Записки научных
семинаров ПОМИ
Том 277, 2001 г.

С. А. Норин

ПОЛИНОМИАЛЬНАЯ НИЖНЯЯ ОЦЕНКА ДЛЯ РАЗМЕРА НЕЗАВИСИМОГО ПО k ОТНОСИТЕЛЬНО МИНИМУМА МНОЖЕСТВА ПЕРЕСТАНОВОК

1. Введение

Понятие независимости (по k) относительно минимума для множеств перестановок было впервые введено в 1998 г. и изучалось в работах Бродера [1], а также Бродера, Чарикара, Митценмахера и Фриза [2].

Их исследования были мотивированы тем фактом, что такие множества необходимы для построения алгоритма, используемого на практике индексирующими программами поисковой системы Altavista для выявления и фильтрования идентичных и почти идентичных документов.

Эффективность алгоритма, фильтрующего идентичные документы в Сети, со временем приобретает все большее значение. С момента своего возникновения Интернет растет с экспоненциальной скоростью и, естественно, в нем все больше возрастает количество идентичных и почти идентичных документов. Исследования показывают (об этом, в частности, упоминается в [2]), что на данный момент около 20% документов в Сети являются дубликатами. Такие дубликаты могут быть локальными копиями популярных файлов, результатом деятельности так называемых "спаммеров", или же просто появляться вследствие ошибок поисковых программ. В любом случае они представляют серьезную проблему для индексирующих программ, поскольку, во-первых, индексация дублирующих друг друга документов затрачивает ресурсы и, во-вторых, пользователи крайне редко бывают заинтересованы в получении в ответ на свои запросы ссылок на одинаковые документы.

Так как применение для определения совпадения документов стандартных расстояний между строками (например, метрики Хэмминга и ее обобщений) крайне неэффективно, был выработан новый подход к проблеме, при котором и используются независимые (по k) относительно минимума множества перестановок.

Независимые по k относительно минимума множества перестановок также используются для решения чисто теоретических вопросов, например, для дерандомизации алгоритмов [3, 7] решения задачи пропозициональной выполнимости. Подробнее см. статью [7] (§3) в этом сборнике.

В процессе исследований Бродер и его коллеги помимо практических задач столкнулись с рядом интересных теоретических вопросов, не на все из которых им удалось найти ответы.

Понятно, что в связи с использованием на практике ключевыми являются вопросы мощности минимальных множеств перестановок независимых (по k) относительно минимума.

Перейдем к основным определениям.

Определение 1.1 ([2]). *Пусть $\mathcal{C} \subseteq S_n$ – множество перестановок, на котором задано некоторое вероятностное распределение. Будем называть множество $\mathcal{C}$ независимым относительно минимума, если для любого $X \subseteq [1, n]$ и любого $x \in X$*

$$Pr_{\pi \in \mathcal{C}}\{\min\{\pi(X)\} = \pi(x)\} = \frac{1}{|X|}.$$

Специально подчеркнем, что свойство независимости относительно минимума зависит от распределения на множестве $\mathcal{C}$.

Уже в [2] была приведена теорема, дающая экспоненциальную по n нижнюю оценку на мощность множеств перестановок, независимых относительно минимума для равномерного распределения.

Теорема 1.2 ([2, теорема 1]). *Пусть $\mathcal{F}$ – независимое относительно минимума (для равномерного распределения на $\mathcal{F}$) множество перестановок чисел от 1 до n. Тогда*

$$|\mathcal{F}| \geq НОК\{1, 2, \ldots, n\},$$

где НОК обозначает наименьшее общее кратное. Следовательно,

$$|\mathcal{F}| \geq e^{n-o(n)}.$$

В 2000 году Ито, Такеи и Шинозаки в работе [5] сконструировали независимое относительно минимума (для равномерного распределения) множество $\mathcal{F}_n \subseteq S_n$, такое, что

$$|\mathcal{F}| = \text{НОК}\{1, 2, \ldots, n\},$$

тем самым сняв вопрос о мощности минимальных множеств перестановок независимых относительно минимума.

