac{\mathbf{m}(f) f(\omega)}{\int f(\omega)\, dP(\omega)}.$$

Преимущество этой формулы в том, что она сохраняет смысл в более общих случаях, чем канторовское пространство, когда никаких выделенных интервалов нет и мы работаем прямо с каким-то классом базисных функций.

Отметим в заключение, что игровая интерпретация теории вероятностей, согласно которой случайность объекта есть не его свойство, а, грубо говоря, тип гарантии, с которой этот объект продаётся, развита в книге Шейфера и Вовка [24].

3 От тестов к сложностям

Формула (1) выражает дефект случайности бесконечной последовательности (значение универсального ограниченного в среднем теста) через сложность её конечных начал. Возникает естественный вопрос: можем ли мы действовать в обратном направлении и связать сложность конечного слова с дефектом случайности его бесконечных продолжений?

Мы уже переходили от бесконечных последовательностей к конечным в предложении 2.7. Это можно сделать и для универсального теста:

Определение 3.1. *Фиксируем вычислимую меру P. Пусть t — ограниченный в среднем тест на Ω. Определим для любого конечного слова z значение $\bar{t}(z)$ как минимум значений t на всех продолжениях:*

$$\bar{t}(z) = \inf_{\omega \sqsupseteq x} t(\omega).$$

Функция $\bar{t}$ однозначно определяется по t, и, напротив, позволяет восстановить t, поэтому её можно считать конечной версией теста t. Интуитивно говоря, слово выглядит неслучайным, если все его бесконечные продолжения имеют большой дефект случайности с точки зрения t.

Вопрос. Колмогоров [13] предлагал аналогичный подход к конечным словам: для каждого слова z можно рассмотреть минимальный дефект случайности (относительно равномерного распределения, понимаемый как разность между длиной и сложностью) его *конечных* продолжений. Есть ли тут какая-то связь с функцией $\bar{t}$?

Покажем, каким образом можно определить функцию $\bar{\mathbf{t}}_P$, соответствующую универсальному тесту, не обращаясь к бесконечным последовательностям.

Определение 3.2 (расширенный тест для вычислимой меры)**.** *Будем называть неотрицательную монотонную перечислимую снизу функцию $T\colon \{0,1\}^* \to [0,+\infty]$* расширенным тестом *для вычислимой меры P, если для любого N среднее значение T на словах длины N не превосходит* 1*:*

$$\sum_{\{x\colon |x|=N\}} P(x)T(x) \leqslant 1.$$

Монотонность означает, что $T(x) \leqslant T(y)$ при $x \sqsubseteq y$. Она гарантирует, что сумму по всем словам данной длины можно заменить на сумму по произвольному конечному (или даже бесконечному) беспрефиксному множеству S:

$$\sum_{x\in S} P(x)T(x) \leqslant 1.$$

(В самом деле, продолжим слова из S до какой-то большой общей длины.)

Предложение 3.3. *Всякий расширенный тест порождает (в смысле определения* 2.5*) некоторый ограниченный в среднем тест на бесконечных последовательностях. Обратно, всякие ограниченный в в среднем тест на бесконечных последовательностях порождается некоторым расширенным тестом.*

Доказательство. Первая часть непосредственно следует из определения (и теоремы о монотонной сходимости под знаком интеграла). В обратную сторону можно положить $T = \bar{t}$ или сослаться на предложение 2.4, если не хотеть использовать компактность.

Но можно рассматривать расширенные тесты и не упоминая бесконечные последовательности. Обычным образом можно доказать, что среди них существует максимальный:

Предложение 3.4. *Пусть P — вычислимая мера. Среди всех расширенных тестов для меры P существует максимальный (с точностью до умножения на константу).*

Определение 3.5. *Будем называть этот максимальный тест* универсальным расширенным тестом *для меры P.*

Предложение 3.6. *Универсальный расширенный тест для меры P совпадает с $\overline{\mathbf{t}}_P$ с точностью до ограниченного множителя.*

Доказательство. Поскольку $\overline{\mathbf{t}}_P$ является расширенным тестом, то он не превосходит универсального (с точностью до константы). С другой стороны, универсальный расширенный тест задаёт тест на бесконечных последовательностях, и остаётся сравнить его с максимальным.

Это построение по существу использует компактность пространства Ω (и потому, например, не проходит для последовательностей натуральных чисел), но и без этого можно построить максимальный расширенный тест, который будем обозначать $\mathbf{t}_P(x)$; использование одного и того же обозначения $\mathbf{t}_P$ не вызовет путаницы, так как в одном случае аргументом являются бесконечные последовательности, а в другом — конечные слова.

Определение расширенного теста позволяет изгнать бесконечные последовательности, сохранив по существу то же понятие универсального теста и даже немного обогатив его: отметим, что не всякий расширенный тест, порождающий универсальный ограниченный в среднем тест, является универсальным расширенным тестом (его значение на каком-то слове может быть малым, что не мешает универсальности на уровне бесконечных последовательностей, поскольку значения на всех продолжениях большие).

Описанный способ перехода от тестов на бесконечных последовательностях к тестам на словах не является единственно возможным.

Определение 3.7. *Предположим, что вычислимая мера P положительна на всех интервалах: $P(x) > 0$ для любого слова x. Обозначим через $\hat{\mathbf{t}}_P(x)$ условное математическое ожидание $\mathbf{t}_P(\omega)$ при условии, что ω начинается на x. Другими словами, $\hat{\mathbf{t}}_P(x)$ есть среднее значение t на интервале $x\Omega$, то есть отношение интеграла*

$$U(x) = \int\limits_{x\Omega} \mathbf{t}(\omega)\, dP(\omega)$$

к $P(x)$.

Функция U является перечислимой снизу полумерой. Более того, она обладает свойствами меры, за исключением того, что мера всего Ω не равна единице (отметим также, что $U(x)$ не обязано быть вычислимым). Эта мера имеет плотность $\hat{\mathbf{t}}$ относительно P. Отсюда следует, что функция $\hat{\mathbf{t}}_P$ является мартингалом в смысле следующего определения:

Определение 3.8. *Функция* $g\colon \{0,1\}^* \to \mathbb{R}$ *называется* мартингалом *относительно распределения вероятностей* P*, если*

$$P(x)g(x) = P(x0)g(x0) + P(x1)g(x1)$$

для любого слова x*. Если заменить знак "=" на "*$\geqslant$*", получим определение* супермартингала.

Будучи мартингалом, функция $\hat{\mathbf{t}}_P(x)$ не является монотонной по x.

Следующая теорема устанавливает соотношение между различными мерами неслучайности двоичных слов:

Теорема 3.9.

$$\frac{\mathbf{m}(x)}{P(x)} \stackrel{\cdot}{<} \mathbf{t}_P(x) \stackrel{\cdot}{<} \hat{\mathbf{t}}_P(x) \stackrel{\cdot}{<} \frac{\mathbf{a}(x)}{P(x)},$$

где $\mathbf{m}(x)$ *— дискретная априорная вероятность слова* x *(см. предложение* 2.20*), а* $\mathbf{a}(x)$ *— непрерывная априорная вероятность (на дереве, см. предложение* 2.28*) того же слова.*

Доказательство. Первое неравенство можно даже усилить, заменив $\mathbf{m}(x)/P(x)$ на сумму $\sum_{t\sqsubseteq x} \mathbf{m}(t)/P(t)$: эта сумма является частью выражения для $\mathbf{t}_P(\omega)$ для любого продолжения ω слова x.

Второе неравенство связывает среднее и наименьшее значения случайной величины.

Последнее неравенство следует из сравнения перечислимой снизу полумеры на дереве U с максимальной.

Отметим ещё, что $\hat{\mathbf{t}}_P(x)$ является мартингалом, а $\mathbf{a}(x)/P(x)$ — лишь супермартингалом (максимальным среди супермартингалов относительно P, с точностью до мультипликативной константы).

Замечания 3.10.

1. *Между первым и вторым членом неравенства последней теоремы можно поместить ещё два:*

$$\stackrel{\cdot}{<} \max_{t\sqsubseteq x} \frac{\mathbf{m}(t)}{P(t)} \stackrel{\cdot}{<} \sum_{t\sqsubseteq x} \frac{\mathbf{m}(t)}{P(t)} \stackrel{\cdot}{<}$$

2. *В логарифмической шкале имеем*

$$-\log P(x) - KP(x) \stackrel{+}{<} \log \mathbf{t}_P(x) \stackrel{+}{<} \log \hat{\mathbf{t}}_P(x) \stackrel{+}{<} -\log P(x) - KA\,(x).$$

3. *Мера* U *зависит от* P *(напомним, что* U *— это максимальная перечислимая снизу мера, имеющая плотность относительно* P*), и для различных мер* P *(например, с различными носителями) меры* U *могут быть разными. Но зависимость эта не так велика: теорема показывает, что возможные колебания ограничены разностью между* $KP(x)$ *и* $KA\,(x)$*.*

4. *Последнее неравенство в теореме (*$\hat{\mathbf{t}}_P(x) \stackrel{\cdot}{<} \mathbf{a}(x)/P(x)$*) нельзя заменить на равенство. Пусть, например, мера* P *равномерна, а в качестве* x *берутся начала возрастающей длины какой-то вычислимой последовательности. Тогда* $U(x)$ *стремится к нулю (область интегрирования сходится к одноэлементному множеству, имеющему меру нуль), а* $\mathbf{a}(x)$ *отделено от нуля.*

5. *Мы использовали компактность (конечность алфавита* $\{0,1\}$*), доказывая предложение* 2.7*. Вместо этого можно было бы использовать предложение* 2.6 *и получить аналогичные результаты для бэровского пространства последовательностей натуральных чисел.*

Все перечисленные в теореме 3.9 величины могут быть использованы для характеризации случайности: последовательность случайна тогда и только тогда, когда любая из этих величин ограничена на её начальных отрезках. В самом деле, теорема Левина – Шнорра гарантирует, что для случайной последовательности последнее отношение ограничено, а первое нет. Поскольку вторая величина монотонна, то для неслучайной последовательности все величины, начиная со второй, стремятся к бесконечности. Как мы уже упоминали, про первую величину этого утверждать нельзя.

Вопрос. Некоторые из величин, упомянутых в теореме 3.9 (вторая слева, а также две промежуточные между первой и второй), монотонны. Первая величина (см. обсуждение критерия случайности), а также величина $\hat{\mathbf{t}}_P(x)$ (мартингал), не монотонны. Что можно сказать про последнюю?

Отметим, что все эти величины если и не монотонны, то близки к монотонным.

4 Бернуллиевы последовательности

Можно стараться определить случайность не относительно конкретной меры, а относительно класса мер. (Интуитивно это означает, что мы готовы поверить, что последовательность получена в результате вероятностного процесса с распределением в этом классе.) Впоследствии мы сделаем это для произвольного *эффективно замкнутого* класса мер, но для наглядности начнём с конкретного примера: класса *бернуллиевых* мер.

4.1 Тесты для бернуллиевых последовательностей

Бернуллиева мера B_p соответствует последовательности независимых бросаний не обязательно симметричной монеты; вероятность появления единицы в каждом испытании равна некоторому $p \in [0,1]$ (одному и тому же во всех испытаниях). Отметим, что p не обязано быть вычислимым.

Определение 4.1 (ограниченный в среднем бернуллиев тест)**.** *Перечислимая снизу функция на бесконечных последовательностях называется ограниченным в среднем бернуллиевым тестом, если её интеграл по любой мере* B_p (*при любом* $p \in [0,1]$) *не превосходит* 1*.*

Предложение 4.2 (универсальный бернуллиев тест)**.** *Среди всех таких тестов существует максимальный (с точностью до мультипликативной константы).*

Доказательство. Перечислимая снизу функция есть предел возрастающей последовательности базисных функций. Для каждой из этих базисных функций её интеграл по мере B_p представляет собой многочлен от p, и легко проверить, что он не больше 1 при всех p (если это так). Соответственно можно фильтровать все негодные функции и перечислять все бернуллиевы тесты. Складывая их с коэффициентами, получаем универсальный.

Определение 4.3. *Фиксируем универсальный бернуллиев тест и обозначим его* $\mathbf{t}_{\mathcal{B}}(\omega)$*. Его логарифм будем называть* дефектом бернуллиевости *и обозначать* $\mathbf{d}_{\mathcal{B}}(\omega)$*. Последовательность называется* бернуллиевой*, если её дефект конечен.*

Как и раньше, можно немного модифицировать определение, чтобы считать дефект неотрицательным целым числом.

Как и для вычислимых мер, можно перейти к конечным последовательностям:

Определение 4.4. *Будем называть монотонную перечислимую снизу неотрицательную функцию* $T\colon \{0,1\} \to [0,+\infty]$ расширенным бернуллиевым тестом*, если для любого натурального* N *и для любого* $p \in [0,1]$ *выполняется неравенство* $\sum_{\{x\colon |x|=N\}} B_p(x)T(x) \leqslant 1$.

Как и для вычислимых мер, тесты на конечных и бесконечных последовательностях связаны:

Предложение 4.5. *Всякий расширенный бернуллиев тест порождает бернуллиев тест на* Ω*. Напротив, всякий бернуллиев тест на* Ω *порождается некоторым расширенным бернуллиевым тестом.*

Среди расширенных бернуллиевых тестов существует максимальный; он порождает универсальный бернуллиев тест на Ω. Как и раньше, мы будем использовать одно и то же обозначение $\mathbf{t}_{\mathcal{B}}$ для максимальных тестов на конечных и бесконечных последовательностях.

4.2 Другие варианты определения бернуллиевости

Как и для случайности относительно вычислимых мер, есть разные эквивалентные варианты определения. Можно рассматривать ограниченные по вероятности тесты (вероятность события $t(\omega) > N$ по любой из мер B_p должна быть не больше $1/N$). Можно, следуя определению Мартин-Лёфа для вычислимых мер, назвать тестом вычислимую последовательность эффективно открытых множеств U_i, для которых $B_p(U_i) \leqslant 2^{-i}$ при любом i и при любом $p \in [0,1]$. Все эти варианты определения эквивалентны (и доказывается это точно так же, как для случайности по вычислимой мере).

Интересно, что первоначальное определение бернуллиевости, данное Мартин-Лёфом в [19], было немного другим. Сейчас мы покажем, что оно также эквивалентно остальным, но это несколько сложнее.

Обозначение 4.6. *Через* $\mathbb{B}(n,k)$ *мы обозначаем множество всех слов длины* n*, содержащих ровно* k *единиц.*

Мартин-Лёф определяет тест бернуллиевости как семейство перечислимых множеств слов $U_1 \supset U_2 \supset U_3 \supset \ldots$; каждое из множеств наследственно вверх, то есть вместе с любым словом содержит все его продолжения. Ограничение на эти множества такое: рассмотрим произвольные целые $n \geqslant 0$ и k от 0 до n; через $\mathbb{B}(n,k)$ обозначим множество всех слов длины n, содержащих k единиц (и $n-k$ нулей); требуется, чтобы при всех i доля слов в $\mathbb{B}(n,k)$, принадлежащих U_i, была бы не больше 2^{-i}.

Для удобства сравнения заменим множества U_i на перечислимую снизу функцию d с целыми значениями, для которой $U_i = \{x \mid d(i) \geqslant i\}$. Наследственность множеств означает монотонность этой функции; помимо этого, требуется, чтобы вероятность события $d \geqslant i$ внутри любого множества $\mathbb{B}(n,k)$ была бы не больше 2^{-i}. Видно, что эти требования соответствуют ограниченным по вероятности расширенным тестам (в логарифмической шкале), но только вместо класса мер B_p на словах длины n рассматривается другой класс мер, а именно класс мер, сосредоточенных на словах данной длины с данным числом единиц. Меры из класса B_p принимают равные значения на словах одинаковой длины с одинаковым числом единиц, поэтому представимы в виде смеси равномерных мер на $\mathbb{B}(n,k)$ с некоторыми коэффициентами, от замены B_p на эти меры условие становится более сильным.

Покажем, что тем не менее класс бернуллиевых последовательностей не меняется от такой замены и, более того, универсальный тест (как функция на бесконечных последовательностях) тоже не меняется (с точностью до ограниченного множителя, как обычно). Мы покажем это для ограниченных в среднем вариантов тестов (соответственно изменив определение Мартин-Лёфа); на класс бернуллиевых последовательностей это не влияет. Рассуждение для ограниченных по вероятности тестов аналогично.

Дадим соответствующие определения.

Определение 4.7. *Функцию* $f\colon \{0,1\}^* \to [0,+\infty]$ *назовём* комбинаторным бернуллиевым тестом*, если*

(а) *она перечислима снизу;*

(б) *она монотонна (увеличивается при добавлении битов в конец слова);*

(в) *для любых целых* n, k *с* $0 \leqslant k \leqslant n$ *среднее значение функции* f *на множестве* $\mathbb{B}(n,k)$ *не превосходит* 1*.*

Можно сравнить эти требования со случаем равномерной меры: тогда мы требовали, чтобы среднее по всему $\{0,1\}^n$ не превосходило единицы; теперь требование сильное: среднее по по каждой из его частей $\mathbb{B}(n,k)$ должно быть не больше 1.

Имея такой тест для слов ограниченной длины, можно продолжать его по монотонности:

Предложение 4.8. *Пусть имеется функция* f*, определённая на словах длины меньше* n *и удовлетворяющая требованиям* (а)–(в)*. Тогда её продолжение по монотонности на слова больших длин также удовлетворяет этим требованиям.*

Доказательство. Будем продолжать её на слова длины n, положив $f(x0)$ и $f(x1)$ равным $f(x)$ для слов x длины $n-1$. Множество $\mathbb{B}(n,k)$ состоит из двух частей: слов, оканчивающихся на нуль, и слов, оканчивающихся на единицу. Первые находятся во взаимно однозначном соответствии с $\mathbb{B}(n-1,k)$, вторые — с $\mathbb{B}(n-1,k-1)$. Функция сохраняет значения при этом соответствии, поэтому среднее по каждой из частей не больше 1. Следовательно, и среднее по всему $\mathbb{B}(n,k)$ не больше 1.

Как обычно, можно определить универсальный комбинаторный бернуллиев тест:

Предложение 4.9 (универсальный комбинаторный бернуллиев тест)**.** *Среди комбинаторных бернуллиевых тестов существует максимальный с точностью до мультипликативной константы.*

Определение 4.10. *Фиксируем универсальный комбинаторный тест* $\mathbf{b}(x)$ *и продолжим его на бесконечные последовательности, положив*

$$\mathbf{b}(\omega) = \sup_{x \sqsubseteq \omega} \mathbf{b}(x).$$

Полученную функцию будем называть универсальным комбинаторным тестом на Ω *и обозначать той же буквой* $\mathbf{b}$*.*

(В силу монотонности точную верхнюю грань в этом определении можно заменить на предел.)

Покажем, что этот тест совпадает (с точностью до ограниченного множителя) с введёнными ранее бернуллиевыми тестами в смысле определения 4.1.

Теорема 4.11.

$$\mathbf{b}(\omega) \doteq \mathbf{t}_{\mathcal{B}}(\omega).$$

Доказательство. Мы уже видели, что комбинаторный бернуллиев тест является расширенным бернуллиевым тестом (из ограничений на среднее по каждой части $\mathbb{B}(n,k)$ следует ограничение на математическое ожидание по мере B_p, так как эта мера постоянна на каждой части). Следовательно, $\mathbf{b}(\omega) \dot{<} \mathbf{t}_{\mathcal{B}}(\omega)$.

Обратное утверждение неверно: расширенный бернуллиев тест может не быть комбинаторным тестом. Однако можно построить комбинаторный тест, который принимает те же значения (с точностью до константы) на бесконечных последовательностях, а только это и утверждается в теореме.

Идея тут состоит в следующем. Рассмотрим расширенный бернуллиев тест t на словах длины n и перенесём его на слова существенно большей длины N (применяя старый тест к их началам длины n). Получим некоторую функцию t'. Нам нужно показать, что t' близка к комбинаторному тесту (превышает его не более чем в константу раз). Для этого надо усреднить t' по множеству $\mathbb{B}(N,K)$ для произвольного K между 0 и N. Другими словами, нам нужно усреднить t по распределению вероятностей на n-битовых началах последовательностей длины N, содержащих K единиц. При $N \gg n$ это распределение будет близко к бернуллиевому с вероятностью $p = K/N$.

В терминах теории вероятностей мы имеем урну с N шарами, из которых K чёрных, и вынимаем из неё (без возвращения) n шаров. Нам надо сравнить распределение вероятностей с бернуллиевым, которое получилось бы при выборке с возвращением. Покажем, что

> *при $N = n^2$ распределение без возвращения не более чем в $O(1)$ раз превосходит распределение с возвращением.*

(Кстати, обратное неравенство не верно: при $K = 1$ без возвращения бы не можем получить слово с двумя единицами, а с возвращением можем. Но нам достаточно неравенства в эту сторону.)

В самом деле, при выборке без возвращения вероятность вытащить шар данного цвета равна отношению

$$\frac{\text{число оставшихся шаров этого цвета}}{\text{число всех оставшихся шаров}}.$$

Оставшихся шаров этого цвета не больше, чем в случае с возвращением, а знаменатель не меньше $N - n$. Поэтому вероятность любой комбинации при выборке без возвращения не больше вероятности же комбинации с возвращением, умноженной на $N/(N-n)$ в степени n. При $N = n^2$ возникает множитель $(1 + O(1/n))^n = O(1)$.

Таким образом, если взять расширенный бернуллиев t и затем определить $t'(x)$ на слове x длины N как t на начале слова x длины $\lfloor\sqrt{N}\rfloor$, то полученная функция t' будет комбинаторным тестом с точностью до константы. (Отметим, что её монотонность следует из монотонности t.)

