Наш семинар: математические сюжеты

Четыре алгоритмических лица случайности

В. А. Успенский

Введение

Если кто-либо скажет нам, что он подбросил «честную» монету двадцать раз и, обозначив герб единицей, а решетку нулем, получил такой результат:

$$10001011101111010000 \qquad \text{(I)}$$

или такой:

$$01111011001101110001, \qquad \text{(II)}$$

мы вряд ли будем удивлены. Однако если нам скажут, что результат бросаний был таким:

$$00000000000000000000 \qquad \text{(III)}$$

или таким:

$$01010101010101010101 \qquad \text{(IV)}$$

— мы будем поражены и вообще не поверим или же усомнимся в корректности эксперимента. Возникает вопрос: почему?

По-видимому, цепочки (I) и (II) воспринимаются как случайные, а цепочки (III) и (IV) — как неслучайные.

Но что означают слова «воспринимается как случайная»? Классическая теория вероятностей не дает ответа на этот важный вопрос. Не столь редко можно услышать следующее объяснение: вероятность каждой из цепочек (III) и (IV) слишком мала, она равна 2^{-20}, что меньше одной миллионной. Но ведь ровно такую же вероятность имеют цепочки (I) и (II).

Математическое просвещение, сер. 3, вып. 10, 2006(71–108)

Три важных замечания.

Во-первых, впечатление случайности зависит от распределения вероятностей. Если одна сторона монеты тяжелее другой или если человек в процессе бросания научается подкидывать монету так, чтобы она падала на нужную сторону, то появление цепочек (III) или (IV) может стать вполне ожидаемым. Вначале — для простоты — мы будем заниматься лишь независимыми бросаниями с вероятностями одна вторая для герба и одна вторая для решетки.

Во-вторых, говорить о случайности имеет смысл лишь в применении к очень длинным цепочкам. Бессмысленно спрашивать, которая из цепочек 00, 01, 10, 11 более случайна, чем другие.

В-третьих, точной границы между случайными и неслучайными (в интуитивном смысле) цепочками нет и не может быть. Ведь если в случайной цепочке заменить один знак, она остается случайной. Но, заменяя много раз, мы от любой цепочки можем прийти к (III) или (IV). Это известный парадокс кучи.

Итак, следует рассматривать только очень длинные цепочки, а в идеале — бесконечные. (Вообще, бесконечность — это такое полезное приближение сверху к очень большому конечному.) Бесконечные цепочки принято называть *последовательностями*. Оказывается, последовательности уже можно довольно осмысленно разделить на случайные и неслучайные. Иными словами, можно не без успеха пытаться найти строгое математическое определение для понятия „случайная последовательность нулей и единиц“. В настоящем очерке мы изложим результаты таких попыток, предпринятых различными авторами. Однако следует честно признать, что для практических приложений интерес представляют именно конечные случайные цепочки и потому идеализация, происходящая при переходе к цепочкам бесконечной длины, неизбежно сопряжена с «отрывом от жизни». Впрочем, аналогичный отрыв возникает и при изучении совокупности *всех* конечных цепочек, поскольку в жизни встречаются лишь цепочки ограниченной длины, а в очень длинных случайных цепочках возникают такие эффекты, которые могут и не соответствовать наивным представлениям о случайности.

Сделав это признание, приступим к изложению.

Нам будут полезны некоторые термины и обозначения.

Всякая конечная цепочка $\langle i_1, \ldots, i_n \rangle$ нулей и единиц называется также *двоичным словом*, а число n — *длиной* этого слова. Длина слова x обозначается так: $|x|$. Длина слова может быть равна нулю; слово длины нуль ничего не содержит и называется *пустым*. Пустое слово обозначается буквой Λ.

Множество всех двоичных слов обозначается буквой Ξ. Множество всех последовательностей нулей и единиц обозначается буквой Ω.

Для последовательности $\langle a_1, a_2, a_3, a_4, \ldots\rangle$ каждое из двоичных слов $\langle a_1, a_2, \ldots, a_n\rangle$ является ее *началом*. Пусть $x \in \Xi$ — двоичное слово. Множество всех тех последовательностей из Ω, для которых x служит началом, обозначим через Ω_x. Каждое множество вида Ω_x, где $x \in \Xi$, принято называть *шаром*. Шару Ω_x приписывается его *объём* $\mathbf{v}(x)$, равный $2^{-|x|}$.

Содержательно, каждую последовательность из Ω мы рассматриваем как серию результатов бросаний монеты. Еще раз подчеркнем, что для наглядности мы ограничиваемся на первых порах ситуацией, когда результаты всех бросаний равновероятны. На математическом языке это означает, что на Ω задано так называемое *равномерное распределение вероятностей* — а это, в свою очередь, означает следующее: для каждого шара вероятность того, что случайно выбранная последовательность из Ω попадет в этот шар, равна объему этого шара.

Наша цель — попытаться выделить из Ω некоторое точно описанное подмножество, претендующее на роль множества всех случайных последовательностей. Традиционная теория вероятностей не только не приближается к решению этой задачи, но даже не может ее сформулировать в своих терминах. На помощь приходит теория алгоритмов. Может показаться парадоксальным, что понятие случайности уточняется на основе такого чуждого случайности понятия, как алгоритм, — тем не менее дело обстоит именно так: все известные до сих пор определения случайности индивидуального объекта (в нашем примере — индивидуальной последовательности нулей и единиц) опираются на понятие алгоритма.

Чтобы найти требуемое определение, поступают так. Формулируют некое характеристическое свойство, которым обладают случайные (в неформальном, интуитивном смысле) последовательности. А затем последовательности, обладающие этим свойством, и объявляют — по определению — случайными.

Какими же свойствами обладает случайная последовательность нулей и единиц?

Во-первых, она *частотоустойчива*. Вот что это означает для того простейшего случая, когда нули и единицы равновероятны: частота нулей, как и частота единиц, стремится к одной второй. (*Частота нулей* — это их доля в начальном отрезке последовательности.) Но более того: в случайной последовательности указанная устойчивость частот выполняется не только для последовательности в целом, но и для любой ее законной, разумной подпоследовательности.

Во-вторых, она *хаотична*. Это означает, что она сложно устроена и не может иметь разумного описания. Психологическому эксперименту, с которого мы начали, Колмогоров дал такое объяснение. Цепочки (I) и (II) потому воспринимаются как случайные, что они сложны, их устройство

нельзя коротко описать. А вот цепочки (III) и (IV) имеют простое, легко описываемое устройство.

В-третьих, она *типична*. Это означает, что она принадлежит любому разумному большинству.

В-четвертых, она *непредсказуема*. Это означает, что играя против нее на деньги (т. е. пытаясь угадать члены последовательности и делая ставки), последовательность невозможно обыграть, какой бы разумной стратегией ни пользоваться.

Слово «разумный», встречающееся в объяснениях перечисленных четырех свойств, конечно, нуждается в уточнении. Теория алгоритмов как раз и предлагает такие уточнения, наполняя это слово точным смыслом — своим для каждого из наших четырех свойств. Тем самым возникают четыре алгоритмических свойства: *частотная устойчивость*, *хаотичность*, *типичность*, *непредсказуемость*. Каждое из них представляет свое собственное алгоритмическое лицо случайности, и каждое из них с большими или меньшими основаниями может претендовать на роль строгого математического определения понятия случайности. Можно сказать и так: возникают четыре точно очерченных класса последовательностей, каждый из которых претендует на то, чтобы служить «истинным» классом случайных последовательностей; некоторые из этих претензий более оправданы, чем другие.

Лицо Первое: Частотоустойчивость и стохастичность

По-видимому, одним из первых поставил вопрос о том, что такое отдельно взятая случайная последовательность, замечательный немецкий математик Рихард фон Мизес в начале XX века — в 1919 г. Во всяком случае, именно он первым предложил сравнительно удачное (хотя и нестрогое) определение, послужившее отправной точкой для дальнейшего развития.

Мизес исходил из того, что случайной последовательности должна быть присуща устойчивость частот. А именно, доля единиц (как и доля нулей) в начальном отрезке случайной последовательности должна стремиться к одной второй при неограниченном увеличении длины начального отрезка. Но этого недостаточно. Например, этим свойством устойчивости частот обладает заведомо неслучайная последовательность

$$0, 1, 0, 1, 0, 1, 0, 1, \ldots .$$

Очевидным образом необходимо, чтобы устойчивостью частот обладала не только вся последовательность целиком, но и ее подпоследовательности. Однако устойчивостью частот не могут обладать *все* подпоследовательности. В самом деле, возьмем самую что ни на есть случайную

последовательность и отберем в подпоследовательность те ее члены, которые равны нулю... Значит, подпоследовательность, в которой должна соблюдаться устойчивость частот, должна быть *разумной*, или *допустимой*. Например, если отобрать в подпоследовательность все те члены, чьи номера суть простые числа, или же все те члены, которые непосредственно следуют за нулевыми, то полученная в каждом из этих двух вариантов подпоследовательность является допустимой.

Последовательности, в которых устойчивость частот соблюдается во всех ее допустимых подпоследовательностях, Колмогоров предложил называть *стохастическими*.

Сам Мизес объяснял, какие последовательности следует считать допустимыми, в расплывчатых терминах. Первое уточнение предложил в 1940 г. американский математик Алонзо Чёрч — один из создателей теории алгоритмов. Чёрч предложил, чтобы допустимые подпоследовательности строились на основе определенных алгоритмов. Последовательности, в которых устойчивость частот наблюдается во всех допустимых по Чёрчу подпоследовательностях, получили название *стохастических по Чёрчу*. Однако оказалось, что определение Чёрча слишком широко: существует, например, стохастическая по Чёрчу последовательность, перестающая быть таковой после вычислимой перестановки ее членов.

В 1963 г. Колмогоров усовершенствовал конструкцию Чёрча, предложив более обширный класс допустимых алгоритмов отбора членов исходной последовательности в подпоследовательность и тем самым заведомо не меньший (а на деле более обширный) класс допустимых подпоследовательностей; в частности, колмогоровский алгоритм разрешает членам подпоследовательности иметь иной порядок следования, чем в исходной последовательности. Поэтому класс всех последовательностей, *стохастических по Колмогорову* (т. е. таких, у которых устойчивость частот наблюдается во всех допустимых по Колмогорову подпоследовательностях), является подклассом класса последовательностей, стохастических по Чёрчу, — на самом деле даже строгим подклассом, т. е. не совпадающим с объемлющим классом. Для дальнейших ссылок класс всех последовательностей, стохастических по Колмогорову, обозначим буквой **S**.

Выяснилось, что класс **S** также слишком широк. Оказалось, например, что может быть построена такая стохастическая по Колмогорову последовательность, в каждом начальном отрезке которой единиц больше, чем нулей; а это противоречит как интуиции, так и законам теории вероятностей (например, закону возвратности состояний при случайном блуждании). Таким образом, даже наиболее продвинутое из известных на сегодняшний день математически строгих уточнений идей Мизеса не дает полного отражения интуитивного представления о случайности (хотя и это неполное отражение может оказаться полезным).

Для большей ясности приведем конструкции Чёрча и Колмогорова. Центральным для обеих конструкций является указание тех правил, согласно которым из членов рассматриваемой последовательности составляются допустимые подпоследовательности. Поскольку каждое такое правило производит отбор членов последовательности для включения их в допустимую подпоследовательность, сами эти правила принято называть *допустимыми правилами отбора*.

Для наглядности представим себе, что члены исследуемой последовательности написаны на картах. Карты лежат одна за другой, лицевой стороной вниз, так что мы не видим, что на них написано. Наша цель — отобрать некоторые из этих карт и составить из них другую последовательность — допустимую подпоследовательность. (Как мы увидим, в случае колмогоровской конструкции термин «подпоследовательность» имеет более широкий смысл, чем это обычно принято.) Допустимое правило отбора представляет собою алгоритм, который на каждом этапе построения подпоследовательности предписывает, во-первых, которую из карт надлежит открыть и, во-вторых, следует или нет включать эту открываемую карту в подпоследовательность. Входом алгоритма служит информация о всех уже открытых к этому моменту картах. Может случиться, что алгоритм отберет лишь конечное число членов исходной последовательности; в этом случае считается, что никакой допустимой подпоследовательности не образовалось. (Напомним, что исходная последовательность признаётся стохастической при наличии устойчивости частот в любой ее допустимой подпоследовательности.)