Другим понятием, тесно связанным с исходным, является понятие множества перестановок независимых по k относительно минимума. Такие множества, как было показано в [2], позволяют проверять документы на наличие в них одинаковых фрагментов, размера не превосходящего k.

Определение 1.3 ([2]). *Пусть $\mathcal{C} \subseteq S_n$ – множество перестановок, на котором задано некоторое вероятностное распределение. Будем называть множество $\mathcal{C}$ независимым по k относительно минимума, если для любого $X \subseteq [1, n]$, такого, что $1 \le |X| \le k$, и для любого $x \in X$*

$$Pr_{\pi \in \mathcal{C}}\{\min\{\pi(X)\} = \pi(x)\} = \frac{1}{|X|}.$$

Свойство независимости по k относительно минимума также зависит от распределения на множестве $\mathcal{C}$.

Лучшая известная верхняя оценка на мощность минимальных множеств перестановок, независимых по k относительно минимума (для равномерного распределения) была получена в работе Ито, Такеи и Таруи [4] в 2000 году и является полиномиальной по n.

Теорема 1.4 ([4, следствие 5.1]). *Для любого целого $n \ge 1$, существует множество перестановок $\mathcal{C} \subseteq S_n$, такое, что для равномерного распределения на нем множество $\mathcal{C}$ является независимым по k относительно минимума, и*

$$|\mathcal{C}| = O(n^{(1+(1/\ln n))k} \text{НОК}\{1, 2, \dots, k-1\}) = O(n^{(1+(2/\ln n))k}).$$

В работе М. А. Всемирнова [6] были получены конструкции, улучшающие оценки Ито, Такеи и Таруи для $k = 3, 4$ и сделано предположение, что множества перестановок независимые по k относительно минимума, можно получать, рассматривая автоморфизмы $(k-2)$-мерных конечных проективных пространств.

Гипотеза 1.5 ([6, гипотеза 2]). *Рассмотрим группу $G = PGL(k-1, \mathbf{F}_q)$, действующую на точках $(k-2)$-мерного проективного пространства $\mathbf{P}^{k-2}(\mathbf{F}_q)$, где $k \ge 3$. Введем на G равномерное распределение. Тогда при некоторой нумерации точек $\mathbf{P}^{k-2}(\mathbf{F}_q)$ группа*

G является независимым по k относительно минимума множеством перестановок.

Первая нетривиальная нижняя оценка на мощность минимальных множеств перестановок независимых по k относительно минимума была получена в [4].

Теорема 1.6 ([4, следствие 2.1]). *Пусть $k \geq 3$ и $\mathcal{C}$ – независимое по k относительно минимума множество перестановок чисел от 1 до n. Тогда*

$$|\mathcal{C}| \geq n - 1.$$

В [6] М. А. Всемирнов улучшает эту оценку при $k \geq 4$.

Теорема 1.7 ([6]). *Пусть $n \geq k \geq 4$ и $\mathcal{C}$ – независимое по k относительно минимума множество перестановок чисел от 1 до n. Тогда*

$$|\mathcal{C}| \geq 2n - 2.$$

Как видно, при том, что лучшая доказаная верхняя оценка на мощность минимальных множеств перестановок независимых по k относительно минимума имеет порядок $O(n^k)$, наилучшие из известных ранее нижних оценок были линейны по n.

Основным результатом настоящей работы является следующая теорема.

Теорема 1.8. *Пусть $\mathcal{C}$ – независимое по k относительно минимума (для некоторого вероятностного распределения) множество перестановок чисел от 1 до n. Тогда*

$$|\mathcal{C}| \geq \binom{n-1}{\lfloor \frac{k-1}{2} \rfloor}.$$

2. Доказательство основной теоремы

С этого момента мы будем считать, что нам дано $\mathcal{C}$ – множество перестановок чисел от 1 до n с заданным на нем некоторым распределением. Причем для данного распределения множество $\mathcal{C}$ является независимым по k относительно минимума. Везде далее вероятности и математические ожидания вычисляются для этого распределения.