4.3 Критерий бернуллиевости

Естественно сравнивать понятие бернуллиевой последовательности (для которой тест бернуллиевости конечен) с понятием случайной по мере B_p последовательности. Однако определение случайности по Мартин-Лёфу предполагало вычислимость меры, и непосредственно не применимо к мере B_p при невычислимом p.

Можно, однако, релятивизировать определения Мартин-Лёфа, разрешив обращаться к оракулу для p. С таким оракулом мера B_p становится вычислимой и определение случайности по Мартин-Лёфу приобретает смысл.

Следующая теорема подтверждает интуитивный смысл бернуллиевых последовательностей как последовательностей, случайных по мере B_p при некотором p:

Теорема 4.12. *Последовательность ω является бернуллиевой тогда и только когда, когда она случайна по мере B_p с оракулом p для некоторого $p \in [0,1]$.*

Говоря об оракуле p, мы имеем в виду возможность получать по i значение i-го бита в двоичном разложении p (которое единственно, за исключением тех случаев, когда p двоично-рационально, а в этих случаях оба разложения вычислимы и оракул тривиален).

Мы будем доказывать эту теорему (и притом в более сильной количественной форме), введя понятие теста случайности по мерам B_p как функции двух аргументов (последовательности и p). Требуемый результат получится как комбинация следующих утверждений:

(а) Среди таких "равномерных" тестов случайности существует максимальный тест $\mathbf{t}(\omega, p)$.

(б) Функция $\omega \mapsto \inf_p \mathbf{t}(\omega, p)$ совпадает (как обычно, с точностью до ограниченного множителя) с универсальным бернуллиевым тестом.

(в) При фиксированном p функция $\omega \mapsto \mathbf{t}(\omega, p)$ совпадает (с той же точностью) с релятивизированным относительно p максимальным тестом случайности относительно (p-)вычислимой меры B_p.

Из этих трёх утверждений легко следует теорема 4.12: последовательность ω бернуллиева, если тест бернуллиевости на ω конечен; он равен точной нижней грани $\mathbf{t}(\omega, p)$, поэтому его конечность означает, что $\mathbf{t}(\omega, p) < \infty$ при некотором p, что равносильно p-релятивизованной случайности по мере B_p.

Нам понадобится некоторая техническая подготовка. Тесты случайности (как функции двух аргументов) тоже будут перечислимыми снизу, но это понятие требует уточнения, поскольку добавился второй аргумент, действительное число. (Впоследствии мы рассмотрим и более общую ситуацию, когда вторым аргументом является мера.) Дадим соответствующие определения.

Определение 4.13. *Назовём* базисными прямоугольниками *в пространстве $\Omega \times [0,1]$ множества вида $x\Omega \times (u, v)$, где x — двоичное слово, а u, v — рациональные числа, причём $u < v$. (Техническая оговорка: числа u, v могут лежать и вне $[0,1]$, в этом случае по второй координате берётся пересечение (u, v) с $[0,1]$.)*

Функция $f\colon \Omega \times [0,1] \to [-\infty, +\infty]$ называется перечислимой снизу, *если существует алгоритм, который получает на вход рациональное r и порождает прямоугольники, в объединении дающие всё множество пар $\langle \omega, p \rangle$ с $r < f(\omega, r)$.*

Это определение, как и раньше, требует, чтобы прообраз $(-\infty, r)$ был эффективно открытым множеством равномерно по r, но только теперь мы рассматриваем эффективно открытые множества в $\Omega \times [0,1]$, определённые естественным образом.

Аналогично определяется и перечислимость сверху, *равносильная перечислимости снизу функции $(-f)$.*

Функцию с конечными действительными значения называют вычислимой, *если она перечислима и снизу, и сверху.*

Поскольку пересечение эффективно открытых множеств эффективно открыто, получаем такую формулировку:

Предложение 4.14. *Функция $f\colon \Omega \times [0,1] \to \mathbb{R}$ вычислима тогда и только тогда, когда для каждого интервала (u, v) с рациональными концами его прообраз есть объединение последовательности базисных прямоугольников, эффективно порождаемой по u и v.*

Интуитивный смысл этого определения можно понять, если иметь в виду, что задача "указывать приближения к α с любой заданной точностью" равносильна задаче "перечислять все интервалы, содержащие α". Поэтому для вычислимой функции f мы можем находить приближения к $f(\omega, p)$, если нам дают приближения к ω и p.

Определение (неотрицательной) перечислимой снизу функции можно переформулировать, введя понятие базисной функции. Нам будет важно, что базисные функции непрерывны, поэтому зависимость от действительного аргумента будет кусочно-линейной, а не скачками.

Определение 4.15 (базисные функции, бернуллиев случай)**.** *Пусть x — двоичное слово, (u, v) — рациональный интервал, а k — натуральное число, для которого $u + 2^{-k} < v - 2^{-k}$. Определим функцию $g_{x,u,v,k}(\omega, p)$ так: если ω не начинается на x, то она равна нулю; если ω начинается на x, то зависимость от p будет кусочно-линейной, причём при $p \notin (u, v)$ функция равна нулю, внутри $(u + 2^{-k}, v - 2^{-k})$ функция равна 1, а в промежутке линейно меняется.*

Теперь рассмотрим наименьший класс функций, содержащий все функции $g_{x,u,v,k}$ и замкнутый относительно линейных комбинаций с рациональными коэффициентами, максимумов и минимумов. Это счётное множество функций, которые можно задавать конструктивно, и эти функции будем называть базисными.

Теперь можно дать эквивалентное описание перечислимости снизу:

Предложение 4.16. *Функция $f\colon \Omega \times [0,1] \to [0, +\infty]$ перечислима снизу тогда и только тогда, когда она представима в виде поточечного предела неубывающей последовательности базисных функций.*

Доказательство. Это было бы совсем ясно, если считать базисными функциями характеристические функции базисных прямоугольников и максимумы конечного числа таких функций. Но мы хотим, чтобы базисные функции были непрерывны по p (это будет важно в дальнейшем). Поэтому надо заметить, что при $k \to \infty$ функция $g_{x,u,v,k}$ стремится к характеристической функции прямоугольника.

Непрерывность базисных функций гарантирует такое важное свойство:

Предложение 4.17. *Пусть $f\colon \Omega \times [0,1] \to \mathbb{R}$ — базисная функция. Тогда значение интеграла $\int f(\omega, p)\, dB_p(\omega)$ является вычислимой функцией от p (и от базисной функции f).*

(Вычислимость понимается в описанном выше смысле; отметим, что всякая вычислимая функция непрерывна. Аналогичное утверждение верно для любой вычислимой функции f, не только базисной, но нам это не понадобится.)

Теперь мы готовы сформулировать и доказать важный технический факт (доказанный в [12]); он не раз нам понадобится (в том числе и в более общей ситуации).

Предложение 4.18 (усечение)**.** *Пусть $\varphi\colon \Omega \times [0,1] \to [0, \infty]$ — перечислимая снизу функция. Тогда можно построить другую перечислимую снизу функцию $\varphi'(\omega, p)$, не превосходящую $\varphi(\omega, p)$ в каждой точке, для которой при любом p:*

(а) $\int \varphi'(\omega, p)\, dB_p(\omega) \leqslant 2$;

(б) *если* $\int \varphi(\omega, p)\, dB_p(\omega) \leqslant 1$, *то* $\varphi'(\omega, p) = \varphi(\omega, p)$ *при всех* ω.

Доказательство. Согласно Предложению 4.16, можно представить φ в виде суммы ряда неотрицательных базисных функций: $\varphi(\omega, p) = \sum_n h_n(\omega, p)$. Предложение 4.17 гарантирует, что интеграл

$$\int \sum_{i \leqslant n} h_i(\omega, p)\, dB_p(\omega)$$

является вычислимой функцией от p (равномерно по n), и поэтому множество S_n тех p, где этот интеграл меньше 2, эффективно открыто (равномерно по n).

Теперь положим $h'_n(\omega,p) = h_n(\omega,p)$, если $p \in S_n$, и $h'_n(\omega,p) = 0$ в противном случае. Функция h'_n будет перечислимой снизу, и интеграл $\int \sum_{i\leqslant n} h'_i(\omega,p)\,dB_p(\omega)$ будет меньше 2 при всех p. Положив $\varphi' = \sum h'_n$, мы получим перечислимую снизу функцию, и по теореме о монотонной сходимости $\int \varphi'(\omega,p)\,dB_p(\omega)$ не больше 2 при всех p.

Остаётся заметить, что если при некотором p интеграл $\int \varphi(\omega,p)\,dB_p(\omega)$ не превосходит 1, то это p войдёт во все S_n и переход от h_n к h'_n, как и переход от φ к φ', ничего не изменит.

После этой подготовки мы можем определить равномерные тесты бернуллиевости и доказать их свойства:

Определение 4.19. *Функцию t от двух аргументов $\omega \in \Omega$ и $p \in [0,1]$ назовём* равномерным тестом бернуллиевости*, если*

(а) *она перечислима снизу (в описанном выше смысле, как функция пары)*;

(б) *для любого $p \in [0,1]$ математическое ожидание $t(\omega,p)$ по мере B_p (то есть интеграл $\int t(\omega,p)\,dB_p(\omega)$) не превосходит* 1.

Нам осталось доказать три обещанных утверждения:

Лемма 4.20. *Существует* универсальный *равномерный тест бернуллиевости $\mathbf{t}(\omega,p)$, который является максимальным в этом классе (с точностью до константы).*

Лемма 4.21. *Для этого теста функция $\mathbf{t}'(\omega) = \inf_p \mathbf{t}(\omega,p)$ совпадает (с точностью до ограниченного в обе стороны множителя) с универсальным тестом бернуллиевости $\mathbf{t}_{\mathcal{B}}(\omega)$ в смысле определения* 4.3.

Из этих двух лемм вытекает, что последовательность ω является бернуллиевой тогда и только тогда, когда $\mathbf{t}'(\omega) < \infty$, то есть $\mathbf{t}(\omega,p) < \infty$ при некотором p, и это позволяет завершить доказательство теоремы 4.12 ссылкой на такое утверждение:

Лемма 4.22. *Для фиксированного p функция $\mathbf{t}_p(\omega) = \mathbf{t}(\omega,p)$ совпадает (с точностью до ограниченного множителя) с релятивизованным относительно p универсальным тестом случайности относительно меры B_p*

Доказательство. (лемма 4.20) Будем перечислять все перечислимые снизу функции двух аргументов. К каждой из них применим предложение 4.18, и полученные суммы сложим с коэффициентами, образующими вычислимо сходящийся ряд с суммой меньше $1/2$.

Доказательство. (лемма 4.21) Покажем, что функция $\mathbf{t}'$ является универсальным бернуллиевым тестом. При любом p математическое ожидание этой функции по мере B_p не больше 1 (поскольку она не превосходит $\mathbf{t}(\omega,p)$ для этого конкретного p).

Кроме того, эта функция перечислима снизу. Это доказывается аналогично предложению 2.7 с использованием компактности. (Аналогичное утверждение в более общей ситуации будет доказано в предложении 7.20.)

Таким образом, $\mathbf{t}'$ является бернуллиевым тестом. Универсальность (максимальность) очевидно следует из того, что любой бернуллиев тест можно рассматривать как функцию двух переменных, которая будет равномерным бернуллиевым тестом.

Аналогичное рассуждение показывает, что для естественным образом определённого универсального расширенного равномерного бернуллиева теста $\mathbf{t}(x,p)$ (первым аргументом которого является двоичное слово) величина $\inf_p \mathbf{t}(x,p)$ будет универсальным расширенным бернуллиевым тестом.

Доказательство. (лемма 4.22) Предположим вначале, что p вычислимо. Тогда мы можем перечислять все интервалы, содержащие p, и функция $\mathbf{t}_p\colon \omega \to \mathbf{t}(\omega, p)$ перечислима снизу (чтобы перечислять те x, где $\mathbf{t}_p(\omega) < r$, мы перечисляем прямоугольники, в которых $\mathbf{t}(\omega, p)$, и отбираем из них те, где вторая проекция содержит p).

Аналогичное рассуждение можно провести для любого p и установить, что функция $\mathbf{t}_p$ перечислима снизу с p-оракулом. Таким образом, $\mathbf{t}_p$ не превосходит универсального релятивизованного теста для B_p.

Обратное рассуждение чуть сложнее. Пусть имеется некоторый тест t для B_p, перечислимый снизу с оракулом p. Мы должны найти равномерный тест $t'(\omega, p)$, который мажорирует t (при данном p). Другими словами, нужно продолжить функцию, первоначально определённую только для одного p, на все значения p, и при этом ещё и гарантировать оценку для интеграла.

Начнём с простого случая, когда p вычислимо. В этом случае оракул не нужен и функция t перечислима снизу. Добавив в неё фиктивный второй аргумент p, мы получим перечислимую снизу функцию двух аргументов. Но тестом эта функция, скорее всего, не будет, так как про её математическое ожидание по мере B_q при $q \neq p$ мы ничего не знаем. Тут нам помогает предложение 4.18: с его помощью мы преобразуем t в перечислимую снизу функцию $t'(\omega, q)$ (которая теперь уже реально зависит от q), для которой $\int t'(\omega, q)\, dB_q(\omega) \leqslant 2$ при всех q, а $t'(\cdot, p) = t(\cdot, p)$. Поделив t' пополам, получим равномерный тест.

Теперь рассмотрим случай невычислимого p. В этом случае p иррационально, поэтому биты его двоичного разложения можно получать, имея перечисление всех содержащих его интервалов (дождавшись, пока появится интервал, однозначно определяющий нужный нам бит). Поэтому машина, использующая оракул p, может быть преобразована в машину, которая перечисляет снизу некоторую функцию $\tilde{t}(\omega, q)$, совпадающую с $t(\omega)$ при $q = p$. Эта функция вовсе не обязана быть равномерным тестом бернуллиевости (поскольку условие на интеграл гарантировано только при $q = p$, но её опять же можно подвергнуть усечению с помощью предложения 4.18.

5 Произвольные меры на Ω

В этом разделе мы по-прежнему ограничиваемся двоичными последовательностями, но меры на Ω могут быть любыми, а не только бернуллиевыми.

Обозначение 5.1. *Будем обозначать множество всех вероятностных распределений на Ω через $\mathcal{M}(\Omega)$.*

(Напомним, что мера всего пространства Ω всегда равна 1.)

5.1 Равномерные тесты случайности

Определение 5.2. *Назовём* равномерным тестом *перечислимую снизу функцию $t(\omega, P)$ двух аргументов (последовательности ω и меры P на Ω), для которой*

$$\int t(\omega, P)\, dP(\omega) \leqslant 1$$

для любой меры P.

Здесь требует уточнения понятие перечислимой снизу функции. Пространство всех мер $\mathcal{M}(\Omega)$ можно рассматривать как замкнутое подмножество бесконечного произведения

$$\Xi = [0,1] \times [0,1] \times [0,1] \times \ldots$$

(мера задаётся своими значениями на интервалах, которых счётное число; эти значения должны удовлетворять соотношениям, выделяющим замкнутое множество). Теперь определим базисные множества, эффективно открытые множества и пр. для пространства мер:

Определение 5.3. *Базисное множество* (*открытый* интервал) *в пространстве мер задаётся конечным множеством условий вида $u < P(y) < v$, где y — двоичное слово, а u, v — рациональные числа. Оно состоит из всех мер P, удовлетворяющих этим условиям.* Базисное открытое множество *в пространстве $\Omega \times \mathcal{M}(\Omega)$ имеет вид $x\Omega \times \beta$ (произведение интервалов в Ω и в $\mathcal{M}(\Omega)$).*

Перечислимость снизу, сверху и вычислимость теперь определяются стандартным образом в терминах базисных открытых множеств, см. определение 4.13.

Мы будем использовать компактность пространств Ω (мы рассматриваем последовательности над конечным алфавитом $\{0,1\}$) и $\mathcal{M}(\Omega)$. Это свойство позволяет выбирать из любого открытого покрытия конечное подпокрытие. Нам понадобится эффективный вариант этого свойства:

Определение 5.4 (эффективная компактность)**.** *Компактное подмножество C пространства $\mathcal{M}$ называется* эффективно компактным*, если множество*

$$\{S \mid S \text{ — конечное семейство базисных множеств, покрывающее } C\}$$

перечислимо.

Само пространство $\mathcal{M}(\Omega)$, как легко видеть, компактно и эффективно компактно. Компактно оно как замкнутое множество в произведении компактных пространств, а эффективность следует из того, что мы можем проверить, что данные базисные множества покрывают всё пространство (речь идёт о линейных равенствах и неравенствах с конечным числом переменных, а там всё алгоритмически разрешимо). Отсюда легко следует такое утверждение:

Предложение 5.5. *Всякое эффективно замкнутое подмножество $\mathcal{M}(\Omega)$ эффективно компактно.*

Доказательство. Пусть эффективно замкнутое множество C имеет дополнение, являющееся объединением базисных открытых множеств $U_1, U_2, \ldots$; семейство S является покрытием C тогда и только тогда, когда вместе с некоторым конечным набором из U_i оно покрывает всё пространство. А это свойство перечислимо.

Верно и обратное — что эффективно компактное подмножество эффективно замкнуто. (Его дополнение есть объединение всех интервалов, которые не пересекаются с некоторым конечным покрытием множества, а такие ситуации можно перечислять.)

Теперь мы можем продолжить построение по аналогии с бернуллиевыми мерами. Определим понятие базисной функции (по аналогии с определением 4.15; конкретный вид базисных функций не имеет большого значения):

Определение 5.6 (базисные функции для равномерных тестов по произвольным мерам в Ω)**.** Базисные функции *на множестве $\Omega \times \mathcal{M}(\Omega)$ определяются аналогично определению* 4.15*, начиная с функций*

$$g_{x,y,u,v,k} \colon \Omega \times \mathcal{M}(\Omega) \to [0,1].$$

Здесь x, y — двоичные слова, u, v — рациональные числа, а k — натуральное число; значение $g_{x,y,u,v,k}(\omega, P)$ равно нулю, если ω не начинается на x; при $x \sqsubseteq \omega$ это значение кусочно-линейно зависит от $P(y)$, и равно 0 при $x \notin (u, v)$ и 1 при $x \in (u + 2^{-k}, v - 2^{-k})$.

Как и в предложении 4.16, всякая перечислимая снизу неотрицательная функция является пределом возрастающей последовательности базисных функций (заметим, что в $g_{x,y,u,v,k}(\omega, P)$ входит значение P только на слове y, но мы затем берём минимумы и максимумы).

Далее, как в предложении 4.17, можно заметить, что для базисной функции f интеграл $\int f(\omega, P)\, dP(\omega)$ является вычислимой функцией меры P (и базисной функции f).

Наконец, остаётся верным (с тем же доказательством) и аналог предложения 4.18:

Теорема 5.7 (усечение)**.** *Пусть $\varphi(\omega, P)$ — перечислимая снизу функция. Тогда существует перечислимая снизу функция $\varphi'(\omega, P)$, для которой для любого P*

(а) $\int \varphi'(\omega, P)\, dP(\omega) \leqslant 2$*;*

(б) *если* $\int \varphi(\omega, P)\, dP(\omega) \leqslant 1$*, то* $\varphi'(\omega, P) = \varphi(\omega, P)$ *при всех* ω*.*

Это позволяет построить универсальный тест как функцию последовательности и произвольной меры на Ω:

Теорема 5.8. *Существует универсальный (максимальный с точностью до постоянного множителя) равномерный тест.*

Доказательство. Как и раньше, будем перечислять все перечислимые снизу функции, подвергать каждую из них усечению (которое её не портит, если функция и так была тестом), и затем сложим получившиеся тесты (или почти-тесты) с подходящими коэффициентами.

Определение 5.9. *Фиксируем один из универсальных равномерных тестов и будем обозначать его $\mathbf{t}(\omega, P)$. Будем говорить, что последовательность является* равномерно случайной *по мере P (не обязательно вычислимой), если $\mathbf{t}(\omega, P) < \infty$.*

Покажем, что для вычислимых мер это определение согласуется с прежним (определение 2.13):

Предложение 5.10. *Пусть P — вычислимая мера, а $\mathbf{t}_P(\omega)$ — универсальный ограниченный в среднем тест для этой меры в смысле определения* 2.13*. Тогда $c_1\, \mathbf{t}_P(\omega) \leqslant \mathbf{t}(\omega, P) \leqslant c_2\, \mathbf{t}_P(\omega)$ для некоторых $c_1, c_2 > 0$ и для всех ω.*

Здесь константы c_1 и c_2 зависят от выбора меры P и от выбора теста $\mathbf{t}_P$ для этой меры (этот выбор был произвольно сделан для каждой вычислимой меры).

Это предложение показывает, в частности, что для вычислимых мер равномерная случайность совпадает со случайностью в смысле Мартин-Лёфа.

Доказательство. Для начала покажем, что $\mathbf{t}(\omega, P) \leqslant c_2\, \mathbf{t}_P(\omega)$. Заметим, что функция $\omega \mapsto \mathbf{t}(\omega, P)$ является перечислимой снизу: поскольку мера P вычислима, мы можем эффективно перечислять все интервалы в пространстве мер, её содержащие. Поэтому она мажорируется максимальным перечислимым снизу P-тестом $\mathbf{t}_P$.

Чтобы доказать обратное неравенство, добавим фиктивный аргумент и рассмотрим функцию

$$t(\omega, Q) = \mathbf{t}_P(\omega).$$

Эта функция, естественно, не является равномерным тестом, так как её интеграл по мере Q (при $Q \neq P$) может быть любым, но после усечения она становится (почти) тестом, не меняясь на P.

