Осенний коллоквиум курса «Теория вероятностей»
ФКН НИУ ВШЭ, 2-й курс ОП ПМИ, 2-й модуль, 2016 учебный год

Билет 1

Дискретное вероятностное пространство. Вероятностный алгоритм проверки на простоту. Универсальная хэш-функция.

Дискретное вероятностное пространство

Рассмотрим некоторый эксперимент, все мыслимые исходы которого описываются конечным числом различных исходов ω_1, ..., ω_N. Несущественна природа этих исходов, важно лишь то, что их число N конечно.

Определение 1. Исходы ω_1, ..., ω_N будем называть *элементарными событиями*, а их совокупность

$$\Omega = \{\omega_1, ..., \omega_N\}.$$

(конечным) пространством элементарных событий или *пространством исходов*.

Замечание 1. Можно также называть Ω *множеством элементарных исходов*. Именно так его называют в кратком конспекте лекций.

Определение 2. Все те подмножества $A \subseteq \Omega$, для которых по условиям эксперимента возможен ответ одного из двух типов: «исход $\omega \in A$» или «исход $\omega \notin A$», — будем называть *событиями*.

Определение 3. Функцию

$$P\colon 2^\Omega \to [0, 1],$$

удовлетворяющую следующим свойствам:

(a) $P(\Omega) = 1$,
(b) $A \cap B = \varnothing \Rightarrow P(A \cup B) = P(A) + P(B)$ (*правило суммы* или *аддитивность*)

называют *вероятностной мерой*, а значение $P(A)$ *вероятностью события* A.

Замечание 2. Вероятностная мера P полностью определяется значениями $P(\omega_1) = p_1$, ..., $P(\omega_N) = p_N$.

Следствие 1. *Из определения вероятностной меры следует, что*

(a) $p_\omega \geqslant 0$,
(b) $\sum\limits_\omega p_\omega = 1$,
(c) *вероятность произвольного события A вычисляется по формуле*

$$P(A) = \sum_{\omega \in A} p_\omega.$$

Определение 4. Если все элементарные исходы *равновозможны*, то полагаем, что

$$p_{\omega_1} = ... = p_{\omega_n} = \frac{1}{n}.$$

Замечание 3. В случае, если все элементарные исходы равновозможны, вероятность события A равна отношению количества исходов из A к числу всех исходов в Ω.

Определение 5. *Дискретное вероятностное пространство* — это пара из множества элементарных событий Ω и определенной для него вероятностной мерой.

Вероятностный алгоритм проверки на простоту

Пусть дано некоторое натуральное число $N > 1$. Мы хотим проверить, является ли это число простым. Можно перебирать все простые делители до $\sqrt{N}$, но это очень долго. Хотелось бы иметь быстрый способ проверки.
Если N — простое число, то по малой теореме Ферма для всякого натурального числа b такого, что $(b, N) = 1$, число $b^{N-1} - 1$ делится на N. Следовательно, если для некоторого b, удовлетворяющего условию $(b, N) = 1$, число $b^{N-1} - 1$ не делится на N, то N не является простым.
Пусть основание b мы выбираем случайно из множества $\mathbb{Z}_N^*$. Предположим, что существует такое основание, для которого N не проходит тест. Какова вероятность выбрать такое основание?
Предположим, что для $a \in \mathbb{Z}_N^*$ число N не проходит тест.

Замечание 4. $\mathbb{Z}_p^*$ - мультипликативная группа поля $\mathbb{Z}_p$, то есть группа, содержащая все ненулевые элементы из $\mathbb{Z}_p$, и операция в ней совпадает с операцией умножения в $\mathbb{Z}_p$.

Если N проходит тест для основания b, то для основания ab число N уже тест не проходит. В противном случае

$$(ab)^{N-1} \equiv 1 \ (mod\ N),\ (b^{-1})^{N-1} \equiv 1 \ (mod\ N).$$

Следовательно,

$$\begin{cases} a^{N-1} \equiv (b^{-1})^{N-1}(ab)^{N-1} \ (mod\ N), \\ (b^{-1})^{N-1}(ab)^{N-1} \equiv 1 \ (mod\ N), \end{cases} \Rightarrow a^{N-1} \equiv 1 \ (mod\ N),$$

что противоречит предположению. Таким образом, каждому основанию b, для которого N не проходит тест, можно сопоставить основание ab, для которого результат теста отрицательный. Значит, оснований, для которых N не проходит тест, не меньше оснований, для которых N проходит тест на простоту. Искомая вероятность не меньше $\frac{1}{2}$. Если независимым образом повторять выбор основания k раз, то вероятность выбрать основание, для которого данное число проходит тест, меньше $\frac{1}{2^k}$.

Замечание 5. Бывают числа, которые проходят тест для всех оснований b. Это числа Кармайкла, например 561.

