УДК 621.391.1

О ВЫПУКЛОСТИ ВВЕРХ ФУНКЦИИ ЭНТРОПИИ БЕРНУЛЛИЕВСКОГО ИСТОЧНИКА

Беленский В.З., проф., Калашникова Н.И., доц.

В работах по теории информации и кодирования встречаются задачи на поиск экстремума функции энтропии. Эта функция широко применяется в задачах оптимального кодирования информации, задачах оптимизации и моделирования [1]. Однако при большом числе генерируемых источником информации слов и, следовательно, слагаемых в выражении для энтропии вычисление последней затруднительно даже при использовании быстродействующих ЭВМ. Особенно это важно при универсальном кодировании бернуллиевских источников информации, т.е. двоичное кодирование, при котором появление двоичных символов не зависит от предшествующих. В случае если $x \in I := [0,1]$ - вероятность появления единицы, то вероятность генерирования слова из n элементов, среди которых k единиц и $(n-k)$ нулей, задается формулой бернуллиевского распределения,

$$p_n^k := C_n^k x^k (1-x)^{n-k}, \quad k = 0, 1, \ldots, n, \tag{1}$$

где C_n^k - биноминальные коэффициенты.

Решить поставленные выше задачи можно, введя дополнительный источник информации, генерирующий признаки классов эквивалентности, на которые предварительно разбивается исходное множество слов [1]. В результате исходный источник информации представляется в виде композиции двух с энтропиями:

$$H_1 = -\sum_{k=0}^{n} p_n^k \log_2 p_n^k \quad ,$$

$$H_2 = -\sum_{k=0}^{n} p_n^k \log_2 C_n^k \quad .$$

Требуется показать, что функции H_1 и $(-H_2)$ достигают максимума при одинаковых вероятностях генерирования нуля и единицы, т.е. при $x = 0,5$.

Заметим, что в теории информации используется двоичный логарифм $\log_2$. Однако переход от одного основания к другому - это просто изменение масштаба, что непринципиально; для нас удобнее натуральный логарифм ln. Поэтому в данной работе мы исследуем функцию

$$H := -\sum_{k=0}^{n} p_n^k \ln p_n^k$$

как функцию от $x \in I$. Априори ясно, что функция H достаточно "хорошая", она обладает свойствами:

а) неотрицательность: $H(x) \geq 0$, причем $H(0) = H(1) = 0$;

б) симметричность: $H(x) = H(1-x) \quad \forall x \in I$. (2)

Профессором А.А.Борисенко была высказана гипотеза, что кроме того функция H выпукла вверх:

$$H''(x) < 0 \quad \forall x \in (0,1) \tag{3}$$

(штрих означает дифференцирование по x); из этой гипотезы следует, что в частности, что энтропия унимодальна, т.е. имеет единственный максимум в точке $x = 1/2$. Настоящая заметка посвящена доказательству гипотезы Борисенко.

1. Записав вероятность (1) в виде $p_n^k = C_n^k \pi_n^k$, где

$$\pi_n^k(x) := x^k (1-x)^{n-k}, \tag{4}$$

представим энтропию в форме $H = G - F$, где

$$G := -\sum_{k=0}^{n} p_n^k \ln \pi_n^k, \quad F := \sum_{k=0}^{n} p_n^k \ln C_n^k. \tag{5}$$

Очевидно, что **обе функции** G и F хорошие в смысле (2); ниже будет показано, что **каждая из них выпукла вверх**, причем G "сильнее", чем F, так что их **разность** H **удовлетворяет** условию (3).