Можно пытаться определить тесты относительно некоторой одной (не обязательно вычислимой) меры. Будем говорить, что функция $f : \Omega \to [0, +\infty]$ *перечислима снизу относительно меры* P, если она получается из перечислимой снизу функции на $\Omega \times \mathcal{M}(\Omega)$ фиксацией второго аргумента равным P. Теперь можно дать такое определение:

Определение 5.11. *Пусть P — некоторая мера на Ω. Будем называть* тестом случайности относительно P *перечислимую снизу относительно P функцию f, для которой $\int f(\omega)\, dP(\omega)$ не превосходит единицы.*

Теорема 5.7 показывает, однако, что нового понятия случайности при этом не получается — мы приходим к тем же самым равномерным тестам:

Теорема 5.12. *Пусть P — некоторая мера, а $t_P(\omega)$ — тест относительно этой меры. Тогда существует равномерный тест $t'(\cdot, \cdot)$, для которого $t_P(\omega) \leqslant 2t(\omega, P)$. Напротив, сужение любого равномерного теста на меру P является P-тестом.*

Для равномерных тестов также ввести понятие расширенного теста:

Определение 5.13 (расширенные равномерные тесты)**.** *Функция $T\colon \{0,1\}^* \times \mathcal{M}(\Omega) \to [0,1]$, перечислимая снизу и монотонная по первому аргументу, называется* расширенным равномерным тестом*, если*

$$\sum_{|x|=n} T(x, P)P(x) \leqslant 1$$

для любого n и любой меры P.

Как и раньше, в силу монотонности можно суммировать не по словам данной длины, а по любому беспрефиксному множеству.

Следующее предложение легко доказывается с использованием аналога предложения 4.16 (представления неотрицательных перечислимых снизу функций как сумм рядов):

Предложение 5.14. *Любой равномерный тест $t(\omega, P)$ порождается некоторым расширенным равномерным тестом в следующем смысле:*

$$t(\omega, P) = \sup_{x \sqsubseteq \omega} T(x, P).$$

Напротив, эта формула по любому расширенному равномерному тесту T даёт равномерный тест t.

Среди расширенных равномерных тестов тоже можно выбрать максимальный (проводя аналогичное усечение и складывая результаты). Фиксируем один из таких тестов; будем обозначать его $\mathbf{t}(x, P)$ (где $x \in \{0,1\}*$, P — мера на Ω). Он порождает максимальный расширенный равномерный тест $\mathbf{t}(\omega, P)$ (с точностью до ограниченного множителя).

Замечание 5.15. *Если мы захотим перенести развитую нами теорию на случай некомпактного пространства последовательностей натуральных чисел (вместо компактного пространства последовательностей нулей и единиц), то можно и нужно определять расширенные тесты непосредственно, а не через тесты на бесконечных последовательностях. При этом можно доказать и существование максимального среди них.*

Предложение 5.10 позволяет обобщить результат, известный нам для бернуллиевых мер:

Теорема 5.16. *Пусть мера P вычислима относительно оракула A и, напротив, оракул A может быть эффективно восстановлен по известным приближения с произвольной точностью к значениям меры P. В этом случае последовательность ω равномерно случайна относительно меры P тогда и только тогда, когда она случайна по Мартин-Лёфу относительно меры P с оракулом A.*

(Поскольку добавление оракула A делает меру P вычислимой, случайность по Мартин-Лёфу имеет смысл.)

Доказательство. Пусть $\mathbf{t}(\omega, P)$ бесконечно. Заметим, что функция $\mathbf{t}(\cdot, P)$ является A-перечислимой снизу, так как мера P вычислима с оракулом A. Поэтому ω не случайна по мере P с оракулом A.

Напротив, пусть ω не случайна по мере P с оракулом A и $t(\cdot)$ — A-перечислимый снизу P-тест, для которого $t(\omega) = \infty$. Поскольку оракул A может быть восстановлен по приближениям к P, то существует перечислимая снизу функция $\bar{t}(\cdot, \cdot)$, для которой $\bar{t}(\cdot, P) = t(\cdot)$. Остаётся преобразовать $\bar{t}$ в равномерный тест с помощью теоремы 5.7.

Отметим, что не всякая мера P удовлетворяет условию теоремы (оно означает, что массовая проблема "указывать приближения к значениям меры P" равносильна проблеме разрешения некоторого множества; про степени таких массовых проблем см. в [21]). Впоследствии (теорема 5.36) мы покажем, как можно характеризовать равномерную случайность для произвольных мер (в терминах случайности по Мартин-Лёфу с оракулом).

Другое применение техники усечения: покажем, что равномерные тесты являются обобщением равномерных бернуллиевых тестов в смысле определения 4.19.

Теорема 5.17. *Пусть $\mathbf{t}(\omega, P)$ — универсальный равномерный тест и $\mathbf{t}(\omega, p)$ — универсальный равномерный тест бернуллиевости из леммы* 4.20. *Тогда $\mathbf{t}(\omega, B_p) \stackrel{+}{=} \mathbf{t}(\omega, p)$.*

Здесь B_p — бернуллиева мера с параметром p.

Доказательство. Для доказательства $\stackrel{+}{<}$-неравенства заметим, что функция $\langle \omega, p \rangle \mapsto \mathbf{t}(\omega, B_p)$ является равномерным бернуллиевым тестом, поскольку функция $p \mapsto B_p$ вычислима (в естественном смысле).

Чтобы доказать обратное неравенство, заметим, что существует вычислимое отображение из пространства мер в $[0, 1]$, которое переводит B_p в p (достаточно взять вероятность однобитового слова). Комбинируя эту функцию с $\mathbf{t}(\omega, p)$, мы получаем перечислимую снизу функцию $f(\omega, P)$, определённую на всех мерах и продолжающую наш тест на бернуллиевых мерах: $f(\omega, B_p) = \mathbf{t}(\omega, p)$. Функция f ещё не является равномерным тестом, но к ней можно применить усечение.

5.2 Априорная вероятность с оракулом и равномерные тесты

Для вычислимой меры у нас было выражение для универсального теста (предложение 2.21) через (дискретную) априорную вероятность. Аналогичное выражение существует и для универсального равномерного теста:

Теорема 5.18.

$$\mathbf{t}(\omega, P) \stackrel{+}{=} \sum_{x \sqsubseteq \omega} \frac{\mathbf{m}(x|P)}{P(x)}.$$

Здесь, правда, нам ещё предстоит определить понятие априорной вероятности относительно меры, то есть величину $\mathbf{m}(x|P)$. Мы сейчас это сделаем, после чего вернёмся к доказательству.

Определение 5.19. *Будем называть неотрицательную функцию $t(x, P)$, аргументами которой являются двоичное слово x и мера P,* равномерной перечислимой снизу полумерой, *если она перечислима снизу и $\sum_x t(x, P) \leqslant 1$ для любой меры P на Ω.*

Предложение 5.20. *Среди всех равномерных перечислимых снизу полумер существует наибольшая с точностью до умножения на константу.*

Это доказывается тем же способом, что и существование универсального теста (и даже немного проще, поскольку здесь ограничение на значения теста не зависит от меры).

Определение 5.21. *Фиксируем одну из таких наибольших полумер и назовём её* априорной вероятностью относительно P. *Будем обозначать её $\mathbf{m}(x|P)$.*

(Мы используем чёрточку вместо запятой, чтобы подчеркнуть родство с рассматриваемой обычно условной априорной вероятностью.)

Доказательство.(теоремы 5.18) Нам нужно проверить две вещи. Во-первых, мы должны убедиться, что правая часть формулы задаёт равномерный тест. Каждый из членов суммы можно рассматривать как функцию двух аргументов, равную 0 вне конуса продолжений x и равную $\mathbf{m}(x|P)/P(x)$ внутри этого конуса. Для данного x функции $\mathbf{m}(x|P)$ и $1/P(x)$ перечислимы снизу (равномерно по x), и их суммирование даёт перечислимую снизу функцию. Интеграл этой функции по какой-либо мере P равен сумме интегралов слагаемых, то есть $\sum_x \mathbf{m}(x|P)$, и потому не превосходит 1.

Здесь есть особый случай, когда $P(x) = 0$ для некоторого x. В этом случае соответствующее слагаемое формулы становится бесконечным для ω, продолжающих x. Но поскольку мера этого конуса равна нулю, то интеграл по нему равен нулю, и потому слагаемое хоть и не равно $\mathbf{m}(x|P)$, но только меньше. Таким образом, правая часть формулы есть равномерный тест и потому не превосходит универсального равномерного теста: мы доказали $\dot{>}$-неравенство.

Вторая часть доказательства не так проста: наблюдая за увеличением значений равномерного теста, мы должны распределить это увеличение между различными членами суммы в правой части, при этом сохранив перечислимость снизу. Трудность в том, что если, скажем, сначала перечислимая снизу функция была равна 1 на некотором эффективно открытом множестве A, а вне него равнялась нулю, а потом это множество заменилось на большее множество B, то разница (характеристическая функция $B \setminus A$), вообще говоря, не будет перечислимой снизу, так как в пространстве мер (как, например, и на отрезке) разность двух интервалов не будет открытым множеством.

Решение этой проблемы состоит в том, что мы переходим к непрерывным функциям. Пусть нам дан произвольный равномерный тест $t(\omega, P)$. Поскольку он перечислим снизу, его можно представить как предел неубывающей последовательности неотрицательных базисных функций, или — переходя к разностям — в виде суммы ряда из неотрицательных базисных функций: $t(\omega, P) = \sum t_i(\omega, P)$.

Будучи базисной, функция ω зависит лишь от некоторого конечного начала последовательности ω; обозначим длину этого начала n_i. Для каждого слова x длины n_i мы получаем некоторую перечислимую снизу функцию $t_{i,x}(P)$, при этом $t_i(\omega, P) = t_{i,x}(P)$, если ω начинается на x. Теперь положим $m_i(x, P) = t_{i,x}(P) \cdot P(x)$, если x имеет длину n_i (для остальных длин нуль). Функция m_i перечислима снизу (как произведение двух перечислимых снизу функций) равномерно по i, поэтому и сумма $m(x, P) = \sum m_i(x, P)$ будет перечислимой снизу.

Покажем, что m является полумерой, то есть что $\sum_x m(x,P) \leqslant 1$ при любом P. В самом деле, $\sum_x m_i(x,P)$ ненулевые члены соответствуют словам длины n_i, и эта сумма равна $\sum_x t_{i,x}(P)P(x)$, то есть в точности интегралу $\int t_i(\omega,P)\,dP(\omega)$, а сумма этих интегралов не превосходит 1 по условию.

Кроме того, если для всех начал x последовательности ω мера $P(x)$ не равна нулю, то

$$\sum_{x\sqsubseteq\omega} \frac{m_i(x,P)}{P(x)} = \frac{t_{i,x_i}(P)\cdot P(x)}{P(x)} = t_i(\omega,P)$$

(здесь x_i — начало ω длины n_i), поэтому после суммирования по i

$$\sum_{x\sqsubseteq\omega} \frac{m(x,P)}{P(x)} = t(\omega,P),$$

и остаётся воспользоваться максимальностью полумеры, чтобы получить $\dot{<}$-неравенство для случая, когда все начала ω имеют ненулевую P-меру. Если же одно из них имеет нулевую P-меру, то правая часть бесконечна, так что и тут неравенство выполнено.

Вопрос. Для универсального теста случайности относительно равномерной меры в этой формуле можно было заменить сумму на максимум. Можно ли это сделать для равномерных тестов? (Применённое тогда рассуждение встречает трудности.) Можно ли разумно определить априорную вероятность на дереве относительно меры, и доказать равномерный вариант теоремы Левина–Шнорра?

Мы вернёмся к определению априорной вероятности с оракулом (и её связи с префиксной сложностью) в разделе 7.4

5.3 Эффективно компактные классы мер

Мы рассматривали бернуллиевы тесты, то есть полунепрерывные снизу функции, интеграл от которых по любой бернуллиевой мере не превосходит 1. В этом определении вместо бернуллиевых мер можно рассматривать произвольный эффективно компактный класс:

Определение 5.22. *Пусть $\mathcal{C}$ — эффективно компактный класс мер на Ω. Неотрицательная полунепрерывная снизу функцию $t\colon \Omega \to [0,\infty]$ называется* $\mathcal{C}$-тестом, *если $\int t(\omega)\,dP(\omega) \leqslant 1$ для любой меры $P \in \mathcal{C}$.*

Теорема 5.23. *Пусть $\mathcal{C}$ — эффективно компактный класс мер.*

(а) *Существует универсальный $\mathcal{C}$-тест $\mathbf{t}_{\mathcal{C}}(\cdot)$.*

(б) $\mathbf{t}_{\mathcal{C}}(\omega) = \inf_{P\in\mathcal{C}} \mathbf{t}(\omega,P)$.

Доказательство. Оба утверждения теоремы доказываются аналогично леммам 4.20 и 4.21.

Замечание 5.24. *Поскольку класс $\mathcal{C}$ компактен, а функция $\mathbf{t}(\omega,P)$ полунепрерывна снизу, то* inf *в утверждении* (б) *можно заменить на* min.

Вопрос. Можно ли найти какие-то критерии случайности по отношению к естественным классам мер (в частности, в терминах сложности)? Например, можно ли охарактеризовать бернуллиевы последовательности в терминах сложности их начальных отрезков? Можно показать, что главный член дефекта бернуллиевости можно записать как

$$\log C_n^k - KP(x|n,k)$$

для начального отрезка x длины n, содержащего k единиц. Подобный критерий приведён в [6], но он выглядит довольно искусственно.

Аналогичные вопросы естественно задать и для других классов (марковские меры, инвариантные относительно сдвигов меры).

5.4 Разреженные последовательности

Бывают ситуации, в которых мы говорим о случайности, но это не сводится к стандартной постановке (случайность данного наблюдения ω относительно данной модели P). Сейчас мы рассмотрим один из таких случаев — понятие разреженной последовательности, введённое в [3]. Другой пример, который можно назвать онлайн-случайностью, рассмотрен в разделе 9.2.

Будем называть p-разреженной последовательность, в которой меньше единиц, чем в случайной по бернуллиевой мере B_p. Другими словами, будем брать произвольные B_p-случайные последовательности и заменять в них некоторые единицы на нули. Всё, что получится таким образом, будет p-разреженным.

Определение 5.25 (разреженные последовательности)**.** *Введём покоординатный порядок на бесконечных последовательностей нулей и единиц (или конечных последовательностей одной длины): $\omega \leqslant \omega'$, если $\omega(i) \leqslant \omega'(i)$ при всех i, то есть ω может быть получена из ω' заменой некоторых единиц на нули.*

Пусть B_p — бернуллиева мера для некоторого вычислимого p. Будем говорить, что последовательность ω является p-разреженной*, если $\omega \leqslant \omega'$ для некоторой B_p-случайной ω'. (В терминах множеств: p-разреженные множества — это подмножества p-случайных множеств.)*

Покажем, что в определении разреженности можно избавиться от квантора существования по ω' и дать критерий в терминах монотонных тестов.

Определение 5.26. *Будем говорить, что функция $f\colon \Omega \to [0,\infty]$ монотонна, если $f(\omega') \geqslant f(\omega)$ при $\omega' \geqslant \omega$.*

Монотонная перечислимая снизу функция $f\colon \Omega \to [0,\infty]$ называется тестом p-разреженности*, если $\int t(\omega)\, dB_p(\omega) \leqslant 1$. Тест разреженности называем* универсальным*, если он максимален среди всех таких тестов (с точностью до умножения на константу, как обычно).*

Монотонность тестов гарантирует, говоря неформально, что закономерностью является наличие единиц на каких-то местах, а не их отсутствие. (Отметим, что раньше мы говорили совсем о другой монотонности, определяя расширенные тесты: там сравнивались значения функции на конечном слове и его продолжении.)

Предложение 5.27. *Рассмотрим универсальный тест $\mathbf{t}(\omega, P)$. Тогда величина*

$$r_p(\omega) = \min_{\omega' \geqslant \omega} \mathbf{t}(\omega, B_p)$$

задаёт универсальный тест p-разреженности.

Доказательство. Тест p-разреженности по определению является тестом по мере B_p. Используя его монотонность и сравнивая с универсальным, получаем, что любой тест разреженности не превосходит r_p (с точностью до константы).

В обратную сторону нужно показать, что минимум в выражении для r_p достигается и что эта функция является тестом p-разреженности. Перечислимость снизу доказывается с использованием того, что свойство $\omega \leqslant \omega'$ задаёт эффективно замкнутое подмножество эффективно

компактного пространства $\Omega \times \Omega$ (ср. ниже предложение 7.20). Монотонность и неравенство для интеграла непосредственно следуют из определения.

Отсюда получаем критерий разреженности в терминах тестов:

Теорема 5.28. *Последовательность является p-разреженной (получается из p-случайной заменой некоторых единиц на нули) тогда и только тогда, когда универсальный тест разреженности $r_p(\omega)$ конечен.*

Разреженность эквивалентна случайности по некоторому классу мер. Чтобы описать этот класс, введём понятие спаривания (coupling) мер.

Определение 5.29. *Пусть P, Q — две меры на Ω. Будем говорить, что мера P может быть спарена с Q (обозначение: $P \preceq Q$), если существует мера R на $\Omega \times \Omega$, для которой:*

(а) *первая проекция R равна P, а вторая равна Q;*

(б) *мера R целиком сосредоточена на парах $\langle \omega, \omega' \rangle$, у которых $\omega \leqslant \omega'$ (вероятность этого события по мере R равна 1).*

Отметим, что отношение спаривания в этом определении несимметрично (хотя из названия этого не видно); более наглядно было бы говорить "P может быть помещена под Q", если $P \preceq Q$.

Следующий критерий спариваемости хорошо известен и восходит к [27]; доказательство можно найти в [3].

Предложение 5.30. *Свойство $P \preceq Q$ равносильно такому: для всякой монотонной базисной функции f выполнено неравенство*

$$\int f(\omega)\, dP(\omega) \leqslant \int f(\omega)\, dQ(\omega).$$

В этом критерии можно допустить произвольные монотонные функции, или, наоборот, ограничится характеристическими функциями множеств.

Определение 5.31. *Пусть $\mathcal{S}_p$ — класс всех мер P, для которых $P \preceq B_p$.*

Предложение 5.32. *Класс мер $\mathcal{S}_p$ является эффективно замкнутым (и, следовательно, эффективно компактным).*

Доказательство. Каждое из неравенств предыдущего предложения (для каждой базисной функции f) задаёт эффективно замкнутое множество, и их пересечение тоже будет эффективно замкнутым.

Теорема 5.33. *Универсальный тест r_p является универсальным тестом для класса мер $\mathcal{S}_p$.*

Отсюда вытекает, что последовательность является p-разреженной тогда и только тогда, когда она равномерно случайна относительно некоторой меры из класса $\mathcal{S}_p$.

Доказательство этой теоремы основано на таком утверждении:

Лемма 5.34 (монотонизация)**.** *Пусть $t\colon \Omega \to \mathbb{R}$ — базисная функция, и $\int t(\omega)\, dQ(\omega) \leqslant 1$ для любой меры $Q \in \mathcal{S}_p$. Определим монотонную базисную функцию $\hat{t}(\omega) = \max_{\omega' \leqslant \omega} t(\omega)$; это определение корректно, так как $t(\omega)$ зависит только от конечного числа позиций в ω. Тогда $\int \hat{t}(\omega)\, dB_p(\omega) \leqslant 1$.*

Доказательство. Пусть функция t зависит только от первых n координат. Для каждого $x \in \{0,1\}^n$ выберем $x' \leqslant x$, где $t(x')$ достигает максимума (среди таких x'). Помимо распределения B_p рассмотрим распределение Q, в котором бернуллиева мера x перенесена на x' (при этом меры из различных x могут быть отнесены к одному x' и тогда складываются). Мы описали поведение Q на первых n битах; следующие биты добавляются независимо и вероятность единицы в каждой позиции равна p; отметим также, что для математических ожиданий функций t и $\hat{t}$ важны только первые n битов.

По построению $Q \preceq B_p$ (мы по существу описали меру на парах), поэтому $\int t(\omega)\, dQ(\omega) \leqslant 1$. Но этот интеграл равен $\int \hat{t}(\omega)\, dB_p(\omega)$. Лемма доказана.

Вернёмся к теореме 5.33.

Доказательство. Всякий тест p-разреженности t является тестом для класса $\mathcal{S}_p$. В самом деле, интеграл t по мере из класса $\mathcal{S}_p$ не превосходит интеграла t по мере B_p в силу монотонности теста и возможности спаривания.

В другую сторону: покажем, что для любого теста t для класса $\mathcal{S}_p$ существует не меньший его тест p-разреженности. В самом деле, тест t может быть представлен в виде предела возрастающей последовательности базисных функций t_n. Применив к ним лемму о монотонизации, получим возрастающую последовательность базисных функций $\hat{t}_n$, которые всюду не меньше t_n и имеют интегралы не больше 1 по мере B_p. Их предел и будет требуемым тестом p-разреженности.

5.5 Варианты определений случайности

Мы уже определили равномерную случайность последовательности относительно произвольной (не обязательно вычислимой) меры. Однако есть и другие варианты такого рода определений.

Оракулы

Мы можем использовать определение случайности по Мартин-Лёфу, добавляя оракул, который делает меру вычислимой. А именно, назовём последовательность ω случайной относительно меры P, если существует оракул A, относительно которого P вычислима, и при этом ω случайна в смысле Мартин-Лёфа с оракулом A относительно P.

(Мы говорим "существует оракул A, делающий меру P вычислимой", а не "для любого оракула A, делающего P вычислимой", поскольку среди таких оракулов есть и оракул, делающий последовательность ω вычислимой. В этом случае она не может быть случайной, если только не является атомом меры P.)