Введем следующие удобное обозначения. Множество целых чисел от 1 до n будем обозначать через $[1,n]$. Пусть $z \in [1,n]$, $X \subseteq [1,n]$, $\pi \in S_n$. Будем писать

$$\pi(X) = \{\pi(x) | x \in X\},$$

$$\pi(z) < \pi(X), \text{ если } \forall x \in X(\pi(z) < \pi(x)),$$

$$\pi(z) > \pi(X), \text{ если } \forall x \in X(\pi(z) > \pi(x)).$$

Следующая ключевая лемма даст нам полезное для работы с независимыми относительно минимума по k множествами перестановок тождество.

Лемма 2.1. *Пусть* $X, Y \subset [1,n], X \cap Y = \varnothing$, $|X| + |Y| \leq k-1$, $z \in [1,n] \backslash (X \cup Y)$. *Тогда*

$$\Pr\bigl(\pi(X) < \pi(z) < \pi(Y)\bigr) = \frac{|X|! \cdot |Y|!}{(|X| + |Y| + 1)!}$$

Доказательство. Индукция по $|X|$.

База: $|X| = 0$. Так как $|Y| + 1 \leq k$, то из определения независимого относительно минимума по k множества перестановок следует, что

$$\Pr\bigl(\pi(z) < \pi(Y)\bigr) = \Pr\bigl(\min\{\pi(\{z\} \cup Y)\} = \pi(z)) = \frac{1}{|Y|+1}. \qquad (1)$$

Доказательство базы завершает тривиальное тождество

$$\frac{1}{|Y|+1} = \frac{0! \cdot |Y|!}{(0 + |Y| + 1)!}.$$

Индукционный переход. Пусть утверждение леммы доказано при $|X| < l$, докажем его при $|X| = l$. Так как для фиксированной перестановки π условие $\pi(z) < \pi(Y)$ равносильно тому, что существует ровно одно подмножество $A \subseteq X$ такое, что $\pi(A) < \pi(z) < \pi(Y \cup (X \backslash A))$, то мы можем произвецти следующую серию преобразований (здесь и далее ind обозначает индикатор события):

$$\begin{aligned}
&\Pr\bigl(\pi(z) < \pi(Y)\bigr) = \\
&\quad = \sum_{\pi\in\mathcal{C}} \Pr(\pi)\cdot \operatorname{ind}(\pi(z) < \pi(Y)) = \\
&\quad = \sum_{\pi\in\mathcal{C}} \Pr(\pi) \sum_{A\subseteq X} \operatorname{ind}(\pi(A) < \pi(z) < \pi(Y\cup(X\backslash A))) = \\
&\quad = \sum_{A\subseteq X}\sum_{\pi\in\mathcal{C}} \Pr(\pi)\cdot \operatorname{ind}(\pi(A) < \pi(z) < \pi(Y\cup(X\backslash A))) = \\
&\quad = \sum_{A\subseteq X} \Pr\bigl(\pi(A) < \pi(z) < \pi(Y\cup(X\backslash A))\bigr).
\end{aligned}$$

Следовательно,

$$\begin{aligned}
&\Pr\bigl(\pi(X) < \pi(z) < \pi(Y)\bigr) = \\
&\quad = \Pr\bigl(\pi(z) < \pi(Y)\bigr) - \sum_{A\subsetneqq X} \Pr\bigl(\pi(A) < \pi(z) < \pi(Y\cup(X\backslash A))\bigr).
\end{aligned} \tag{2}$$

Пусть $|Y| = m$. Согласно (1)

$$\Pr\bigl(\pi(z) < \pi(Y)\bigr) = \frac{1}{m+1}. \tag{3}$$

По предположению индукции для любого $A \subsetneqq X$

$$\Pr\bigl(\pi(A) < \pi(z) < \pi(Y\cup(X\backslash A))\bigr) = \frac{|A|!\cdot(m+l-|A|)!}{(m+l+1)!} \quad .$$