2. Начнем с функции G, она вычисляется в явном виде. Имеем

$$G(x) = -\sum_{k=0}^{n} p_n^k [k \ln x + (n-k)\ln(1-x)] =$$

$$= -\left\{ \ln x \sum_{k=0}^{n} k C_n^k \pi_n^k(x) + \ln(1-x) \sum_{k=0}^{n} (n-k) C_n^k \pi_n^k(x) \right\}.$$

Введем для удобства обозначения для сумм:

$$S_0 = \sum_{k=0}^{n} k C_n^k \pi_n^k(x), \quad S_1 = \sum_{k=0}^{n} (n-k) C_n^k \pi_n^k(x).$$

Используя равенства $\pi_n^k(x) = x\pi_{n-1}^{k-1}(x) = (1-x)\pi_{n-1}^k(x)$, соотношения

$$kC_n^k = \begin{cases} 0, & k = 0, \\ nC_{n-1}^{k-1}, & k \geq 1; \end{cases} \qquad (n-k)C_n^k = \begin{cases} 0, & k = n, \\ nC_{n-1}^k, & k \leq n-1; \end{cases} \tag{6}$$

и тождество $\sum_{k=0}^{n} p_n^k \equiv 1$, преобразуем суммы S_0, S_1. Для S_0 получим, полагая $j = k-1$:

$$S_0 = n \sum_{k=1}^{n} C_{n-1}^{k-1} \pi_n^k(x) = nx \sum_{j=0}^{n-1} p_{n-1}^j(x) = nx;$$

аналогично, $S_1 = n(1-x)$. Таким образом,

$$G(x) = -n[x \ln x + (1-x)\ln(1-x)]. \tag{7}$$

Дифференцируя дважды, находим

$$G''(x) = -\frac{n}{x(1-x)} < 0. \tag{8}$$

Отметим, что в силу (7) функцию G можно записать в виде $G = nh$, где h - энтропия отдельной частицы. Это можно интерпретировать как свойство аддитивности функции G относительно количества наблюдений.

3. Введем удобное для дальнейшего понятие биноминальных многочленов. Биноминальный многочлен $Q_n(\tilde{L})$ ранга задается выражением

$$Q_n(L) := \sum_{k=0}^{n} L_k p_n^k, \tag{9}$$

где p_n^k рассматриваются как базисные многочлены ранга n переменной x, определенные формулой (1); здесь $\tilde{L} := (L_0, \ldots, L_n)$ есть $(n+1)$ - мерный вектор коэффициентов многочлена. Заметим, что степень многочлена $Q_n(\tilde{L})$ относительно аргумента x не превосходит его ранга n.

Пусть, далее, Δ - оператор правой разности, который произвольный вектор $\tilde{L}$ переводит в вектор на единицу меньшей размерности по формуле

$$\Delta\tilde{L} := (\Delta L_0, \ldots, \Delta L_{n-1}), \quad \Delta L_k := L_{k+1} - L_k, ; k = 1, \ldots, n-1.$$

Двукратное применение оператора $\Delta(\Delta^2\tilde{L} := \Delta(\Delta\tilde{L}))$ понижает размерность на две единицы и т.д.

Имеет место следующая замечательно простая формула биномиального дифференцирования (по x):

$$Q_n'(\tilde{L}) = nQ_{n-1}(\Delta\tilde{L}). \tag{10}$$

Замечание. В многочлене $Q_{n-1}(\Delta\tilde{L})$ ранг базисных многочленов тоже уменьшен на единицу.

Доказательство формулы (10) проводится аналогично п.2 Имеем

$$\left[p_n^k(x)\right]' = C_n^k [kx^{k-1}(1-x)^{n-k} - (n-k)x^k(1-x)^{n-k-1}] =$$

$$= C_n^k \left[k\pi_{n-1}^{k-1}(x) - (n-k)\pi_{n-1}^k(x) \right].$$

Дифференцируя многочлен (9) и применяя соотношения (6), получим

$$Q_n'(\tilde{L}) = \sum_{k=1}^{n} L_k k C_n^k \pi_{n-1}^{k-1} - \sum_{k=0}^{n-1} L_k (n-k) C_n^k \pi_{n-1}^k =$$

