Алгоритмы и структуры данных

Конспекты лекций

Лектор: Г.О. Евстропов

Под редакцией Валерии Маликовой и Глеба Новикова

НИУ ВШЭ, 2016-2017

Немного теории вероятностей

Disclaimer 1. Это первая из двух глав теории вероятностей. В них мы будем говорить исключительно о дискретной (с конечными вероятностными пространствами) вероятности, так как в алгоритмах ничего другого особо не понадобится. Кроме того, эти две главы не заслуживают даже названия «Начала теории вероятностей», так как являются совсем частным случаем общей теории вероятностей. Несмотря на это, изложенный в них кусочек начала теорвера достаточен для того, чтобы понимать, что происходит.

Disclaimer 2. Иногда (примерно всегда) мы будем ставить знак «=» между числами тогда, когда его ставить в строгом смысле не очень честно. Однако, поскольку нас всюду интересует ассимптотическое поведение величин, то этот знак будет стоять вполне себе честно.

Определение 1.1. *Вероятностное пространство* $\left(\Omega, 2^\Omega, \mathsf{P}\right)$ – структура, состоящая из:

- Ω – множество элементарных исходов (конечное);
- 2^Ω – множество всевозможных событий (наборов исходов);
- P – функция вероятности $\mathsf{P} : 2^\Omega \to [0,1]$ такая, что $\mathsf{P}(\Omega) = 1$.

Теперь договоримся о некоторых обозначениях и лексике. Вместо $\mathsf{P}(\{\omega\})$ мы всегда будем писать $\mathsf{P}(\omega)$ и называть это *вероятностью исхода* ω. Если вероятность какого-то события $B \in 2^\Omega$ равна нулю, то есть $\mathsf{P}(B) = 0$, то событие B называется *невозможным*. Если же $\mathsf{P}(B) = 1$, то B называется *достоверным* событием.

Через $\mathsf{P}(A|B)$ будем обозначать вероятность события при условии наступления события B. Такая вероятность считается по формуле

$$\mathsf{P}(A|B) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)}$$

Определение 1.2. *Независимыми* называются события A и B, если $\mathsf{P}(A|B) = \mathsf{P}(A)$ или $\mathsf{P}(A) \cdot \mathsf{P}(B) = \mathsf{P}(A \cap B)$.

Несовместными называются события, для которых $\mathsf{P}(A|B) = 0$.

Определение 1.3. $A_1, \ldots, A_n$ - множство попарно несовместных событий, то есть $\forall i, j\ \mathsf{P}(A_i|A_j) = 0$. Если $\mathsf{P}(\bigcup\limits_{i=1}^{n} A_i) = 1$, тогда набор $\{A_i\}_{i=1}^n$ называется *полной группой событий.*

Тогда если есть событие B и полная группа событий $\{A_i\}_{i=1}^n$, тогда $B = \bigcup\limits_{i=1}^{n} B \cap A_i$ и

$$\mathsf{P}(B) = \sum_{i=1}^{n} \mathsf{P}(B \cap A_i) = \sum_{i=1}^{n} \mathsf{P}(B|A_i) \cdot \mathsf{P}(A_i)$$

Больше теории вероятностей

Рассмотрим функцию ξ, определенную на вероятностном пространстве Ω:

$$\xi : \Omega \to \mathbb{R}$$

Математическим ожиданием ξ называется такая величина:

$$\mathsf{E}\xi = \sum_{\omega \in \Omega} \xi(\omega)\,\mathsf{P}(\omega)$$

Пример 2.1. Пусть вы - опытный продавец мороженного. Вам нужно заранее забронировать место в парке развелечений, поэтому вы хотите узнать, в какие дни стоит это делать. Вам известен прогноз погоды и то, сколько вы зарабатываете за сутки в определенную погоду.

Вероятностное пространство событий: $\Omega = \{$жара, облачно, дождь, ураган$\}$.

$\xi($жара$) = 100\$$, $\xi($облачно$) = 50\$$, $\xi($дождь$) = 10\$$, $\xi($ураган$) = 1\$$.

Аренда места обходится в 50\$ за сутки.

Пусть будет жарко с вероятностью 0.9 и облачно с вероятностью 0.1. Тогда:

$$\mathsf{E}\xi = 0{,}9 \cdot 100\$ + 0{,}1 \cdot 50\$ - 50\$ = 45\$$$

Если все исходы равновероятны, то:

$$\mathsf{E}\xi = 100 \cdot 0.25 + 50 \cdot 0.25 + 10 \cdot 0.25 + 1 \cdot 0.25 = 50\$$$

Введем определение индикаторной случайно величины.