Преобразуем сумму в правой части тождества (2)

$$\begin{aligned}
&\sum_{A\subsetneqq X} \Pr\bigl(\pi(A) < \pi(z) < \pi(Y\cup(X\backslash A))\bigr) = \\
&\qquad = \sum_{i=0}^{l-1}\sum_{A:|A|=i} \frac{i!\cdot(m+l-i)!}{(m+l+1)!} = \\
&\qquad = \sum_{i=0}^{l-1}\binom{l}{i}\frac{i!\cdot(m+l-i)!}{(m+l+1)!} =
\end{aligned}$$

$$= \sum_{i=0}^{l-1} \frac{l!}{i! \cdot (l-i)!} \cdot \frac{i! \cdot (m+l-i)!}{(m+l+1)!} =$$

$$= \sum_{i=0}^{l-1} \frac{(m+l-i)!}{m! \cdot (l-i)!} \cdot \frac{m! \cdot l!}{(m+l+1)!} =$$

$$= \frac{m! \cdot l!}{(m+l+1)!} \sum_{i=0}^{l-1} \binom{m+l-i}{m}.$$

Воспользуемся хорошо известным комбинаторным тождеством

$$\sum_{i=0}^{l} \binom{m+l-i}{m} = \binom{m+l+1}{m+1}$$

и продолжим преобразования:

$$\frac{m! \cdot l!}{(m+l+1)!} \sum_{i=0}^{l-1} \binom{m+l-i}{m} =$$

$$= \frac{m! \cdot l!}{(m+l+1)!} \left(\sum_{i=0}^{l} \binom{m+l-i}{m} - \binom{m}{m} \right) =$$

$$= \frac{m! \cdot l!}{(m+l+1)!} \left(\binom{m+l+1}{m+1} - 1 \right) =$$

$$= \frac{m! \cdot l!}{(m+l+1)!} \cdot \frac{(m+l+1)!}{(m+1)! \cdot l!} - \frac{m! \cdot l!}{(m+l+1)!} =$$

$$= \frac{1}{m+1} - \frac{m! \cdot l!}{(m+l+1)!}.$$

Итак,

$$\sum_{A \subseteq X, A \neq X} \Pr\bigl(\pi(A) < \pi(z) < \pi(Y \cup (X \backslash A))\bigr) =$$

$$= \frac{1}{m+1} - \frac{m! \cdot l!}{(m+l+1)!}. \quad (4)$$

Подставляя (3) и (4) в (2), получаем

$$\Pr\bigl(\pi(X) < \pi(z) < \pi(Y)\bigr) =$$

$$= \frac{1}{m+1} - \left(\frac{1}{m+1} - \frac{m! \cdot l!}{(m+l+1)!} \right) =$$

$$= \frac{m! \cdot l!}{(m+l+1)!},$$

завершая доказательство перехода и леммы. □.

Мы будем рассматривать два случая – четное и нечетное k.

Сначала рассмотрим случай нечетного k. Пусть $k = 2r+1$. Построим семейство случайных величин, отображающих $\mathcal{C}$ в $\mathbb{C}$.

Для любых $i_1, i_2, \ldots, i_r \in \mathbb{N}$, таких, что $1 < i_1 < i_2 < \cdots < i_r \leq n$, будем отождествлять (комплексное) пространство векторов, координаты которых занумерованы подмножествами $I \subseteq \{i_1, \ldots, i_r\}$, с пространством $\mathbb{C}^{2^r}$. Для любого $a \in \mathbb{C}^{2^r}$ определим отображение

$$X^a_{i_1,\ldots,i_r} : \quad \mathcal{C} \longrightarrow \mathbb{C},$$

следующим образом:

$$X^a_{i_1,\ldots,i_r}(\pi) = a_I, \text{ где } I = \{i \in \{i_1, \ldots, i_r\} | \pi(i) < \pi(1)\},$$
$$\text{т. е. } \pi(I) < \pi(1) < \pi(\{i_1, \ldots, i_r\} \backslash I).$$