Переходим к более точному описанию.

Функция называется *вычислимой*, коль скоро существует *вычисляющий* ее алгоритм. При этом говорят, что алгоритм *вычисляет* функцию f, коль скоро он перерабатывает всякий ее аргумент x, на котором функция определена, в соответствующее значение $f(x)$ и не выдает никакого результата в применении ко всякому такому аргументу, на котором функция не определена.

Исследуемую на устойчивость последовательность обозначим $a_1, a_2, a_3, \ldots$, так что на n-й карте записана цифра a_n, равная 0 или 1.

Допустимое по Чёрчу правило отбора является произвольной вычислимой функцией G, определенной на множестве Ξ всех двоичных слов и принимающей одно из двух значений: *Да* и *Нет*. Карты открываются последовательно, одна за другой, начиная с первой, и каждый раз — до открытия очередной карты — решается вопрос, включать ли эту карту в подпоследовательность. Вопрос этот решается следующим образом. Пусть уже открыты n карт, на которых записаны, соответственно, цифры $a_1, \ldots, a_n$. Если $G(a_1, \ldots, a_n)$ есть *Нет*, то следующая, $(n+1)$-я карта не включается в подпоследовательность. Если $G(a_1, \ldots, a_n)$ есть *Да*,

то следующая, $(n+1)$-я карта включается в подпоследовательность в качестве очередного члена. Включать или не включать в подпоследовательность член a_1, зависит от значения $G(\Lambda)$. Таким образом, правилу G отвечает бесконечная подпоследовательность (в случае, если таковая образовалась) $a_{n(1)}, a_{n(2)}, a_{n(3)}, \ldots$, где числа $n(1)$, $n(2)$, $n(3)$, ... образуют возрастающую последовательность, составленную из всех чисел n, для которых $G(a_1, \ldots, a_{n-1}) = \textit{Да}$.

Изложению конструкции Колмогорова предпошлем следующее обобщение понятия подпоследовательности. *Обобщенной подпоследовательностью* последовательности $a_1, a_2, \ldots, a_n, \ldots$ назовем всякую последовательность вида

$$a_{\varphi(1)}, a_{\varphi(2)}, \ldots, a_{\varphi(k)}, \ldots,$$

для которой выполнено условие

$$i < j \implies \varphi(i) \neq \varphi(j).$$

Для обычной подпоследовательности (которая является частным случаем обобщенной) это условие заменено на более сильное — с правой частью $\varphi(i) < \varphi(j)$.

Каждое допустимое по Колмогорову правило отбора имеет целью создать некоторую обобщенную подпоследовательность исходной последовательности. Мы потому написали «имеет целью создать», а не «создает», что процесс создания может прерваться, и тогда в результате применения правила возникает не бесконечная последовательность, а *кортеж* (т. е. конечный набор), составленный из членов исходной последовательности. Для стохастичности по Колмогорову требуется устойчивость частот в каждой из допустимых, т. е. образованных каким-либо допустимым по Колмогорову правилом, обобщенных подпоследовательностей.

Само колмогоровское (т. е. допустимое по Колмогорову) правило состоит из двух вычислимых функций: F и G. Первая служит для образования некой вспомогательной обобщенной подпоследовательности. Окончательная же обобщенная подпоследовательность исходной последовательности строится при помощи функции G в качестве обычной подпоследовательности вспомогательной обобщенной подпоследовательности. Аргументами функций F и G служат двоичные слова, так что каждая из этих функций определена на своем подмножестве множества Ξ. Значениями функции F служат целые положительные числа, у функции G два возможных значения: *Да* и *Нет*. Сперва строится последовательность натуральных чисел $n(1), n(2), n(3), \ldots$:

$$n(1) = F(\Lambda),\ n(2) = F(a_{n(1)}),\ \ldots,\ n(k+1) = F(a_{n(1)}, \ldots, a_{n(k)}).$$

Построение этой последовательности прекращается (и возникает конечный кортеж), как только наступает один из трех случаев:

▷ значение $F(a_{n(1)}, \ldots, a_{n(k)})$ не определено;

▷ значение $G(a_{n(1)}, \ldots, a_{n(k)})$ не определено;

▷ значение $F(a_{n(1)}, \ldots, a_{n(k)})$ совпадает с одним из чисел $n(1), \ldots, n(k)$.

Если же остановки в построении не произошло и возникла бесконечная последовательность номеров $n(1), n(2), n(3), \ldots$, то далее строится вспомогательная обобщенная подпоследовательность $a_{n(1)}, a_{n(2)}, a_{n(3)}, \ldots$. Из нее, наконец, отбираются — в порядке возрастания k — те ее члены $a_{n(k)}$, для которых выполнено равенство $G(a_{n(1)}, \ldots, a_{n(k-1)}) = Да$.

Лицо Второе: Хаотичность

Вернемся к примерам конечных цепочек из Введения и вспомним объяснение, предложенное Колмогоровым: цепочки (I) и (II) воспринимаются как случайные, потому что они *сложны*; цепочки (III) и (IV) воспринимаются как неслучайные, потому что они *просты*. По-видимому, мы ожидаем, что результат случайного эксперимента окажется сложным, и удивляемся, когда получаем что-то простое.

Некоторые объекты можно квалифицировать как большие или маленькие, другие — как тяжелые или легкие, третьи — как сложные или простые. В начале 60-х годов Колмогоров наметил математическую теорию, позволяющую оценивать сложность объектов. Сейчас эта теория называется теорией колмогоровской сложности.

В основе теории колмогоровской сложности лежит следующая простая и естественная идея:

сложность объекта измеряется длиной его кратчайшего описания.

В самом деле, каждый объект может иметь сколь угодно сложные описания, но сложный объект невозможно описать коротко.

Пусть Y — множество всех объектов, которые мы рассматриваем, а X — множество всех мыслимых описаний этих объектов. Через $|x|$ обозначим длину описания x. Сложность объекта y обозначим $\mathrm{Comp}\,(y)$. В соответствии со сказанным,

$$\mathrm{Comp}\,(y) = \min_x \{|x| : x \text{ есть описание для } y\}.$$

Если объект y неописуем, т. е. для него не существует описания, то, естественно, его сложность равна бесконечности.

Разумеется, длины надо мерить единообразным способом, чтобы не получилось, что описание одного и того же объекта имеет на китайском языке длину единица (один иероглиф), а на русском — сорок (сорок букв). Поэтому все описания кодируются в двоичном алфавите — в виде двоичных цепочек. Вот эти-то двоичные коды мы и будем отныне считать описаниями, так что отныне $X = \Xi$.

Множество всех таких пар $\langle x, y\rangle$, что x служит описанием для y, естественно называть *языком описания*.

Заметим, что и один и тот же объект может иметь в данном языке много описаний, и одно и то же описание может служить описанием многих объектов — таково, например, описание: «двоичное слово, состоящее только из нулей» (а, скажем, выражение «двоичное слово» годится в качестве описания для любого двоичного слова).

Всё сказанное носило предварительный характер, чтобы оправдать нижеследующее формальное изложение.

В декартовом произведении $\Xi \times Y$ рассматриваем произвольное подмножество E, которое называем *языком описания*. Если $\langle x, y\rangle \in E$, будем говорить, что x является *описанием* объекта y. *Сложность* Comp_E объекта y *относительно* языка E определяется так:

$$\mathrm{Comp}_E(y) = \min_x\{|x| : \langle x, y\rangle \in E\}.$$

Напомним, что минимум по пустому множеству принято считать равным бесконечности.

Для языка $E = \Xi \times Y$, в котором каждое x служит описанием для каждого y, сложность любого объекта y равна нулю, поскольку одним из описаний этого y является пустое слово; такой язык мыслим, но не будет встречаться в тех семействах языков, которые мы будем рассматривать.

Представим себе два языка, причем описание какого-либо объекта на втором языке происходит путем удвоения описания этого же объекта на первом языке. Ясно, что второй язык «хуже» первого. Предпочтительнее те языки, которые в состоянии давать более короткие описания.

Будем говорить, что язык A *не хуже* языка B и писать $A \leqslant B$, если существует такая константа c, что для всякого y справедливо неравенство $\mathrm{Comp}_A(y) < \mathrm{Comp}_B(y) + c$, т. е. если

$$\exists c\, \forall y\ \mathrm{Comp}_A(y) < \mathrm{Comp}_B(y) + c.$$

Если принять, что для так называемых естественных языков, т. е. для тех, которыми пользуется человечество, могут быть сформулированы правила перевода с одного языка на другой, то становится очевидным, что каждый из этих языков не хуже другого. Ведь, скажем, турецкое описание какого-либо объекта можно составить так: взять произвольное японское описание этого объекта и дополнить его правилами японо-турецкого перевода. И турецкое, и японское описания, и правила перевода считаются закодированными в виде двоичных слов. Поэтому длина такого турецкого описания равна длине японского описания плюс не зависящая от выбора объекта длина правил перевода. В качестве японского

описания выберем кратчайшее. В итоге получаем, что турецкий язык не хуже японского.

Встает вопрос о выборе оптимального языка — такого, который не хуже любого другого. Пусть $\mathcal{L}$ — некоторое языковóе семейство, т. е. попросту некоторое множество языков. Язык A из этого семейства $\mathcal{L}$ называется *оптимальным* (для $\mathcal{L}$), если он не хуже любого другого языка из этого семейства, т. е. если

$$\forall B \in \mathcal{L}\ A \leqslant B.$$

Если оптимальный язык существует, то именно с его помощью следует измерять сложность. Сложность объекта относительно одного из оптимальных языков называется *алгоритмической энтропией* этого объекта. Энтропию и рассматривают как окончательную искомую меру сложности — в рамках заданного языковóго семейства.

Для некоторых важных языковы́х семейств имеет место теорема о существовании оптимального языка, а тем самым и энтропии. Эта теорема называется *теоремой Соломонова – Колмогорова*.

Для данного семейства может существовать много оптимальных языков и тем самым много энтропий. Однако, в силу определения оптимальности, любые две энтропии (взятые для одного и того же фиксированного языковóго семейства) различаются не более чем на аддитивную константу. Иными словами, если A и B суть два оптимальных для $\mathcal{L}$ языка, то существует такая константа c, что для всех y

$$|\operatorname{Comp}_A(y) - \operatorname{Comp}_B(y)| \ < \ c.$$

Замечание. Конечно, неприятно, что алгоритмическая энтропия, претендующая на роль «истинной сложности» определяется не однозначно, а всего лишь с точностью до аддитивной добавки, не превышающей константы. Однако попытки найти среди энтропий «наиболее оптимальную» пока что ни к чему хорошему не привели. С другой стороны, ведь, скажем, и длина, и вес также определены не однозначно, а с точностью до мультипликативной константы, зависящей от выбранной единицы измерения (10 см=100 мм; 2 кг=2000 г; и т. п.).

Из уважения к Колмогорову алгоритмическую энтропию обозначают обычно буквой K; иногда к этой букве K добавляют еще вторую букву, указывающую то языковóе семейство, к которому относится рассматриваемая энтропия. Если K' и K'' суть две энтропии, относящиеся к одному и тому же языковóму семейству, то, как отмечалось,

$$|\mathrm{K}' - \mathrm{K}''| \ < \ c.$$

Колмогоров не только сформулировал понятие алгоритмической энтропии объекта, но и применил его к исследованию случайных последовательностей. Наблюдение Колмогорова состояло в том, что в случайной последовательности энтропия начального отрезка растет достаточно быстро при неограниченном увеличении длины этого отрезка. (Ничего не поделаешь, случайная последовательность может начинаться с миллиона нулей — но и этот миллион ничто перед бесконечностью, и если взять все начальные отрезки в совокупности, то их энтропия неизбежно будет расти.)