$$= n \sum_{j=0}^{n-1} L_{j+1} p_{n-1}^j - n \sum_{k=0}^{n-1} L_k p_{n-1}^k =$$

$$= n \sum_{k=0}^{n-1} \Delta L_k p_{n-1}^k = n Q_{n-1}(\Delta \tilde{L}).$$

4. Применяя двукратно формулу (10) к функции F, определенной в (5), имеем:

$$F = Q_n(\tilde{L}), \quad \tilde{L} = \{L_k := C_n^k \ln C_n^k, ; k = 0, 1, \ldots, n\};$$

$$F'' = n(n-1) Q_{n-2}(\Delta^2 \tilde{L}). \tag{11}$$

Убедимся, что при заданных L_k, $(n-1)$ - мерный вектор $\Delta^2 \tilde{L}$ отрицателен, т.е. все $\Delta^2 L_k < 0$; отсюда следует неравенство $F''=(x) < 0$, $x \in I$ при $n \geq 2$ (при $n = 0, 1$, очевидно, $F'' \equiv 0$.) Последнее означает выпуклость вверх функции F. В самом деле,

$$\Delta L_k = \ln \frac{C_n^{k+1}}{C_n^k} = \ln \frac{n-k}{k+1},$$

$$\Delta^2 L_k = \ln \left(\frac{n-k-1}{k+2} \cdot \frac{k+1}{n-k} \right) =$$

$$= \ln \left(1 + \frac{n+1}{(k+1)(n-k-1)} \right) < 0, \quad k = 0, 1, \ldots, n-2. \tag{12}$$

Таким образом, выпуклость вверх функции F, и следовательно, и функции $(-H_2)$ доказана. В силу ее симметричности относительно прямой $x = 1/2$ можно заключить, что она достигает максимума в единственной точке $x = 0,5$.

Осталось сделать последний шаг - доказать неравенство

$$H'' = G'' - F'' < 0 \quad ; \tag{13}$$

этому посвящен следующий пункт.

5. Запишем, исходя из (11), (12),

$$F''(x) = n(n-1) \sum_{k=0}^{n-2} \Delta^2 L_k C_{n-2}^k x^k (1-x)^{n-2-k} =$$

$$= \sum_{k=0}^{n-2} (k+1)(n-k-1) \Delta^2 L_k C_n^{k+1} x^k (1-x)^{n-2-k} =$$

$$= \sum_{j=1}^{n-1} j(n-j) \Delta^2 L_{j-1} C_n^j x^{j-1} (1-x)^{n-j-1} =$$

$$= -\frac{n}{x(1-x)} \sum_{j=1}^{n-1} a_j p_n^j, \tag{14}$$

где

$$a_j := -\frac{j(n-j)}{n} \Delta^2 L_{j-1} =$$

$$= \frac{j(n-j)}{n} \ln \left(1 + \frac{n+1}{j(n-j)} \right), \quad j = 1, \ldots, n-1. \tag{15}$$

В п.6 будет показано, что все коэффициенты (15) не превосходят единицы, откуда, с учетом (8), следует неравенство

$$F''(x) \geq -\frac{n}{x(1-x)} \sum_{j=1}^{n-1} p_n^j \geq -\frac{n}{x(1-x)} \sum_{j=1}^{n} p_n^j = -\frac{n}{x(1-x)} = G''(x),$$

которое равносильно (13) и (3). Тем самым доказательство гипотезы Борисенко будет завершено.

6. Рассмотрим функцию

$$f(\varepsilon) := \varepsilon - \frac{\varepsilon^2}{4} - \ln(1+\varepsilon), \quad \varepsilon \geq 0.$$

Имеем, $f(0) = 0$,

$$f'(\varepsilon) = 1 - \frac{\varepsilon}{2} - \frac{1}{1+\varepsilon} = \frac{\varepsilon(1-\varepsilon)}{2(1+\varepsilon)} \begin{cases} > 0 & \text{при } 0 < \varepsilon < 1; \\ < 0 & \text{при } \varepsilon > 1. \end{cases}$$

