

УДК 621.3.037.37

**СРЕДНЯЯ ДЛИНА ДВОИЧНЫХ БИНОМИАЛЬНЫХ ЧИСЕЛ
ПРОИЗВОЛЬНОГО ДИАПАЗОНА**

И.А. Кулик

Сумский государственный университет

В статье определяются средние длины двоичных неравномерных биномиальных чисел для неполного и произвольного диапазонов. Полученные оценки раскрывают одно из важных свойств биномиальных чисел – свойство неравномерности. Результаты работы предоставляют возможность разработать строгие методы оценки и сравнительного анализа алгоритмов биномиального кодирования и декодирования.

ПОСТАНОВКА ПРОБЛЕМЫ

В статье [1] производится оценка средней длины двоичных неравномерных биномиальных чисел полного диапазона, генерируемых биномиальной системой счисления с параметрами n и k . Но практика биномиальных счета и кодирования требует постановки задачи в более общем виде, а именно нахождение средней длины двоичных неравномерных биномиальных чисел, принадлежащих произвольному диапазону.

Решение такой более общей задачи позволит более строго подойти к оценке стоимости биномиального кодирования и сложных характеристик алгоритмов, оперирующих неравномерными биномиальными числами. Например, методы биномиального сжатия двоичных кодов [2] будут иметь более высокую эффективность и более широкую область использования, если применяемые при преобразовании биномиальные числа брать из произвольного числового диапазона, выбор которого осуществляется исходя из минимизации кодовой длины. Но при таком гибком подходе к биномиальному сжатию существенную трудность представляет собой оценка среднего времени преобразования исходных двоичных кодов в биномиальные числа [3]. Рассматриваемое время преобразования напрямую связано с числом отбрасываемых разрядов от двоичной комбинации, что, в свою очередь, тесно связано со средней длиной двоичных неравномерных биномиальных чисел, принадлежащих произвольному диапазону. Точно такая же проблема стоит и при обратном преобразовании биномиального числа к соответствующему двоичному коду.

Далее биномиальное сжатие длинных последовательностей с переменными значениями числа k единиц (например, двоичные последовательности, соответствующие телевизионным строкам изображения [4]) объективно будет связано с биномиальными числами неполного или произвольного диапазонов. При значительных n и k число двоичных биномиальных чисел огромно и, как правило, намного превышает используемые на сегодняшний день размеры информационного поля телевизионного экрана. Но чтобы строго

определить степень биномиального сжатия изображений для этого случая, также необходимо знание средней длины биномиальных чисел, принадлежащих неполному или произвольному диапазонам.

Таким образом, задача по определению средней длины двоичных неравномерных биномиальных чисел произвольного диапазона является актуальной, а ее решение имеет определенное научно-практическое значение.

1 СОСТОЯНИЕ ПРОБЛЕМЫ И ПОДХОД К ЕЕ РЕШЕНИЮ

Пусть имеется полное множество X двоичных неравномерных биномиальных чисел, генерируемых двоичной биномиальной системой счисления с параметрами n и k . Мощность такого полного множества $|X| = C_n^k$ [5]. Длина r двоичных неравномерных биномиальных чисел $X_j \in X$ изменяется в пределах $\min(k, n-k) \leq r \leq n-1$, $X_j = x_1 x_2 \dots x_i \dots x_r$, $j = 1, \overline{C_n^k}$. В соответствии с числовой функцией десятичный количественный эквивалент двоичного биномиального числа определяется как

$$F_j = \text{dec } X_j = \sum_{i=1}^r x_i C_{n-i}^{k-q_i},$$

где $q_i = \sum_{t=1}^{i-1} x_t$ [5].