Итак, в качестве описываемых объектов мы будем рассматривать двоичные цепочки — такие, как (I), (II), (III), (IV) и т. п. Поэтому отныне $Y = \Xi$.

Если язык содержит пару $\langle z, z\rangle$, то это означает попросту, что цепочка z описывает самоё себя. Язык D, состоящий в точности из пар такого вида, называется *диагональным* (термин из математики), или *автонимным* (термин из лингвистики). Для такого языка $\mathrm{Comp}_D(y) = |y|$. Ограничимся языковы́ми семействами, содержащими автонимный язык (таким будет, в частности, семейство монотонных языков, описанное ниже). Тогда для каждой энтропии K, соответствующей этому семейству, и подходящей константы c будет выполнено неравенство:

$$\mathrm{K}(y) < |y| + c.$$

Таким образом, если пренебречь аддитивной константой, максимально возможное значение для энтропии цепочки равно длине цепочки. Колмогоров предположил, что у случайной последовательности энтропия начального отрезка достигает этого максимума, то есть равна длине этого отрезка, — опять-таки, при пренебрежении аддитивной константой. В этом состояла основная идея Колмогорова относительно хаотичности.

Итак, фиксируем некоторое языково́е семейство и одну из соответствующих этому семейству энтропий K. Назовем последовательность

$$a_1, a_2, \ldots, a_n, \ldots$$

хаотической, если существует такая константа c, что

$$\mathrm{K}(a_1, a_2, \ldots, a_n) > n - c.$$

Очевидно, свойство хаотичности не зависит от выбранной энтропии, а только от выбранного семейства.

Оказалось, что при подходяще выбранном языковом семействе строгое понятие хаотичности хорошо отражает интуитивное представление о случайности.

Создавая теорию сложности объектов, Колмогоров придал соотношению „описание – объект“ алгоритмический характер. Следуя Колмогорову, ограничимся *перечислимыми* языками. Понятие *перечислимого*

множества — это одно из основных понятий теории алгоритмов (да и математики в целом). Этому понятию можно дать такое наглядное объяснение. Представим себе безостановочно работающий принтер, последовательно печатающий слова. После напечатания слова принтер делает пробел, так что слова отделены одно от другого. Перед напечатанием очередного слова принтер берёт время на размышление. Это время может оказаться бесконечным, и тогда слово не печатается вовсе; в таком случае напечатанным будет лишь конечное множество слов — в частности, пустое множество, если принтер в самом начале своей работы задумался навсегда. Так вот, множество всех когда-либо напечатанных таким принтером слов окажется перечислимым — и всякое перечислимое множество может быть получено таким образом при подсоединении принтера к «идеальному компьютеру». Перечислимо и множество теорем любой формальной теории. Здесь нет места разъяснять, ни что такое идеальный компьютер, ни что такое формальная теория. Но определение понятия „перечислимое множество“ мы сейчас приведем.

Для наглядности начнем с термина «счетное множество». Термин этот имеет два варианта значения. В первом, более узком (и более распространенном) значении счетное множество — это такое множество, которое можно поставить во взаимно однозначное соответствие с натуральным рядом. Во втором, более широком значении счетное множество — это такое множество, которое можно поставить во взаимно однозначное соответствие с каким-либо начальным отрезком натурального ряда. Напомним, что *начальным отрезком натурального ряда* называется произвольное множество M натуральных чисел, содержащее вместе со всяким своим элементом и все меньшие числа. Таким образом, и весь натуральный ряд, и пустое множество являются начальными отрезками. Поэтому при втором понимании счетности все конечные множества, в том числе пустое множество, оказываются счетными. Именно это второе, более широкое понимание счетности удобно для наших целей; его мы и примем. А тогда счетному множеству можно дать и такое определение: *множество называется счетным, если оно либо пусто, либо может быть расположено в последовательность* (*т. е. является множеством членов какой-либо последовательности*). Например, конечное множество $\{a, b, c\}$ можно расположить в последовательность $a, b, c, c, c, c, \ldots$. Заменяя в приведённом определении счетного множества термин «последовательность» на термин «вычислимая последовательность», мы получаем определение перечислимого множества. Что касается определения термина «вычислимая последовательность», то оно сейчас будет дано.

Последовательность $w_1, w_2, \ldots, w_n, \ldots$ называется *вычислимой*, коль скоро существует алгоритм, вычисляющий ее n-й член w_n по его номеру n.

Поэтому понятие вычислимой последовательности называют *алгоритмическим аналогом* или *эффективным аналогом* понятия последовательности, а понятие перечислимого множества — алгоритмическим аналогом или эффективным аналогом понятия счетного множества. Перечислимые множества называются также *эффективно счетными*. Повторим определение: *множество называется перечислимым, или эффективно счетным, если оно либо пусто, либо может быть расположено в вычислимую последовательность* (*т. е. является множеством членов какой-либо вычислимой последовательности*).

Все рассматриваемые нами языки суть подмножества декартова произведения $\Xi \times \Xi$ и, следовательно, счетны. Идеология Колмогорова состояла в том, чтобы рассматривать только эффективно счетные (они же — перечислимые) языки.

Окончательно надлежащий выбор языково́го семейства произвел колмогоровский ученик Леонид Левин. Именно, в 1973 г. он ввел в рассмотрение семейство монотонных языков и изучил соответствующее этому семейству понятие хаотичности. Приведем необходимые определения.

Будем говорить, что двоичные слова u и v *согласованы* и писать $u \approx v$, если одно из этих слов есть начало другого.

Язык E называется *монотонным*, если он перечислим и выполнено условие:

$$[\langle x_1, y_1\rangle \in E \;\&\; \langle x_2, y_2\rangle \in E \;\&\; (x_1 \approx x_2)] \implies [y_1 \approx y_2].$$

Всякую последовательность, являющуюся хаотической для семейства монотонных языков, будем называть просто *хаотической*.

Алгоритмическая энтропия, соответствующая семейству монотонных языков, называется *монотонной энтропией* и обозначается КМ. Условие хаотичности последовательности $\langle a_1, a_2, a_3, \ldots\rangle$ можно записать так:

$$\exists c\, \forall n\, \mathrm{KM}(a_1, a_2, \ldots, a_n) > n - c.$$

Класс всех хаотических последовательностей обозначается буквой $\mathbf{C}$.

Считается, что строгое понятие хаотической последовательности может служить хорошим отражением интуитивного понятия случайной последовательности. К тому есть два основания.

Во-первых, для каждой отдельно взятой хаотической последовательности выполняются основные законы теории вероятностей.

Во-вторых, класс хаотических последовательностей совпадает с другим «претендентом» на роль строгого аналога расплывчатого класса случайных последовательностей — а именно, с классом $\mathbf{T}$ всех типических последовательностей (о нём будет рассказано далее):

$$\mathbf{C} = \mathbf{T}.$$

Поэтому хаотические последовательности можно было бы называть *типическо-хаотическими* или *хаотическо-типическими*, а сам класс обозначать **TC** или **CT**. Этот класс у́же класса **S** всех последовательностей, стохастических по Колмогорову (последний, как уже́ отмечалось, слишком широк):

$$\mathbf{TC} \subset \mathbf{S}, \quad \mathbf{TC} \neq \mathbf{S}.$$

Лицо Третье: Типичность

Сказать про какой-либо объект, что он типичен — это значит сказать, что он принадлежит к любому подавляющему большинству. Например, типичный человек, встреченный на московской улице, имеет рост менее двух метров (т. е. принадлежит к подавляющему большинству людей, имеющих рост менее двух метров), возраст более трех лет (т. е. принадлежит к подавляющему большинству людей, имеющих возраст более трех лет) и т. д. Конечно, любое из большинств, о которых идет речь, должно быть разумным: ведь какой объект ни возьми, он заведомо не принадлежит к подавляющему большинству, образованному всеми остальными, отличными от данного, объектами.

Мы исходим из интуитивного представления о том, что случайный объект должен обладать свойством типичности. Наша цель — дать строгое математическое определение этого свойства для последовательностей нулей и единиц при равномерном распределении вероятностей на пространстве Ω таких последовательностей. Последовательности, удовлетворяющие этому строгому определению, мы будем называть *типическими*, оставляя слово *типичный* для употребления на интуитивном уровне. В силу сказанного выше, нам надо определить сперва, что есть подавляющее большинство последовательностей, а затем — какие большинства следует считать разумными. А тогда класс типических последовательностей автоматически определится как пересечение всех разумных большинств.

Оставляя термин «подавляющее большинство» для интуитивного употребления, мы будем говорить о *больших* множествах последовательностей. Дополнение к большому множеству до всего Ω будем называть *малым*. Очевидно, достаточно определить, что такое малое множество, — а тогда большие множества определятся как дополнения к малым.

Итак, пусть дано какое-то множество Q последовательностей, $Q \subset \Omega$. Мы хотим определить, когда его следует считать *малым*. В терминах теории вероятностей мы сказали бы, что Q малое, коль скоро вероятность попадания в Q равна нулю. В терминах теории меры мы сказали бы, что Q малое, коль скоро оно имеет меру нуль. Мы, однако, постараемся определить, что такое малое множество в более простых терминах.

Множество Q называется *малым*, если его можно покрыть шарами, сумма объемов которых сколь угодно мала. Определению малого множества можно дать такую формулировку: множество Q называется *малым*, если для каждого натурального числа m найдется такая последовательность двоичных слов $\langle x(1), x(2), \ldots, x(n), \ldots\rangle$, что

$$Q \subset \bigcup_n \Omega_{x(n)};$$

$$\sum_n \mathbf{v}(\Omega_{x(n)}) = \sum_n 2^{-|x(n)|} < \frac{1}{m}.$$

Очевидно, каждая отдельная последовательность из Ω образует малое множество, а потому понятие большого множества не может претендовать на роль уточнения понятия *разумного* большинства. Пересечение всех больших множеств пусто.

«Разумность» вводится путем следующей корректировки определения.

Во-первых, потребуем, чтобы упоминаемая в определении последовательность двоичных слов $\langle x(1), x(2), \ldots, x(n), \ldots\rangle$ была вычислимой. Другими словами, мы требуем, чтобы существовал алгоритм, вычисляющий ее n-й член x_n по его номеру n.

Во-вторых, потребуем, чтобы такая вычислимая последовательность не просто существовала для каждого m, но строилась бы по этому m *эффективно*. Слово «эффективно» означает „с помощью алгоритма“. Здесь необходимы разъяснения. Дело в том, что алгоритм, получающий на вход m и выдающий на выходе требуемую последовательность, невозможен просто потому, что последовательность — это бесконечный объект, а алгоритмы оперируют лишь с конечными объектами. Однако в нашем случае, поскольку последовательность вычислима, то у нее есть алгоритм, который ее вычисляет — даже очень много таких алгоритмов. Алгоритмы (в другой системе терминов — программы алгоритмов) являются конечными объектами и потому вполне осмысленно говорить об алгоритме, который по числу m, поступившему на его вход, выдает на выходе один из алгоритмов (в другой системе терминов — одну из программ), вычисляющих последовательность $\langle x(1), x(2), \ldots, x(n), \ldots\rangle$.

Внеся в определение малого множества эти два добавления, мы получаем определение *эффективно малого* множества — а тем самым и определение *эффективно большого* множества. Пересечение всех эффективно больших множеств оказывается непустым. Более того, оно само является эффективно большим. Это наименьшее среди эффективно больших множеств и есть искомое множество $\mathbf{T}$ всех *типических* последовательностей.

Типические последовательности называют также *случайными по Мартин-Лёфу* — по имени ученика Колмогорова, замечательного шведского математика Пера Мартин-Лёфа, который в 1966 г. сформулировал только

что изложенное определение типичности в качестве строгого уточнения понятия случайности.

