Игровой подход к определению случайности бесконечных последовательностей

Ан.А. Мучник

Как определить случайность бесконечной последовательности? Общий способ был предложен A.Wald'ом в 1938г. Если дан язык, то среди счетного числа определимых на нем множеств бесконечных последовательностей можно выделить те, которые имеют меру 0. Случайными относительно данного языка естественно называть последовательности, не лежащие ни в одном из выделенных множеств. Ясно, что почти все последовательности будут случайными. Этот подход связан с неопределенностью: какой язык выбрать?

Другой способ — это определять случайность бесконечной последовательности через случайность ее конечных начал. На этом пути Л.А.Левиным и C.P.Schnorr'ом независимо в 1973г. [1,2] было введено удовлетворительное определение. Соответствующее понятие случайности мы примем за эталон. В этом определении неслучайность конечной последовательности имеет меру, называемую дефектом случайности.

В этой работе мы рассмотрим третий подход. Представим себе упрощенную схему Казино-Игрок. Казино предлагает для игры карточную колоду (бесконечную). На лицевой стороне каждой карты изображена двоичная цифра. Игрок выбирает любую карту (лежащую вниз лицевой стороной) и делает ставку на одну из двух цифр. После этого карта переворачивается и определяется кто выиграл. Инициатива в игре все время исходит от Игрока. Например, он всегда может уйти. Поэтому неограниченность ставок сверху или равенство выигрыша проигрышу давало бы возможность Игроку выиграть у Казино любую сумму, после чего выйти из игры. Для этого надо случайно (с помощью монеты) выбирать цифру и в первом случае делать на нее экспоненциально растущую ставку, а во втором просто ждать пока по закону больших чисел выигрыш не достигнет нужной величины. Итак, ставки не должны превышать некоторого потолка, и при ставке равной s проигрыш Игрока равен s, а выигрыш — ϕs для фиксированного $\phi < 1$. Ясно, что если колода образует случайную (полученную с помощью монеты) последовательность цифр, то по закону больших чисел с ростом продолжительности игры (не обязательно с одним игроком) выигрыш Казино будет стремиться к бесконечности. Было бы очень приятно, если бы последнее свойство можно было принять за определение случайности. Возможно для этого пришлось бы обогатить правила игры.

Если поведение Игрока регулируется алгоритмически, то возникает формальное определение случайности. Первое точное определение такого типа (с помощью правил выбора) было предложено А.Н.Колмогоровым в 1963г. [3]. В 1968г. А.Н.Колмогоров сформулировал [§] без доказательства утверждение о том, что существует последовательность с логарифмическим ростом сложности начальных отрезков, случайная в предыдущем смысле. Это, как будет продемонстрировано ниже, ошибка. В 1987г. M.Lambalgen рассмотрел последовательность, случайно возникающую в процессе бросания несимметричных монет, в котором вероятность выпадения нуля на n-ом шагу равна $1/2 + y_n$, где $\{y_n\}$ — вычислимая последовательность рациональных чисел, стремящаяся к 0. Интуитивно очевидно, что эта последовательность удовлетворяет закону больших чисел. M.Lambalgen доказал тоже самое для последовательностей выбираемых из данной монотонными правилами. А.Х.Шень [5] распространил это на немонотонные правила. При условии, что Ы$y_n^2 = \infty$, неслучайность рассмотренной последовательности относительно равномерной меры доказана Solomonoff'ым в 1978г. [6]. При этом дефект случайности начала длины N по порядку равен Ы$_{n<=N}y_n^2$. В 1987г. В.Г.Вовк [7] показал, что эту оценку нельзя улучшить. Назовем последовательность $\{d_n\}$ натуральных чисел сублинейной, если $\forall \phi > 0 \exists N \forall n > N d_n < \phi n$.

ЛЕММА 1. По вычислимой сублинейной последовательности $\{d_n\}$ натуральных чисел можно построить такую вычислимую последовательность $\{y_n\}$ рациональных чисел из [0,1/2], что $y_n \to 0$ и $\exists N_0 \forall N > N_0$ Ы$_{n<=N}y_n^2 > d_n$.