Введем обозначения:

$$I_0 = \{i_1, \ldots, i_r\},$$
$$J_0 = \{j_1, \ldots, j_r\}.$$

Лемма 2.2. *Для любых* $i_1, i_2, \ldots, i_r, j_1, j_2, \ldots, j_r \in \mathbb{N}$, *таких, что* $1 < i_1 < i_2 < \cdots < i_r \leq n, 1 < j_1 < j_2 < \cdots < j_r \leq n$, *и для любых* $a, b \in \mathbb{C}^{2^r}$ *выполняется равенство*

$$\mathbf{E}\big[X^a_{i_1,\ldots,i_r} \cdot X^b_{j_1,\ldots,j_r}\big] = \sum_{\substack{I \subseteq I_0,\, J \subseteq J_0 \\ J \cap (I_0 \backslash I) = \varnothing \\ I \cap (J_0 \backslash J) = \varnothing}} \frac{|I \cup J|! \cdot |(I_0 \backslash I) \cup (J_0 \backslash J)|!}{(|I_0 \cup J_0| + 1)!} a_I b_J.$$

Доказательство.

$$\mathbf{E}\big[X^a_{i_1,\ldots,i_r} \cdot X^b_{j_1,\ldots,j_r}\big] = \sum_{\pi \in \mathcal{C}} \Pr(\pi) \cdot X^a_{i_1,\ldots,i_r}(\pi) \cdot X^b_{j_1,\ldots,j_r}(\pi) =$$

$$= \sum_{\substack{I \subseteq I_0, \\ J \subseteq J_0}} \Pr\{(\pi(I) < \pi(1) < \pi(I_0 \backslash I)) \wedge (\pi(J) < \pi(1) < \pi(J_0 \backslash J)\} \cdot a_I b_J =$$

$$= \sum_{I \subseteq I_0, J \subseteq J_0} \Pr\{\pi(I \cup J) < \pi(1) < \pi((I_0 \backslash I) \cup (J_0 \backslash J))\} \cdot a_I b_J. \quad (5)$$

Если $(I \cup J)$ пересекается с $((I_0 \backslash I) \cup (J_0 \backslash J))$, то очевидно, что $\Pr\{\pi(I \cup J) < \pi(1) < \pi((I_0 \backslash I) \cup (J_0 \backslash J))\}$ равняется нулю. Если же

$(I\cup J)$ и $((I_0\backslash I)\cup(J_0\backslash J))$ не пересекаются, то мы можем вычислить эту вероятность по лемме 2.1. А именно, так как $|(I\cup J)|\cup(I_0\setminus I)\cup(J_0\setminus J)|=|I_0\cup J_0|\leqslant 2t=k-1$, то

$$\Pr\{\pi(I\cup J)<\pi(1)<\pi((I_0\backslash I)\cup(J_0\backslash J))\}= \\ =\frac{|I\cup J|!\cdot|(I_0\backslash I)\cup(J_0)\backslash J)|!}{(|I_0\cup J_0|+1)!} \qquad (6)$$

Так как $(I\cup J)$ не пересекается с $((I_0\backslash I)\cup(J_0\backslash J))$ тогда и только тогда, когда $J\cap(I_0\backslash I)=\varnothing$ и $I\cap(J_0\backslash J)=\varnothing$, то подставляя (6) в (5), получаем необходимое равенство. □

Определим вектор $d\in\mathbb{C}^{2^r}$ следующим образом:

$$d_\varnothing=1,$$
$$d_I=0,\quad \text{для всех } I\subseteq I_0,\quad \text{таких, что}\quad I\neq\varnothing.$$

Лемма 2.3. *Для любых $j_1,j_2,\dots,j_r\in\mathbb{N}$, таких, что $1<j_1<j_2<\dots<j_r\leq n$, существует вектор $\gamma\in\mathbb{C}^{2^r}$, такой, что для любых $i_1,i_2,\dots,i_r\in\mathbb{N}$, $1<i_1<\dots<i_r\leq n$ выполняется следующее равенство*

$$\mathbf{E}\big[X^d_{i_1,\dots,i_r}\cdot X^\gamma_{j_1,\dots,j_r}\big]=\begin{cases}1, & \text{если } I_0=J_0,\\ 0, & \text{если } I_0\neq J_0,\end{cases}$$

где, как и выше, $I_0=\{i_1,\dots,i_r\}$, $J_0=\{j_1,\dots,j_r\}$.