Как уже говорилось, класс **T** всех типических последовательностей совпадает с классом **C** всех хаотических последовательностей:

$$\mathbf{T} = \mathbf{C}.$$

Поэтому, как уже говорилось, типические последовательности можно именовать также *хаотическо-типическими* или *типическо-хаотическими*, а сам класс всех таких последовательностей обозначать **CT** или **TC**.

Как мы уже знаем,

$$\mathbf{CT} \subset \mathbf{S}, \quad \mathbf{CT} \neq \mathbf{S}.$$

Лицо Четвертое: Непредсказуемость

Вот начало «Двенадцати стульев» Ильфа и Петрова: «В уездном городе N было так много парикмахерских заведений и бюро похоронных процессий, что казалось, жители города рождаются лишь затем, чтобы побриться, остричься, освежить голову вежеталем и умереть». Суждение остается верным, если заменить N на Москву, парикмахерские на залы игровых автоматов, похоронные бюро на казино, а цель рождения на «играть». Этот печальный факт, однако, позволяет наглядно объяснить то свойство случайных последовательностей, которое мы называем непредсказуемостью.

Интуитивно ясно, что всякая случайная последовательность является *непредсказуемой* в том смысле, что в каком бы порядке мы ни выбирали ее члены, знание значений уже выбранных членов не позволяет предсказать значение того следующего члена, который мы намереваемся выбрать. Таким образом, Казино, обладающее такой последовательностью и предлагающее Игроку угадывать ее члены и делать при этом денежные ставки, не разорится; говоря более точно, Казино уверено, что никакой Игрок не может обладать такой стратегией игры, которая приведет к разорению Казино, каким бы капиталом оно ни обладало.

Непредсказуемость какой-либо последовательности, таким образом, определяется в терминах игры, которую Игрок ведет против обладающего этой последовательностью Казино, или, для краткости — *против данной последовательности*.

Итак, представим себе что Игрок приходит в Казино. Каждая из сторон — и Казино, и Игрок — обладает своим начальным капиталом. Казино располагает некоторой фиксированной, но неизвестной Игроку последовательностью нулей и единиц, и предлагает Игроку предсказывать ее члены — необязательно в монотонном порядке их следования и даже необязательно все ее члены.

Для наглядности представим себе, что члены последовательности написаны на картах, которые лежат рубашками вверх, так что Игрок не видит, что там написано. Последовательность предстает перед Игроком в виде бесконечного ряда таких карт. Игра состоит в том, что Игрок на каждом своем ходу указывает ту карту, которая должна быть открыта, одновременно предсказывая значение, которое обнаружится на этой карте, и объявляя размер денежной ставки. Если предсказание окажется правильным, Казино выплачивает Игроку сумму ставки, если неправильным — Игрок выплачивает эту сумму Казино. Считается, что Игрок *выиграл*, если он сумел разорить Казино. Разумеется, если Игроку открыт неограниченный кредит, он всегда может разорить Казино, удваивая ставки. Но Игра идет на наличные, так что величина ставки ограничена текущим капиталом Игрока.

Последовательность называется *предсказуемой*, если существует выигрывающий алгоритм игры. *Выигрывающим* мы называем алгоритм со следующим свойством: каким бы начальным капиталом ни обладало Казино, оно рано или поздно будет разорено, если Игрок применит этот алгоритм. Последовательность называется *непредсказуемой*, если она не является предсказуемой.

На математическом языке ситуация описывается так.

Рассматривается бесконечная последовательность нулей и единиц

$$\mathbf{a} = \langle a_1, a_2, a_3, \ldots \rangle.$$

При каждом ходе Игрока возникает тройка чисел

$$\langle n, i, v \rangle,$$

где

$$n \in \mathbb{N}, \quad i \in \{0,\ 1\}, \quad v \in \mathbb{Q}, \quad v \geqslant 0.$$

Содержательно: натуральное число n есть номер того члена последовательности, на который делается ставка; i есть предсказываемое значение этого члена; неотрицательное рациональное число v есть размер ставки. Ходы делаются друг за другом, начиная с первого; тройка, возникающая на k-м ходу, обозначается $\langle n(k), i(k), v(k) \rangle$. Более математически грамотно было бы сказать, что каждый ход *есть* тройка чисел, и что ходы не делаются, а *предъявляются*.

Капитал Игрока перед k-м ходом обозначается $V(k-1)$. Без ограничения общности можно считать, что начальный капитал Игрока равен единице: $V(0) = 1$.

После каждого хода капитал Игрока меняется (если только ставка не была нулевой). А именно:

- ▷ если $i(k) = a_{n(k)}$ (Игрок угадал), то $V(k) = V(k-1) + v(k)$;
- ▷ если $i(k) \neq a_{n(k)}$ (Игрок не угадал), то $V(k) = V(k-1) - v(k)$.

И еще два прибавления к сказанному.

Во-первых, ходы бывают *корректные* и *некорректные*, и чтобы игра продолжалась, необходимо, чтобы ход был корректным. А именно, k-й ход считается *корректным*, если выполнены оба нижеследующие требования:

1) номер открываемой карты *корректен*; это значит, что подлежащая открытию карта не была уже открыта ранее, т. е. $n(k)$ не совпадает ни с одним из чисел $n(1), \ldots, n(k-1)$;

2) делаемая Игроком ставка *корректна*; это значит, что она меньше его текущего капитала, т. е. $v(k) < V(k-1)$.

Если же хотя бы одно из этих требований не выполнено, ход считается *некорректным*.

Основное правило остановки. Если Игрок совершает некорректный ход, игра останавливается. При этом Игрок остается при имеющемся у него к данному моменту капитале — и тем самым заведомо не выигрывает.

Во-вторых, не исключается возможность того, что Игрок вообще не делает очередного хода (даже самого первого хода!), и в этом случае Игрок также остается при имеющемся у него к данному моменту капитале и, как и в случае некорректного хода, заведомо не выигрывает. Однако в этом случае — в отличие от случая некорректного хода — мы избегаем выражения «игра останавливается». Дело в том, что ситуацию неделания хода можно наглядно представить себе следующим образом. Перед каждым своим ходом Игрок решает, какой ход ему следует сделать. Решение требует обдумывания, и Игрок берет время на обдумывание. Время это ничем не ограничено, и процесс размышления может затянуться до бесконечности. В течение времени обдумывания хода капитал Игрока не меняется. Поэтому если Игрок ни за какое конечное время не принимает решения о своем ходе, его капитал застывает. В этом случае выигрыша Игрока произойти не может. Однако в этом случае (в отличие от случая некорректного хода) не наступает момента остановки игры — никогда не поступает сигнала об остановке. Таким образом, у игры есть три возможных сценария развития: 1) Игрок делает бесконечное число ходов; 2) Игрок делает лишь конечное число ходов, и причиной этого служит то, что был сделан некорректный ход; 3) Игрок делает лишь конечное число ходов, и причиной этого служит то, что на каком-то этапе игры Игрок не в состоянии прийти к решению об очередном ходе. Разумеется, сказанное носит иллюстративный характер, и математическое описание не включает в себя ссылку на такие понятия, как „решать“, „обдумывать“ и т. п.

По определению, Игрок *выигрывает* (при игре против **a**), если

$$\sup_k V(k) = +\infty,$$

т. е. если

$$\forall W\ \exists k\ V(k) > W.$$

Содержательно это означает, что Игрок разоряет Казино, каким бы исходным капиталом W Казино ни обладало. Очевидно, что Игрок в состоянии выиграть лишь при условии, что всякий раз, когда ему предстоит делать ход, он его делает и этот ход оказывается корректным.

Как выглядит игра, мы описали. Перейдем теперь к понятию системы игры, или *стратегии*. Смысл стратегии в том, чтобы избавить Игрока от необходимости самостоятельно принимать решения: стратегия берет эту функцию на себя. Стратегия есть правило, для каждого хода указывающее Игроку, какой на этом ходу он должен сделать ход (т. е. какую тройку предъявить). Разумеется, стратегия выдает такое указание лишь в том случае, если ход должен быть сделан. Выше уже отмечалась возможность того, что никакого хода не делается; в этом случае, естественно, стратегия не выдает никакого указания. При указании хода стратегия опирается на всю предшествующую историю игры. История же игры состоит из всех уже сделанных к рассматриваемому моменту ходов и из всех ставших уже известными членов последовательности. Мы лишь потому не включаем в историю игры информацию о капитале Игрока на каждый момент, что эта информация легко вычисляется из только что перечисленных сведений.

Таким образом, историю игры перед k-м ходом можно записать в виде таблицы

$$\begin{array}{cccc} n(1) & n(2) & \ldots & n(k-1) \\ i(1) & i(2) & \ldots & i(k-1) \\ v(1) & v(2) & \ldots & v(k-1) \\ a_{n(1)} & a_{n(2)} & \ldots & a_{n(k-1)} \end{array}$$

(Если $k = 0$, таблица пуста.)

На математическом языке стратегия есть функция, которая каждой подобной таблице (в том числе пустой) либо ничего не ставит в соответствие, либо ставит в соответствие некоторый ход, т. е. тройку $\langle n, i, v\rangle$. Под «каждой подобной таблицей» мы понимаем отнюдь не только такую таблицу, которая отражает реальное течение игры, а произвольную таблицу, в которой в первой строке стоят положительные целые числа, в третьей — неотрицательные рациональные числа, а во второй и четвертой — нули и единицы.

Если таблица реально встретилась в процессе игры (в качестве истории игры на каком-то этапе) и если задана стратегия, то первые три

строки таблицы однозначно восстанавливаются по ее четвертой строке. В самом деле, применение стратегии к пустой таблице дает нам первый ход $\langle n(1), i(1), v(1)\rangle$. Тем самым — поскольку четвертая строка предполагается известной — мы получаем историю игры перед вторым ходом в виде таблицы

$$\begin{array}{c} n(1) \\ i(1) \\ v(1) \\ a_{n(1)} \end{array}$$

Теперь к этой таблице снова применяем стратегию, получаем второй ход $\langle n(2), i(2), v(2)\rangle$ и таблицу

$$\begin{array}{cc} n(1) & n(2) \\ i(1) & i(2) \\ v(1) & v(2) \\ a_{n(1)} & a_{n(2)} \end{array}$$

И так далее.

Сказанное дает нам право при определении стратегии брать в качестве аргумента не всю таблицу в целом, а лишь ее четвертую строку. Заметим, что в этой четвертой строке стоит двоичное слово, т. е. элемент множества Ξ. Стратегия должна, имея этот элемент, или не выдавать ничего, или выдавать ход, который есть тройка, т. е. элемент декартова произведения $\mathbb{N}\times\{0,1\}\times\mathbb{Q}^{\geqslant 0}$. Здесь символом $\mathbb{Q}^{\geqslant 0}$ обозначено множество всех неотрицательных рациональных чисел.

Мы приходим к окончательному определению понятия стратегии: *стратегия* есть отображение некоторого подмножества множества Ξ всех двоичных слов в декартово произведение $\mathbb{N}\times\{0,1\}\times\mathbb{Q}^{\geqslant 0}$:

$$\Xi \longrightarrow \mathbb{N}\times\{0,1\}\times\mathbb{Q}^{\geqslant 0}.$$

Для наших целей особый интерес представляет случай, когда указанное отображение задается каким-то алгоритмом. Поясним, что это значит. Пусть **A** — алгоритм, на вход которого могут подаваться элементы из множества X, а на выходе получаются элементы из множества Y. В множестве X выделяется *область результативности* алгоритма **A**, состоящая из тех и только тех элементов, в применении к которым **A** дает результат. Алгоритм **A** *задает* следующее отображение: область определения этого отображения совпадает с областью результативности алгоритма, и для каждого элемента этой области значение отображения на этом элементе совпадает с тем результатом, который получается при применении алгоритма к этому элементу.