Необходимо определить среднюю длину $L_{n,k}[\alpha, \beta]$ двоичных неравномерных биномиальных чисел X_j в более общем виде, когда

$$\alpha \leq X_j \leq \beta, \quad \alpha \geq 0, \quad \beta \leq C_n^k - 1.$$

В этой связи можно рассматривать три случая для средней длины $L_{n,k}[\alpha, \beta]$:

- 1) $\alpha = 0$, $\beta = C_n^k - 1$, то есть полный диапазон чисел с параметрами n и k ;
- 2) $\alpha = 0$, $\beta < C_n^k - 1$, то есть диапазон чисел произвольный по значению конечного числа β ;
- 3) $\alpha > 0$, $\beta < C_n^k - 1$, то есть диапазон чисел произвольный по значению и начального, и конечного чисел α и β соответственно.

В работе [1] автором был рассмотрен случай полного диапазона двоичных неравномерных биномиальных чисел, когда $\alpha = 0$ и $\beta = C_n^k - 1$. Полученная оценка средней длины полного диапазона биномиальных чисел для параметров n и k имеет следующий вид:

$$L_{n,k} = L_{n,k}[\mathbf{0}, C_n^k - \mathbf{1}] = \frac{k(n-k)(n+2)}{(k+1)(n-k+1)}. \quad (1)$$

Принимая во внимание известное соотношение для биномиальных коэффициентов

$$C_n^k = C_{n-1}^k + C_{n-1}^{k-1}, \quad (2)$$

можно предположить, что исходное множество X двоичных биномиальных чисел с параметрами n и k можно разбить на два подмножества X' и X'' биномиальных чисел, но уже с параметрами $n-1$ и k для X' и $n-1$ и $k-1$ для X'' . Границей такого разбиения будет являться значение старшего разряда x_1 биномиальных чисел $X_j \in X$. В подмножестве X' содержатся биномиальные числа, в двоичной записи которых $x_1 = 0$, а в подмножестве X'' – биномиальные числа, у которых $x_1 = 1$. Продолжая разбиение согласно (2) получаемых подмножеств по значениям $x_i, i = \overline{1, r}$, получаем диапазоны двоичных неравномерных биномиальных чисел с параметрами $n-i$ и $k-q_i$, для которых применима оценка (1) средней длины, где q_i – сумма единичных значений биномиальных разрядов от x_1 до x_{i-1} включительно. Значения биномиальных разрядов $x_1 x_2 \dots x_i \dots x_r$, в совокупности составляющие биномиальное число, будут указывать на те оценки (1) средней длины, которые и будут в конечном итоге определять результирующее значение средней длины биномиальных чисел произвольного диапазона.

2 АНАЛИЗ СТРУКТУРЫ НЕРАВНОМЕРНОГО БИНОМИАЛЬНОГО КОДА

С целью четкого определения целей проводимого анализа и разграничения решаемых задач введем следующие определения.

Определение 1 Диапазон $[X_1, X_2]$ двоичных неравномерных биномиальных чисел с параметрами n и k , расположенных в лексикографическом порядке, называется полным, если количество P чисел диапазона $P = C_n^k$.

Определение 2 Диапазон $[X_1, X_2]$ двоичных неравномерных биномиальных чисел с параметрами n и k , расположенных в лексикографическом порядке, называется неполным, если $X_1 = 0$, а $X_2 < C_n^k - 1$.

Определение 3 Диапазон $[X_1, X_2]$ двоичных неравномерных биномиальных чисел с параметрами n и k , расположенных в лексикографическом порядке, называется произвольным, если $X_1 > 0$, а $X_2 < C_n^k - 1$.

Таким образом, формула (1) позволяет вычислить среднюю длину двоичных неравномерных биномиальных чисел полного диапазона.

Выражение

$$C_{n-i+1}^{k-q_i} = C_{n-i}^{k-q_i} + C_{n-i}^{k-q_i-1}$$

является математическим обоснованием i -го разбиения исходного множества двоичных неравномерных биномиальных чисел полного диапазона с параметрами $n-i+1$ и $k-q_i$ на два полнодиапазонных подмножества чисел с параметрами $n-i, k-q_i$ и $n-i, k-q_i-1$.

Схема разбиения биномиальных чисел с параметрами $n = 5$ и $k = 3$ показана на рисунке 1.