Оказывается, что это определение (как доказали Адам Дей и Джозеф Миллер), этот вариант определения равносилен равномерной случайности. Доказательство этой эквивалентности требует некоторых приготовлений.

Прежде всего зададим себе вопрос, почему нельзя взять в качестве оракула саму меру (как это делалось для случая бернуллиевых мер, когда мы в качестве оракула брали двоичное разложение числа p). Дело в том, что выбор такого разложения не однозначен ($0.01111\ldots = 0.10000\ldots$). Когда речь идёт об одном числе p, то это не важно, поскольку неоднозначность возникает только для рациональных p, и в этом случае оба представления вычислимы. Однако для мер это уже не так: мера задаётся счётным количеством действительных чисел (скажем, вероятностями отдельных слов, или условными вероятностями), и произвол в выборе представления может не сводиться к конечному число вариантов.

Определение 5.35. *Зафиксируем некоторый способ кодирования мер на Ω двоичными последовательностями, то есть вычислимое отображение $\pi \mapsto R_\pi$ множества Ω в пространство мер. Например, можно разбить последовательность π на счётное число частей и каждую из них считать двоичной записью условной вероятности единицы после некоторого начала (начал тоже счётное число). При этом возникает неоднозначность (если вероятность какого-то начала равна нулю, то условные вероятности после него не играют роли), но она и так была.*

Определим теперь r-тест (*representation test, тест случайности по данному представлению меры*) *как перечислимую снизу неотрицательную функцию $t(\omega, \pi)$, для которой неравенство $\int t(\omega, \pi)\, dR_\pi(\omega) \leqslant 1$ выполняется при всех π.*

Как мы уже обсуждали, одна и та же мера P может иметь много представлений (может быть много различных π, для которых $R_\pi = P$), и значения теста для различных представлений одной и той же меры могут быть разными.

Как обычно, легко доказать, что

(а) всякая перечислимая снизу функция может быть эффективно усечена (сделана не более чем вдвое превосходящей r-тест), при этом если она уже была r-тестом, то она не изменится;

(б) существует универсальный (максимальный с точностью до константы) r-тест $\mathbf{t}(\omega, \pi)$.

При фиксированном π функция $\mathbf{t}(\cdot, \pi)$ совпадает с универсальным π-вычислимым ограниченным в среднем тестом случайности относительно меры R_π. В самом деле, она является таким тестом; с другой стороны, любой такой тест перечисляется снизу машиной с оракулом, и эта машина может быть применена к любому оракулу (но может не давать теста); мы получаем перечислимую снизу функцию $t'(\omega, \pi)$, которая совпадает с исходным тестом при данном π; остаётся применить свойство (а).

Как следствие этого простого рассуждения мы получаем, что величина $\mathbf{t}(\omega, \pi)$ конечна тогда и только тогда, когда последовательность ω случайна с оракулом π относительно меры R_π.

Теорема 5.36 (Дей–Миллер)**.** *Последовательность ω равномерно случайна по мере P тогда и только тогда, когда существует оракул π, делающий меру P вычислимой, а последовательность ω — случайной в смысле Мартин-Лёфа по мере P с оракулом π.*

Кроме того,

$$\mathbf{t}(\omega, P) \doteq \inf_{\{\pi | R_\pi = P\}} \mathbf{t}(\omega, \pi).$$

Доказательство. Докажем указанное в теореме равенство. Заметим, что если t — равномерный тест, то $t(\omega, R_\pi)$ как функция от ω и π представляет собой r-тест, и потому мажорируется универсальным r-тестом.

Обратное утверждение несколько сложнее. Нам нужно доказать, что функция в правой части перечислима снизу как функция последовательности ω и меры P. (Условие на интеграл после этого получается легко, поскольку мера P имеет хотя бы одно представление π.) Это можно доказать, используя эффективную компактность множества пар $\langle P, \pi \rangle$, для которых $P = R_\pi$. В общем виде (для произвольных конструктивных метрических пространств) это утверждение составит содержание леммы 7.21 на с. 45, и мы не будем приводить отдельного доказательства для рассматриваемого частного случая, поскольку оно ничем не отличается от общего.

Осталось объяснить, как связаны доказанное равенство и случайность с оракулом. Если $\mathbf{t}(\omega, P)$ конечно, то по доказанному равенству существует π, при котором $R_\pi = P$ и $\mathbf{t}(\omega, \pi)$ конечно. Как мы видели, это в свою очередь означает, что ω случайна по мере R_π, то есть по мере P, с оракулом π, который делает меру P вычислимой.

Напротив, если $t(\omega, P)$ бесконечно, а оракул A делает меру P вычислимой, то функция $t(\cdot, P)$ будет A-перечислимой снизу, и её интеграл не превосходит 1 по мере P, так что последовательность ω не будет случайной с оракулом A по мере P.

Слепая (безоракульная) случайность

Можно использовать определение эффективно нулевого множества (или перечислимого снизу теста) и в ситуации невычислимой меры. При этом может не существовать максимального эффективно нулевого множества. Например, если мера P сосредоточена на единственной невычислимой последовательности π, то все интервалы, не содержащие π, будут эффективно нулевыми множествами, а их объединение (дополнение к синглетону $\{\pi\}$) таковым не будет, иначе π была бы вычислимой.

Тем не менее мы можем определить понятие случайной последовательности как последовательности, не лежащей *ни в одном* эффективно нулевом множестве (что эквивалентно тому, что все тесты на ней конечны). Кьёс-Хансен предложил называть такую случайность "гиппократовой" (ссылаясь на легенду о враче Гиппократе), но мы предпочитаем говорить о "слепой" (blind) или "безоракульной" случайности.

Определение 5.37 (слепые тесты)**.** *Перечислимые снизу (без оракула) функции $t(\omega)$, для которых $\int t(\omega)\, dP(\omega) \leqslant 1$, будем называть* слепыми, *или* безоракульными, *тестами для меры P. Последовательность ω будем называть* безоракульно случайной *по мере P, если $t(\omega)$ конечно для любого такого теста t.*

Как мы видели, может не существовать максимального слепого теста.

Это понятие безоракульной случайности можно характеризовать во введённых ранее терминах:

Теорема 5.38. *Последовательность ω является безоракульно случайной относительно меры P тогда и только тогда, когда она случайна относительно любого эффективно компактного класса мер, содержащего меру P.*

Доказательство. Предположим, что ω не случайна относительно некоторого эффективно компактного класса мер, содержащих меру P. Тогда перечислимый снизу (безо всякого оракула) тест для этого класса будет безоракульным тестом для P, так что ω не будет безоракульно случайной.

С другой стороны, предположим, что существует некоторый безоракульный тест t для меры P, для которого $t(\omega)$ бесконечно. Тогда можно попросту рассмотреть класс всех мер Q, для которых t является тестом, то есть для которых $\int t(\omega)\, dQ(\omega) \leqslant 1$. Этот класс эффективно замкнут, и потому эффективно компактен. В самом деле, если для некоторой меры Q интеграл больше 1, то уже для некоторого базисного приближения t_n к t (снизу) этот интеграл больше 1, а последнее свойство задаёт эффективно открытое множество в пространстве мер. Перечисляя все t_n и объединяя все эти множества, получаем эффективно открытое множество.

Из определения (или из последней теоремы) легко вывести, что из равномерной случайности последовательности ω относительно P следует безоракульная случайность. Обратное утверждение неверно:

Теорема 5.39. *Существует последовательность ω, безоракульно случайная по некоторой мере P, но не являющаяся равномерно случайной по этой мере.*

Доказательство. Заметим, что безоракульная случайность не изменится, если мы чуть-чуть изменим меру P (так, чтобы мера любого множества изменилась не более чем в $O(1)$ раз). С другой стороны, с вычислительной точки зрения новая мера может быть гораздо сильнее. Например, начнём с равномерной бернуллиевой меры $B_{1/2}$, в которой все испытания независимы и имеют вероятность успеха $1/2$, и возьмём некоторую случайную по ней последовательность $\omega = \omega(1)\omega(2)\ldots$. Затем рассмотрим чуть сдвинутую меру B', в которой вероятности успеха равна $1/2 + \omega(1)\varepsilon_1, 1/2 + \omega(2)\varepsilon_2, \ldots$; здесь ε_i настолько малы и так быстро сходятся к нулю, что B' отличается от B на любом множестве не более чем в константу раз. Тогда B' содержит информацию об ω, и несложно построить равномерный тест t, для которого $t(\omega, B') = \infty$.

Однако бывают некоторые классы мер, для которых понятия равномерной и безоракульной случайности совпадают. (В частности, таковы бернуллиевы меры.) Чтобы сформулировать достаточные условия такого совпадения, начнём с определения.

Определение 5.40. *Обозначим через* $\mathrm{Randoms}(P)$ *множество последовательностей, равномерно случайных относительно меры* P*. Назовём класс мер* эффективно ортогональным*, если*

$$\mathrm{Randoms}(P) \cap \mathrm{Randoms}(Q) = \varnothing$$

для любых двух различных мер P *и* Q *из этого класса.*

Теорема 5.41. *Пусть* $\mathcal{C}$ *— эффективно компактный и эффективно ортогональный класс мер. Тогда для любой меры* P *из этого класса понятия равномерной и безоракульной случайности относительно* P *совпадают.*

Утверждение этой теоремы выглядит парадоксально: мы утверждаем нечто о случайности по одной мере P, а в условии стоит возможность вложить P в класс мер с некоторыми свойствами. (Было бы естественно найти более явные достаточные условия на P.)

Из этой теоремы следует, что построенная при доказательстве теоремы 5.39 мера не может быть вложена в эффективно замкнутый эффективно ортогональный класс.

Доказательство. Мы уже отмечали, что в одну сторону утверждение прямо следует из определения. Докажем обратное. Предположим, что последовательность ω безоракульно случайна относительно меры P. По теореме 5.38 она случайна относительно класса мер $\mathcal{C}$. Следовательно, она равномерно случайна относительно некоторой меры P' из класса $\mathcal{C}$. Остаётся доказать, что $P = P'$.

Пусть это не так. Покажем, что существует компактный класс мер $\mathcal{C}'$, который содержит P, но не содержит P'. В самом деле, $P \neq P'$ означает, что некоторое слово имеет различные меры относительно P и P', и можно найти замкнутое условие на меру этого слова, которое отделит P от P'. Теперь рассмотрим эффективно компактный класс $\mathcal{C} \cap \mathcal{C}'$. Он содержит P и потому последовательность ω будет случайной относительно этого класса. Значит, и в нём есть мера P'', относительно которой последовательность ω равномерно случайна. Но $P' \neq P''$ (одна мера лежит в $\mathcal{C}'$, а другая — нет), так что получаем противоречие с эффективной ортогональностью класса $\mathcal{C}$.

Замечание 5.42. *Доказанная теорема применима, в частности, к классу бернуллиевых мер. Может показаться, что её можно доказать проще: если последовательность* ω *случайна (безоракульно или тем более равномерно) по мере* B_p*, то предел частоты единиц в ней равен* p *и тем самым он определяется самой последовательностью (и дополнительный оракул для* p *ничего нового не даёт). Это рассуждение, однако, неправильно: хотя* p *и определяется*

последовательностью, но не является вычислимой (или хотя бы непрерывной) функцией от неё. В самом деле, никакой начальный отрезок последовательности не гарантирует, что её предельная частота будет в заданном интервале. Аналогичное рассуждение, однако, можно применить к последовательностям, у которых дефект случайности ограничен заранее известной константой. (См. подробнее в [11]*, где введено относящееся к этому понятие послойной вычислимости.) В частности, можно показать, что если последовательность ω безоракульно случайна по мере B_p, то двоичное разложение p вычислимо с оракулом ω.*

6 Нейтральная мера

Следующая теорема, опубликованная в [16] и затем в [10], указывает на парадоксальное свойство равномерной случайности, которая отличает её от случайности с оракулом.

Определение 6.1. *Назовём меру на* Ω нейтральной*, если всякая последовательность является равномерно случайной относительно этой меры.*

Теорема 6.2. *Существует нейтральная мера; более того, существует мера N, для которой $\mathbf{t}(\omega, N) \leqslant 1$ при всех $\omega \in \Omega$.*

Прежде чем доказывать эту теорему, отметим, что все вычислимые последовательности обязаны быть атомами нейтральной меры (иметь положительную вероятность). В самом деле, можно построить тест, который ищет длинные отрезки вычислимой последовательности, имеющие малую меру (этого можно дождаться, имея последовательность и меру в качестве аргументов), и присваивает им большое значение дефекта.

Отсюда следует, что нейтральная мера не может быть вычислимой. В самом деле, для вычислимой меры легко построить вычислимую последовательность, не являющуюся атомом (надо из двух продолжений слова выбирать то, которое имеет меньшую — или хотя бы не сильно бóльшую — меру). Аналогичное рассуждение показывает, нейтральная мера не эквивалентна никакому оракулу (нет оракула, который делал бы её вычислимой и, напротив, мог бы быть восстановлен по приближениям к ней). В самом деле, если A — такой оракул, то (как мы видели) равномерная случайность равносильна случайности с оракулом A, и можно повторить то же рассуждение.

Нейтральная мера не может быть также перечислимой сверху или снизу, но при нашем определении (когда мера всего пространства должна равняться единице) это не даёт ничего нового. Некоторые более осмысленные (и менее тривиальные) варианты этого утверждения доказаны в [17].

Доказательство. Рассмотрим универсальный тест $\mathbf{t}(\omega, P)$. Мы утверждаем, что существует мера N, для которой $\mathbf{t}(\omega, N) \leqslant 1$ для любой последовательности N. Другими словами, для каждой последовательности ω мы имеем условие на N, состоящее в том, что $\mathbf{t}(\omega, N) \leqslant 1$, и надо доказать, что все эти условия совместны (пересечение их непусто). Каждое условие задаёт замкнутое множество в компактном пространстве всех мер (вспомним, что $\mathbf{t}$ перечислимо снизу, и тем более полунепрерывно снизу), поэтому достаточно доказать совместность любого конечного числа условий .

Итак, возьмём k произвольных последовательностей $\omega_1, \ldots, \omega_k$. Нам надо доказать, что существует мера N, для которой $\mathbf{t}(\omega_i, N) \leqslant 1$ при всех $i = 1, \ldots, k$. Эту меру мы будем искать в виде выпуклой комбинации мер, сосредоточенных в $\omega_1, \ldots, \omega_k$. Таким образом, нам надо показать, что k замкнутых подмножеств k-мерного симплекса (соответствующих k условиям) имеют общую точку. Это делается с помощью известной топологической леммы (используемой в стандартном доказательстве теоремы Брауэра о неподвижной точке):

Лемма 6.3. *Пусть симплекс с вершинами $1, \ldots, n$ покрыт k замкнутыми множествами $A_1, \ldots, A_k$. Пусть при этом вершина i всегда принадлежит множеству A_i, ребро i–j целиком лежит в объединении $A_i \cup A_j$, и так далее (грань $(i_1, \ldots, i_s)$ целиком лежит в $A_{i_1} \cup \ldots \cup A_{i_s}$). Тогда пересечение $A_1 \ldots, A_k$ непусто.*

Приведём для полноты стандартное доказательство этой леммы. Рассмотрим симплициальное разбиение данного симплекса на мелкие симплексы и пометим каждую их вершину каким-то числом i от 1 до k, с тем условием, чтобы эта вершина лежала в A_i. Более того, можно предполагать что вершина i помечена числом i, вершины на отрезке i–j помечены либо числом i, либо числом j, и так далее. Комбинаторная лемма Шпернера говорит, что есть симплекс разбиения, у которого все вершины помечены по-разному. Устремляя размер максимального симплекса разбиения к нулю, и выбирая предельную точку последовательности получившихся разноцветных симплексов, находим искомую точку в пересечении всех A_i. Лемма доказана.

Применим теперь доказанную лемму. Согласно этой лемме, нам достаточно доказать, что любая точка на грани симплекса покрывается объединением соответствующих множеств. Пусть, скажем, есть точка (мера) X, являющаяся смесью вершин ω_1, ω_5 и ω_7; это значит, что мера X сосредоточена в множестве $\{\omega_1, \omega_5, \omega_7\}$. Нам нужно показать, что точка X покрыта объединением множеств A_1, A_5 и A_7, в нашем случае это означает, что одно из чисел $\mathbf{t}(\omega_1, X)$, $\mathbf{t}(\omega_3, X)$ и $\mathbf{t}(\omega_7, X)$ не превосходит единицы. Но мы знаем, что $\int \mathbf{t}(\omega, X)\, dX(\omega) \leqslant 1$ согласно определению теста, а этот интеграл является взвешенным средних этих трёх величин, так что хотя бы одна из них не превосходит 1.

7 Случайные точки метрического пространства

Большая часть сформулированных нами результатов о случайных последовательностях битов переносится на случай последовательностей натуральных чисел, а часто и на более общий случай метрических пространств. Мы сейчас изложим такие обобщения (а также и некоторые новые — даже и для пространства Ω — результаты).

7.1 Конструктивные метрическое пространства

Определяя конструктивные метрические пространства и перечислимые снизу функции на них, мы следуем [10, 12] (см. также[6]).

Определение 7.1. *Конструктивное метрическое пространство $\mathbf{X} = (X, d, D, \alpha)$ состоит из полного сепарабельного метрического пространства (X, d), в котором выделено счётное плотное множество D и его нумерация $\alpha\colon \mathbb{N} \to D$ (определённая на всём $\mathbb{N}$). Требуется, чтобы расстояние $d(\alpha(m), \alpha(n))$ было бы эффективно вычислимо (с любой требуемой точностью) по m и n.*

Открытые шары с центрами в точках из D и рациональными радиусами будем называть базисными шарами*; множество всех таких шаров образует* канонический базис *топологии X как метрического пространства.*

Будем называть последовательность $s_1, s_2, \ldots$ точек из D сильно фундаментальной*, если $d(s_m, s_n) \leqslant 2^{-m}$ при $n > m$. По предположению пространство X полно, поэтому всякая такая последовательность имеет единственный предел, и точки пространства можно отождествить с классами эквивалентности сильно фундаментальных последовательностей.*

Часто мы будем обозначать конструктивное метрическое пространство $\mathbf{X}$ просто X (если d, D и α ясны из контекста).

Примеры 7.2.

1. В *дискретном* метрическом пространстве $\Sigma = \{s_1, s_2, \ldots\}$ расстояние между любыми двумя различными точками равно 1; множество D содержит все точки и $\alpha(i) = s_i$.

2. Можно добавить к натуральным числам бесконечный элемент, положив $\overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$. Расстояние при этом можно определить как обычное расстояние между обратными величинами (при этом, естественно, $1/\infty = 0$). Это метрическое пространство можно назвать *одноточечной компактификацией* пространства натуральных чисел с дискретной метрикой предыдущего примера.

3. Вещественная прямая $\mathbb{R}$ с обычным расстоянием $d(x, y) = |x - y|$ также является конструктивным метрическим пространством (с каким-либо естественным выбором множества D и его нумерации). То же самое можно сказать о её положительной части $\mathbb{R}_+ = [0, \infty)$; можно добавить и бесконечный элемент $+\infty$, но тогда надо изменить метрику (например, перенеся её с отрезка).

4. Имея два конструктивных метрических пространства $\mathbf{X}$ и $\mathbf{Y}$, можно рассмотреть их произведение $\mathbf{Z} = \mathbf{X} \times \mathbf{Y}$ с одной из естественных метрик (например, сумма расстояний по обеим координатам); множество $D_{\mathbf{Z}}$ также является произведением $D_{\mathbf{X}} \times D_{\mathbf{Y}}$.

5. Пусть X — конечный или счётный алфавит (с фиксированной нумерацией). Тогда множество $X^{\mathbb{N}}$, состоящее из бесконечных последовательностей $x = (x(1), x(2), \ldots)$ с $x(i) \in X$, превращается в метрическое пространство, если положить $d(x, y) = 2^{-n}$, где n — минимальный индекс i, где $x(i)$ и $y(i)$ различаются. Это пространство является обобщением рассмотренного нами двоичного канторовского пространства; шары в нём являются цилиндрами: для данной конечной последовательности точек $z \in X^*$ мы берём все продолжения z.

Замечание 7.3. *Каждая точка x конструктивного метрического пространства $\mathbf{X}$ может рассматриваться как "массовая проблема" в смысле Медведева* [20]*: по данному рациональному $\varepsilon > 0$ указать номер точки из D, приближающей x с погрешностью не более ε. Легко проверить, что эта проблема эквивалентна (в смысле Медведева) проблеме перечисления всех базисных шаров, содержащих точку x.*

Замечание 7.4. *Конструктивное метрическое пространство является частным случаем более общего (и часто полезного) понятия конструктивного топологического пространства.*

Конструктивное топологическое пространство $\mathbf{X} = (X, \tau, \nu)$ *состоит из топологического пространства X, базиса открытых множеств τ и его нумерации ν: это значит, что $\tau = \{\nu(1), \nu(2), \ldots\}$ и что открытыми множествами в X являются объединения множеств из τ.*

Если Z является непустым подмножеством эффективного топологического пространства, то оно само становится эффективным топологическим пространством: мы пересекаем все базисные множества с Z, не меняя их нумерацию. В этом состоит важное преимущество понятия конструктивного топологического пространства (по сравнению с конструктивными метрическими пространствами): там мы можем перенести метрику на подмножество, но никакого естественного способа выделить в нём счётное плотное множество с нумерацией не видно.

В этой статье мы, однако, не будем стараться обобщить наши результаты на конструктивные топологические пространства (для чего нужно было бы сформулировать точно, какой класс конструктивных топологических пространств мы хотим рассматривать), а в качестве компромисса ограничимся подмножествами метрических пространств.