Доказательство. Используя лемму 2.2 и тот факт, что при подсчете математического ожидания ненулевыми будут лишь слагаемые содержащие $d_\varnothing$, (т. е. те, где $I=\varnothing$), имеем:

$$\mathbf{E}\big[X^d_{i_1,\dots,i_r}\cdot X^\gamma_{j_1,\dots,j_r}\big]=\sum_{\substack{J\subseteq J_0\\ J\cap I_0=\varnothing}}\frac{|J|!\cdot|I_0\cup(J_0\backslash J)|!}{(|I_0\cup J_0|+1)!}\cdot\gamma_J. \qquad (7)$$

Положим $K_0=I_0\cap J_0$. Легко видеть, что

$$J\cap I_0=J\cap K_0. \qquad (8)$$

Кроме того,

$$|I_0\cup J_0|=|I_0|+|J_0|-|K_0|=2r-|K_0|. \qquad (9)$$

Также для любого $J\subseteq J_0$, такого, что $J\cap I_0=\varnothing$, имеем

$$|I_0\cup(J_0\backslash J)|=|I_0\cup J_0|-|J|=2r-|K_0|-|J|. \qquad (10)$$

Используя (8), (9) и (10), преобразуем правую часть (7):

$$\sum_{\substack{J\subseteq J_0\\ J\cap \overline{I}_0=\varnothing}} \frac{|J|!\cdot|I_0\cup(J_0\backslash J)|!}{(|I_0\cup J_0|+1)!}\cdot\gamma_J = \sum_{\substack{J\subseteq J_0\\ J\cap \overline{K}_0=\varnothing}} \frac{|J|!\cdot(2r-|K_0|-|J|)!}{(2r+1-|K_0|)!}\cdot\gamma_J .$$

Теперь легко видеть, что γ находится из системы уравнений следующего вида:

$$\sum_{\substack{J\subseteq J_0\\ J\cap \overline{K}_0=\varnothing}} \frac{|J|!\cdot(2r-|K_0|-|J|)!}{(2r+1-|K_0|)!}\cdot\gamma_J = \begin{cases} 1, & \text{если } K_0=J_0,\\ 0, & \text{если } K_0\neq J_0,\end{cases} \tag{11}$$

и наша задача сводится к доказательству разрешимости этой системы.

Эти уравнения не зависят от I_0, а лишь от K_0. K_0 может быть любым подмножеством J_0. Таким образом, имеется система из 2^r линейных уравнений относительно 2^r (комплексных) неизвестных γ_J, где $J\subseteq J_0$.

Занумеруем уравнения подмножествами из J_0 таким образом, чтобы уравнение (11) имело номер (K_0). Будем говорить, что переменная γ_J является переменной ранга $|J|$, а уравнение с номером (K_0) – уравнением ранга $r-|K_0|$.

Докажем индукцией по s ($s\in[0,r]$) следующее утверждение: существует набор $\{\gamma_J\in\mathbb{C}\,|\,J\subseteq J_0, |J|\le s\}$ удовлетворяющий всем тем уравнениям из системы (11), ранг которых не превосходит s. Легко видеть, что в уравнения ранга s с ненулевыми коэффициентами входят только переменные ранга не превосходящего s, так что наше утверждение осмысленно.

База: $s=0$. Единственное уравнение ранга 0 – это уравнение (J_0):

$$\frac{1}{r+1}\cdot\gamma_\varnothing = 1$$

Число $\gamma_\varnothing = r+1$ удовлетворяет условию базы.

Индукционный переход. Найдем набор, удовлетворяющий сформулированному условию при $s=s_0$, если для $s=s_0-1$ такой набор уже найден.