Если стратегия задана каким-то алгоритмом, она называется *вычислимой*. (Если подать на вход задающего стратегию алгоритма такую

историю игры, для которой очередной ход не определен, то алгоритм работает на этом входе бесконечно долго, не приходя ни к какому результату, но и не выдавая сообщение об отсутствии такового. Эта бесконечная работа алгоритма как раз и происходит за то взятое Игроком на обдумывание бесконечное время, о котором говорилось выше.)

Стратегия называется *выигрывающей для последовательности* **a**, если Игрок, применяющий эту стратегию в игре против **a**, выигрывает.

Последовательность называется *предсказуемой*, если для нее существует выигрывающая вычислимая стратегия, и *непредсказуемой* в противном случае.

Класс всех непредсказуемых последовательностей обозначается буквой **U**.

Известно, что всякая непредсказуемая последовательность является стохастической по Колмогорову (принадлежит классу **S**) и что всякая типическо-хаотическая последовательность (последовательность класса **CT**) непредсказуема:

$$\mathbf{CT} \subset \mathbf{U} \subset \mathbf{S}.$$

Известно также, что класс стохастических по Колмогорову последовательностей существенно шире класса непредсказуемых:

$$\mathbf{S} \neq \mathbf{U}.$$

Открытым остается вопрос о совпадении классов хаотических (он же класс типических) и непредсказуемых последовательностей:

$$\mathbf{CT} = \mathbf{U}??$$

Эта важная проблема ждет своего решения.

О БЕЗОСТАНОВОЧНЫХ СТРАТЕГИЯХ. Если игра никогда не останавливается, она называется *безостановочной*. Стратегия называется *безостановочной*, если какова бы ни была последовательность, применение против нее этой стратегии приводит к безостановочной игре. В определении предсказуемости можно ограничиться безостановочными стратегиями и дать такую равносильную формулировку: последовательность называется *предсказуемой*, если для нее существует выигрывающая вычислимая безостановочная стратегия. Чтобы убедиться в равносильности, достаточно объяснить, как можно алгоритм **A**, задающий выигрывающую стратегию, переделать в алгоритм **B**, задающий выигрывающую безостановочную стратегию. Такая переделка осуществляется весьма просто. Сперва, по поступившему на вход алгоритма **A** двоичному слову восстанавливается история игры, что позволяет знать как номера всех открытых ранее карт, так и величину текущего капитала Игрока. Затем всякий ход, предписываемый исходным алгоритмом **A**, проверяется на

корректность, и если он оказывается некорректным, то новый алгоритм **B** никакого хода не выдает, а объявляется не определенным на указанном двоичном слове.

Обобщение на вычислимые распределения вероятностей

Предварительные замечания

До сих пор, для простоты и большей наглядности, мы ограничивались ситуацией, когда на пространстве Ω всех двоичных последовательностей было задано *равномерное распределение вероятностей*. Все основные идеи были видны на примере этой простейшей ситуации. Для полноты картины мы намерены рассмотреть теперь общую ситуацию, когда на пространстве Ω задано произвольное *вычислимое распределение вероятностей*. Что это такое, будет разъяснено позже. А сейчас мы хотим сказать несколько слов, направленных на то, чтобы сделать изложение доступным и такому Читателю, который не знаком с общим понятием *распределения вероятностей*, или *вероятностной меры*.

Говорят, что на множестве M задана *мера* μ, коль скоро, во-первых, выделен некоторый класс подмножеств множества M, называемых *измеримыми*, и, во-вторых, каждому измеримому подмножеству A отнесено неотрицательное число $\mu(A)$, называемое мерой этого подмножества. При этом требуется, чтобы выполнялись некоторые аксиомы, приводить которые мы здесь не будем; отметим только, что следствием этих аксиом является такой факт: объединение непересекающихся измеримых множеств измеримо и его мера равна сумме мер этих множеств. Если $\mu(M) = 1$, мера называется *вероятностной*, или *распределением вероятностей*. В этом случае $\mu(A)$ содержательно трактуется как вероятность того, что случайно выбранный элемент множества M попадает в A.

Каждая мера на Ω характеризуется мерами шаров. Так, для равномерного распределения (и только для него) $\forall x \in \Xi\ \mu(\Omega_x) = 2^{-|x|}$.

Равномерное распределение вероятностей на пространстве Ω отвечает тому сценарию, когда нули и единицы возникают в последовательности с равными вероятностями. Ближайшим обобщением равномерного распределения является *распределение Бернулли*, или *бернуллиевское распределение* (называемое также *биномиальным*), при котором нули и единицы возникают с вероятностью p для единицы и вероятностью $1-p$ для нуля; это число p условимся называть *параметром* бернуллиевского распределения. Если параметр равен одной второй, получаем равномерное распределение. Говоря формально, распределение Бернулли с параметром p задается формулой $\mu(\Omega_x) = p^k(1-p)^{|x|-k}$, где k — количество единиц в слове x.

Следующим обобщением служит класс распределений, которые мы будем называть *квазибернуллиевскими*. Пусть дана последовательность действительных чисел $\mathbf{p} = \langle p(1), p(2), \ldots, p(k), \ldots\rangle$, $0 \leqslant p(k) \leqslant 1$. Про распределение μ будем говорить, что оно *квазибернуллиевское с параметром* $\mathbf{p}$, коль скоро для всякого двоичного слова $x = \langle x_1, \ldots, x_n\rangle$ имеет место равенство $\mu(\Omega_x) = \prod_{i=1}^{i=n} r_i$, где $r_i = p(i)$ при $x_i = 1$ и $r_i = 1 - p(i)$ при $x_i = 0$. На содержательном уровне это означает, что единицы и нули появляются с переменной вероятностью, зависящей только от номера члена. Для ясности: бернуллиевское распределение с параметром p является квазибернуллиевским с параметром $\langle p, p, \ldots, p, \ldots\rangle$.

В этой главке мы укажем как определения стохастичности, хаотичности, типичности и непредсказуемости должны быть расширены для общей ситуации *вычислимого распределения вероятностей* (что это такое, будет разъяснено ниже). Объявим наперёд, что при этих расширенных определениях для произвольного вычислимого распределения μ сохраняются те же соотношения, которые были выписаны выше для случая равномерного распределения:

$\mathbf{C}(\mu) = \mathbf{T}(\mu) \subset \mathbf{U}(\mu) \subset \mathbf{S}(\mu)$;

$\mathbf{S}(\mu) \neq \mathbf{U}(\mu)$ (при условии, что мера любого шара положительна).

Здесь через $\mathbf{C}(\mu)$, $\mathbf{T}(\mu)$, $\mathbf{U}(\mu)$, $\mathbf{S}(\mu)$ соответственно обозначены те классы последовательностей, хаотических, типических, непредсказуемых, стохастических по Колмогорову относительно распределения μ, кои будут определены ниже. (В наших прежних обозначениях $\mathbf{C} = \mathbf{C}(\eta)$, $\mathbf{T} = \mathbf{T}(\eta)$, $\mathbf{U} = \mathbf{U}(\eta)$ и $\mathbf{S} = \mathbf{S}(\eta)$, где η есть равномерное распределение.)

Эта главка — для любителей обобщений, и мы призываем уважаемого Читателя подумать, стоит ли ее читать. Во-первых, она несколько труднее предыдущих главок. Во-вторых, сама задача о формулировке строгого определения случайной последовательности делается тем менее ясной, чем более обширным становится класс рассматриваемых распределений. Ведь само представление о случайности индивидуальной последовательности имеет сколько-нибудь ясный интуитивный смысл лишь для простейших распределений вероятностей — таких, как равномерное и его ближайшие обобщения.

Вычислимые меры и вычислимые распределения

Хотелось бы называть меру μ на пространстве Ω вычислимой, если существует такой алгоритм, который по каждому двоичному слову x дает меру $\mu(\Omega_x)$ шара Ω_x. Однако алгоритмы не могут иметь дело с действительными числами, а только с числами целыми, рациональными

и т. п. — строго говоря, даже не с числами, а с их именами в виде слов в каком-либо фиксированном конечном алфавите. (*Словом в* данном алфавите называется любая конечная цепочка, составленная из букв этого алфавита.) Дать же имена всем действительным числам невозможно, поскольку для каждого конечного алфавита все слова в нём образуют лишь счетное множество. Поэтому можно требовать лишь наличие алгоритма, дающего не самоё меру шара, а сколь угодно точное рациональное приближение к ней.

Окончательное определение таково. Меру μ называют *вычислимой*, коль скоро существует алгоритм, который для каждого двоичного слова x и каждого положительного рационального числа ε выдает такое рациональное число, которое отличается от меры $\mu(\Omega_x)$ шара Ω_x не более, чем на ε.

Иногда в определение вычислимости включают еще и требование о существовании такого алгоритма, который для каждого двоичного слова x определяет, имеет ли место равенство $\mu(\Omega_x) = 0$. Вычислимые меры, удовлетворяющие этому дополнительному требованию (которое мы не включаем в определение вычислимости), будем называть *сильно вычислимыми*. Заметим. что из вычислимости меры не вытекает, как могло бы показаться, ее сильная вычислимость.

Важным подклассом класса сильно вычислимых мер является класс вычислимо-рациональных мер. Условимся называть меру *вычислимо-рациональной*, если мера каждого шара есть рациональное число и существует алгоритм, вычисляющий эту меру по заданному шару, т. е., на более точном языке, алгоритм, который для каждого двоичного слова x выдает дробь, выражающую меру $\mu(\Omega_x)$ шара Ω_x. Заметим для ясности, что если мера вычислима, а ее значения на шарах рациональны, то отсюда еще не следует, что она вычислимо-рациональна: умение находить для данного рационального числа сколь угодно точные рациональные приближения к нему еще не означает умения выписать само это число в виде отношения двух целых чисел!

Теперь нет нужды отдельно определять, что такое вычислимое распределение вероятностей, — это просто-напросто вычислимая вероятностная мера. Именно на вычислимые распределения естественно обобщаются многие конструкции и факты, изложенные нами для равномерного распределения. В частности, для произвольного вычислимого распределения верны как *теорема Мартин-Лёфа*, утверждающая, что пересечение всех эффективно больших множеств само является эффективно большим, так и *теорема Левина*, утверждающая, что типичность равносильна хаотичности, определенной на основе монотонной энтропии.

Стохастичность

Для какой-либо последовательности **e** ее k-й член обозначаем e_k или, дабы избежать многоэтажных индексов, $e(n)$.

В случае равномерного распределения стохастичность последовательности понималась как *глобальная устойчивость частот*, т. е. как устойчивость частот в *каждой* из допустимых подпоследовательностей. Допустимые же подпоследовательности возникали путем применения допустимых по Колмогорову правил отбора. В ситуации произвольной вероятностной меры общая схема сохраняется, только устойчивость частот заменяется на некоторое более общее свойство, а именно — на так называемый *закон больших чисел*.

Определение стохастичности для бернуллиевского распределения с параметром p очевидно: надо потребовать, чтобы в каждой допустимой подпоследовательности соблюдалась устойчивость частот с одной и той же предельной частотой p. Иными словами, доля единиц в начальном отрезке всякой допустимой подпоследовательности должна, при неограниченном удлинении отрезка, стремится к p. Случаи, когда $p = 0$ или $p = 1$ являются вырожденными; для первого из них не считается стохастической последовательность, в которой встречается хотя бы одна единица, для второго — последовательность, в которой встречается хотя бы один ноль.

Фон Мизес понимал вероятность как предельную частоту. Поэтому его идеология не простирается за пределы бернуллиевских распределений. Мостом к общему случаю произвольного распределения служат распределения квазибернуллиевские. Ими и займемся.