Определив структуру конструктивного топологического пространства, мы можем теперь определить некоторые эффективные варианты топологических понятий (для конструктивных метрических пространств):

Определение 7.5. *Открытое подмножество U конструктивного метрического пространства X называется* эффективно открытым, *если оно является объединением перечислимого семейства базисных шаров. Дополнение эффективно открытых множеств мы называем* эффективно замкнутыми.

Пусть A — некоторое подмножество X. Будем говорить, что множество $U \subseteq X$ открыто на A, *если найдётся эффективно открытое в X множество V, для которого $U \cap A = V \cap A$.*

Отметим, что в последнем определении U не обязано быть частью A, но реально играет роль лишь пересечение U с A.

Теперь можно определить вычислимость в терминах эффективно открытых множеств.

Определение 7.6 (вычислимые функции)**.** *Пусть $f\colon X \to Y$ — отображение метрических пространств, определённое на всём X. Функция f* непрерывна *тогда и только тогда, когда прообраз любого базисного открытого шара $B \subset Y$ является открытым подмножеством X; назовём функцию f* вычислимой, *если этот прообраз является эффективно открытым множеством (равномерно по B).*

Частичная функция f из X в Y, область определения которой содержит некоторое подмножество $A \subset X$, вычислима на A, *если прообраз любого базисного открытого шара B является эффективно открытым на A (равномерно по B). Из этого определения видно, что поведение функции вне A на вычислимость (на A) не влияет.*

Частным случаем вычислимости функций можно считать вычислимость точек: точку $x \in X$ будем называть вычислимой, если функция на одноэлементном пространстве $f\colon \{0\} \to X$ с $f(0) = x$ вычислима.

Если функция f, определённая в единственной точке $x_0 \in X$ и принимающая значение $y_0 \in Y$, вычислима на своей области определения, то мы говорим, что $y_0 = f(x_0)$ является x_0-вычислимым. *Если функция $f\colon Y \times Z \to Y$, определённая на $X \times \{z_0\}$, вычислима на своей области определения, мы называем функцию $g\colon X \to Y$, отображающую x в $g(x) = f(x, z_0)$,* z_0-вычислимой, *или* вычислимой относительно z_0.

Несложно проверить, что наше определение вычислимой точки эквивалентно более привычным:

Предложение 7.7. *Следующие свойства точки x конструктивного метрического пространства $\mathbf{X} = (X, d, D, \alpha)$ равносильны:*

(i) *x вычислима;*

(ii) *множество базисных шаров, содержащих x, перечислимо;*

(iii) *существует вычислимая последовательность $z_1, z_2, \ldots$ элементов D (заданных своими α-номерами), для которой $d(x, z_n) \leqslant 2^{-n}$ при всех n.*

Следующее предложение даёт более привычную переформулировку определения вычислимости:

Предложение 7.8. *Пусть $f\colon X \to Y$ — отображение метрических пространств. Функция f вычислима тогда и только тогда, когда существует вычислимое преобразование (задаваемое машиной с оракулом), которое переводит любую сильно фундаментальную последовательность точек из $D_{\mathbf{X}}$ с пределом x в сильно фундаментальную последовательность точек из $D_{\mathbf{Y}}$ с пределом $f(x)$.*

Если f — частичная функция с областью определения $Z \subset X$ и значениями в Y, то её вычислимость на Z равносильна существованию вычислимого преобразования, которое применимо к любой сильно фундаментальной последовательности с пределом $x \in Z$ и преобразует её в сильно фундаментальную последовательность с пределом $f(x)$.

Замечание 7.9. *В некотором смысле x_0-вычислимость аналогична вычислимости с оракулом (и в канторовском пространстве совпадает с ней), но в общем случае надо иметь в виду, что определение имеет более сложную природу: в этом определении машина имеет доступ к некоторой последовательности $s_1, s_2, \ldots$, сходящейся к x_0, но эта последовательность не фиксирована.*

Определение 7.10 (перечислимость снизу)**.** *Пусть дано конструктивное метрическое пространство $\mathbf{X} = (X, d, D, \alpha)$. Функция $f\colon X \to [-\infty, \infty]$* полунепрерывна снизу*, если множества $\{x \mid f(x) > r\}$ открыты при любом рациональном r (отсюда следует, что они открыты при любом r, не обязательно рациональном). Она называется* перечислимой снизу*, если эти множества эффективно открыты при любом рациональном r (равномерно по r). Частичная функция f из X в Y, определённая (по крайней мере) на некотором подмножестве A пространства X, называется* перечислимой снизу на A*, если множества $\{x \mid f(x) > r\}$ эффективно открыты на A равномерно по r.*

Аналогично определяется и перечислимость сверху; она равносильна перечислимости снизу функции $-f$.

Легко проверить, что функция $f\colon X \to \mathbb{R}$ вычислима тогда и только тогда, когда она перечислима сверху и снизу.

Как и раньше, можно дать эквивалентное определение перечислимости снизу с помощью базисных функций.

Определение 7.11 (базисные функции в конструктивном метрическом пространстве)**.** *Определим счётное множество* базисных *функций $\mathcal{E} = \{e_1, e_2, \ldots\}$ в конструктивном метрическом пространстве $\mathbf{X} = (X, d, D, \alpha)$ следующим образом. Для каждой точки $d \in D$ и для любых положительных рациональных чисел r и ε определим функцию $g_{d,r,\varepsilon}$: её значение в точке x определяется расстоянием от x до d и равно 1, если это расстояние не больше r, равно нулю, если расстояние не меньше $r + \varepsilon$, и линейно меняется, когда расстояние пробегает $[r, r + \varepsilon]$. См. рис.* 1.

Затем мы рассматриваем все функции, получаемые из семейства $g_{d,r,\varepsilon}$ замыканием относительно рациональных линейных комбинаций и операций максимума и минимума. Множество таких функций и обозначается $\mathcal{E}$; они имеют естественную нумерацию.

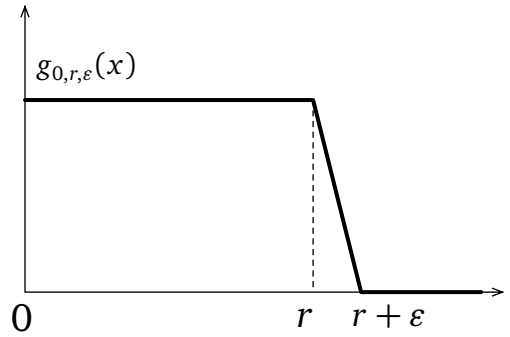

Figure 1: Функции, из которых строятся базисные.

Предложение 7.12. *Функция $f\colon X \to [0, +\infty]$, определённая на конструктивном метрическом пространстве, перечислима снизу тогда и только тогда, когда она является поточечным пределом неубывающей вычислимой последовательности базисных функций.*

Если в этом предложении отказаться от требования вычислимости, то вместо перечислимости снизу получится полунепрерывность снизу.

Определение 7.13. *Аналогичным образом определяется перечислимость снизу* относительно точки z_0 *некоторого метрического пространства* (*как перечислимость снизу на произведении области определения и* $\{z_0\}$).

Выбор метрики в конструктивном метрическом пространстве часто не влияет на понятие вычислимости. В частности, мы можем не различать эквивалентные метрики в смысле следующего определения (конструктивный вариант равномерной непрерывности тождественного отображения пространства в себя с другой метрикой):

Определение 7.14 (равномерная непрерывность, эквивалентность)**.** *Пусть X и Y — конструктивные метрические пространства, а $f\colon X \to Y$ — вычислимая функция. Мы говорим, что f* равномерно непрерывна*, если для любого рационального $\varepsilon > 0$ существует рациональное $\delta > 0$, для которого $d_{\mathbf{X}}(x_1, x_2) \leqslant \delta$ гарантирует $d_{\mathbf{Y}}(f(x_1), f(x_2)) \leqslant \varepsilon$. Мы говорим, что f* эффективно равномерно непрерывна*, если по ε можно эффективно найти соответствующее δ.*

Две метрики на одном и том же пространстве называются (эффективно) эквивалентными*, если тождественная функция, рассматриваемая как отображение пространства с одной метрикой в пространство с другой метрикой,* (*эффективно*) *равномерно непрерывна* (*в обе стороны*).

Например, в $\mathbb{R}^2$ евклидова метрика и L_1-метрика эффективно эквивалентны.

Понятие эффективной компактности (определение 5.4) очевидно переносится на произвольные конструктивные метрические пространства. Есть эффективный аналог и у более слабого понятия локальной компактности.

Определение 7.15 (эффективная компактность и локальная компактность в метрических пространствах)**.** *Компактное подмножество C эффективного метрического пространства X называется* эффективно компактным*, если можно перечислять все конечные его покрытия базисными множествами пространства X.*

Подмножество C метрического пространства называется локально компактным*, если его можно покрыть шарами B, у которых пересечения $\overline{B} \cap C$ компактны.* (*Здесь $\overline{B}$ — замыкание шара B.*) *Подмножество C будем называть* эффективно локально компактным*, если существует вычислимая последовательность базовых шаров B_k, покрывающих C, для которых $\overline{B}_k \cap C$ эффективно компактны равномерно по k.*

Примеры 7.16.

1. Конечное дискретное пространство компактно, а бесконечное — локально компактно.

2. Отрезок $[0, 1]$ является эффективно компактным. Прямая $\mathbb{R}$ является эффективно локально компактным пространством.

3. Для конечного алфавита X пространство бесконечных последовательностей $X^{\mathbf{N}}$ является эффективно компактным. (Для бесконечного алфавита это пространство не будет локально компактным.)

4. Пусть $\alpha \in [0, 1]$ — перечислимое снизу действительное число, не являющееся вычислимым. (Такие числа существуют; например, число $\sum 2^{-KP(x)}$ является таковым.) Рассмотрим отрезок $[0, \alpha]$ как конструктивное метрическое пространство: перечислимость α снизу позволяет перенумеровать рациональные числа, меньшие α. Это конструктивное метрическое пространство будет компактным, но не эффективно компактным.

Следующее предложение даёт полезный критерий эффективной компактности в конструктивных метрических пространствах.

Предложение 7.17. (а) *Компактное подмножество C конструктивного метрического пространства X эффективно компактно тогда и только тогда, когда по любому рациональному $\varepsilon > 0$ можно эффективно указать конечное покрытие множества C базисными шарами радиуса ε.*

(б) *Пусть C — эффективно компактное подмножество конструктивного метрического пространства. Тогда из любого перечислимого семейства базисных шаров, покрывающих C, можно эффективно выбрать конечное покрытие.*

Доказательство. Пусть для каждого ε мы умеем указывать конечное покрытие S_ε множества C шарами радиуса ε. Вместе с таким покрытием будем перечислять и все покрытия *гарантированно* большими шарами (это значит, что для каждого шара $B(x, \varepsilon)$ из покрытия S_ε в этом новом покрытии найдётся шар $B(y, \sigma)$ с $\sigma > \varepsilon + d(x, y)$). Компактность C гарантирует, что при этом (если сделать это для всех ε) будут перечислены все покрытия C. (В самом деле, если есть какое-то покрытие S', которое в перечисление не попадёт, то для каждого ε есть один из шаров покрытия S_ε, которые не попадает гарантированно внутрь одного из шаров покрытия S'. Применив компактность и взяв предельную точку центров этих непопадающих шаров, получим противоречие.)

Остальные утверждения доказываются совсем просто.

Следующее предложение обобщает предложение 5.5 и доказывается аналогичным рассуждением.

Предложение 7.18. *Всякое эффективно замкнутое подмножество эффективно компактного множества эффективно компактно.*

(Вспоминая определения, это утверждение можно переформулировать так: пересечение эффективно компактного множества в пространстве X с эффективно замкнутым подмножеством пространства X является эффективно компактным.)

Как и раньше, верно и обратное: всякое эффективно компактное подмножество конструктивного метрического пространства эффективно замкнуто. В самом деле, мы можем рассмотреть всевозможные покрытия этого множества базисными шарами, а также шары, заведомо (по соотношению расстояния и радиусов) не пересекающиеся с множествами покрытия. Все такие шары вместе в объединении дадут дополнение нашего эффективно компактного множества.

Образ компактного пространства при непрерывном отображении компактен. Это утверждение имеет эффективный аналог (с аналогичным доказательством):

Предложение 7.19. *Пусть C — эффективно компактное подмножество конструктивного метрического пространства X, а f — вычислимое на C отображение C в другое конструктивное метрическое пространство. Тогда $f(C)$ эффективно компактно.*

Утверждение о том, что полунепрерывная снизу функция на компактном множестве достигает минимума, тоже имеет вычислимый аналог (мы приведём сразу параметрический вариант):

Предложение 7.20 (минимум с параметром)**.** *Пусть Y, Z — конструктивные метрические пространства, $f\colon Y \times Z \to [0, \infty]$ — перечислимая снизу функция, а C — эффективно компактное подмножество $Y \times Z$. Тогда функция*

$$g(y) = \inf_{\{z \mid (y,z) \in C\}} f(y, z)$$

(*где* inf *можно заменить на* min *в силу компактности, только надо считать, что минимум пустого множества равен* $+\infty$) *перечислима снизу.*

Вместо эффективной компактности C *достаточно предполагать, что проекция* $C_Y = \{y \mid \exists z\,(y,z) \in C\}$ *эффективно замкнута и покрыта вычислимой последовательностью шаров* B_k*, для которых* $(\overline{B}_k \times Z) \cap C$ *эффективно компактно равномерно по* k*.*

Последнее условие заведомо выполняется, если Y эффективно локально компактно, а Z эффективно компактно.

Доказательство. Для начала воспроизведём классическое доказательство полунепрерывности снизу. Нам надо проверить, что для любого r множество $\{y \mid r < g(y)\}$ открыто. Это множество можно представить в виде объединения, заметив, что условие $r < g(y)$ равносильно условию

$$(\exists r' > r)\,\forall z\,\left[(y,z) \in C \Rightarrow f(y,z) > r'\right],$$

и достаточно проверить, что множество

$$U = \left\{y \mid \forall z\,\left[(y,z) \in C \Rightarrow f(y,z) > r'\right]\right\}$$

открыто. Множество U можно представить в виде

$$U = (Y \setminus C_Y) \cup \bigcup_k (B_k \cap U).$$

Множество $Y \setminus C_Y$ открыто по условию, поэтому достаточно показать, что каждое из множеств $B_k \cap U$ открыто. Положим $F_k = \overline{B}_k \times Z$; по предположению, $F_k \cap C$ компактно. Условие $f(y,z) > r'$ в силу полунепрерывности задаёт некоторое открытое множество пар V, следовательно, $F_k \cap C \setminus V$ — замкнутое подмножество компактного множества и потому компактно. Поэтому его проекция

$$\{y \in \overline{B}_k \mid \exists z\,(y,z) \in F_k \cap C \setminus V\}$$

является непрерывным образом компактного множества и потому компактна (тем самым замкнута), а её дополнение в B_k (то есть $B_k \cap U$) открыто.

Теперь надо перевести это рассуждение на эффективный язык. Прежде всего заметим, что можно ограничиться рациональными r и r'. Затем надо заметить, что множество V эффективно открыто (равномерно по r'), а множества $F_k \cap C \setminus V$ равномерно эффективно замкнуты (и будучи подмножествами эффективно компактного множества, эффективно компактны). Поэтому их проекции (как образы) эффективно компактны и эффективно замкнуты, а дополнения эффективно открыты.

Следствием этого предложения является такая лемма:

Лемма 7.21. *Пусть* X, Z, Z' *— метрические пространства, причём* X *локально компактно, а* Z *компактно. Пусть* $f\colon Z \to Z'$ *— непрерывная функция, множество значений которой совпадает с* Z'*, а* $t\colon X \times Z \to [0, +\infty]$ *— полунепрерывная снизу функция. Тогда функция* $t_f\colon X \times Z' \to [0, +\infty]$*, определённая формулой*

$$t_f(x, z') = \inf_{\{z \mid f(z) = z'\}} t(x, z),$$

является полунепрерывной снизу.

Если X*,* Z*,* Z' *— конструктивные метрические пространства,* X *эффективно локально компактно,* Z *эффективно компактно, функция* f *вычислима, а функция* t *перечислима снизу, то функция* t_f *перечислима снизу.*

Доказательство. Приведём рассуждение сразу для эффективного варианта. Применим предыдущее предложение с $Y = X \times Z'$ и

$$C = X \times \{(f(z), z) \mid z \in Z\}.$$

Тогда

$$t_f(x, z') = \inf_{(x,z',z)\in C} t(x, z).$$

Множество Y эффективно локально компактно как произведение эффективно локально компактного и эффективно компактного множеств. Проекция C на Y совпадает со всем Y и потому замкнута. Поэтому можно применить предыдущее предложение.

7.2 Меры на конструктивном метрическом пространстве

В метрическом пространстве выделяются борелевские множества (минимальная σ-алгебра, содержащая открытые множества), и можно говорить о мерах на борелевских множествах. Они обладают следующим свойством *регулярности*:

Предложение 7.22 (регулярность)**.** *Пусть P — мера на полном сепарабельном метрическом пространстве. Тогда любое измеримое множество A можно приблизить большими открытыми множествами:*

$$P(A) = \inf_{G\supseteq A} P(G),$$

где G открыто.

На мерах можно ввести расстояние:

Определение 7.23 (расстояние Прохорова)**.** *Определим расстояние от точки x до множества A в метрическом пространстве как $d(x, A) = \inf_{y\in A} d(x, y)$. Определим ε-окрестность множества A как $A^\varepsilon = \{x \mid d(x, A) < \varepsilon\}$.*

Расстояние Прохорова $\rho(P, Q)$ между двумя мерами P и Q определяется как точная нижняя грань тех $\varepsilon > 0$, для которых $P(A) \leqslant Q(A^\varepsilon) + \varepsilon$, а также $Q(A) \leqslant P(A^\varepsilon) + \varepsilon$, при всех борелевских A.

Известно, что определённая таким образом функция расстояния действительно является метрикой; тем самым на множестве всех вероятностных мер в метрическом пространстве возникает структура метрического пространства. Есть и другие способы ввести расстояние на пространстве мер, дающие эквивалентные метрики (в том смысле, что тождественное отображение со сменой метрики равномерно непрерывно в обе стороны).

Определение 7.24 (пространство мер)**.** *Пусть $\mathbf{X}$ — конструктивное метрическое пространство. Введём в пространстве мер на нём структуру конструктивного метрического пространства $\mathbf{M} = \mathcal{M}(\mathbf{X})$. В качестве счётного всюду плотного множества $D_{\mathbf{M}}$ возьмём множество мер, сосредоточенных на некотором конечном подмножестве множества $D_{\mathbf{X}}$, и принимающих рациональные значения. Такие меры имеют естественное описание как конструктивные объекты, нужно указать номера элементов этого конечного подмножества и их меры. Таким образом мы получаем нумерацию $\alpha_{\mathbf{M}}$ точек в $D_{\mathbf{M}}$.*

Вычислимые точки этого конструктивного метрического пространства называют вычислимыми *мерами на $\mathbf{X}$.*

Определённое таким образом понятие вычислимости является обобщением введённого нами ранее понятия вычислимой меры на канторовском пространстве (а также естественного понятия вычислимой меры на бэровском пространстве последовательностей натуральных чисел).

Имеет место аналог предложения 4.17: интеграл $\int f(\omega, P)\, dP(\omega)$ базисной функции f по мере P является вычислимой функцией от P и f.

Вот ещё один родственный результат:

Предложение 7.25. *Пусть f — ограниченная эффективно равномерно непрерывная функция. Тогда её интеграл по мере P, рассматриваемый как функция от меры P, является эффективно равномерно непрерывной функцией.*

Доказательство. Без ограничения общности можно считать, что f неотрицательна (добавим константу). Пусть меры P и P' близки. Тогда $P'(A) \leqslant P(A_\varepsilon) + \varepsilon$, где A_ε обозначает ε-окрестность множества A. Тогда

$$\int f\, dP' \leqslant \int f_\varepsilon\, dP + \varepsilon,$$

где $f_\varepsilon(x)$ есть точная верхняя грань f на ε-окрестности точки x. (Интеграл неотрицательной функции g определяется мерами множеств $G_t = \{x \mid g(x) \geqslant t\}$, согласно теореме Фубини об изменении порядка интегрирования эту меру как функцию от t надо проинтегрировать по t. При этом если $f(x) \geqslant t$, то $f_\varepsilon(x) \geqslant t$ в ε-окрестности точки x.) Остаётся воспользоваться эффективной равномерной непрерывностью функции f, чтобы узнать, с какой точностью надо задавать меру, чтобы получить данную точность в интеграле.

С другой стороны, мера $P(B)$ базисного шара не обязана быть вычислимой, но она перечислима снизу (равномерно по B). Как показано в [12], это свойство (равномерная перечислимость снизу) также является критерием вычислимости меры P.

Известно, что для компактного сепарабельного метрического пространства X пространство мер на X с описанной метрикой также компактно. Это утверждение имеет конструктивный вариант, который доказывается стандартным образом:

Предложение 7.26. *Если конструктивное метрическое пространство $\mathbf{X}$ эффективно компактно, то и пространство $\mathcal{M}(X)$ вероятностных мер на X эффективно компактно.*

Для пространства Ω это упоминалось в предложении 5.5.

Примеры 7.27.