Пусть $\{\gamma_J\in\mathbb{C}\,|\,J\subseteq J_0, |J|\le s_0-1\}$ – найденный по предположению индукции набор, удовлетворяющий всем уравнениям всех рангов, не превосходящих s_0-1. Найдем $\{\gamma_J\in\mathbb{C}\,|\,J\subseteq J_0, |J|=s_0\}$

так, чтобы удовлетворялись также и уравнения ранга s_0. Отметим, что в уравнение (K_0) с ненулевыми коэффициентами входят те и только те γ_J, для которых $J \cap K_0 = \varnothing$. Т. е. в каждом уравнении ранга s_0 присутствует с ненулевым коэффициентом ровно одна переменная ранга s_0, а именно $\gamma_{J_0 \setminus K_0}$, а все остальные входящие в это уравнение с ненулевыми коэффициентами переменные имеют ранг меньший s_0. Кроме того, для любого $J \subseteq J_0, |J| = s_0$ переменная γ_J присутствует с ненулевым коэффициентом ровно в одном уравнении ранга s_0, а именно в уравнении $(J_0 \setminus J)$.

Таким образом, мы можем найти требуемый набор по следующим формулам:

$$\gamma_J = -\Bigg(\sum_{\substack{J' \subseteq J_0 \\ J' \cap (J_0 \setminus J) = \varnothing \\ J' \neq J}} \frac{|J'|! \cdot (r + s_0 - |J'|)!}{(r+1+s_0)!} \cdot \gamma_{J'} \Bigg) \cdot \left(\frac{|J|! \cdot (r + s_0 - |J|)!}{(r+1+s_0)!} \right)^{-1},$$

что завершает доказательство перехода. Остается заметить, что при $s = r$ доказанное нами по индукции утверждение эквивалентно утверждению леммы. □

Следствие 2.4. *Пусть*

$$S = \sum_{\substack{i_1, i_2, \dots, i_r \in \mathbb{N} \\ 1 < i_1 < i_2 < \dots < i_r \leq n}} c_{i_1 \dots i_r} X^d_{i_1 \dots i_r}.$$

Тогда для любых $j_1, j_2, \dots, j_r \in \mathbb{N}$, $1 < j_1 < j_2 < \dots < j_r \leq n$, *существует вектор* $\gamma \in \mathbb{C}^{2^r}$, *такой, что*

$$\mathbf{E}\big[S \cdot X^\gamma_{j_1, \dots, j_r}\big] = c_{j_1 \dots j_r}.$$

Доказательство. Найдем вектор $\gamma \in \mathbb{C}^{2^r}$, удовлетворяющий лемме 2.3. Он будет нам подходить. Действительно, для него выполняется следующее равенство:

$$\mathbf{E}\big[c_{i_1 \dots i_r} X^d_{i_1, \dots, i_r} \cdot X^\gamma_{j_1, \dots, j_r}\big] = \begin{cases} c_{j_1 \dots j_r}, & \text{если } I_0 = J_0, \\ 0, & \text{если } I_0 \neq J_0. \end{cases}$$

Остается лишь сложить такие равенства для всех наборов $1 < i_1 < i_2 < \dots < i_r \leq n$. □

Следствие 2.5. *Множество случайных величин* $\{X^d_{i_1,\ldots,i_r} \mid 1 < i_1 < i_2 < \cdots < i_r \le n\}$ *является линейно независимым над* $\mathbb{C}$.

Доказательство. Пусть

$$S = \sum_{1<i_1<i_2<\cdots<i_r\le n} c_{i_1\ldots i_r} X^d_{i_1\ldots i_r} = 0.$$

Нам необходимо доказать, что для любых $i_1, i_2, \ldots, i_r \in \mathbb{N}$, таких, что $1 < i_1 < i_2 < \cdots < i_r \le n$, коэффициент $c_{i_1\ldots i_r}$ равен 0.