Для доказательства надо для каждого n равномерно размазать d_n по целым точкам из $[n/2, n]$. После этого в каждой точке взять максимум полученных значений. Начиная с этого места под случайностью будем иметь ввиду понятие из [1,2]. Игровое же понятие будем называть стохастичностью. Итак, случайная последовательность всегда стохастична, а стохастичная последовательность не обязательно является случайной. На сколько она может быть неслучайной? Утверждение А.Н.Колмогорова из [§] давало бы максимальную неслучайность.

ТЕОРЕМА 1. Для вычислимой последовательности $\{d_n\}$ натуральных чисел стохастичная двоичная последовательность, дефект n-го начала которой начиная с некоторого N больше d_n, существует тогда и только тогда, когда $\{d_n\}$ сублинейна.

ДОКАЗАТЕЛЬСТВО. "Тогда" вытекает из примера Lambalgen'а и леммы 1.

"Только тогда". Пусть, $\phi > 0$ и $d_n > \oint n$ для бесконечно многих n. Двоичную последовательность, дефект начал которой мажорирует $\{d_n\}$, назовем $\{d_n\}$-неслучайной. Мы предъявим такое конечное множество вычислимых стратегий Игрока, что для любой $\{d_n\}$-неслучайной колоды какая-нибудь из стратегий даст отклонение последовательности выигрышей и проигрышей от закона больших чисел. Количество стратегий будет зависеть от ϕ.

Мы будем пользоваться следующим простым свойством.

ЛЕММА 2. Если у конечной двоичной последовательности начало ограниченной заранее длины заменить на другое начало ограниченной длины, то изменение дефекта последовательности тоже будет ограничено.

Построим вычислимую последовательность $\{t_k\}$ попарно непересекающихся отрезков натурального ряда быстрорастущей длины $\{l_k\}$, так чтобы для каждого k соответствующий t_k отрезок из $\{d_n\}$-неслучайной последовательности имел дефект больше μl_k ($\mu = \phi/2$).

Игрок делит игру на этапы. На k-ом этапе он имеет дело с отрезками колоды t_{2k} и t_{2k+1}. Стратегии игрока разбиваются на две группы. Стратегии первой группы на k-ом этапе читают (ставка равна 0) карты из отрезка t_{2k}, после чего смотрят сколько времени требуется для выявления того факта, что прочитанный отрезок имеет дефект больше μl_{2k} (множество последовательностей с большим дефектом перечислимо). Затем они высказывают гипотезу, что на выявление аналогичного факта про t_{2k+1} ушло бы не больше времени, получая тем самым некоторое множество двоичных последовательностей длины l_{2k+1}, в котором (в случае верности гипотезы) лежит соответствующий отрезок колоды. Стратегии второй группы наоборот читают нечетные отрезки и гадают о четных. Хотя бы одна из двух групп стратегий бесконечно много раз выдвигает верную гипотезу. Таким образом мы свели задачу к конечной игре с последовательностью карт, длины (скажем) l. Про нее заранее известно множество двоичных последовательностей длины l, мощность которого меньше $2^{(1-\mu)l}$, содержащее последовательность цифр, изображенных на лицевой стороне карт. Крупный (размера μl) выигрыш в этой игре давал бы крупный выигрыш Игрока на очередном этапе, и в виду быстрого роста $\{l_k\}$ происходило бы отклонение от закона больших чисел.

Описанную конечную игру рассматривал R.P.Daley. В 1971г. он естественно пришел к ней, занимаясь соотношением между игровым определением случайности и определением типа [1,2] с ограничением на вычислительные ресурсы. Будем считать, что на множестве малой мощности, в котором заведомо лежит неизвестная последовательность цифр, дано равномерное распределение вероятностей. R.P.Daley показал, что если $\mu < 1/2$, то годится монотонная стратегия ставящая на более вероятную цифру. Он же заметил, что если очередная вероятность близка к $1/2$, то карту осмысленно просто прочесть.