Бесконечное дискретное метрическое пространство ($\mathbb{N}$) не компактно, и множество мер на нём (которое можно отождествить с множеством функций $P\colon \mathbb{N} \to [0,1]$, для которых $\sum_i P(i) = 1$) тоже не компактно. С другой стороны, если перейти к полумерам, заменив условие на $\sum_i P(i) \leqslant 1$, то в естественной метрике получится компактное пространство. В самом деле, в отличие от равенства неравенства достаточно проверять для конечных сумм, и каждое неравенство задаёт замкнутое множество в компактном произведении $[0,1]^{\mathbb{N}}$, так что пересечение таких множеств тоже будет компактным. Легко понять, что оно будет и эффективно компактным.

Переход к полумере соответствует компактификации пространства $\mathbb{N}$: недостающая часть суммы ряда переносится на бесконечную точку.

7.3 Случайность в метрическом пространстве

В пространстве Ω мы определяли

- случайность относительно вычислимых мер (в смысле Мартин-Лёфа; см. определение 2.9 на языке тестов);
- равномерную случайность относительно произвольных мер (тест является функцией последовательности и меры, определение 5.2);
- случайность относительно эффективно компактного класса мер, определение 5.22;
- слепую (безоракульную) случайность, определение 5.37.

Все эти понятия с небольшими изменениями переносятся на произвольное конструктивное метрическое пространство. В этом разделе мы обсудим эти обобщения и их свойства, а затем более подробно рассмотрим случайность относительно ортогональных классов мер.

Для вычислимых мер тест определяется как перечислимая снизу функция на метрическом пространстве, интеграл от которой не превосходит 1. Среди таких тестов существует максимальный с точностью до константы. Как и раньше, это доказывается с помощью усечения: мы перечисляем все перечислимые снизу функции, принудительно превращая их в тесты или почти тесты, и затем складываем их с коэффициентами, образующими сходящийся ряд.

Это делается как и раньше, при этом мы рассматриваем перечислимые снизу функции как возрастающие пределы базисных. Важно, что интеграл от базисной функции по вычислимой мере вычислим. Более того, интеграл от базисной функции по произвольной мере вычислимо зависит от этой функции и от этой меры. Для базисных точек в пространстве мер это ясно, далее надо воспользоваться предложением 7.25.

Легко обобщить на случай конструктивных метрических пространств и понятие равномерного теста (см. определение 5.2 для случая канторовского пространства). Такой тест представляет собой перечислимую снизу функцию двух аргументов $t(x, P)$, где x — точка нашего метрического пространства, а P — мера на этом пространстве. Условие на интеграл, как и раньше, имеет вид $\int t(x, P)\, dP(x) \leqslant 1$.

Как и раньше, существует универсальный тест, и это можно доказать с помощью техники усечения:

Теорема 7.28 (усечение в метрических пространствах)**.** *Пусть $u(x, P)$ — перечислимая снизу функция, первый аргумент которой — точка конструктивного метрического пространства, а второй — мера на этом пространстве. Тогда существует равномерный тест $t(x, P)$, для которого $u(x, Q) \leqslant 2t(x, Q)$ при всех Q, для которых функция $u_Q\colon x \mapsto u(x, Q)$ является тестом по мере Q, то есть $\int u(x, Q)\, dQ(x) \leqslant 1$.*

Доказательство повторяет рассуждение из теоремы 5.7, при этом используется тот факт, что для базисной функции $b(x, P)$ на произведении пространств интеграл $\int b(x, P)\, dP(x)$ является вычислимой (непрерывной) функцией от P (что доказывается аналогично приведённому нами рассуждению про вычислимость интеграла).

Универсальный тест мы будем обозначать $\mathbf{t}(x, P)$. Вообще-то для каждого конструктивного метрического пространства он свой, но обычно понятно, какое пространство имеется в виду, так что в обозначение оно не входит.

Помимо существования универсального теста, из возможности усечения следует возможность "униформизации" в следующем смысле. Определим тест относительно меры P (не обязательно вычислимой) на конструктивном метрическом пространстве X:

Определение 7.29. *Пусть* $\mathbf{X} = (X, d, D, \alpha)$ *— конструктивное метрическое пространство, а* P *— мера на нём. Назовём* P-тестом случайности *функцию* $f\colon X \to [0, +\infty]$*, если она перечислима снизу относительно* P *и если* $\int f(x)\, dP(x) \leqslant 1$*.*

В этом определении фигурирует только мера P. Ясно, что из равномерного теста можно получить тест относительно P. Оказывается, что (с точностью до константы) так получаются все тесты относительно P:

Теорема 7.30 (униформизация)**.** *Пусть* P *— некоторая мера на конструктивном метрическом пространстве* X*, и дан некоторый* P*-тест* $t_P(x)$*. Тогда существует равномерный тест* $t'(\cdot, \cdot)$*, для которого* $t_P(x) \leqslant 2t'(x, P)$*.*

Доказательство. Из определения перечислимой снизу относительно P функции сразу же следует, что она является сужением некоторой перечислимой снизу функции двух аргументов. Остаётся применить предыдущую теорему к этому продолжению.

Многие результаты (например, теорема 5.36) обобщаются на произвольные метрические пространства. Вот ещё один пример такого обобщения (равномерный вариант так называемой "случайности по Курцу", Kurtz randomness):

Предложение 7.31. *Пусть* X *— конструктивное метрическое пространство, а* S *— эффективно открытое подмножество пространства* $X \times \mathcal{M}(X)$*. Если множество* $S_P = \{x \mid (x, P) \in S\}$ *имеет* P*-меру* 1 *для некоторой меры* $P \in \mathcal{M}(X)$*, то* S_P *содержит все равномерно* P*-случайные точки.*

Доказательство. Характеристическая функция $1_S(x, P)$ множества S, равная единице внутри множества и нулю снаружи, перечислима снизу и потому есть предел вычислимой возрастающей последовательности базовых функций $g_n(x, P)$ с $0 \leqslant g_n(x, P) \leqslant 1$. Последовательность функций $G_n\colon P \mapsto \int g_n(x, P)\, dP(x)$ представляет собой неубывающую последовательность непрерывных вычислимых (равномерно по n) функций. По теореме о монотонной сходимости значения $G_n(P)$ стремятся к единице для тех мер P, для которых $P(S_P) = 1$. Определим для каждой меры P числа $n_k(P)$ как минимальные значения n, для которых $G_n(P) > 1 - 2^{-k}$. Эти числа перечислимы сверху как функции от P (в естественном смысле; заметим, что для мер P, при которых $P(S_P) < 1$, некоторые из $n_k(P)$ бесконечны). Соответственно функции $1 - g_{n_k(P)}(x, P)$ как функции от x и P (такую функцию мы считаем нулём при бесконечном $n_k(P)$, независимо от x) перечислимы снизу, равномерно по k. Положим теперь $t(x, P) = \sum_{k>0}(1 - g_{n_k(P)}(x, P)$. Эта функция является равномерным тестом, поскольку при данном P её k-е слагаемое равно нулю, если $n_k(P)$ бесконечно, и имеет интеграл по мере P не больше 2^{-k} при конечном $n_k(P)$.

В условии теоремы говорится о мере P, для которой $P(S_P) = 1$. Тогда все $n_k(P)$ конечны, а $g_{n_k(P)}(x, P) = 0$ для любого x вне S_P. Поэтому все слагаемые в сумме, образующей тест, равны единице, и x не является равномерно P-случайной точкой. Следовательно, S_P включает в себя все равномерно P-случайные точки.

7.4 Априорная вероятность с оракулом

В разделе 5.2 мы определили априорную вероятность с условием, роль которого играла мера на Ω. Теперь, введя понятие конструктивного метрического пространства, мы можем заметить, что это определение естественно обобщается на любое пространство $\mathbf{X}$: мы рассматриваем неотрицательные перечислимые снизу функции $m\colon \mathbb{N} \times X \to [0, +\infty]$, для которых $\sum_i m(i, x) \leqslant 1$ при любом $x \in X$.

Среди таких функций существует максимальная с точностью до константы. Это доказывается методом усечения: мы не будем повторять рассуждение подробно, отметим лишь, что перечислимую снизу функцию $m(i,x)$ можно получить как сумму ряда из базисных функций, каждая из которых отлична от нуля только для одного i.

Максимальную из таких функций будем называть *априорной вероятностью с условием* x и обозначать $\mathbf{m}(i|x)$.

Мы считали первый аргумент натуральным числом, но это не существенно: можно рассматривать слова (или любые другие дискретные конструктивные объекты). Частным случаем этого определения является определение априорной вероятности относительно меры (раздел 5.2), а также стандартные понятия априорной вероятности с оракулом (что соответствует $\mathbf{X} = \Omega$) и условной априорной вероятности (что соответствует $\mathbf{X} = \mathbb{N}$).

По аналогии с теоремой Дея–Миллера, можно выразить априорную вероятность с условием в произвольном эффективно компактном метрическом пространстве $\mathbf{X}$ через априорную вероятность с оракулом.

Предложение 7.32. *Пусть $F\colon \Omega \to X$ — вычислимое отображение, образом которого является всё пространство X. Тогда*

$$\mathbf{m}(i|x) \mathrel{\dot=} \min_{\{\pi|F(\pi)=x\}} \mathbf{m}(i|\pi).$$

Доказательство. Рассуждаем как в доказательстве теоремы 5.36. Функция $(i,\pi) \mapsto \mathbf{m}(i|F(\pi))$ является перечислимой снизу на $\mathbb{N}\times\Omega$, откуда получается $\dot<$-неравенство.

Чтобы получить обратное неравенство, мы пользуемся 7.21 и замечаем, что функция в правой части корректно определена (минимум достигается) и перечислима снизу.

Заметим, что априорная вероятность (с оракулом) в правой части предложения 7.32 может быть выражена через префиксную сложность (с оракулом). Для случая условий в метрических пространствах не ясно, как определять префиксную сложность с таким условием (можно говорить о функциях с перечислимым относительно точки x графиком, но неясно, как строить универсальную). Можно формально определить $KP(i|x)$ как $\max_{\{\pi|F(\pi)=x\}} KP(i|\pi)$, тогда $KP(i|x) \stackrel{+}{=} -\log \mathbf{m}(i|x)$, но вряд ли это можно считать удовлетворительным определением префиксной сложности (скажем, обычные рассуждения, где используется самоограниченность программы, при таком определении уже не применимы, хотя многие результаты остаются верными; например, формулу $KP(i,j|x) \stackrel{+}{<} KP(i|x) + KP(j|x)$ можно доказать, не приписывая друг к другу самоограниченные программы, а рассуждая с вероятностями) — честнее просто говорить о логарифме априорной вероятности.

Замечание 7.33. *Аналогичным образом можно добавлять точки конструктивных метрических пространств в качестве условий и в другие наши определения. Например, можно рассматривать равномерные тесты на Ω с условиями в произвольном конструктивном метрическом пространстве X: это будут перечислимые снизу функции $t(\omega,P,x)$, для которых $\int t(\omega,P,x)\,dP(\omega) \leqslant 1$ при всех x. Можно также фиксировать вычислимую меру P, например, равномерную, и определить тесты относительно этой меры с условиями в X.*

8 Классы ортогональных мер

Как и в случае пространства Ω, для мер в произвольном метрическом пространстве можно определить понятие эффективно компактного класса и универсального теста случайности

относительно этого класса. Сохраняется (с тем же доказательством) и формула для универсального теста относительно класса (теорема 5.23).

Класс бернуллиевых мер (как и многие другие часто используемые классы) обладает важным свойством: случайная по одной из мер этого класса последовательность однозначно определяет меру, по которой она случайна. Именно это обстоятельство по существу было использовано в теореме 5.41. В этом разделе мы рассмотрим тот же вопрос в более общей ситуации, когда речь идёт о мерах на конструктивном метрическом пространстве $\mathbf{X} = (X, d, D, \alpha)$.

Это обобщение включает в себя естественные примеры: конечные и бесконечные марковские цепи, стационарные эргодические процессы (см. ниже).

Сначала приведём определение ортогональности мер (которое можно считать классическим аналогом эффективной ортогональности (определение 5.40).

Определение 8.1. *Пусть P, Q — две меры на $(X, \mathcal{A})$, где X — некоторое пространство, а $\mathcal{A}$ — некоторая σ-алгебра подмножеств X. Говорят, что меры P и Q* ортогональны, *если пространство можно разбить на два непересекающихся множества U и V из $\mathcal{A}$, для которых $P(V) = Q(U) = 0$.*

Говорят, что класс $\mathcal{C}$ является ортогональным, *если существует измеримая функция $\varphi\colon X \to \mathcal{C}$, для которой $P(\varphi^{-1}(P)) = 1$ для любой меры $P \in \mathcal{C}$.*

Когда мы говорим об измеримости функции φ, имеется в виду, что в метрическом пространстве $\mathcal{M}(X)$ определены борелевские множества.

Примеры 8.2.

1. В ортогональном классе мер любые две (различные) меры P и Q ортогональны. В самом деле, множества $\{P\}$ и $\{Q\}$ борелевские (замкнутые), и потому их прообразы измеримы (и, очевидно, не пересекаются). Обратное утверждение неверно: класс $\mathcal{C}$ попарно ортогональных мер не обязан быть ортогональным, даже если он эффективно компактен. Пусть λ — равномерная мера на отрезке $[0, 1]$. Для каждой точки $x \in [0, 1]$ рассмотрим меру δ_x, сосредоточенную в точке x. Тогда класс $\{\lambda\} \cup \{\delta_x \mid x \in [0, 1]\}$ эффективно компактен, и его элементы попарно ортогональны. Однако он не является ортогональным классом: условие ортогональности требует, чтобы мера δ_x была значением φ на x, а тогда $\varphi^{-1}(\lambda)$ будет пустым.

2. Пусть P и Q — две меры. Если $\mathrm{Randoms}(P)$ и $\mathrm{Randoms}(Q)$ не пересекаются, то эти меры ортогональны (в качестве U и V можно взять, скажем, случайные и неслучайные по мере P последовательности). Обратное, вообще говоря, неверно: меры λ и δ_x ортогональны, но множества случайных последовательностей пересекаются, если x взять случайным по мере λ.

В качестве примера рассмотрим стационарные эргодические процессы.

Определение 8.3. *Рассмотрим на пространстве Ω бесконечных двоичных последовательностей преобразование левого сдвига:*

$$T\colon \omega(1)\omega(2)\ldots \mapsto \omega(2)\omega(3)\ldots.$$

Распределение вероятностей P на Ω назовём стационарным, *если*

$$P(T^{-1}(A)) = P(A)$$

для любого борелевского множества A. Легко проверить, что это эквивалентно требованию

$$P(x) = P(0x) + P(1x)$$

для всех слов x.

Борелевское множество $A \subset \Omega$ назовём инвариантным *относительно сдвига, если $T(A) \subset A$. Например, множество последовательностей, в которых частота единиц стремится к $1/2$, является инвариантным. Стационарное распределение называется* эргодическим, *если любое инвариантное борелевское множество имеет меру 0 или 1.*

Вот пример стационарного процесса.

Пример 8.4. *Пусть $Z_1, Z_2, \ldots$ — последовательность независимых одинаково распределённых случайных величин, принимающих значения 0 и 1 с вероятностями соответственно 0.9 и 0.1. Определим $X_0, X_1, X_2, \ldots$ так: X_0 принимает значения $0, 1, 2$ с равными вероятностями и независима от всех Z_i, $X_n = X_0 + \sum_{i=1^n} Z_n \bmod 3$. Наконец, пусть $Y_n = 0$ при $X_n = 0$ и $Y_n = 1$ при $X_n \neq 0$. Легко видеть, что процесс $Y_0, Y_1, \ldots$ является стационарными; можно доказать, что он эргодический. Поскольку он является функцией марковской цепи X_n, его называют* скрытой марковской цепью (*hidden Markov chain*).

Следующее утверждение является следствием эргодической теоремы Биркгофа о поточечной сходимости. Через g_x будем обозначать индикатор события $x \sqsubseteq \omega$, то есть $g_x(\omega)$ равно 1 при $x \sqsubseteq \omega$ и 0 в противном случае.

Предложение 8.5. *Пусть P — стационарное распределение вероятностей на пространстве Ω.*

(а) *Для почти всех по мере P последовательностей ω последовательность*

$$A_{x,n}(\omega) = \tfrac{1}{n}(g_x(\omega) + g_x(T\omega) + \ldots + g_x(T^{n-1}\omega))$$

сходится.

(б) *Для эргодического процесса этот предел равен $P(x)$.*

(Для неэргодических процессов предел может зависеть от ω.)

Общая теорема Биркгофа касается произвольных пространств и сохраняющих меру преобразований, а в качестве g_x можно взять произвольную интегрируемую функцию, и гарантировать поточечную сходимость (в эргодическом случае — к математическому ожиданию).

Утверждение (б) показывает, что класс $\mathcal{C}$ эргодических мер является ортогональным. В самом деле, будем называть последовательность ω "стабильной", если для неё существуют пределы из п. (а) при любом x. Легко понять, что в этом случае эти пределы задают некоторую меру Q_ω. Определим функцию $\varphi\colon \Omega \to \mathcal{C}$, положив $\varphi(\omega) = Q_\omega$, если мера Q_ω является эргодической, и выбрав в качестве значения произвольную эргодическую меру, если Q_ω не является эргодической или ω не является стабильной. Пункт (б) гарантирует, что $P(\varphi^{-1}(P)) = 1$ для любой эргодической меры P.

Тут используется, что множество стабильных последовательностей борелевское. Заметим, что класс эргодических мер незамкнут, но это в определении не предполагается.

Мы видели (пример 8.2.2), что две меры могут быть ортогональны, но иметь общие случайные последовательности. Однако для вычислимых мер, как мы сейчас докажем, это невозможно.

Будем говорить, что две меры эффективно ортогональны, если классы равномерно случайных относительно них последовательностей не пересекаются. (Это позволяет переформулировать определение 5.40 так: класс мер эффективно ортогонален, если любые две меры в этом классе эффективно ортогональны.)

Теорема 8.6. *Две вычислимые меры на конструктивном метрическом пространстве ортогональны тогда и только тогда, когда они эффективно ортогональны.*

Доказательство. Как мы уже говорили, в одну сторону это верно для любых мер. Докажем обратное утверждение. Пусть даны две вычислимые ортогональные меры P, Q; по определению ортогональности существует измеримое множество A, для которого $P(A) = 1$, $Q(A) = 0$. В силу регулярности (предложение 7.22) найдётся последовательность открытых множеств G_n, содержащих A, для которых $Q(G_n) < 2^{-n}$. Поскольку G_n содержит A, то $P(G_n) = 1$. Множество G_n открыто, поэтому найдётся конечное объединение $H_n \subset G_n$ базисных шаров, для которого $P(H_n) > 1-2^{-n}$; для H_n мера Q тоже меньше 2^{-n}. Перебором можно найти вычислимую последовательность множеств H_n с такими свойствами (P-мера большая, Q-мера маленькая).

Рассмотрим теперь $\limsup H_n$, то есть множество $\bigcap_m U_m$, где $U_m = \bigcup_{n>m} H_n$. Согласно предложению 7.31, каждое из множеств U_m, а значит, и их пересечение, содержит все P-случайные точки. С другой стороны, множества U_n образуют тест в смысле Мартин-Лёфа относительно меры Q, поэтому это пересечение не содержит ни одной Q-случайной точки.

Мы видели, что эргодические меры образуют ортогональный класс. Более детальный анализ показывает, что они эффективно ортогональны.

Теорема 8.7. *Эргодические меры на канторовском пространстве образуют эффективно ортогональный класс.*

Доказательство. В статье [28] приведено доказательство эффективной эргодической теоремы, из которого следует, что

(а) Равномерно случайные последовательности по стационарной мере стабильны (в том смысле, что для них существует указанный выше предел частот);

(б) Для равномерно случайных по эргодической мере последовательностей этот предел совпадает с мерой $P(x)$.

Опишем коротко схему доказательства. Для любых рациональных чисел $0 < \alpha < \beta$ рассмотрим перечислимую снизу функцию $\omega \mapsto \sigma(\omega, \alpha, \beta)$, которая считает, сколько раз величина $A_{x,n}(\omega)$ с ростом n пересекла промежуток (α, β) слева направо (была меньше α и стала больше β). Затем можно доказать, что $(1 + \alpha^{-1})(\beta - \alpha) \int \sigma(\omega, \alpha, \beta)\, dP(\omega) \leqslant 1$, то есть функция $(1 + \alpha^{-1})(\beta - \alpha)\sigma(\alpha, \beta)$ является ограниченным в среднем тестом. Следовательно, для равномерно случайных (и даже для безоракульно случайных) последовательностей число таких пересечений интервала конечно.

Чтобы доказать (б), если (а) уже гарантировано, достаточно для каждого x установить, что

$$\liminf_n A_{x,n}(\omega) \leqslant P(x) \leqslant \limsup_n A_{x,n}(\omega)$$

для всех случайных ω. Рассмотрим, например, первое неравенство (второе аналогично). Достаточно показать для любых k и m, что

$$\inf_{n \geqslant m} A_{x,n}(\omega) \leqslant P(x) + 2^{-k}$$

для случайной ω. Множество

$$S_{x,k,m} = \{(\omega, P) \mid (\exists n \geqslant m) A_{x,i}(\omega) < P(x) + 2^{-k}\}$$

является эффективно открытым, и по теореме Биркгофа множество $S_{x,k,m}(P) = \{x \mid (x, P) \in S_{x,k,m}\}$ имеет P-меру 1, если мера P эргодическая. Предложение 7.31 гарантирует, что множество $S_{x,k,m}(P)$ содержит все равномерно P-случайные точки.

Другой подход к доказательству состоит в том, чтобы установить сходимость частот к $P(x)$ для случайных по Мартин-Лёфу последовательностей относительно вычислимых эргодических мер (при этом доказательство должно выдерживать релятивизацию), не используя явного теста, как это сделано в [2]. После этого можно сослаться на теорему 5.36 и получить сходимость для равномерно случайных последовательностей.