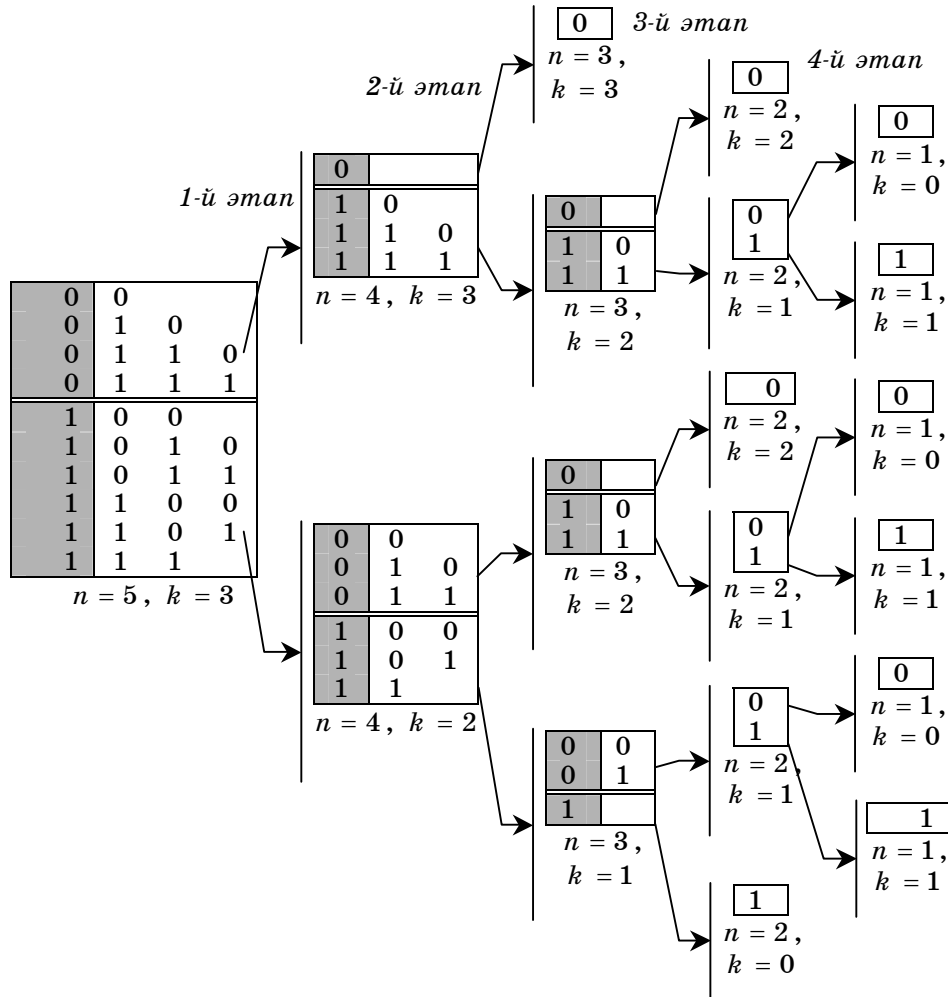


Рисунок 1 – Разбиение исходного множества биномиальных чисел на полнодиапазонные подмножества

3 СРЕДНЯЯ ДЛИНА БИНОМИАЛЬНЫХ ЧИСЕЛ НЕПОЛНОГО И ПРОИЗВОЛЬНОГО ДИАПАЗОНОВ

С целью определения средней длины двоичных неравномерных биномиальных чисел неполного и произвольного диапазонов сформулируем следующие леммы, приведенные в виду их очевидности без доказательства.

Лемма 1 Если значения переменной $l = l' + C$ для любых l и l' , где C – постоянная, то их средние значения находятся в следующей зависимости:

$$L_{cp} = L'_{cp} + C.$$

Лемма 2 Среднее значение L_{cp} переменной l на числовом диапазоне $N = N' + N''$

$$L_{cp} = \frac{N'}{N} L'_{cp} + \frac{N''}{N} L''_{cp},$$

где L'_{cp} – среднее значение переменной l на числовом диапазоне N' ;

L''_{cp} – среднее значение переменной l на числовом диапазоне N'' .