Определение 2.2. Индикаторная случайная величина I_A события A - величина, принимающая значение 1, если событие A произошло, и 0 в обратном случае.

$$I_A = \begin{cases} 1, & \mathsf{P}(A) \\ 0, & 1 - \mathsf{P}(A) \end{cases}$$

Тогда $\mathsf{E}I_A = \mathsf{P}(A)$.

Заметим, что для любого $r \subset R$ верно, что $\mathsf{P}(\xi \in r) = \sum_{\xi(\omega) \in r} \mathsf{P}(\omega)$.

Мы уже знакомы с понятием «независимые события». Теперь мы хотим понять, какие случайные величины являются независимыми.

Определение 2.3. Случайные величины ξ и ψ независимы, если для любых $r_1 \in N(\xi)$ и $r_2 \in N(\psi)$, где $N(\ldots)$ - область значений величины, верно:

$$\begin{cases} A = \xi \in r_1 \\ B = \psi \in r_2 \end{cases} \Rightarrow \mathsf{P}(A)\,\mathsf{P}(B) = \mathsf{P}(A \cap B)$$

Определение 2.4. Случайные величины ξ и ψ независимы, если для $\forall x, y \in \mathbb{R}$, верно:

$$\begin{cases} A = (\xi \equiv x) \\ B = (\psi \equiv y) \end{cases} \Rightarrow \mathsf{P}(A)\,\mathsf{P}(B) = \mathsf{P}(A \cap B)$$

Докажем, что эти определения эквивалентны.

###
Я вообще не верю в то, что логика доказательства верна. Когда доказывается эквивалентность двух утверждений $A \Rightarrow B$ и $C \Rightarrow D$, то доказывается $A \Leftrightarrow C$.
###

Доказательство. Пусть $A_1 \ldots A_n$, где n - число элементов в r_1. $A_i = (\xi = x_i), x_i \in r_1$.

$$\mathsf{P}(A) = \mathsf{P}(A_1) + \ldots + \mathsf{P}(A_n) \qquad \mathsf{P}(B) = \mathsf{P}(B_1) + ... + \mathsf{P}(B_m)$$

$$(\mathsf{P}(A_1) + \mathsf{P}(A_2) + \ldots + \mathsf{P}(A_n))(\mathsf{P}(B_1) + \mathsf{P}(B_2) + \ldots + \mathsf{P}(B_m)) =$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{m} \mathsf{P}(A_i)\,\mathsf{P}(B_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathsf{P}(A_i \cap B_j) = \mathsf{P}(A \cap B)$$

□

Теперь рассмотрим самое главное свойство случайной величины. Мы будем использовать это свойство на протяжении всего курса алгоритмов.

Теорема 2.5.
$$\mathsf{E}(\alpha\xi + \beta\psi) = \alpha\mathsf{E}\xi + \beta\mathsf{E}\psi$$

Доказательство.
$$\begin{aligned}\mathsf{E}(\alpha\xi + \beta\psi) =& \sum_{\omega\in\Omega}(\alpha\xi(\omega) + \beta\psi(\omega))\,\mathsf{P}(\omega) =\\ =&\alpha\sum \xi(\omega)\,\mathsf{P}(\omega) + \beta\sum \psi(\omega)\,\mathsf{P}(\omega) =\\ =&\alpha\mathsf{E}\xi + \beta\mathsf{E}\psi\end{aligned}$$

□

Пример 2.6. Найти математическое ожидание количества статичных точек($p_i = i$) в перестановке из n элементов.
$$\mathsf{E}\xi = \mathsf{E}(I_1 + ...I_n) = \sum_{i=1}^{n}\mathsf{E}I_i = \sum_{i=1}^{n}\mathsf{P}(I_i) = \sum_{i=1}^{n}\frac{1}{n} = 1$$

Мы теперь знаем, что $\mathsf{E}(\xi + \psi) = \mathsf{E}\xi + \mathsf{E}\psi$. Верно ли, что $\mathsf{E}(\xi\psi) = \mathsf{E}(\xi)\mathsf{E}(\psi)$?

Теорема 2.7. *Если ξ и ψ - независимые величины, то* $\mathsf{E}(\xi\psi) = \mathsf{E}(\xi)\mathsf{E}(\psi)$.