Ясно, что для квазибернуллиевского распределения предельной частоты в подпоследовательности может и не быть, а если она и есть, то, вообще говоря, в каждой подпоследовательности — своя. Приходится поэтому говорить о стохастичности относительно данного правила отбора. Пусть **p** — параметр квазибернуллиевского распределения. Прежде всего заявим, что не считается стохастической последовательность, в которой значение хотя бы одного ее члена имеет нулевую вероятность. Или, говоря на формальном языке: последовательность **a** не считается стохастической, если существует такое k, что либо $a(k) = 0$ и $p(k) = 1$, либо $a(k) = 1$ и $p(k) = 0$. Для остальных последовательностей определение таково. Последовательность **a** будем называть *стохастической относительно правила отбора* Θ, коль скоро ее обобщенная подпоследовательность

$$\mathbf{b} = \langle a(m_1), a(m_2), \ldots, a(m_k), \ldots \rangle,$$

полученная по этому правилу, удовлетворяет требованию закона больших

чисел:

$$\frac{a(m_1)+\cdots+a(m_k)}{k} - \frac{p(m_1)+\cdots+p(m_k)}{k} \to 0$$

при $k \to \infty$. Далее, последовательность $\mathbf{a}$ называется *стохастической* (*по Колмогорову*), если для любого колмогоровского правила, приводящего к бесконечной обобщенной подпоследовательности, она является стохастической относительно этого правила. (Замечание. Слово «бесконечной» в предыдущей фразе избыточно, так как всякая обобщенная подпоследовательность является последовательностью, а следовательно, бесконечна по определению. Тем не менее мы будем иногда использовать это формально избыточное упоминание о бесконечности, дабы подчеркнуть, что имеется в виду именно последовательность, а не кортеж.)

Чтобы объяснить, как понятие стохастичности по Колмогорову обобщается на более широкий класс вероятностных распределений, введем несколько обозначений.

Пусть $n(1), n(2), \ldots, n(k)$ суть натуральные числа и пусть $i(1), i(2), \ldots, i(k) \in \{0,1\}$. Через $A^{n(1),\ldots,n(k)}_{i(1),\ldots,i(k)}$ обозначим множество всех таких последовательностей $\mathbf{a} \in \Omega$, у которых

$$a_{n(1)} = i(1),\ a_{n(2)} = i(2),\ \ldots,\ a_{n(k)} = i(k). \qquad (*)$$

Дробь

$$\mu(A^{n(1),\ \ldots,\ n(k),\ m}_{i(1),\ \ \ldots,\ i(k),\ \ 1}) / \mu(A^{n(1),\ \ldots,\ n(k)}_{i(1),\ \ \ldots,\ i(k)})$$

обозначим символом $\mu\left(\begin{array}{c|ccc} m & n(1), & \ldots, & n(k) \\ 1 & i(1), & \ldots, & i(k) \end{array}\right)$; это есть условная вероятность того, что m-й член последовательности $\mathbf{a}$ будет равен 1 при условии $(*)$; эта величина не определена, коль скоро знаменатель дроби обращается в нуль.

Рассмотрим два произвольных, но фиксированных объекта: последовательность $\mathbf{a} \in \Omega$ и допустимое по Колмогорову правило отбора Θ. Наша цель — придать смысл высказыванию «последовательность $\mathbf{a}$ стохастична относительно правила Θ». Процесс, посредством которого правило Θ отбирает из последовательности $\mathbf{a}$ члены обобщенной подпоследовательности $\mathbf{b}$, состоит из двух этапов. На первом этапе строится вспомогательная обобщенная подпоследовательность $\mathbf{c}$, из некоторых членов которой на втором этапе и строится $\mathbf{b}$. Более подробно,

$$\mathbf{c} = \langle a_{n(1)}, a_{n(2)}, \ldots, a_{n(k)}, \ldots \rangle,$$

где номер $n(k)$ вычисляется алгоритмически по кортежу $\langle a_{n(1)}, a_{n(2)}, \ldots, a_{n(k-1)} \rangle$. Имея на входе этот же кортеж, правило Θ решает, включать или нет член $a_{n(k)}$ в окончательную последовательность $\mathbf{b}$. Таким образом,

$$\mathbf{b} = \langle a(n(k_1)), a(n(k_2)), \ldots, a(n(k_j)), \ldots \rangle.$$

На каждом из обоих этапов может наступить момент, когда правило Θ не даст никакого результата, т. е. не выдаст номера на первом этапе или не примет решения на втором этапе. Если такое произойдет, последовательность **b** окажется конечной — т. е., говоря более строго, окажется не последовательностью, а кортежем. В этом случае никакого требования к **b** не выдвигается и, говоря формально, **a** признаётся стохастической относительно Θ. Если же **b** оказалось бесконечной, то для признания последовательности **a** стохастической относительно Θ требуется выполнение нижеследующего свойства последовательности **a**.

Обозначим через r_j величину

$$\mu\left(\begin{array}{c|ccc} n(k_j) & n(1), & n(2)\ldots, & n(k_j-1) \\ 1 & a(n(1)), & a(n(2))\ldots, & a(n(k_j-1)) \end{array}\right).$$

Рассмотрим разность

$$\delta_j \;=\; \frac{r_1+r_2+\cdots+r_j}{j} - \frac{a(n(k_1))+a(n(k_2))+\cdots+a(n(k_j))}{j}.$$

Величина δ_j не определена, если не определена хотя бы одна из величин $r_1,\ldots,r_j$. Скажем, что **b** *подчиняется закону больших чисел*, если все величины δ_j определены и $\delta_j \to 0$ при $j\to\infty$. Последовательность **a** назовем *стохастической относительно правила* Θ, если обобщенная подпоследовательность **b**, полученная из **a** согласно правилу Θ, подчиняется закону больших чисел.

Наконец, последовательность **a** называется *стохастической* (*по Колмогорову*), если для любого допустимого по Колмогорову правила отбора она является стохастической относительно этого правила — при условии, что это правило строит бесконечную обобщенную подпоследовательность (а не конечный кортеж).

Заметим, что в самом определении вычислимость меры не используется. Однако без предположения о вычислимости едва ли возможно сравнивать частотоустойчивость (она же стохастичность) с другими алгоритмическими лицами случайности.

Хаотичность

Определение последовательности, хаотической относительно меры μ, таково. Последовательность $\mathbf{a} = \langle a_1, a_2, a_3, \ldots\rangle$ называется *хаотической относительно* μ, если существует такая (зависящая от меры μ) константа c, что для всех n выполняется неравенство

$$\mathrm{KM}(a_1, a_2, \ldots, a_n) > -\log\mu(\Omega_{a_1,a_2,\ldots,a_n}) - c,$$

где, как всегда, знак логарифма без нижнего индекса означает логарифм по основанию два.

Мотивировка этого определения, имеющего содержательный смысл для вычислимых вероятностных мер, такая же, как и в случае равномерного распределения вероятностей. Дело в том, что для всякой вычислимой вероятностной меры выполняется соотношение:

$$\exists c\ \forall x \in \Xi\ \ \mathrm{KM}(x) < -\log \mu(\Omega_x) + c.$$

Типичность

Понятие типичности обобщается на произвольную меру достаточно очевидным образом — надо просто в том определении типической последовательности, которое давалось для случая равномерного распределения, заменить объём шара на его меру.

Сперва определяется понятие эффективно малого множества. Множество $Q \subset \Omega$ называется *эффективно малым относительно меры* μ, коль скоро существует алгоритм $\mathbf{A}$, удовлетворяющий нижеследующему требованию. При поступлении на вход алгоритма $\mathbf{A}$ натурального числа m на его выходе возникает алгоритм построения такой последовательности двоичных слов $\langle x(1), x(2), \ldots, x(n), \ldots\rangle$ (т. е. алгоритм вычисления такой функции $n \mapsto x(n)$), что

$$Q \subset \bigcup_n \Omega_{x(n)};$$

$$\sum_n \mu(\Omega_{x(n)}) < \frac{1}{m}\,.$$

Далее, *эффективно большое относительно меры* μ множество определяется как дополнение (до Ω) к эффективно малому.

Для вычислимых мер имеет место *теорема Мартин-Лёфа*: объединение всех эффективно малых множеств является эффективно малым, а пересечение всех эффективно больших — эффективно большим. Существующее согласно этой теореме наименьшее эффективно большое множество называют *конструктивным носителем меры* μ. В случае, если μ — вычислимое распределение вероятностей, последовательности, принадлежащие конструктивному носителю распределения μ, и называют *типическими* относительно этого распределения. Таким образом, $\mathbf{T}(\mu)$ есть не что иное как конструктивный носитель распределения μ.

Непредсказуемость

Здесь мы укажем, что́, при переходе к произвольным распределениям, нужно добавить к определениям, сформулированным для равномерного случая. Таких добавлений будет два: некий мультипликативный коэффициент (для равномерного распределения он равен единице и потому не

нужен) и дополнительное правило остановки (в случае равномерного распределения оно потому не нужно, что не возникает такой ситуации, при которой оно могло бы быть применено).

Распределение вероятностей влияет на правило изменения капитала Игрока после очередного хода. Если Игрок не угадал, его капитал уменьшается на сумму сделанной им ставки — так же, как и в случае равномерного распределения. Но если Игрок угадал, то прирост его капитала равняется ставке, умноженной на некоторый коэффициент. Этот коэффициент сравнительно велик, если вероятность угадать была низка, и сравнительно мал, если вероятность угадать была высока. Для равномерного распределения вероятность угадать всегда равна одной второй, а коэффициент всегда равен единице. Точная формулировка правила прироста капитала будет сейчас изложена.

Для последовательности $\mathbf{a}$ ее k-й член обозначаем a_k, для последовательности $\mathbf{a}'$ ее k-й член обозначаем a'_k, и т. д.

Ход, делаемый на j-м ходу, есть тройка $\langle n(j), i(j), v(j)\rangle$.

Пусть последовательность, которой располагает Казино, есть $\mathbf{a}$. Положим

$$A(k-1) = \{\mathbf{a}' \in \Omega : a'_{n(j)} = a_{n(j)} \text{ при } j = 1, 2 \ldots, k-1\}$$

(так что $A(0) = \Omega$),

$$A_i(k) = \{\mathbf{a}' \in A(k-1) : a'_{n(k)} = i\} \text{ для } i = 0, 1.$$

Разумеется, эти обозначения имеют смысл в предположении, что все участвующие в их определениях номера $n(l)$ определены. Полезно заметить, что

$$\Omega = A(0) \supset A(1) \supset A(2) \supset \ldots; \tag{1}$$

$$1 = \mu(A(0)) \geqslant \mu(A(1)) \geqslant \mu(A(2)) \geqslant \ldots; \tag{2}$$

$$\text{если Игрок на } k\text{-м ходу угадал, то } \ i(k) = a_{n(k)},\ A_{i(k)}(k) = A(k); \tag{3}$$

$$\text{если Игрок на } k\text{-м ходу не угадал, то } \ i(k) \neq a_{n(k)},\ A_{1-i(k)}(k) = A(k); \tag{4}$$

$$A(k-1) = A_0(k) \cup A_1(k). \tag{5}$$

В случае, если $i(k) = a_{n(k)}$ (прогноз Игрока на его k-м ходе оправдался — он угадал), капитал Игрока увеличивается по формуле

$$V(k) = V(k-1) + v(k)\mu(A_{1-i(k)}(k))/\mu(A_{i(k)}(k)). \tag{6}$$

Наша формула (6) гарантирует «честность» игры: математическое ожидание выигрыша, то есть прироста капитала, за один ход равно нулю. Однако на пути применения этой формулы нас подстерегает неприятность, невозможная для равномерного распределения — как и для любой

позитивной меры (мера называется *позитивной*, если мера любого шара положительна). Неприятность заключается в том, что стоящая в знаменателе вероятность $\mu(A_{i(k)}(k))$ может оказаться равной нулю. Эта неприятность устраняется путем введения Дополнительного правила остановки, которое мы сейчас приведем. В случае равномерного распределения это правило было излишним, поскольку предусмотренная в нём ситуация не могла встретиться.