Основываясь на вышеуказанных леммах, сформулируем и приведем доказательство следующей теоремы.

Теорема 1 Средняя длина $L_{n,k}[0, X]$ двоичных неравномерных биномиальных чисел неполного диапазона от 0 до X равна

$$L_{n,k}[0, X] = \frac{\sum_{i=1}^r x'_i [L(n-i, k-q_i) + i] N(n-i, k-q_i)}{X+1}, \quad (3)$$

где x'_i – двоичные разряды биномиального числа $X' = X + 1$;

$L(n-i, k-q_i)$ – средняя длина двоичных неравномерных биномиальных чисел полного диапазона с параметрами $n-i$ и $k-q_i$;

$N(n-i, k-q_i)$ – диапазон двоичных неравномерных биномиальных чисел полного диапазона с параметрами $n-i$ и $k-q_i$.

Доказательство. Обратимся к числовой функции двоичной биномиальной системы счисления с параметрами n и k :

$$F(n, k) = \sum_{i=1}^r x_i C_{n-i}^{k-q_i}, \text{ где } q_i = \sum_{t=1}^{i-1} x_t,$$

и представим ее в следующем виде, используя замену индексной переменной i на $\alpha + 1$ ($i = \alpha + 1$):

$$F(n, k) = x_1 C_{n-1}^{k-q_1} + \sum_{i=2}^r x_i C_{n-i}^{k-q_i} = x_1 C_{n-1}^k + \sum_{\alpha=1}^{r'} x'_\alpha C_{n'-\alpha}^{k'-q'_\alpha}, \quad (4)$$

где $q_1 = 0$, $n' = n - 1$, $k' = k - x_1$, $\min(k', n' - k') \leq r' \leq n' - 1$,

$$q'_\alpha = \sum_{t'=1}^{\alpha-1} x'_{t'} = \sum_{t=2}^{i-1} x_t = q_i - x_1.$$

Второе слагаемое выражения (4) представляет собой числовую функцию биномиальной системы счисления $F'_{x_1}(n', k')$, но уже с параметрами $n' = n - 1$ и $k' = k - x_1$, то есть

$$F(n, k) = x_1 C_{n-1}^k + F'_{x_1}(n-1, k-x_1).$$

При этом, очевидно, в зависимости от значения x_1 в полном диапазоне исходного множества биномиальных чисел можно выделить два полных диапазона биномиальных систем счисления:

$$F'_0(n-1, k) \text{ при } x_1 = 0 \text{ и } F'_1(n-1, k-1) \text{ при } x_1 = 1.$$

Таким образом, исключение старшего разряда x_1 из биномиальных чисел с параметрами n и k приводит к рассмотрению двух меньших (по диапазону) биномиальных систем счисления с параметрами $n-1$, k и $n-1$, $k-1$, числа которых сохраняют исходный порядок расположения.

Исключим теперь x_2 исходных биномиальных чисел с параметрами n и k (или, что тоже самое, x'_1 биномиальных чисел с параметрами $n-1$, k и $n-1$, $k-1$). Получаем

$$\begin{aligned} F(n, k) &= x_1 C_{n-1}^k + x'_1 C_{n'-1}^{k'-q'_1} + \sum_{\alpha=2}^{r'} x'_\alpha C_{n'-\alpha}^{k'-q'_\alpha} = \\ &= x_1 C_{n-1}^k + x_2 C_{n-2}^{k-x_1} + \sum_{\alpha=1}^{r''} x''_\alpha C_{n''-\alpha}^{k''-q''_\alpha}, \end{aligned} \quad (5)$$

где $q'_1 = 0$, $x'_1 = x_2$, $n'' = n' - 1 = n - 2$, $k'' = k' - x'_1 = k - x_1 - x_2$,

$$\min(k'', n'' - k'') \leq r'' \leq n'' - 1, \quad q''_\alpha = \sum_{t''=1}^{\alpha-1} x''_{t''} = q'_\alpha - x'_1 = q_i - x_1 - x_2.$$