Для работы с ортогональными классами мер полезно понятие сепаратора. В этом определении, говоря об измеримости функций, мы имеем в виду их измеримость по Борелю (прообраз борелевского множества является борелевским); меры мы тоже считаем определёнными на борелевских множествах.

Определение 8.8 (функция-сепаратор)**.** *Пусть $\mathcal{C}$ — класс мер на конструктивном метрическом пространстве $\mathbf{X}$. Измеримую функцию*

$$s\colon X \times \mathcal{M}(\mathbf{X}) \to [0, +\infty]$$

назовём сепаратором *для класса $\mathcal{C}$, если $\int s(x,P)\,dP(x) \leqslant 1$ для любой меры P, а для любых двух различных мер $P, Q \in \mathcal{C}$ и для любой точки x хотя бы одно из значений $s(x,P)$ и $s(x,Q)$ бесконечно.*

Сепаратор называется тестом-сепаратором*, если он перечислим снизу как функция x и P.*

В определении сепаратора мы требуем $\int s(x,P)\,dP(x) \leqslant 1$ для всех мер (а не только для мер из класса P), но это не очень существенно, поскольку к тест-сепаратору можно применить усечение.

Следующая теорема связывает понятие ортогонального класса мер с сепараторами, а также устанавливает, что для случая эффективно ортогонального класса каждая мера может быть восстановлена по случайной (по этой мере) последовательности.

Теорема 8.9. *Пусть $\mathcal{C}$ — класс мер на конструктивном метрическом пространстве.*

(а) *Если борелевский класс мер $\mathcal{C}$ ортогонален, то для него существует сепаратор.*

(б) *Класс $\mathcal{C}$ является эффективно ортогональным тогда и только тогда, когда существует тест-сепаратор.*

(Что касается обратного к (а) утверждения, то авторы не знают, верно ли оно.)

Доказательство. Докажем сначала (а). Пусть $\varphi(x)$ — функция, которая для каждого $x \in X$ указывает меру в соответствии с определением ортогональности. По предположению эта функция борелевская, поэтому (см. [14]) её график является борелевским множеством. Положим $s(x,P) = 1$ при $P \notin \mathcal{C}$, а также при $P \in \mathcal{C}$ и $\varphi(x) = P$, и положим $s(x,P) = \infty$ при $P \in \mathcal{C}$ и $\varphi(x) \neq P$.

Докажем теперь утверждение (б). Если класс $\mathcal{C}$ эффективно ортогонален, то универсальный равномерный тест и будет тест-сепаратором для класса $\mathcal{C}$. С другой стороны, пусть имеется тест-сепаратор для класса $\mathcal{C}$. Пусть P и Q — две различные меры для класса $\mathcal{C}$, и точка x равномерно случайна по обеим мерам. Поскольку s является равномерным тестом случайности, то $s(x,P)$ и $s(x,Q)$ конечны, что противоречит определению теста-сепаратора.

Следующий результат менее ожидаем; он показывает, что для случая эффективно компактного класса мер из существования полунепрерывного снизу сепаратора следует существование и перечислимого снизу сепаратора (то есть теста-сепаратора).

Теорема 8.10. *Пусть для эффективно компактного класса мер существует полунепрерывный снизу сепаратор $s(x,P)$. Тогда этот класс эффективно ортогонален.*

Доказательство. Пусть $\mathcal{C}$ — эффективно компактный класс мер на конструктивном метрическом пространстве. Мы должны показать, что в предположениях теоремы для любых двух различных мер $P_1, P_2 \in \mathcal{C}$ множества случайных последовательностей не пересекаются:

$$\mathrm{Randoms}(P_1) \cap \mathrm{Randoms}(P_2) = \varnothing.$$

Возьмём в пространстве мер два непересекающихся базисных замкнутых шара B_1 и B_2, содержащих меры P_1 и P_2, и рассмотрим классы мер $\mathcal{C}_1 = \mathcal{C} \cap B_1$ и $\mathcal{C}_2 = \mathcal{C} \cap B_2$. Это непересекающиеся эффективно компактные классы мер, содержащие P_1 и P_2. Рассмотрим теперь функции

$$t_i(x) = \inf_{P \in \mathcal{C}_i} s(x, P).$$

Для любого x хотя бы одно из значений $t_1(x)$ и $t_2(x)$ бесконечно. Функции t_1 и t_2 полунепрерывны снизу (первая часть доказательства предложения 7.20) и являются $\mathcal{C}_1$- и $\mathcal{C}_2$-тестами.

Теперь мы можем применить рассуждение, аналогичное доказательству предложения 7.31. Пусть $k > 1$ — целое число. Рассмотрим открытое множество $S_k = \{x \mid t_1(x) > 2^k\}$. Поскольку t_1 является $\mathcal{C}_1$-тестом, то $P(S_k) < 2^{-k}$ для всех $P \in \mathcal{C}_1$. С другой стороны, поскольку для каждого x одно из значений $t_1(x)$ и $t_2(x)$ бесконечно, то $P(S_k) = 1$ для всех $P \in \mathcal{C}_2$. Характеристическая функция 1_{S_k} множества S_k полунепрерывна снизу, и потому может быть представлена как поточечный предел неубывающей последовательности (не обязательно вычислимой!) базисных функций $g_{k,n}$. Рассуждая как в предложении 7.31, мы заключаем, что найдётся $n = n_k(P)$, при котором $\int g_{k,n}\, dP > 1 - 2^{-k}$ для любого $P \in \mathcal{C}_2$. Эффективная компактность класса $\mathcal{C}_2$ позволяет выбрать n общим для всех $P \in \mathcal{C}_2$.

Зафиксируем достигнутое: для всякого k найдётся базисная функция h_k, для которой

$$\int h_k\, dP < 2^{-k} \text{ при всех } P \in \mathcal{C}_1;$$
$$\int h_k\, dP > 1 - 2^{-k} \text{ при всех } P \in \mathcal{C}_2.$$

Такую функцию можно найти эффективно по k перебором.

Теперь можно построить перечислимую снизу функцию

$$t_1'(x) = \sum_k h_k(x).$$

Она является тестом для класса $\mathcal{C}_1$. Функция $t_2'(x) = \sum_k (1 - h_k(x))$ по аналогичным причинам будет тестом для класса $\mathcal{C}_2$. Эти тесты должны быть конечны для случайных по мерам P_1 и P_2 последовательностей, а одновременно для обоих тестов это быть не может.

Смысл введённого нами понятия теста-сепаратора можно пояснить следующим образом. Универсальный тест $\mathbf{t}(\omega, P)$ в силу эффективной ортогональности позволяет разделить последовательности, случайные по разным мерам из класса $\mathcal{C}$: глядя на последовательность ω, равномерно случайную по одной из мер этого класса (=случайную относительно класса $\mathcal{C}$), мы ищем $P \in \mathcal{C}$, для которого $\mathbf{t}(\omega, P)$ конечно. Такая мера P в классе $\mathcal{C}$ единственна (согласно определению эффективной ортогональности).

Последнее, однако, может выполняться и для неуниверсального теста, и такие тесты мы назвали тестами-сепараторами. Неуниверсальный тест менее требователен к идее случайности, и описывает её, так сказать, в первом приближении: может оказаться, что та последовательность, которую он считает случайной (на которой значение $t(\omega, P)$ конечно), более серьёзный

тест уже отбракует. (Обратное невозможно, так как универсальный тест максимален.) Важно только, чтобы уже эта предварительная грубая отбраковка позволяла разделить меры из класса $\mathcal{C}$, то есть чтобы ни одна последовательность не казалась "в первом приближении случайной" сразу по двум мерам.

Определение 8.11. *Для данного тест-сепаратора $s(x, P)$ будем называть элемент x* случайным в первом приближении относительно P, *если значение этого тест-сепаратора конечно: $s(x, P) < \infty$.*

В качестве примера рассмотрим класс бернуллиевых мер. В качестве такого "теста в первом приближении" можно вспомнить слова фон Мизеса, который самым первым свойством случайной последовательности (*коллектива*, как он говорил) называл устойчивость частот. Свойство устойчивости частот (усиленный закон больших чисел в современной терминологии) состоит в том, что $S_n(\omega)/n \to p$. Здесь $S_n(\omega)$ — количество единиц в начальном отрезке последовательности ω длины n, а p — параметр бернуллиевой меры B_p.

Можно пытаться использовать это свойство для построения сепаратора разными способами, вот несколько возможных требований:

(1) $S_n(\omega)/n \to p$ с некоторой фиксированной скоростью сходимости.

(2) $S_n(\omega)/n \to p$ без указания конкретной скорости сходимости.

(3) Можно вспомнить доказательство теоремы 8.7 для класса всех эргодических стационарных мер на Ω, и получить тест, гарантирующий сходимость всех частот $A_{x,n}(\omega)$ к соответствующим вероятностям $P(x)$.

Наиболее простое и естественное (с математической точки зрения) требование (2) не записывается в виде перечислимого снизу теста, но чтобы поправить дело, можно перейти к (1) и фиксировать скорость сходимости. Вот один из возможных вариантов. (Для простоты мы будем использовать только неравенство Чебышёва, и получим сходимость частот не на всех отрезках, а только по степеням двойки. Более аккуратная оценка позволила бы получить сходимость частот по всем начальным отрезкам.)

Неравенство Чебышёва гарантирует, что

$$B_p(\{x \in \mathbb{B}^n \colon |S_n(x) - np| > \lambda n^{1/2}(p(1-p))^{1/2}\}) \leqslant \lambda^{-2}.$$

Здесь $S_n(x)$ — частота единиц в слове длины n. Поскольку $p(1-p) \leqslant 1/4$, отсюда следует, что

$$B_p(\{x \in \mathbb{B}^n \colon |S_n(x) - np| > \lambda n^{1/2}/2\}) \leqslant \lambda^{-2}.$$

Положив, скажем, $\lambda = n^{0.1}$ (и опуская множитель $1/2$, что лишь ослабляет утверждение), получаем

$$B_p(\{x \in \mathbb{B}^n \colon |S_n(x) - np| > n^{0.6}\}) \leqslant n^{-0.2}.$$

Чтобы ряд сходился, ограничимся лишь членами вида $n = 2^k$:

$$B_p(\{x \in \mathbb{B}^{2^k} \colon |S_{2^k}(x) - 2^k p| > 2^{0.6k}\}) \leqslant 2^{-0.2k}.$$

Теперь для бесконечной последовательности ω и для $p \in [0, 1]$ положим

$$g(\omega, B_p) = \sup\{k \colon |S_{2^k}(\omega) - 2^k p| > 2^{0.6k}\}.$$

При этом

$$\int g(\omega, B_p)\, dB_p(\omega) \leqslant \sum_k k \cdot 2^{-0.2k} = c < \infty$$

и поделив на c, получаем тест. Это тест-сепаратор, так как $g(\omega, B_p) < \infty$ влечёт сходимость последовательности $2^{-k}S_{2^k}(\omega)$ к p и для двух разных p такого случиться не может.

Теорема 5.41 обобщается на случай произвольного метрического пространства (доказательство остаётся практически тем же, надо использовать базисные шары вместо начальных отрезков): для всякого эффективно компактного эффективно ортогонального класса мер и всякой меры в этом классе слепая (безоракульная) случайность равносильна равномерной случайности. В связи с этим естественно спросить, нельзя ли произвольную эргодическую меру поместить в некоторый эффективно компактный класс. Оказывается, что нет.

Теорема 8.12. *Рассмотрим стационарные (инвариантные относительно сдвига) меры на пространстве Ω. Среди них существует эргодическая мера, не содержащаяся ни в каком эффективно компактном классе стационарных эргодических мер.*

Прежде чем доказывать эту теорему, приведём некоторые вспомогательные утверждения.

Предложение 8.13. *Как эргодические, так и неэргодические меры плотны в классе стационарных мер.*

Доказательство. Для начала покажем, что всякую стационарную меру можно приблизить эргодической. Без ограничения общности можно считать, что вероятность $P(x)$ появления любого слова x по этой мере строго положительна. (Если нет, можно подмешать немного равномерной меры.) Фиксируем какое-либо n и рассмотрим значения меры $P(x)$ на строках длины не более n. Существует марковский процесс с таким же распределением вероятностей, в котором вероятность следующего бита определяется $n-1$ предыдущими битами: для любого $x \in \mathbb{B}^{n-2}$ и любых битов b, b' вероятность перехода от bx к xb' равна $P(bxb')/P(bx)$. В этом процессе все вероятности перехода положительны, и поэтому он является эргодическим. С ростом n он стремится к исходной стационарной мере.

С другой стороны, всякую стационарную меру P можно приблизить и неэргодической мерой. Очевидно, достаточно рассмотреть случай, когда сама мера P эргодическая. Тогда, согласно эргодической теореме, можно найти последовательность, в которой предельные частоты всех подслов соответствуют мере. (Почти все — в смысле этой меры — последовательности таковы.) Взяв длинный кусок этой последовательности и зациклив его, можно найти периодическую последовательность, в которой частоты слов длины не больше n отличаются от меры P не более чем на ε (для любых данных n и $\varepsilon > 0$). (При зацикливании образуются новые слова на месте склейки, но при большой длине это не играет роли.) Теперь можно рассмотреть меру, соответствующую случайным сдвигам этой последовательности (она сосредоточена на конечном множестве последовательностей — их столько, каков минимальный период). Эта мера не эргодична, но близка к P.

Нам понадобится ещё одно утверждение:

Предложение 8.14. *Множество эргодических мер образует G_δ-подмножество в метрическом пространстве всех мер на Ω.*

Доказательство. Мы можем ограничиться (замкнутым) множеством стационарных мер. Рассмотрим функцию $A_{x,n}$ на Ω, положив $A_{x,n}(\omega)$ равным доле вхождений слова x среди первых n возможных позиций (мы прикладываем x к ω, начиная с первой, второй,..., n-ой позиции и смотрим долю совпадений). Эргодическая теорема гарантирует, что при каждом x последовательность функций $A_{x,1}, A_{x,2}, \ldots$ сходится в смысле L_1 (на самом деле имеет место даже и сходимость почти всюду). При этом стационарная мера P будет эргодической тогда и только тогда, когда пределом этой последовательности будет константа $P(x)$.

В силу существования предела для стационарных мер достаточно проверять, что константа $P(x)$ является предельной точкой. Для любых x, N и ε множество $S_{x,N,\varepsilon}$ тех P, для которых существует $n \geqslant N$ с

$$\int |A_{x,n}(\omega) - P(x)|\, dP(\omega) < \varepsilon,$$

является открытым, а пересечение этих множеств по всем x, N, ε и даёт указанный выше критерий эргодичности стационарной меры.

Теперь мы можем доказать теорему 8.12.

Доказательство. Объединение всех эффективно компактных классов эргодических мер является F_σ-множеством. Предположим, что оно включает в себя все эргодические меры. Тогда множество неэргодических мер является G_δ-множеством, которое плотно в силу предложения 8.13. С другой стороны, как мы видели в предложениях 8.14 и 8.13, множество неэргодических мер также является плотным G_δ-множеством. Но пересечение этих двух множеств пусто, что противоречит теореме Бэра о категории. Теорема 8.12 доказана.

Тем не менее вопрос, с которого мы начали, остаётся открытым:

Вопрос. Существует ли эргодическая мера, для которой понятия равномерной и слепой (безразличной) случайности не совпадают?

Возвращаясь к произвольным эффективно компактным эффективно ортогональным классам, мы можем связать универсальные тесты с тестами для класса (см. теорему 5.23) и тестами-сепараторами.

Теорема 8.15. *Пусть $\mathcal{C}$ — эффективно компактный класс эффективно ортогональных мер. Пусть $\mathbf{t}_{\mathcal{C}}(x)$ — универсальный тест случайности для этого класса, а $s(x,P)$ — некоторый тест-сепаратор для $\mathcal{C}$. Тогда универсальный равномерный тест $\mathbf{t}(x,P)$ для мер этого класса можно выразить так:*

$$\mathbf{t}(x,P) \doteq \max(\mathbf{t}_{\mathcal{C}}(x), s(x,P))$$

для всех $P \in \mathcal{C}$ и для всех x.

Доказательство. Заметим прежде всего, что $t_{\mathcal{C}}(x)$ и $s(x,P)$ не превосходят универсального равномерного теста $\mathbf{t}(x,P)$ (что следует из его универсальности).

С другой стороны, покажем, что если $\mathbf{t}_{\mathcal{C}}(x)$ и $s(x,P)$ конечны, то $\mathbf{t}(x,P)$ не превосходит наибольшего из них (с точностью до константы). Конечность первого теста гарантирует, что $\min_{Q\in\mathcal{C}} \mathbf{t}(x,Q)$ конечен: этот минимум равен $\mathbf{t}_{\mathcal{C}}(x)$ с точностью до константы. Если этот минимум достигался бы на какой-то мере $Q \neq P$, то оба значения $s(x,Q)$ и $s(x,P)$ были бы конечны, что противоречит определению сепаратора. (Заметим, что мы доказали чуть более сильное утверждение, чем обещали: вместо "наибольшего из них" можно написать "первого из них, если второй конечен".)

Утверждение этой теоремы позволяет разбить проверку случайности по некоторой мере P из класса $\mathcal{C}$ на две части (указывает две возможные причины неслучайности). Во-первых, мы должны убедиться, что x случайно относительно класса $\mathcal{C}$. Например, в случае меры B_p из класса $\mathcal{B}$ бернуллиевых мер мы вначале должны убедиться, что $\mathbf{t}_{\mathcal{B}}(\omega)$ конечно. После этого мы знаем, что наша последовательность бернуллиева и достаточно какой-то простой проверки типа закона больших чисел, чтобы выяснить, по какой именно бернуллиевой мере она случайна: B_p или какой-то другой. Вторую часть можно рассматривать как аналогичную параметрическому тестированию в статистике.

С количественной точки зрения (если нас интересует не просто случайность и неслучайность, но и значение теста) вторая часть тестирования не важна: про сепаратор нам надо знать лишь, конечно или бесконечно его значение.

9 Равномерные тесты: слишком сильные требования?

9.1 Монотонность и квази-выпуклость

С интуитивной точки зрения равномерные тесты случайности (в наиболее общей форме см. определение 7.29) могут казаться слишком сильными, если мера P не вычислима. И действительно, они не обладают некоторыми интуитивно желательными свойствами, которыми обладает понятие случайности относительно вычислимых мер (в смысле Мартин-Лёфа или равномерное, для вычислимых мер это одно и то же). Одним из таких свойств является монотонность: большая (с точностью до константы) мера имеет больше случайных объектов.

Предложение 9.1. *Пусть P и Q — две вычислимые меры, а $\lambda > 0$ — рациональное число, причём $\lambda P(A) \leqslant Q(A)$ для всех A. Тогда*

$$\mathbf{m}(\lambda) \cdot \lambda \cdot \mathbf{t}(x, Q) \stackrel{\cdot}{<} \mathbf{t}(x, P).$$

Здесь $\mathbf{m}(\cdot)$ — дискретная априорная вероятность рационального числа λ; константа в $\stackrel{\cdot}{<}$ определяется сложностью пары программ, задающих P и Q.

Доказательство. Функция $\lambda\, \mathbf{t}(\cdot, Q)$ является P-тестом, так как

$$\int \lambda\, \mathbf{t}(x, Q)\, dP(x) \leqslant \int \mathbf{t}(x, Q)\, dQ(x) \leqslant 1.$$

Используя усечение, мы заключаем, что сумма

$$\sum_{\{\lambda \mid \lambda \cdot \int \mathbf{t}(x,Q)\, dP(x) < 2\}} \mathbf{m}(\lambda) \cdot \lambda \cdot \mathbf{t}(x, Q)$$

с точностью до константы является P-тестом, и потому не превосходит $\mathbf{t}(x, P)$. Тем более это верно и для всех членов этой суммы. (Упомянутая константа обратно пропорциональна $\mathbf{m}(P, Q)$.)

Интуитивная мотивировка свойства монотонности такова: если есть два устройства с внутренними датчиками случайности, генерирующие объекты с выходным распределением P и Q, и $\lambda P \leqslant Q$, то можно представить себе, что с вероятностью λ второе устройство моделирует первое, а в остальных случаях делает что-то своё. Тогда всякий объект, который с интуитивной точки зрения правдоподобен на выходе первого устройства, должен считаться правдоподобным и на выходе второго: вдруг оно-таки промоделировало первое? (Численное значение дефекта, конечно, может быть немного больше, так как мы дополнительно должны поверить, что произошло событие с вероятностью λ.)

Для равномерной случайности, увы, это свойство не выполнено: если мера Q больше, но вычислительно сложнее, то тесты случайности относительно Q могут использовать эту дополнительную информацию, чтобы сделать неслучайными некоторые объекты, которые относительно P были случайными (см. доказательство теоремы 5.39). Именно в этом причина

отличия равномерной случайности от слепой (безоракульной), для которой аналогичное свойство выполнено по очевидным причинам.

Другая ситуация, в которой у нас есть некоторая интуиция случайности — это смесь (выпуклая комбинация) мер. Представим себе два устройства с выходными мерами P и Q, и внешнюю оболочку, которая с какими-то вероятностями λ и $1-\lambda$ запускает одно из них. В целом мы получаем систему, выход которой распределён по мере $\lambda P + (1-\lambda)Q$. Про какие объекты мы готовы поверить, что они случайно получены в результате такого эксперимента? ясно, что это должны быть случайные по мере P объекты, а также случайные по мере Q объекты (при этом если коэффициент мал, то должно добавляться дополнительное удивление, но конечное). И других объектов быть не должно. Количественное уточнение этого результата (который в одну сторону следует из монотонности) даётся в следующем предложении.