По следствию 2.4 для любых $1 < i_1 < i_2 < \cdots < i_r \le n$ существует вектор $\gamma \in \mathbb{C}^{2^r}$, такой, что

$$\mathbf{E}\big[S \cdot X^\gamma_{j_1,\ldots,j_r}\big] = c_{j_1\ldots j_r}.$$

Но $\mathbf{E}\big[S\cdot X^\gamma_{j_1,\ldots,j_r}\big] = 0$, так как $S = 0$. Т. е. для любых $j_1, j_2, \ldots, j_r \in \mathbb{N}$, таких, что $1 < j_1 < j_2 < \cdots < j_r \le n$, коэффициент $c_{j_1\ldots j_r}$ равен 0, что и требовалось. □

Следствие 2.6. *Пусть k нечетно, а множество $\mathcal{C}$ такое же, как и в теореме* 1.8. *Тогда* $|\mathcal{C}| \ge \binom{n-1}{r}$.

Доказательство. Действительно, в наборе $\{X^d_{i_1,\ldots,i_r} | 1 < i_1 < i_2 < \cdots < i_r \le n\}$ всего $\binom{n-1}{r}$ случайных величин отображающих $\mathcal{C}$ в $\mathbb{C}$. И так как $X^d_{i_1,\ldots,i_r}$ можно считать векторами с $|\mathcal{C}|$ комплексными координатами, они не могли бы быть линейно независимыми, если бы мощность $\mathcal{C}$ была бы меньше $\binom{n-1}{r}$. □

Так как $\binom{n-1}{r} = \binom{n-1}{\lfloor\frac{k-1}{2}\rfloor}$, то это доказывает теорему 1.8 в случае нечетного k. Случай же четного k тривиально из него следует, так как $\mathcal{C}$, являясь независимым по k относительно минимума, является также и независимым по $k-1$ относительно минимума множеством перестановок. Поэтому можно взять $r = \frac{k-2}{2} = \lfloor\frac{k-1}{2}\rfloor$ и применить следствие 2.6 к $k-1$, что дает требуемую оценку и для четного k. Таким образом мы завершаем доказательство теоремы 1.8.

Литература

1. A. Z. Broder, *On the Resemblance and Containment of Documents.* Proc. of Compression and Complexity of Sequences (1998), pp. 21–29.
2. A. Z. Broder, M. Charikar, A. M. Frieze, and M. Mitzenmacher, *Min-Wise Independent Permutations.* Proc. of the 30th Annual ACM Symposium on Theory of Computing (1998), pp. 327–336.

3. A. Z. Broder, M. Charikar, and M. Mitzenmacher, *A Derandomization Using Min-Wise Independent Permutations.* Proc. of RANDOM'98, Lecture Notes in Computer Science 1518, Springer (1998), pp. 15–24.
4. T. Itoh, Y. Takei, J. Tarui, *On Permutations with Limited Independence.* Proc. of SODA'2000 (2000), pp. 137–146.
5. T. Itoh, Y. Takei, T. Shinozaki, *Constructing an Optimal Family of Min-Wise Independent Permutations.* IEICE Trans. Fundamentals, vol. E83-A, No. 4 (2000), pp. 747–755.
6. M. Vsemirnov, *Automorphisms of Projective Spaces and Min-Wise Independent Families of Permutations.* preprint (2001).
7. М. А. Всемирнов, Э. А. Гирш, Е. Я. Данцин, С. В. Иванов, *Алгоритмы для пропозиционной выполнимости и верхние оценки их сложности.* Зап. научн. семин. ПОМИ, **277** (2001) (Предварительная английская версия доступна как ECCC technical report по ftp://ftp.eccc.uni-trier.de/pub/eccc/reports/2001/TR01-012/index. html).

Norin S. A. A polynomial lower bound for the size of any k-min-wise independent set of permutations.

The notion of k-min-wise independent set of permutations, which was introduced by A. Broder et al., is considered. A nontrivial polynomial lower bound for the smallest size of such sets is obtained.

Санкт-Петербургский
государственный университет

Поступило 1 октября 2001