Третье слагаемое выражения (5) представляет собой числовую функцию биномиальной системы счисления $F''_{x_1 x_2}(n'', k'')$, но уже с параметрами $n'' = n' - 1 = n - 2$ и $k'' = k' - x'_1 = k - x_1 - x_2$, то есть

$$F(n, k) = x_1 C_{n-1}^k + x_2 C_{n-2}^{k-x_1} + F''_{x_1 x_2}(n-2, k-x_1-x_2).$$

При этом, очевидно, в зависимости от значений x_1 и x_2 в полном диапазоне исходного множества биномиальных чисел можно выделить четыре полных диапазона биномиальных систем счисления:

$$F''_{00}(n-2, k) \text{ при } x_1 = x_2 = 0, \quad F''_{01}(n-2, k-1) \text{ при } x_1 = 0 \text{ и } x_2 = 1,$$

$$F''_{10}(n-2, k-1) \text{ при } x_1 = 1 \text{ и } x_2 = 0, \quad F''_{11}(n-2, k-2) \text{ при } x_1 = x_2 = 1.$$

Продолжая исключение разрядов $x_3, x_4, \dots, x_i, \dots, x_r$ из разрядной сетки исходных биномиальных чисел с параметрами n и k вплоть до получения одноэлементных биномиальных систем счисления, приходим к следующему заключению, что в любом неполном диапазоне исходных биномиальных чисел от 0 до X с параметрами n и k можно выделить полнодиапазонные множества биномиальных чисел, соответствующие биномиальным системам счисления с параметрами $n-i$ и

$k - x_1 - x_2 - \dots - x_{i-1}$. При этом числовую функцию в общем виде можно представить как

$$\begin{aligned}
 F(n, k) &= x_1 C_{n-1}^k + x_2 C_{n-2}^{k-x_1} + \dots + x_d C_{n-d}^{k-x_1-x_2-\dots-x_{d-1}} + \dots \\
 &+ F_{x_1 x_2 \dots x_d}^{(d)}(n-d, k-x_1-x_2-\dots-x_{d-1}) = x_1 N(n-1, k) + \\
 &+ x_2 N(n-2, k-x_1) + \dots + x_d N(n-d, k-x_1-x_2-\dots-x_{d-1}) + \quad (6) \\
 &+ F_{x_1 x_2 \dots x_d}^{(d)}(n-d, k-x_1-x_2-\dots-x_d) = \\
 &= \sum_{i=1}^d x_i N(n-i, k-q_i) + F_{x_1 x_2 \dots x_d}^{(d)}(n-d, k-q_{d+1}).
 \end{aligned}$$

Из выражения (6) следует, что разряды исходного биномиального числа $x_1 x_2 \dots x_i \dots x_r$ служат признаками включения биномиальных чисел полного диапазона биномиальных систем счисления с параметрами $n-i$ и $k-q_i$. Каждая из включенных биномиальных систем счисления характеризуется своей средней длиной $L(n-i, k-q_i)$ принадлежащих ей неравномерных биномиальных чисел. Биномиальные числа включаемых биномиальных систем счисления есть составные части исходных биномиальных чисел, от которых по длине они отличаются на i разрядов. Следовательно, согласно лемме 1 средняя длина исходных неравномерных биномиальных чисел с параметрами n и k , составными частями которых являются биномиальные числа с параметрами $n-i$ и $k-q_i$, равна

$$L'_i = L(n-i, k-q_i) + i, \quad (7)$$

где i – фиксированная разность между длиной биномиального числа с параметрами $n-i$, $k-q_i$ и длиной соответствующего ему числа с параметрами n и k .

Тогда на основании леммы 2 и общего выражения (6) для среднего значения $L_{n,k}[0, X]$ можно записать

$$L_{n,k}[0, X] = \frac{\sum_{i=1}^r x'_i L'_i \cdot N(n-i, k-q_i)}{N},$$

где N – длина неполного диапазона исходных биномиальных чисел;

x'_i – двоичные разряды биномиального числа $X' = X + 1$.