Доказательство.
$$\begin{aligned}\mathsf{E}\xi =& \sum_{\omega\in\Omega}\xi(\omega)\,\mathsf{P}(\omega) = \sum_{x\in N(\xi)} x\,\mathsf{P}(\xi = x)\\ \mathsf{E}(\xi\psi) =& \sum z(\xi\psi = z) = \sum_{x\in N(\xi)}\sum_{y\in N(\psi)}\mathsf{P}((\xi = x)\cap(\psi = y)) =\\ =& \sum_{x}\sum_{y} xy\,\mathsf{P}(\xi = x)\,\mathsf{P}(\psi = y) = \sum_{x} x\,\mathsf{P}(\xi = x)\sum_{y} y\,\mathsf{P}(\psi = y) = \mathsf{E}(\xi)\mathsf{E}(\psi)\end{aligned}$$

□

Определение 2.8. Дисперсией величины ξ называют такую величину $\mathsf{D}\,\xi$, что

$$\mathsf{D}\,\xi = \mathsf{E}(\xi - \mathsf{E}\xi)^2 = \mathsf{E}\xi^2 - (\mathsf{E}\xi)^2$$

Теорема 2.9. Неравенство Маркова (без док-ва)

$$\xi > 0,\ \mathsf{P}(\xi \geqslant a) \leqslant \frac{\mathsf{E}\xi}{a}$$

Теорема 2.10. Неравенство Чебышёва (без дока-ва)

$$\mathsf{P}(|\xi - \mathsf{E}\xi| \geqslant a) \leqslant \frac{\mathsf{D}\,\xi}{a^2}$$

Пример 2.11. Найти максимальную возрастающую подпоследовательность в случайной перестановке $a_1...a_n$, то есть найти $j_1,,j_k$ такие, что $a_{j_1} < a_{j_2} < ... < a_{j_k}$.

$cur \leftarrow -1$
for $i : 1 \to n$ **do**
  if $a_i > cur$ **then**
    $cur \leftarrow a_i$

###
Что такое A_k?
###

Пусть величина ξ – длина максимальной возрастающей последовательности. Тогда

$$\mathsf{E}\xi = \mathsf{E}(I_1 + ... + I_n) = \sum \mathsf{E}I_k = \sum \mathsf{P}(A_k) = \sum_{k=1}^{n} \frac{1}{k} = \log n.$$

Пример 2.12. Дан взвешенный граф $G(V,E), w(e) \geq 0$. Найти максимальный разрез. (разрез в графе – нетривиальное подмножество $A \subset V$ его вершин, величина разреза - сумма весов ребер, один конец которых лежит в A, другой - в $\overline{A}$). Тогда величина разреза это

$$w(A) = \sum_{u \in A v \not\in A} s(uv),$$

где U_A - множество ребер, лежащих на разрезе, индуцированном множеством A.

Решим задачу таким алгоримом: переберем все вершины, будем добавлять каждую из них в A с вероятностью $\frac{1}{2}$. Докажем, что математическое ожидание величины $w(A)$ это хотя бы половина оптимального ответа.

Вероятность того, что ребро лежит на оптимальном разрезе - $\frac{1}{2}$, потому либо оба конца лежат в A или $\overline{A}$, либо один из концов лежит в A, другой - в $\overline{A}$.

$$\mathsf{E}\xi = \sum \mathsf{E}I(e \in U_A)w(e) = \sum w(e)\,\mathsf{P}(e \in U_A) = \sum \frac{1}{2}w(e) = \frac{1}{2}\sum w(e)$$

Пример 2.13. Приведем пример алгоритма построения случайного дерева. Пусть 1-я вершина - это корень.

---

for $i : 2 \to n$ **do**
  if $a_i > cur$ **then**
    $\mathrm{parent}(i) \leftarrow \mathrm{rand}(1, i-1)$

Докажем, что $\forall u,v$ $\mathsf{E}g(v,u) = O(\log n)$.

$$\mathsf{E}g(v,u) \leqslant \mathsf{E}(d(v) + d(u) = \mathsf{E}d(v) + \mathsf{E}d(u)$$

$$\mathsf{E}d(i) = \log i(i \geqslant 2, d(1) = 0)$$

$$\mathsf{E}d(i) = 1 + \sum_{j=1..i-1} \mathsf{E}d(i)\frac{1}{i-1}$$

$$\exists c : Ed(i) \leqslant c \log n$$

$$\mathsf{E}d(i) \leqslant 1 + \frac{1}{i-1}\sum c\log i =$$

$$= 1 + \frac{\sum_{j=1}^{\frac{i}{2}} c\log j}{i-1} + \frac{\sum_{j=\frac{i}{2}+1}^{i} c\log j}{i-1}$$

$$\mathsf{E}d(i) \leqslant 1 + \frac{1}{i-1}\sum c\log(i-1) + \frac{1}{i-1}\sum c\log i = 1 + \frac{1}{i-1}\sum c\log i$$