Дополнительное правило остановки. Оно применяется тогда, когда в ходе игры *впервые* наступает такая ситуация, что $\mu(A(k)) = 0$ (просим Читателя взглянуть на формулу (2)). Пусть $\mu(A(k-1)) \neq 0$, $\mu(A(k)) = 0$. Последний ход, который был сделан, имел номер k. Делая этот ход, Игрок объявил свой прогноз $i(k)$. Если прогноз оказался верным, то есть оказалось, что $i(k) = a_{n(k)}$, игра останавливается, а капитал Игрока считается возросшим до бесконечности: $V(k) = +\infty$. Таким образом, в этом случае Игрок объявляется выигравшим. Если же прогноз оказался неверным, т. е. оказалось, что $i(k) \neq a_{n(k)}$ (или, что то же самое, $1 - i(k) = a_{n(k)}$), игра опять-таки останавливается, но с тем капиталом Игрока $V(k)$, который у него был к этому моменту. Таким образом, в этом случае Игрок не выигрывает.

Ясно, что проблема с нулевым знаменателем в формуле (6) устраняется этим правилом. Действительно, формула (6) применяется лишь в случае, когда $i(k) = a_{n(k)}$. В этом случае, согласно (3), $A_{i(k)}(k) = A(k)$. Поэтому интересующий нас знаменатель может оказаться нулевым лишь в ситуации, когда $\mu(A(k)) = 0$. Но именно эта ситуация как раз и регулируется не формулой (6), а нашим Дополнительным правилом. (Можно, впрочем, считать, что и формулой (6), — если разрешить деление на нуль положительного числа $\mu(A_{1-i(k)}(k))$ и принять бесконечность в качестве результата такого деления.)

Определения стратегии, вычислимой стратегии, безостановочной стратегии, выигрывающей стратегии остаются, с учетом сделанных добавлений, для общего случая теми же, как и для частного случая равномерного распределения.

О безостановочных стратегиях. В заключение прокомментируем те предусмотренные правилами ситуации, когда игра останавливается. Заметим, что остановки не включены в понятие стратегии. Стратегия лишь указывает очередной ход или не указывает ничего, если таковой не определен; но в последнем случае игра не останавливается, а просто Игрок задумывается навечно — для вычислимой стратегии это означает, в математических терминах, что алгоритм указания хода работает безостановочно, не приходя ни к какому результату. Остановки же регулируются

своими собственными правилами, «внешними» по отношению к стратегии. К сожалению, в общем случае вычислимого распределения наши правила остановки не содержат алгоритма их применения. Более того, алгоритм, позволяющий в любой момент определять, следует или не следует останавливать игру, в общем случае невозможен. Конкретно, для Основного правила речь идет о проверке неравенства, связывающего ставку и капитал, а для Дополнительного правила — о проверке положительности меры некоторого множества. Разумеется, если мера сильно вычислима, сложность со второй из указанных проверок отсутствует. Более того, для всякой вычислимо-рациональной (в частности, равномерной) меры очевидны алгоритмы обеих проверок, но, повторяем, подобные алгоритмы могут и отсутствовать в случае более сложно устроенной вычислимой меры.

Ввиду сказанного представляет интерес утверждение о том, что при определении понятия предсказуемости можно ограничиться одними только безостановочными стратегиями. Именно для обоснования этого утверждения нам и потребуется предположение о вычислимости меры (отметим, что до сих пор оно не использовалось). В этом предположении мы сейчас укажем, каким способом алгоритм **A**, задающий выигрывающую стратегию, переделывается в алгоритм **B**, задающий выигрывающую безостановочную стратегию.

Итак, нам задан алгоритм **A**, на вход которого поступает двоичное слово $x \in \Xi$, а на выходе, в качестве результата, либо не появляется ничего, либо появляется очередной ход. Содержательно, двоичные цифры, составляющие это слово, суть значения уже открытых к этому моменту членов последовательности. Как мы знаем, по слову x восстанавливается вся предыдущая история игры, т. е. все числа $n(k)$, $i(k)$, $v(k)$, $V(k)$ при $k \leqslant s = |x|$. Разумеется, эти числа восстанавливаются, если только они определены (кстати, можно всегда предполагать, что все они определены, если только определен и корректен s-й ход, каковой есть тройка $\langle n(s), i(s), v(s)\rangle$). Чтобы подчеркнуть зависимость указанных чисел от x, будем писать $n(x,k)$, $i(x,k)$, $v(x,k)$, $V(x,k)$; ведь когда мы ранее писали просто $n(k)$, $i(k)$ и т. д., мы предполагали фиксированной ту последовательность, против которой идет игра и «частью» которой является слово x. Вместе с указанными числами возникают, при всех $k \leqslant s$, и те множества, которые для фиксированной последовательности мы обозначали ранее как $A(k)$; теперь мы будем обозначать их $A(x,k)$.

Наша цель — переделать алгоритм **A** в алгоритм **B**, никогда не приводящий к ситуации, требующей остановки игры и такой, что если **A** был выигрывающим, то и **B** — выигрывающий. Для вычислимо-рациональной меры переделка очевидна: получив на вход двоичное слово x, проверяем,

не приводит ли ход, указанный для этого слова алгоритмом **A**, к ситуации остановки; если не приводит, объявляем, что **B** указал нам этот ход; если приводит, объявляем, что **B** на x не дает результата.

В общем же случае произвольной вычислимой меры переделка **A** в **B** требует более изощренного приёма. Новый алгоритм **B** будет работать так. В случае, если ход, указанный алгоритмом **A**, не приводит к ситуации остановки, этот ход — как и для вычислимо-рациональной меры — объявляется результатом алгоритма **B**. Однако в случае, если в результате применения **A** к x возникает ситуация остановки, новый алгоритм **B** не будет давать никакого результата в применении к x. Тем самым окажется, что **B** задает безостановочную стратегию.

Для наглядности представим себе, что **A** и **B** работают параллельно. Напомним, что **A** нам задан, **B** мы строим. Пусть на вход обоих алгоритмов поступило слово x, $|x| = s$. Если **A** не дает результата в применении к этому слову, то и **B** объявляется не дающим результата. Если же **A** в применении к x выдает очередной ход, то алгоритм **B** пытается прежде всего проверить, не возникла ли ситуация, подпадающая под действие одного из правил остановки. Проверка корректности номера не вызывает затруднений. Если номер $n(x, s)$ открываемой карты оказался некорректным, алгоритм **B** объявляется не дающим результата на входе x; капитал Игрока в этом случае перестает меняться, он застывает. Если же номер корректен, включается проверка корректности ставки.

С проверкой корректности ставки дело обстоит сложнее. Как уже отмечалось, может и не существовать алгоритма, правильно отвечающего на вопрос, меньше ли делаемая ставка текущего капитала Игрока. Однако не всё так плохо. Прежде чем сделать ставку, Игроку следует убедиться, что намеченная ставка меньше текущего капитала, и только в этом случае ее объявить. Изложим новый алгоритм **B** поведения Игрока более точно. Мы опираемся на следующий факт: если мера μ вычислима, то можно предъявить такой вспомогательный алгоритм **C** (зависящий от **A**), который для каждого x пытается проверить, выполняется ли неравенство $v(x, s) < V(x, s)$, и достигает при этом следующего эффекта: алгоритм приходит к какому-то результату (скажем, к результату *Да*), если указанное неравенство выполняется, и работает бесконечное время, не приходя ни к какому результату, если выполняется противоположное неравенство $v(x, s) \geqslant V(x, s)$. Собираясь сделать свой s-й ход и получив от алгоритма **A** рекомендацию о ставке $v(x, s)$, Игрок, прежде чем сделать ход, включает алгоритм **C** и ждет результата. Никакого хода не делается, пока **C** не предъявит ему свой результат. Таким образом, алгоритм **B** не выдаст никакого хода, если **A** предложил некорректный ход; в этом случае капитал Игрока застывает. Если же работа алгоритма **C** заканчивается, алгоритм **B** рекомендует ту же ставку, что и **A**, и приступает

к проверке того, не следует ли применить Дополнительное правило остановки.

Алгоритмическая ситуация с Дополнительным правилом похожа на ситуацию с правилом остановки, делаемой по причине некорректности ставки. Алгоритм, распознающий, выполняется или нет равенство $\mu(A(x,s)) = 0$, легко строится в случая сильно вычислимой меры, поскольку множество $A(x,s)$ получается в результате объединения конечного числа шаров, указываемых алгоритмически по паре $\langle x, s\rangle$. В общем же случае вычислимой меры такого алгоритма может и не быть. Существует, однако, более слабый алгоритм, которого оказывается достаточно. Именно, если мера μ вычислима, то можно предъявить такой вспомогательный алгоритм **D** (зависящий от **A**), который для каждого x пытается проверить, выполняется ли неравенство $\mu(A(x,s)) \neq 0$, и достигает при этом следующего эффекта: алгоритм приходит к какому-то результату (скажем, к результату *Да*), если указанное неравенство выполняется, и работает бесконечное время, не приходя ни к какому результату, если выполняется равенство $\mu(A(x,s)) = 0$. Действия алгоритма **B** в связи с проверкой Дополнительного правила описываются — в терминах поведения Игрока — следующим образом. Игрок включает алгоритм **D** и ждет результата. Дождавшись, он делает ход, применяя алгоритм **A**. Но он не делает никакого хода, пока **D** не предъявит ему свой результат. Таким образом, алгоритм **B** выдаст тот же ход, что и **A**, если не возникло ситуации, подпадающей под Дополнительное правило, и не выдаст никакого хода, если такая ситуация возникла. Если ход происходит, то капитал Игрока меняется на общих основаниях — уменьшается на величину ставки или увеличивается согласно формуле (6), причем в этом случае опасный знаменатель заведомо не равен нулю. В случае же неделания хода капитал Игрока, согласно нашим правилам, застывает, если последний прогноз был неверен, и объявляется бесконечным, если этот прогноз был верен.

Только что описанное поведение капитала Игрока можно следующим образом описать в терминах работы алгоритма **D**. Если прогноз неверен ($i(x) \neq a_{n(x)}$), то пока **D** работает, капитал Игрока остается застывшим. Если же прогноз верен ($i(x) = a_{n(x)}$), то с каждым шагом работы алгоритма **D** капитал Игрока возрастает на единицу. Таким образом, капитал будет возрастать неограниченно (и, следовательно, Игрок выиграет), если **D** никогда не закончит работу. Но если и когда алгоритм **D** достигает результата, рост капитала Игрока прекращается. Более того, в этом случае весь произошедший за время работы алгоритма **D** прирост капитала Игрока аннулируется, после чего этот капитал прирастает по формуле (6). И еще: никакое значение капитала, промежуточное между делаемыми ходами (т. е. возникающее во время *завершающейся* работы

алгоритма **D**), не должно включаться в множество тех значений капитала, бесконечность супремума коих влечет выигрыш Игрока (в противном случае у Игрока появилась бы дополнительная возможность выигрыша за счет того, что длительности междуходовых задержек могут оказаться неограниченными в их совокупности).

Наше новое описание поведения капитала Игрока, предполагающее рост капитала во время работы алгоритма **D** с возможным последующим обнулением прироста, может вызвать у просвещенного Читателя два законных замечания.

Первое замечание состоит в том, что мы фактически изменили правила игры — в той их части, в которой говорится о порядке изменения капитала. Согласившись с этим упреком Читателя, ответим следующее. Действительно, есть две системы правил, старая и новая — а тем самым и два определения игры. Но в отношении предсказуемости обе игры равносильны: всякая последовательность, предсказуемая при одной из наших двух систем правил, предсказуема и при другой.