С учетом (7) и того, что $N = X + 1$, приходим к искомому выражению для средней длины двоичных неравномерных биномиальных чисел неполного диапазона от 0 до X :

$$L_{n,k}[0, X] = \frac{\sum_{i=1}^r x'_i [L(n-i, k-q_i) + i] N(n-i, k-q_i)}{X+1},$$

что и требовалось доказать.

Таким образом, выражение (3) позволяет вычислить среднюю длину неравномерных биномиальных чисел с параметрами n и k , начальной точкой диапазона которых является число 0.

Пример 1 Определить среднюю длину $L_{6,3}[0,12]$ неравномерных биномиальных чисел с параметрами $n = 6$ и $k = 3$ для неполного диапазона от 0 до 12.

Решение. Следующим за последним числом диапазона является 5-разрядное биномиальное число 10100, количественным эквивалентом которого есть десятичное число 13. Отличными от 0 в кодовой записи биномиального числа 10100 будут разряды $x'_1 = 1$ и $x'_3 = 1$, а соответствующие им частичные суммы разрядов равны $q_1 = 0$ и $q_3 = 1$. Следовательно, согласно выражению (3) для указанного случая

$$L_{6,3}[0,12] = \frac{[L(5,3) + 1]N(5,3) + [L(3,2) + 3]N(3,2)}{13}.$$

Значения $L(5,3) = L_{5,3}$ и $L(3,2) = L_{3,2}$ вычисляются в соответствие с формулой (1) для полного диапазона биномиальных чисел:

$$L_{5,3} = \frac{3(5-3)(5+2)}{(3+1)(5-3+1)} = \frac{42}{12} = 3,5 \quad \text{и} \quad L_{3,2} = \frac{2(3-2)(3+2)}{(2+1)(3-2+1)} = \frac{10}{6} \approx 1,7.$$

$N(5,3)$ и $N(3,2)$ есть общее количество биномиальных чисел, определяемое как число сочетаний 3 из 5 в первом случае и 2 из 3 во втором:

$$N(5,3) = C_5^3 = 10 \quad \text{и} \quad N(3,2) = C_3^2 = 3.$$

Подставляя найденные значения в общую формулу, получаем искомый результат:

$$L_{6,3}[0,12] = \frac{[3,5 + 1] \cdot 10 + [1,67 + 3] \cdot 3}{13} = \frac{45 + 14,01}{13} \approx 4,54.$$

Имея более расширенную задачу – определение средней длины произвольного диапазона, введем следующие обозначения: X_1 – начальное число, а X_2 – последнее число произвольного диапазона неравномерных биномиальных чисел.

Теорема 2 Средняя длина $L_{n,k}[X_1, X_2]$ двоичных неравномерных биномиальных чисел произвольного диапазона от $X_1 \neq 0$ до $X_2 < C_n^k$, $X_1 \leq X_2$, равна

$$L_{n,k}[X_1, X_2] = \frac{(X_2 + 1)L_{n,k}[0, X_2] - (X_1 + 1)L_{n,k}[0, X_1]}{X_2 - X_1 + 1}. \quad (8)$$

Доказательство Согласно лемме 2 среднюю длину $L_{n,k}[0, X_2]$, где $X_1 \leq X_2$, можно представить как

$$L_{n,k}[0, X_2] = \frac{N[0, X_1]}{N[0, X_2]} L_{n,k}[0, X_1] + \frac{N[X_1, X_2]}{N[0, X_2]} L_{n,k}[X_1, X_2], \quad (9)$$

где $N[0, X_1]$ – количество биномиальных чисел диапазона от 0 до X_1 ;

$N[0, X_2]$ – количество биномиальных чисел диапазона от 0 до X_2 ;

$N[X_1, X_2]$ – количество биномиальных чисел диапазона от X_1 до X_2 .