Следовательно, с ростом ε функция f достигает максимума в точке $\varepsilon = 1$ и далее убывает. Так как $f(3/2) \approx 0,02 > 0$, то имеет место оценка

$$f(\varepsilon) > 0 \Leftrightarrow \frac{\ln(1+\varepsilon)}{\varepsilon} < 1 - \frac{\varepsilon}{4} \quad \text{при } 0 < \varepsilon \le 3/2 . \qquad (16)$$

Применим ее к (15), отметив предварительно, что

$$a_j = \frac{\ln(1+\varepsilon_j)}{\varepsilon_j} \cdot \frac{n+1}{n} , \text{ где } \varepsilon_j := \frac{n+1}{j(n-j)}, \ j = 1, \ldots, n-1 .$$

При фиксированном n величина ε_j достигает максимума по j при $j = 1$ ($\varepsilon_{max} = (n+1)/(n-1)$) и минимума - при $j = n/2$ ($\varepsilon_{min} = 4(n+1)/n^2 > 4/n$) . При $n \ge 5$, очевидно, $\varepsilon_{max} \le 3/2$, поэтому применима оценка (16); это дает

$$a_j \le \left(1 - \frac{\varepsilon_j}{4}\right) \cdot \frac{n+1}{n} \le \left(1 - \frac{\varepsilon_{min}}{4}\right) \cdot \frac{n+1}{n} \le$$

$$\le \left(1 - \frac{1}{n}\right) \cdot \frac{n+1}{n} = 1 - \frac{1}{n^2} < 1, \ j = 1, \ldots, n-1 .$$

При $n = 2, 3, 4$ требуемые условия $a_j < 1$ проверяются непосредственно по формуле (15):

$$a_j = \begin{cases} \frac{1}{2}\ln 4 < 1 & при \quad n = 2, j = 1; \\ \frac{2}{3}\ln 3 < 1 & при \quad n = 3, j = 1, 2; \\ \frac{3}{4}\ln\frac{8}{3} < 1 & при \quad n = 4, j = 1, 3; \\ \ln\frac{9}{4} < 1 & при \quad n = 4, j = 2. \end{cases}$$

Так как $a_j < 1$, $j = 1, \ldots, n-1$, то соотношение (13) доказано. Из последнего вытекает, что функция H (а значит, и функция H_1) выпукла вверх и, следовательно, достигает глобального максимума в единственной точке $x = 0,5$.

SUMMARY

A complex Bernoulli information source arising in problems of optimal coding is considered. The source's enthropy function is shown to be concave and to have a (unique) maximum point.

СПИСОК ЛИТЕРАТУРЫ

1. Борисенко А.А. О некоторых аспектах современной теории информации //Весник Сумского государственного университета.-1994.-№1, с.93-96.
2. Борисенко А.А. О преобразовании источников информации //Тез.докл.научно-технической конференции "Техника и физика электронных систем и устройств", ч.2,1995 - Сумы,Сумский государственный университет.

Поступила в редколлегию 8 сентября 1995г.

УДК 621.391.1

ПРЕОБРАЗОВАНИЕ ПОЗИЦИОННЫХ КОДОВ В БИНОМИАЛЬНЫЕ С МНОГОЗНАЧНЫМ АЛФАВИТОМ

Оканченко Е.Л., ст. преп., Протасова Т.А., асп.

Биномиальные коды с многозначным алфавитом эффективно решают ряд теоретических и практических задач информатики. К таким задачам относятся формирование комбинаторных кодов, помехоустойчивое кодирование, сжатие информации [1]. Кроме того, с помощью биномиальных кодов можно эффективно преобразовывать информацию и соответственно использовать при построении вычислительных цифровых устройств [2]. При этом очень часто возникают задачи преобразования позиционных кодов в биномиальные и обратно.

Многозначные биномиальные коды характеризуются следующим выражением [3]:

$$A = \sum_{l=0}^{x_{k-1}-1} C_{n-l-1}^{k-1} + \sum_{l=0}^{x_{k-2}-1} C_{n-l-2-x_{k-1}}^{k-2} + \ldots +$$