Предложение 9.2. *Пусть P и Q — две вычислимые меры.*

(а) $\mathbf{m}(\lambda) \cdot \mathbf{t}(x, \lambda P + (1-\lambda)Q) \dot< \max(\mathbf{t}(x,P), \mathbf{t}(x,Q))$;

(б) $\mathbf{t}(x, \lambda P + (1-\lambda)Q) \dot> \min(\mathbf{t}(x,P), \mathbf{t}(x,Q))$.

В первом утверждении λ — рациональное число в $(0,1)$, а $\mathbf{m}(\lambda)$ — его дискретная априорная вероятность. Во втором утверждении λ может быть любым. Константы в $\dot<$ не зависит от λ (определяются сложностью пары мер P и Q).

Первое утверждение можно назвать *квази-выпуклостью* тестов случайности (с точностью до константы). Для тестов, обладающих свойством квази-выпуклости в уточнённом варианте, без умножения на константу, можно построить нейтральную перечислимую снизу полумеру (в некотором точном смысле этого слова, при надлежащем обобщении понятия теста на полумеры, см.[16, 8]).

Второе утверждение можно назвать *квази-вогнутостью*; оно показывает, что никаких новых случайных объектов относительно смеси P и Q не появляется.

Доказательство. Первое утверждение является ослаблением предложения 9.1. Если $\lambda \geqslant 1/2$, то из этого предложения следует, что $\mathbf{m}(\lambda) \cdot \mathbf{t}(x, \lambda P + (1-\lambda)Q) \dot< \mathbf{t}(x,P)$ (множитель $1/2$ можно включить в $\dot<$). При $\lambda \leqslant 1/2$ верно аналогичное неравенство $\mathbf{m}(1-\lambda) \cdot \mathbf{t}(x, \lambda P + (1-\lambda)Q) \dot< \mathbf{t}(x,Q)$; при этом $\mathbf{m}(1-\lambda) \doteq \mathbf{m}(\lambda)$.

Второе утверждение следует из того, что правая часть (как функция от x) является тестом относительно любой смеси мер P и Q, и можно воспользоваться усечением, чтобы сделать её равномерным тестом.

Легко понять, что все эти утверждения существенно используют вычислимость мер и коэффициентов в смеси. Соответствующие контрпримеры легко построить, если осознать, что смесь мер может быть как более сильным с вычислительной точки зрения оракулом, чем каждая из них (если пропорции смешивания невычислимы), так и наоборот. Например, разделим отрезок $[0,1]$ на две половины и рассмотрим две меры P и Q, равномерно распределённые на этих половинах. Их смесь с коэффициентами λ и $1-\lambda$ делает число λ заведомо неслучайным (поскольку оно может быть вычислено относительно этой меры), хотя по одной из мер оно вполне может быть случайным. (Взяв вместо P и Q их смеси, скажем, с коэффициентами $1/3$ против $2/3$ и наоборот, можно сделать λ случайным по обеим мерам.)

В этом примере смесь содержит больше информации, чем каждая из мер. Может быть и наоборот: свернём отрезок $[0,1]$ с равномерной мерой в окружность и разобьём его на две полуокружности точками p и $p+1/2$. Тогда равномерные меры на этих полуокружностях делают p вычислимым относительно них и потому неслучайным, а среднее этих мер есть равномерная мера на окружности, относительно которой p вполне может быть случайным.

Заметим, что для слепой (безоракульной) случайности мы можем безо всяких ограничений гарантировать, что множество случайных относительно смеси P и Q точек будет объединением

множеств точек, случайных относительно P и относительно Q. (В одну сторону это следует из монотонности, которую мы уже отмечали. В другую: если точка не случайна относительно P и не случайна относительно Q, то есть два теста, это доказывающих, и их минимум будет перечислимым снизу тестом, доказывающим её неслучайность относительно смеси.) Было бы интересно модифицировать понятие теста случайности, чтобы восстановить эти свойства, сохранив другие желательные свойства (скажем, существование универсального теста и тем самым понятие дефекта случайности). Некоторые предложения такого рода имеются в [16, 8, 17].

9.2 Локальность

Представим себе, что последовательность ω случайна по равномерной мере и начинается с нуля. Теперь изменим эту меру на последовательностях, начинающихся с единицы. Может оказаться, что последовательность перестанет быть случайной, так как значения меры теперь могут быть использованы как оракул (например, последовательность может стать вычислимой относительно новой меры). Но это выглядит странным, так как изменение меры происходит не в той части, где лежит наша последовательность.

Для слепой (безоракульной) случайности конкретно этот пример, как легко видеть, невозможен (тест можно принудительно обнулить на последовательностях, начинающихся с единицы), но в принципе понятие теста зависит от меры не только вдоль последовательности (не только от вероятности появления нуля и единицы после её начал).

Опять же для вычислимых мер ситуация лучше.

Предложение 9.3 (преквенциальное свойство)**.** *Пусть P и Q — две вычислимые меры на пространстве Ω бесконечных последовательностей нулей и единиц, совпадающие на всех началах некоторой последовательности ω. Тогда эта последовательность одновременно случайна или не случайна по мерам P и Q.*

Доказательство. Это немедленно следует из критерия случайности в терминах сложности начальных отрезков (теорема Левина–Шнорра) в любом из его вариантов (теорема 2.24, предложение 2.30 и следствие 2.32).

Для невычислимых мер это (как показывают примеры, аналогичные рассмотренным в предыдущем разделе) неверно.

В случае вычислимых мер в произвольном пространстве имеет место аналогичное утверждение, правда, с более сильным требованием: мы предполагаем, что две меры совпадают на любых множествах, содержащихся в некоторой окрестности последовательности ω. (В этом случае можно умножить тест на базисную функцию, не изменив его в ω и сделав нулевым вне окрестности совпадения.)

Вот ещё один способ, позволяющий получить заведомо преквенциальное определения случайности, в котором дефект случайности является функцией от самой последовательности и от мер её начальных отрезков. Для данной последовательности ω и для данной последовательности $\{q(i)\}$ действительных чисел, для которых $1 = q(0) \geqslant q(1) \geqslant q(2) \geqslant \ldots \geqslant 0$, положим

$$\mathbf{t}'(\omega, q) = \inf \mathbf{t}(\omega, P),$$

где минимум берётся по всем мерам P, для которых $P(\omega(1 : n)) = q(n)$. Соответствующие множества (для случай двоичных последовательностей) эффективно компактны, так что этот минимум будет перечислимой снизу функцией от ω и последовательности q. Если для последовательности ω и мер $q(i)$ её начальных отрезков значение $\mathbf{t}'(\omega, q)$ конечно, то последовательность ω можно назвать *преквенциально случайной*.

Другими словами, последовательность ω преквенциально случайна по мере P, если существует (вообще говоря, другая) мера Q, относительно которой ω случайна и которая совпадает с P на всех начальных отрезках ω.

Требование преквенциальности связано с попытками перенести понятия теории вероятностей и статистики в ситуацию последовательных предсказаний членов последовательности, ср. [5, 26]. Рассмотрим, например, прогноз погоды, в котором $\omega(n)$ означает, что в день n идёт дождь. Метеобюро перед каждым днём указывает число $p(n)$, которое оно называет вероятностью дождя в день n. (При этом на следующие дни никакого распределения вероятностей не указывается. Другими словами, вместо глобального распределения вероятностей бюро прогнозов указывает лишь условные вероятности вдоль пути, соответствующего фактической погоде.)

Можно ли оценить качество прогноза? Кажется, что в некоторых ситуациях да: если, скажем, все предсказания близки к нулю (скажем, меньше 10%), а большинство дней (скажем, более 90%) были дождливые. (Говорят, что такой прогноз плохо *калиброван*.) Но, естественно, возможны и какие-то другие виды несоответствий, не только частотные: общий вопрос состоит в том, можно ли воспринимать данную последовательность как полученную случайно с предсказанными вероятностями. (Другая ситуация, где возникает такой вопрос, это оценка качества датчика случайных битов, который выдаёт бит с заказанным распределением, на каждом шаге своим.)

Дополнительным обстоятельством при оценке качества предсказания является то, что предсказатель может использовать разнообразную информацию, доступную на момент предсказания (скажем, вечер предыдущего дня), а не только предыдущие члены последовательности ω. Наличие такой информации должно учитываться и при оценке качества предсказания.

В статье [26] обсуждаются подобные вопросы и предлагаются различные варианты определений, в частности, связанные с понятием мартингала, и доказывается эквивалентность некоторых из них. Интересно было бы установить связь и с равномерными тестами случайности в духе приведённого выше преквенциального определения дефекта (правда, вместо вероятностей начальных отрезков тут возникают условные вероятности, что не совсем то же самое, если они не отделены от нуля).

10 Вопросы

Мы уже отмечали некоторые вопросы, которые (на наш взгляд) было бы интересно изучить. В этом разделе мы собрали ещё несколько таких вопросов.

1. Рассмотрим следующий метод порождения последовательности $\xi \in \Omega$, распределённой в соответствии с данным распределением P на Ω, при котором вероятности всех слов ненулевые. Возьмём случайную последовательность ρ независимых равномерно распределённых на $[0,1]$ случайных чисел. После того как $\xi(1:n-1)$ уже построена, мы полагаем $\xi(n)=1$ тогда и только тогда, когда

$$\rho(n) < P(\xi(1:n-1)1)/P(\xi(1:n-1)).$$

Если рассматривать это как вероятностный процесс, то выходное распределение будет в точности P. Спрашивается, какие последовательности можно получить на выходе, если начинать со случайных по Мартин-Лёфу последовательностей действительных чисел. (Можно проверить, что для вычислимых мер P получаются в точности случайные по Мартин-Лёфу относительно P последовательности.)

2. Вспомним формулу для дефекта случайности по вычислимой мере:

$$\mathbf{t}(\omega, P) \doteq \sum_{x \sqsubseteq \omega} \frac{\mathbf{m}(x)}{P(x)}.$$

Обе части имеют смысл при произвольном P, но они могут быть различны. Причину этого мы уже обсуждали: небольшое изменение меры почти не влияет на правую часть, но может изменить её вычислительную силу как оракула и существенно изменить левую.

Обозначим правую часть этого равенства через $t'(\omega, P)$. Может быть, имеет смысл считать конечность t' определением случайности по невычислимым мерам? По крайней мере она будет монотонной (от увеличения меры случайность будет только расти). Относительно смеси мер она будет квази-выпуклой, более того, в [7] доказано, что $1/t'(\omega, P)$ является вогнутой функцией от P.

Другое возможное определение дефекта случайности для бесконечной последовательности ω по мере P таково: $\log \sup_{x \prec \omega} [\mathbf{a}(x)/P(x)]$. Для вычислимых мер мы вновь получаем определение, равносильное стандартному определению Мартин-Лёфа. В работе [10] показано, что определённые таким образом тесты случайности (а также аналогичные, использующие монотонную сложность вместо априорной вероятности) не обладают некоторым естественным свойством (*сохранения случайности*) — в отличие от равномерных тестов. В [7] показано, что, с другой стороны, что если в определении t' в правой части рассмотреть сумму по всем вычислимым функциям с конечным числом рациональных значений, а не только характеристические функции множеств $x\Omega$, то свойство сохранения информации выполняется.

3. Можно ли разумно определить дефект случайности последовательности относительно произвольных мер, добившись его монотонности (в каком-нибудь естественном смысле)? Например, можно было бы потребовать

$$P \leqslant c \cdot Q \Rightarrow t(\omega, P) \geqslant t(\omega, Q)/c.$$

(отметим в качестве мотивировки, что правая часть приведённой выше формулы для дефекта обладает этим свойством). Можно ли рассчитывать при этом на свойство квази-выпуклости? Некоторые попытки такого рода предприняты в [16, 8], а также в [17]

Свойство квази-вогнутости, видимо, обеспечить труднее (в этой ситуации приведённые контрпримеры выглядят более устойчивыми).

4. Стандартной процедурой в теории вычислимости является релятивизация: некоторое множество A объявляется разрешимым по определению, и алгоритмам разрешается использовать "оракул" для этого множества (отвечающий на вопросы о принадлежности ему). Это увеличивает класс вычислимых функций, но большинство результатов теории вычислимости остаются верными. Более сложная ситуация возникает, когда мы объявляем некоторое множество E перечислимым по определению. Проблема тут в том, что его можно перечислять в разном порядке, и разные перечисляющие его "оракулы" могут дать разные результаты. Тем не менее существует естественное понятие *перечислимого относительно* E множества (как ещё говорят, множества, *сводящегося по перечислимости* к E). Приведём соответствующее определение. Пусть W есть некоторое множество пар вида $\langle x, S\rangle$, где x — натуральное число, а S — конечное множество натуральных чисел. Будем считать, что множество W перечислимо. Тогда для любого множества E можно рассмотреть множество $S(E, W)$, состоящее из всех x, при которых $\langle x, S\rangle \in W$ при некотором $S \subset E$. (Неформально говоря, пара $\langle x, S\rangle$ понимается как инструкция: выдавать на выход x, обнаружив в перечислении E все элементы из списка S.) Добавление стандартного разрешающего оракула для множества A можно рассматривать как частный случай такой сводимости, положив $E = \{2n \mid n \in A\} \cup \{2n+1 \mid n \notin A\}$.

В некоторых ситуациях можно пытаться обойтись такого рода оракулом. Скажем, вполне можно говорить о перечислимой снизу относительно E функции, поскольку её можно определить в терминах перечислимых множеств. Но не очевидно, что определённые таким образом понятия обладают привычными для нас свойствами.

Можно ли утверждать (для произвольного E), что существует максимальная перечислимая снизу относительно E полумера? Можно ли определить префиксную сложность с оракулом E и будет ли она совпадать с логарифмом максимальной полумеры (если таковая существует)? Что будет, если дополнительно предположить, что E есть множество всех базисных шаров в конструктивном метрическом пространстве, содержащих некоторую точку?

(Для сравнения напомним, что можно определить E-вычислимую функцию как функцию, график которой E-перечислим. При этом выполняются некоторые знакомые свойства, скажем, композиция двух E-вычислимых функций является E-вычислимой. Однако, скажем, утверждение о том, что всякое непустое E-перечислимое множество является областью значений всюду определённой E-вычислимой функции, уже не гарантировано: при некоторых E это не так.)

5. Можно пытаться обобщать понятие случайности в другом направлении, рассматривая не вычислимые меры, а перечислимые полумеры, то есть выходные распределения вероятностных машин, выдающих бит за битом (и, возможно, выдающих конечную последовательность с положительной вероятностью). Это предложил Левин, имея в виду определить независимость двух последовательностей α и β как случайность пары (α, β) относительно полумеры $\mathbf{a} \times \mathbf{a}$. (Такой подход предполагает, что все последовательности случайны относительно полумеры $\mathbf{a}$.) Соответственно дефект случайности пары (α, β) относительно $\mathbf{a} \times \mathbf{a}$ можно было бы назвать количеством общей информации в последовательностях α и β, по аналогии с конечными словами, где взаимная информация слов x и y, определяемая как

$$KP(x) + KP(y) - KP(x, y) = -\log(\mathbf{m}(x) \times \mathbf{m}(y)) - KP(x, y),$$

выглядит как дефект случайности относительно $\mathbf{m} \times \mathbf{m}$.

Одна из возможностей для определения случайности последовательности относительно полумеры Q такая: потребовать ограниченности отношения $\mathbf{a}(x)/Q(x)$ для начальных отрезков x последовательности ω. Другой вариант: рассматривать случайные по Мартин-Лёфу относительно равномерной меры последовательности случайных битов, использовать их как исходы датчика случайных битов в вероятностной машине и смотреть, какие последовательности могут получиться на выходе. Неизвестно, совпадают ли это определения. Неизвестно также, корректно ли второе из них (в том смысле, что двум вероятностным машинам с одним и тем же выходным распределением соответствует одно и то же множество образов случайных последовательностей). Для вычислимых мер (машин, выдающих бесконечные последовательности с вероятностью 1) это действительно так.

6. (Этот вопрос задал С. Симпсон) Можно ли предложить естественное понятие тестов случайности для, скажем, 2-случайных последовательностей? (Обычное определение на языке тестов соответствует тестам, не являющимся полунепрерывными.)

Благодарности

Авторы благодарны своим коллегам, с котороми они обсуждали рассматриваемые в статье вопросы, в первую очередь Л. Левину, к которому восходят многие из понятий этой статьи, А. Буфетову и А. Клименко, а также В. Вьюгину и другим участникам колмогоровского семинара (на мехмате МГУ в Москве). Статья написана при финансовой поддержке грантов NAFIT ANR-08-EMER-008-01, RFBR 0901-00709-а.

References

[1] Jeremy Avigad, Philipp Gerhardy, and Henry Towsner. Local stability of ergodic averages. *Transactions of the American Mathematical Society*, 362(1):261–288, 2010.

[2] Laurent Bienvenu, Adam Day, Mathieu Hoyrup, Ilya Mezhirov, and Alexander Shen. A constructive version of Birkhoff's ergodic theorem for Martin-Löf random points. http://arxiv.org/abs/1007.5249

[3] Laurent Bienvenu, Andrei Romashchenko, Alexander Shen. Sparse sequences. Journees Automates Cellulaires, 2008 (Uzes). Moscow: MCCME publishers, 2008. P.18–28. http://hal.archives-ouvertes.fr/hal-00274010/en.

[4] Gregory J. Chaitin. A theory of program size formally identical to information theory. *Journal of the ACM*, 22:329–340, 1975.

[5] Alexey Chernov, Alexander Kh. Shen, Nikolai Vereshchagin, and Vladimir G. Vovk. On-line probability, complexity and randomness. In Yoav Freund, Laszlo Gyorfi, Gyorgy Turan, and Thomas Zeugmann, editors, *Proceedings of the nineteenth international conference on algorithmic learning theory*, pages 138–153, Tokyo, 2008.

[6] Peter Gács. Lecture notes on descriptional complexity and randomness. Technical report, Boston University, Computer Science Dept., Boston, MA 02215. www.cs.bu.edu/~gacs/papers/ait-notes.pdf.

[7] Peter Gács. *Complexity and Randomness.* PhD thesis, J.W. Goethe University, Frankfurt, W.Germany, 1978. In German.

[8] Peter Gács. Exact expressions for some randomness tests. *Z. Math. Log. Grdl. M.*, 26:385–394, 1980. Short version: Springer Lecture Notes in Computer Science 67 (1979) 124-131.

[9] Peter Gács. On the relation between descriptional complexity and algorithmic probability. *Theoretical Computer Science*, 22:71–93, 1983. Short version: Proc. 22nd IEEE FOCS (1981) 296-303.

[10] Peter Gács. Uniform test of algorithmic randomness over a general space. *Theoretical Computer Science*, 341(1-3):91–137, 2005.

[11] Mathieu Hoyrup and Cristóbal Rojas. An application of Martin-Löf randomness to effective probability. In *Cie2009, LNCS 5635*, pages 260–269, 2009.

[12] Mathieu Hoyrup and Cristóbal Rojas. Computability of probability measures and Martin-Löf randomness over metric spaces. *Information and Computation*, 207(7):830–847, 2009.

[13] Andrei N. Kolmogorov. On the logical foundations of information theory and probability theory. *Problems of Information Transmission*, 5(3):1–4, 1969.

[14] K. Kuratowski. *Topology.* Academic Press, New York, 1966.

[15] Leonid A. Levin. On the notion of a random sequence. *Soviet Math. Dokl.*, 14(5):1413–1416, 1973.

[16] Leonid A. Levin. Uniform tests of randomness. *Soviet Math. Dokl.*, 17(2):337–340, 1976.

[17] Leonid A. Levin. Randomness conservation inequalities: Information and independence in mathematical theories. *Information and Control*, 61(1):15–37, 1984.

[18] Ming Li and Paul M. B. Vitányi. *Introduction to Kolmogorov Complexity and its Applications (Third edition)*. Springer Verlag, New York, 2008.

[19] Per Martin-Löf. The definition of random sequences. *Information and Control*, 9:602–619, 1966.

[20] Y. T. Medvedev. Degrees of difficulty of mass problems. *Doklady Akademii Nauk SSSR. N.S.*, 104:501–504, 1955. In Russian. Mathematical Reviews (MathSciNet): MR0073542.

[21] Joseph Miller. Degrees of unsolvabiility of continuous functions. *The Journal of Symbolic Logic*, 69(2), 555-584, 2004.

[22] Claus Peter Schnorr. Process complexity and effective random tests. *J. Comput. Syst. Sci*, 7(4):376–388, 1973. Conference version: STOC 1972, pp. 168-176.

[23] Glenn Shafer, Alexander Shen, Nikolai Vereshchagin, and Vladimir Vovk. Test martingales, Bayes factors, and p-values.
`arxiv.org`, 0912.4269v2.

[24] Glenn Shafer, Vladimir Vovk. Probability and Finance: It's Only a Game! Wiley, 2001. ISBN: 978-0-471-40226-8.

[25] Alexander Shen. Algorithmic information theory and Kolmogorov complexity. Technical Report TR2000-34, Uppsala University. 31pp. Available at
`http://www.it.uu.se/research/publications/reports/2000-034/`.

[26] Vladimir G. Vovk and Alexander Shen. Prequential randomness and probability. *Theoretical Computer Science*, 411:2632–2646, 2010.

[27] Volker Strassen. The existence of probability measures with given marginals. *Annals of Mathematical Statistics*, 36:423–439, 1965.

[28] V. V. V'yugin. Ergodic theorems for individual random sequences. *Theoretical Computer Science*, 207(2):343–361, 1998.

[29] Alexander K. Zvonkin and Leonid A. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Math. Surveys*, 25(6):83–124, 1970.