Сначала мы построим одну вероятностную стратегию (монотонную), дающую в среднем крупный выигрыш. На i-ом шагу с вероятностью p_i мы будем ставить на более вероятную цифру, а с вероятностью $1 - p_i$ просто читать i-ую карту. Подберем наиболее выгодное значение p_i. Пусть вероятность более вероятной цифры равна $(1 + b_i)/2$. Если $b_i = 0$, то p_i естественно приравнять 0. В качестве точки отсчета рассмотрим ситуацию, когда b_i равно 0 максимально долго. При этом каждый раз число возможных продолжений сокращается вдвое. Поэтому за μl шагов до конца последовательность определится однозначно, и мы выиграем μl ставок. С этой ситуацией мы будем соотносить произвольную другую. Если на i-ой карте изображена более вероятная цифра, то с вероятностью p_i мы приобретаем выигрыш в одну ставку, зато за счет меньшего (по сравнению с $b_i = 0$) сокращения числа продолжений мы в конце теряем $\log_2(1 + b_i)$ предопределенных цифр; с вероятностью же $1 - p_i$ происходит только последняя потеря. Если на i-ой карте изображена менее вероятная цифра, то с вероятностью p_i мы проигрываем одну ставку, зато за счет большего сокращения продолжений приобретаем $-\log_2(1 - b_i)$ выигрышей; с вероятностью же $1 - p_i$ происходит только последнее приобретение. Итак в случае более вероятной карты средний выигрыш (дополнительный по отношению к базовой ситуации) равен $p_i(1 - \log_2(1 + b_i)) + (1 - p_i)(-\log_2(1 + b_i))$, т.е. $p_i - \log_2(1 + b_i)$. В случае же менее вероятной карты средний выигрыш равен $p_i(-1 - \log_2(1 - b_i)) + (1 - p_i)(-\log_2(1 - b_i))$, т.е. $-p_i - \log(1 - b_i)$. Мы хотим максимизировать средний выигрыш в худшем случае, для чего надо сравнять выигрыши; т.е. p_i определяется из уравнения $p_i - \log_2(1 + b_i) = -p_i - \log_2(1 - b_i)$.

При $b_i < 3$:
$p_i = (\log_2(1 + b_i) - \log_2(1 - b_i))/2$, а гарантированный средний выигрыш $-(\log_2(1 + b_i) + \log_2(1 - b_i))/2$, что всегда неотрицательно.

При $b_i \geq 3$:
$p_i = 1$, а средние выигрыши равны соответственно $1 - \log_2(1 + b_i)$ и $-1 - \log_2(1 - b_i)$. Первый из них неотрицателен всегда, а второй при $b_i \geq 1/2$ Таким образом, суммарный средний выигрыш будет не меньше μl.

С рациональным числом к из (0,1] свяжем детерминированную монотонную стратегию ыr, читающую карту когда $b_i < r$, и ставящую на более вероятную цифру в ином случае. Одна из таких стратегий окажется подходящей. При этом значение r можно брать из достаточно густого множества $\{r_1, r_2, \ldots, 1\}$, зависящего от μ. Пусть p(ы) — значение вероятности p, соответствующей b в построенной вероятностной стратегии. Сложим общие выигрыши стратегий ыr_j с коэффициентами $p(r_j) - p(r_{j-1})$ $(r_0 = 0)$. Меняя порядок слагаемых, приходим к рассмотрению взвешенной суммы выигрышей на i-ом шагу, которая получается близкой к среднему выигрышу вероятностной стратегии на i-ом шагу. При достаточно плотном множестве $\{r_j\}$ последняя близость имеет место с точностью до $\mu/2$. В результате отклонение взвешенной суммы общих выигрышей детерминированных стратегий от общего выигрыша вероятностной стратегии будет меньше $\mu l/2$. Итак, взвешенная сумма выигрышей детерминированных стратегий больше $\mu l/2$. Поскольку сумма весов равна $p(1) - p(0) = 1 - 0 = 1$, то одна из детерминированных стратегий выигрывает больше $\mu l/2$. Теорема 1 доказана.