Второе замечание состоит в том, что по новым правилам поведение капитала зависит от вспомогательного алгоритма **D**, а сам этот алгоритм, как показывает анализ, зависит не от меры μ как таковой, а от задающего ее алгоритма — т. е. от того алгоритма, присутствующего в определении вычислимой меры, который для любого шара находит сколь угодно точные приближения к мере этого шара. Возникает естественный вопрос, не может ли случиться, что при новых правилах и понятие предсказуемости окажется зависящим не от самой меры, а от задающего ее алгоритма. Ответ: нет, не может. Ведь течение игры при одном задающем меру алгоритме может отличаться от течения игры при другом алгоритме, задающем ту же меру, лишь длительностью интервалов между ходами; что же касается капитала Игрока, то разница может быть лишь в том, сколь велик окажется тот прирост капитала, который подлежит аннулированию.

История и библиография

[1] А. Н. Колмогоров, В. А. Успенский. *Алгоритмы и случайность* // Теория вероятностей и ее применения, 1987. Т. 32, вып. 3, с. 425–455.

[2] В. А. Успенский, А. Л. Семёнов. *Теория алгоритмов: основные открытия и приложения.* — М.: Физматлит, 1987. — 288 с.

[3] В. А. Успенский, А. Л. Семёнов, А. Х. Шень. *Может ли индивидуальная последовательность нулей и единиц быть случайной?* // Успехи математических наук, 1990. Т. 45, вып. 1, с. 105–162.

[4] V. A. Uspensky, A. Shen. *Relations between varieties of Kolmogorov complexities* // Mathematical Systems Theory, 1996. Vol. 29, no.3, p. 271–292.

[5] An. A. Muchnik, A. L. Semenov, V. A. Uspensky. *Mathematical metaphysics of randomness* // Theoretical Computer Science, 1998. Vol. 207, p. 263–317.

[6] А. Шень. *О соотношениях между различными алгоритмическими определениями случайности* // Доклады Академии наук СССР, 1988. Т. 302, № 3, с. 548–552.

[7] В. В. Вьюгин. *Алгоритмическая энтропия (сложность) конечных объектов и ее применение к определению случайности и количества информации* // Семиотика и информатика. Выпуск 16. — М: ВИНИТИ, 1981. — С. 14–43.

[8] M. Li, P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications.* — New York e. a.: Springer-Verlag, 1993. — xx+546 pp., 38 illustrations.;
Second edition. — New York e. a.: Springer-Verlag, 1997. — xx+637pp., 41 illustrations.

Сам этот список из восьми названий никоим образом не претендует на полноту. Однако в этих публикациях (особенно в [8]) можно найти дальнейшие ссылки, и вкупе с этими ссылками некое подобие полноты уже достигается.

В книге [2] следует обратить внимание на § 2.6 «Приложения к теории вероятностей: определения случайной последовательности». Следует также иметь в виду, что терминология этой книги несколько архаична: хаотические последовательности называются там *случайными по Колмогорову*, а типические — *случайными по Мартин-Лёфу*. Частотоустойчивые, они же стохастические, последовательности называются в [2] *случайными по Мизесу*, стохастические по Чёрчу — *случайными по Мизесу – Чёрчу*, стохастические по Колмогорову — *случайными по Мизесу – Колмогорову – Лавлэнду* (а в [3] — *стохастическими по Колмогорову – Лавлэнду*; Д. Лавлэнд ⟨D. Loveland⟩ открыл эти последовательности независимо от Колмогорова, хотя и позже: соответствующая статья Колмогорова была опубликована в 1963 г., статья Лавлэнда — в 1966 г.) Что касается непредсказуемых последовательностей — в том точном понимании, как было сформулировано выше, — то они в указанной книге отсутствуют, поскольку появились в литературе лишь в 1998 г. в статье [5].

Пример стохастической по Чёрчу последовательности, перестающей быть таковой после подходяще выбранной вычислимой перестановки ее членов, опубликовал в 1966 г. Д. Лавлэнд. Значение этого примера состоит не только в том, что он показывает неадекватность определения Чёрча, но и в том, что он открывает новое качество случайности, интуитивно очевидное, но ранее не замеченное: случайная последовательность должна оставаться случайной после любой вычислимой (т. е. задаваемой каким-то алгоритмом) перестановки ее членов.

Теория сложности объектов (в дополнение к уже развивавшейся в то время теории сложности вычислений) была основана Колмогоровым на его семинарах в Московском университете в начале 60-х годов XX века и имела своей главной целью перестройку понятийной базы теории информации (на основе представления о том, что чем сложнее объект, тем больше информации он содержит) с последующим приложением к теории случайности. В опубликованной в 1969 г. статье Колмогоров писал:

> 1) основные понятия теории информации должны и могут быть обоснованы без помощи обращения к теории вероятностей и так, что понятия «энтропия» и количество информации оказываются применимы к индивидуальным объектам;
>
> 2) введенные таким образом понятия теории информации могут лечь в основу новой концепции случайного, соответствующей естественной мысли о том, что случайность есть отсутствие закономерности.

Предложение измерять сложность объекта длиной его кратчайшего описания и понятие оптимального языка были изложены Колмогоровым в статье 1965 г.; за год до того сходные идеи были опубликованы американским исследователем Рэем Соломоновым ⟨Ray Solomonoff⟩, о работах которого Колмогоров узнал лишь позже. Ввиду этого теорему о существовании оптимального языка мы называем *теоремой Соломонова – Колмогорова*. В эти же годы (т. е. в середине 60-х годов XX в.) Колмогоров высказывает на своих семинарах предположение о том, что быстрота роста энтропии начальных отрезков последовательности может служить критерием случайности рассматриваемой последовательности. Однако введенное им в рассмотрение языковóе семейство приводило к энтропии, непригодной для поставленной цели. Как уже говорилось выше, в главке о хаотичности, годное для этой цели семейство обнаружил в 1973 г. Леонид Левин, введя в рассмотрение монотонную энтропию.

Типические последовательности (под названием “random”, т. е. «случайные») были, как уже отмечалось в главке о типичности, открыты в 1966 г. Пером Мартин-Лёфом ⟨Per Martin-Löf⟩.

Утверждение о том, что класс последовательностей, стохастических по Колмогорову, шире класса типических (они же — хаотические) последовательностей, доказал Александр Шень (см. [6], а также [3, п. 6.2.4]).

Пусть K — энтропия в каком-то из вариантов этого понятия. (В литературе исследованы не менее шести таких вариантов: простая, априорная, монотонная, процессная, префиксная и энтропия разрешения; все эти варианты существенно различны в том точном смысле, что разность энтропий, принадлежащих любым двум из перечисленных вариантов, неограничена.) Скажем, что последовательность $a_1, a_2, a_3, \ldots, a_n \ldots$ является *хаотической относительно* K (при равномерном распределении!), если

$$\exists c\, \forall n\, \mathrm{K}(a_1, a_2, a_3, \ldots, a_n) > n - c.$$

(Заметим для ясности, что ни для простой энтропии, ни для энтропии разрешения последовательностей с таким свойством нет вовсе, а для каждой из остальных четырех энтропий хаотичность оказывается равнообъемной типичности.)

В той же статье Левина 1973 г., в которой была введена монотонная энтропия, содержалось и доказательство того факта, что понятие хаотичности относительно монотонной энтропии равнообъемно понятию типичности. Независимо от Левина в том же 1973 г. Клаус-Петер Шнорр ⟨Claus-Peter Schnorr⟩ ввел в рассмотрение свой вариант энтропии, так называемую *процессную энтропию* (у Шнорра — *process complexity*), и доказал (независимо от Левина, но очень похожим способом), что понятие хаотичности относительно процессной энтропии также равнообъемно понятию типичности. Хотя процессная энтропия и монотонная энтропия существенно различаются (как показал Владимир Вьюгин [7, с. 35, строки 6–4 снизу], их разность не ограничена никакой константой) и хотя впоследствии сам Шнорр перестал пользоваться своей процессной энтропией, фактически от нее отказавшись, вышеуказанную *теорему Левина* иногда называют *теоремой Левина – Шнорра*.

Префиксную энтропию ввел в 1974 г. Левин и годом позже (но независимо от Левина) Грегори Чэйтин ⟨Gregory J. Chaitin⟩ — см. его статью A theory of program size formally identical to information theory // Journal of the Association of Computing Machinery, 1975, v. 22, no. 3, p. 329–340). В той же статье Чэйтин ввел понятие хаотичности относительно префиксной энтропии и объявил (без доказательства), что этот вариант хаотичности равносилен типичности; первое опубликованное доказательство этого факта появилось в статье Вьюгина [7]: следствие 3.2 на с. 38. *Префиксная энтропия* может быть определена как энтропия для семейства префиксных языков. Язык E называется *префиксным*, если он перечислим и выполнено условие:

$$[\langle x_1, y_1\rangle \in E \;\&\; \langle x_2, y_2\rangle \in E \;\&\; (x_1 \approx x_2)] \Longrightarrow [y_1 = y_2].$$

Заметим еще, что в литературе термин «сложность» (“complexity”) часто употребляется в смысле „энтропия“, т. е. в смысле „сложность относительно оптимального языка“.

Непредсказуемые последовательности (в смысле настоящей статьи) впервые возникли весной 1991 г. в совместном докладе под названием “Randomness and Lawlessness” («Случайность и беззаконность»), который Андрей Мучник, Алексей Семёнов и автор этих строк сделали на конференции в Калифорнии. Конференция была посвящена основаниям теории случайности и проходила с 4 по 7 марта в Институте математических исследований в социальных науках (Institute for Mathematical Studies in the Social Sciences) Станфордского университета. Содержание доклада было опубликовано в 1998 г. в виде статьи [5]. В этой статье приведены, в частности, теоремы о соотношении непредсказуемости с другими алгоритмическими лицами случайности.

Отметим, что определение непредсказуемости в нашем настоящем очерке отличается одной деталью от определения в [5]. Именно, в [5] для корректности ставки требовалось выполнение не строгого неравенства $v(k) < V(k-1)$, а более слабого нестрогого неравенства $v(k) \leqslant V(k-1)$. Класс непредсказуемых последовательностей остается одним и тем же при обоих пониманиях корректности. Замена в определении корректности нестрогого неравенства на строгое вызвана двумя причинами. Во-первых, сама игра становится более содержательной: ведь в случае равенства ставки текущему капиталу стоит Игроку ошибиться в своем предсказании, как его капитал обнулится, и он будет вынужден в дальнейшем

делать лишь нулевые ставки. Во-вторых, именно строгое неравенство позволяет при переносе определений с рационально-вычислимых мер (каковые только и рассматривались в [5]) на любые вычислимые меры, сохранить эффективность действий игрока. Ведь каждый раз Игрок должен удостоверяться в корректности хода. Получить же такое удостоверение алгоритмическим путем возможно лишь в варианте строгого неравенства: существует алгоритм, в том и только в том случае дающий положительный ответ на вопрос «$v(k) < V(k-1)$?», когда и в самом деле $v(k) < V(k-1)$; не существует алгоритма, в том и только в том случае дающего положительный ответ на вопрос «$v(k) \leqslant V(k-1)$?», когда и в самом деле $v(k) \leqslant V(k-1)$.

Сама идея о связи случайности с невозможностью гарантированного выигрыша достаточно очевидна: еще фон Мизес, не давая точных формулировок, говорил о «невозможности системы игры». Впоследствии встречались и строгие определения, однако воспроизводимая здесь (с косметическим ремонтом) формулировка из [5] кажется более близкой к интуитивному представлению о случайности. Дело в том, что предшествующие игровые определения либо использовали стратегии, в которых вычислимость (означающая наличие алгоритма, указывающего игроку его очередной ход) заменялась на другое (хотя и связанное с понятием алгоритма, но, видимо, менее естественное) требование, либо заведомо не приводили к классу последовательностей, равнообъемному классу типическо-хаотических последовательностей. Для определения непредсказуемости из [5] остается надежда на указанную равнообъемность. Ясно, что чем бóльшим количеством «лиц случайности» характеризуется какой-то точно очерченный класс последовательностей, тем обоснованнее право этого класса служить формальным аналогом расплывчатого интуитивного представления о случайности.

В. А. Успенский, мехмат МГУ