Замечание 6. Докажем, что $(b^{-1})^{N-1} \equiv 1 \ (mod\ N)$.

Доказательство. Число $b^{N-1} - 1$ делится на N, так N проходит тест для основания b. А значит,

$$b^{N-1} \equiv 1 \ (mod\ N) \mid \cdot (b^{-1})^{N-1},$$
$$1 \equiv (b^{-1})^{N-1} \ (mod\ N).$$

□

Универсальная хэш-функция

Определение 6. Пусть H - конечное множество хэш-функций, которые отображают пространство ключей U ($|U| = n$) в диапазон $\{0, 1, ..., m-1\}$. Такое множество называется *универсальным*, если для каждой пары ключей $k, l \in U, (k \neq l)$, количество хэш-функций $h \in H$, для которых $h(k) = h(l)$ не превышает $\frac{|H|}{m}$.

Иными словами, при случайном выборе хэш-функции из множества H вероятность коллизии между двумя различными ключами k, l не превышает вероятности совпадения двух случайным образом выбранных хэш-значений из множества $\{0, 1, ..., m-1\}$, которая равна $\frac{1}{m}$.
Далее будем считать, что $U = \{0, 1, ..., n-1\}$.

Теорема 1. *Множество хэш-функций* $H_{p,m} = \{h_{a,b} : a \in \mathbb{Z}_p^*, b \in \mathbb{Z}_p\}$, *где*

$$h_{a,b}(k) = ((ak + b) \ mod \ p) \ mod \ m,$$

$\mathbb{Z}_p^* = \{1, 2, ..., p-1\}, \mathbb{Z}_p = \{0, 1, ..., p-1\}$, *p - простое число,* $p > n$, *является универсальным.*

Доказательство. Рассмотрим $k, l \in \mathbb{Z}_p : k \neq l$. Пусть для данной хэш-функции $h_{a,b}$

$$r = (ak + b) \ mod \ p,$$
$$s = (al + b) \ mod \ p.$$

Заметим, что $r \neq s$, так как $r - s \equiv a(k - l) \ (mod \ p)$, а p - простое число, a и $(k - l)$ не равны нулю по модулю p, а значит и разность r и s также отлична от нуля по модулю p. Таким образом, коллизии "по модулю p" отсутствуют. Более того, каждая из $p(p-1)$ возможных пар (a, b) приводит к различным $(r, s) : r \neq s$. Чтобы доказать это, достаточно рассмотреть возможность однозначного определения a и b по заданным r и s:

$$a = ((r - s) \cdot (k - l)^{-1}) \ mod \ p,$$
$$b = (r - ak) \ mod \ p.$$

(Доказательство приведено ниже, в Замечании под номером 7).
Поскольку имеется только $p(p-1)$ возможных пар (r, s): $r \neq s$, то имеется взаимооднозначное соответствие между парами (a, b) и парами (r, s): $r \neq s$. Таким образом, для любых k, l при равномерном случайном выборе пары (a, b) из $\mathbb{Z}_p^* \times \mathbb{Z}_p$ получаемая в результате пара (r, s) может быть с равной вероятностью любой из пар с отличающимися значениями по модулю p.
Отсюда следует, что вероятность того, что различные ключи k, l приводят к коллизии, равна вероятности того, что $r \equiv s \ (mod \ m)$ при произвольном выборе отличающихся по модулю p значений r и s. Для данного r имеется $p - 1$ возможное значение s. При этом число значений $s : s \neq r$ и $s \equiv r \ (mod \ m)$, не превышает

$$\left\lceil \frac{p}{m} \right\rceil - 1 \leqslant \frac{p + m - 1}{m} - 1 = \frac{p - 1}{m}.$$

Вероятность того, что s приводит к коллизии с r при приведении по модулю m, не превышает $\frac{p-1}{m} \cdot \frac{1}{p-1} = \frac{1}{m}$.
Значит, $\forall k \neq l \in \mathbb{Z}_p \ P(h_{a,b}(k) = h_{a,b}(l)) \leqslant \frac{1}{m}$, что означает, что множество хэш-функций $H_{p,m}$ является универсальным. □

Замечание 7. Докажем, что $a = ((r - s) \cdot (k - l)^{-1}) \ mod \ p$.

Доказательство.

$$r = (ak + b) \ mod \ p,$$
$$s = (al + b) \ mod \ p.$$
$$r - s = (a(k - l)) \ mod \ p,$$
$$a \equiv (r - s) \cdot (k - l)^{-1} \ (mod \ p)$$

Так как $a \in \mathbb{Z}_p$, то верно равенство

$$a = ((r - s) \cdot (k - l)^{-1}) \ mod \ p$$

□

Докажем, что $b = (r - ak) \ mod \ p$.

Доказательство. Достаточно вспомнить, что

$$\begin{cases} r = (ak + b) \ mod \ p, \\ b \in \mathbb{Z}_p \end{cases}$$

□