Если Игрок на каждом этапе будет случайно выбирать одну из детерминированных стратегий, то с вероятностью 1 он бесконечно много раз выберет нужную. Таким образом получается

ТЕОРЕМА 2. По вычислимой последовательности $\{n_k\}$ натуральных чисел строится вероятностная стратегия Игрока, побеждающая Казино, если оно пользуется колодой, у которой дефект n_k-го начала больше ϕn_k (ϕ определено во введении).

Эта теорема имеет уже практический смысл.

ЗАМЕЧАНИЕ. Говоря о правилах выбора из [3], можно потребовать, чтобы они из каждой (или почти каждой) последовательности выбирали распределение вероятностей на выбранных подпоследовательностях будет равномерным. Как известно, при этом всевозможные вероятности отклонений будут вычислимы. Следовательно, как доказал в 1969г. C.P.Schnorr [8], существует двоичная последовательность логарифмического роста сложности начал, стохастичная относительно указанных правил выбора. Возможно, это и послужило причиной ошибки в [§].

В реальном казино игроки могут ставить не только на отдельные события, но и на их комбинации. Введем соответствующее обобщение в правила нашей математической игры. Игрок может отмечать конечную последовательность (длины l) еще не перевернутых карт, указывая множество двоичных последовательностей той же длины, мощность которого не больше 2^{l-1}. Затем карты переворачиваются и выясняется, лежит ли последовательность изображенных на них цифр в указанном множестве. Игрок выигрывает, когда принадлежность имеет место.

Если l ограничено, то сохраняет силу пример Lambalgen'a. Неограниченная ситуация, как оказывается, существенно усиливает возможности Игрока. Случайность относительно новой игры будем называть m-стохастичностью. Случайная последовательность всегда m-стохастична (стандартный факт).

ТЕОРЕМА 3. Для вычислимой последовательности $\{d_n\}$ натуральных чисел m-стохастичная двоичная последовательность, дефект n-го начала которой начиная с некоторого N больше d_n, существует тогда и только тогда, когда d_n ограничено.