Выражаем искомую среднюю длину $L_{n,k}[X_1, X_2]$ из соотношения (9):

$$L_{n,k}[X_1, X_2] = \frac{N[0, X_2]L_{n,k}[0, X_2] - N[0, X_1]L_{n,k}[0, X_1]}{N[X_1, X_2]}.$$

Принимая во внимание, что $N[0, X_1] = X_1 + 1$, $N[0, X_2] = X_2 + 1$ и $N[X_1, X_2] = X_2 - X_1 + 1$, в результате получаем

$$L_{n,k}[X_1, X_2] = \frac{(X_2 + 1)L_{n,k}[0, X_2] - (X_1 + 1)L_{n,k}[0, X_1]}{X_2 - X_1 + 1},$$

что и требовалось доказать.

Пример 2 Определить среднюю длину $L_{6,3}[5, 12]$ неравномерных биномиальных чисел с параметрами $n = 6$ и $k = 3$ для произвольного диапазона от 5 до 12.

Решение. Таким образом, $X_1 = 5$, а $X_2 = 12$. Воспользуемся формулой (8). Значение $L_{6,3}[0, 12]$ известно из примера 1: $L_{6,3}[0, 12] = 4,54$. Значение $L_{6,3}[0, 5]$ находим согласно (3) по аналогии с предыдущим примером. С учетом того, что биномиальный вид числа 6 есть 01011 и отличными от 0 являются разряды $x'_2 = 1$, $x'_4 = 1$ и $x'_5 = 1$, то

$$\begin{aligned} L_{6,3}[0, 5] &= \frac{[L(4, 3) + 2]N(4, 3) + [L(2, 2) + 4]N(2, 2) + [L(1, 1) + 5]N(1, 1)}{6} = \\ &= \frac{4,25 \cdot 4 + 4 + 5}{6} \approx 4,33. \end{aligned}$$

Тогда подставляя соответствующие значения в выражение (8), определяем искомую среднюю длину произвольного диапазона:

$$L_{6,3}[5, 12] = \frac{13 \cdot 4,54 - 6 \cdot 4,33}{8} = \frac{59,02 - 25,98}{8} = 4,13.$$

ЗАКЛЮЧЕНИЕ

Выражения (3) и (8) для вычисления среднего значения длины двоичных неравномерных биномиальных чисел неполного и произвольного диапазонов наряду с ранее найденным соотношением (1) для вычисления средней длины полного диапазона всесторонне и в достаточно полной мере раскрывают одно из значимых свойств биномиальных чисел – свойство неравномерности. Обнаруженные автором зависимости для средней кодовой длины биномиальных чисел позволяют

разрабатывать математически более строгие способы оценки сложностных характеристик алгоритмов биномиального кодирования и декодирования, которые оперируют неравномерными биномиальными числами произвольного диапазона, а также получать эффективные способы решения задач информационного характера таких, как определение степени биномиального сжатия, нахождение информационной избыточности биномиальных чисел и т.д.

SUMMARY

AVERAGE LENGTH OF BINARY BINOMIAL NUMBERS OF ARBITRARY RANGE

I.A. Kulik

In the paper the average lengths of binary uneven binomial numbers are determined for incomplete and arbitrary ranges. The obtained estimations open one of important properties of binomial numbers – the property of unevenness. The results of the paper give an opportunity to develop rigorous methods of estimation and comparative analysis of algorithms of binomial coding and decoding.

СПИСОК ЛИТЕРАТУРЫ

1. Кулик И.А. О средней длине двоичных биномиальных чисел // Вісник СумДУ. – 2004. – № 12(71). – С. 106. – 112.
2. Чередниченко В.Б. Метод сжатия двоичных кодов на основе биномиальных чисел // Вісник СумДУ. – 2006. – № 4(88). – С. 61-68.
3. Чередниченко В.Б. Оценка эффективности сжатия равновесных кодов на основе биномиальных чисел // Вісник СумДУ. – 2005. – № 9(81). – С. 26-31.
4. Зубарев Ю.Б., Глоризов Г.Л. Передача изображений. – М.: Радио и связь, 1989. – 336 с.
5. Борисенко А.А. Биномиальный счет. Теория и практика: Монография. – Сумы: ИТД "Университетская книга", 2004. – 170 с.

И.А. Кулик, канд. техн. наук, СумГУ,
г. Сумы

Поступила в редакцию 12 сентября 2007 г.