Теорема §. Пусть даны непересекающиеся вычислимые последовательности натуральных чисел. Если соответствующие им подпоследовательности двоичной последовательности неслучайны, то она не m-стохастична. Доказательство теоремы 3. "Тогда" вытекает из того фундаментального факта, что случайность эквивалентна ограниченности дефекта начал. "Только тогда". Пользуясь вычислимостью $\{d_n\}$, эффективно разрежем с помощью леммы 2 нашу двоичную последовательность на две неслучайные подпоследовательности, после чего дело сводится к теореме §. Доказательство теоремы §. Будем представлять себе две последовательности карт, про каждую из которых заранее известно, что выписанная на ней двоичная последовательность неслучайна. Назовем участком конечную двоичную последовательность, если у нее есть начало с достаточно большим дефектом, и этот факт выявляется за время меньшее длины всей последовательности. Ясно, что свойство быть участком разрешимо. Бесконечная неслучайная последовательность непременно начинается с участка. Рассмотрим три стратегии Игрока. Все они будут делиться на этапы. Опишем какие карты просматриваются на первом этапе. Отложим на первой последовательности кратчайшее начало b_1^1 , являющееся участком, вслед за ним отложим кратчайшее продолжение b_1^2 , являющееся участком. Аналогично на второй последовательности отложим участки b_2^1 и b_2^2 . Пусть l_i — сумма длин b_i^1 и b_i^2 . На первом этапе стратегии просматривают начала обеих последовательностей длины $\max(l_1, l_2)$. Второй этап так же поступает с непросмотренными хвостами первой и второй последовательностей карт и т.д. I стратегия. Один этап. Игрок читает первую последовательность карт, пока ее начало не окажется участком. После этого он еще раз откладывает на продолжении первой последовательности участок. Затем Игрок выдвигает гипотезу, что длина первого участка на второй последовательности меньше суммы длин первых двух участков на первой последовательности. Соответствующие карты второй последовательности переворачиваются, и гипотеза проверяется. В случае необходимости переворачиваются еще карты в соответствии с делением на этапы, описанным выше. Переход к следующему этапу. II стратегия симметрична первой стратегии относительно перемены ролей первой и второй последовательностей карт. III стратегия. Начиная этап, Игрок читает первую последовательность карт до появления участка. Затем он читает вторую последовательность до появления на ней участка. Если участок на первой последовательности длиннее чем на второй, то Игрок выдвигает гипотезу, что сумма длин двух первых участков на второй последовательности не больше длины прочитанного участка на первой последовательности. Если прочитанный участок на второй последовательности не короче, чем на первой, то Игрок выдвигает симметричную гипотезу. После этого гипотеза проверяется, и, перевернув при необходимости дополнительные карты в соответствии с делением на этапы, Игрок переходит к следующему этапу. Стратегии корректны, ибо мало последовательностей имеют начало с большим дефектом. Легко убедиться, что на каждом этапе выигрывают две стратегии из трех. Поэтому в каждый момент времени частота выигрышей какой-нибудь стратегии не меньше 2/3. Теорема § доказана. Как известно важную роль в теории алгоритмов играют эффективные варианты различных понятий. Что было бы естественной эффективизацией неслучайности бесконечной двоичной последовательности? Наверное требование, чтобы дефект случайности ее начал мажорировал вычислимую неограниченную функцию (которую назовем регулятором неслучайности). Теорема :. По вычислимой неограниченной функции строится вероятностная m-стратегия Игрока, побеждающая любую колоду, для которой эта функция является регулятором неслучайности (стратегия зависит от значения ϕ из введения). Доказательство. С помощью регулятора неслучайности эффективно разрежем, пользуясь леммой 2, колоду на группы по a отрезков (a определяется по ϕ), так чтобы все они были достаточно неслучайными. Опишем один этап. В очередной группе отрезков Игрок случайно выбирает (пользуясь равномерно распределенным датчиком) один из a отрезков. Прочтя остальные отрезки, Игрок ждет пока выявится факт неслучайности каждого из них. Выдвигаемая гипотеза состоит в том, что на выявление факта неслучайности выбранного отрезка ушло бы времени не больше чем на предыдущее ожидание. Это дает малое по мощности множество возможных значений выбранного отрезка. Гипотеза оказывается верной во всех случаях кроме одного, т.е. с вероятностью $1 - 1/a$. По закону больших чисел частота выигрышей Игрока будет стремиться к $1 - 1/a$. Теорема : доказана. В этой теореме можно было бы говорить в стиле теоремы § об a неслучайных (без условия эффективности) подпоследовательностях. В заключение сформулируем два вопроса. Нельзя ли избавиться от требования вычислимости $\{d_n\}$ в теоремах 1 и 3? Сужает ли класс стохастичных (m-стохастичных) последовательностей разрешение Игроку пользоваться вероятностным датчиком?

ЛИТЕРАТУРА

1.Левин Л.А. — ДАН, 1973, т.212, N3, с.:§8-::0.

2.Schnorr C.P. — Journal Comp. System. Sci., 1973, vol.7, N4, p.376-388.

3.Kolmogorov A. — Sankhya, Ser.A, 1963, vol.25, N4, p.36-376. (Семиотика и информатика, 1(82, вып.18, с.3-13)).

4.Kolmogoroff A. — IEEE Transactions Info. Theor., 1968, IT-14, N5, p.662-664. (Колмогоров А.Н. — Проблемы передачи информации, 19,9, т.:, N3, с.3-7).

5. Шень А.Х. — ДАН, 0000, т.000, N0, с.000-000.

6. Solomonoff R.J. — IEEE Transactions on Info. Theory, 1978, vol.IT-24, N4, p.422-432,

7. Вовк В.Г. — ДАН, 1987, т.29§, N,, с.1298-1302.

8.Schnorr C.P. — Z. Wahrscheinlichkeits-theorie Verw. Gebiete, 1969, vol.14, p.27-35.