10 Префиксное и разделимое кодирование. Коды Фано и Хаффмена

• Дан источник Бернулли $A = \{a_1, \ldots, a_k\}$ с вероятностями букв $P = \{p_1, \ldots, p_k\}$. *Энтропией* источника Бернулли A называется величина $\mathcal{H}(A) = -\sum_{i=1}^{k} p_i \cdot \log p_i$. Пусть $B = \{0, 1\}$ — *кодирующий алфавит.* Через B^* обозначим множество всех слов конечной длины в алфавите B. Вложение $\Sigma : A \to B^*$ называется *алфавитным кодированием.* Пусть $L = \{l_1, \ldots, l_k\}$ — множество длин кодовых слов. *Стоимостью кодирования* Σ называется величина $C(\Sigma) = \sum_{i=1}^{k} p_i \cdot l_i$. Кодирование Σ *оптимально*, если его стоимость кодирования является наименьшей. Кодирование Σ называется *разделимым*, если из равенства $\Sigma(a_{i_1}) \ldots \Sigma(a_{i_m}) = \Sigma(a_{j_1}) \ldots \Sigma(a_{j_n})$ следует, что $m = n$ и $i_t = j_t$ для всех $t = 1 \ldots, m$ (другими словами любая последовательность кодовых слов единственным образом разделима на кодовые слова). Кодирование Σ называется *префиксным*, если никакое кодовое слово не является началом (префиксом) никакого другого кодового слова. Любой префиксный код, очевидно, является разделимым. Обратное неверно.
• Выполнение *неравенства Крафта-Макмиллана*:

$$\sum_{i=1}^{k} q^{-l_i} \leq 1$$

является необходимым и достаточным условием для существования q-значных префиксного и разделимого кодов с заданным набором длин L.
• Пусть функция $f(x)$ выпукла вверх в области I. Тогда для произвольных чисел $\alpha_i > 0$, $i = 1, \ldots, k$ таких, что $\sum_{i=1}^{k} \alpha_i = 1$, и любых $x_1, \ldots, x_k \in I$ справедливо *неравенство Йенсена*:

$$\sum_{i=1}^{k} \alpha_i \cdot f(x_i) \leq f(\sum_{i=1}^{k} \alpha_i \cdot x_i).$$

10.1 Пусть кодом каждого из данных чисел является его двоичное представление наименьшей возможной длины (без нулей слева), например $1 \to 1, 2 \to 10, 4 \to 100$ и т.д. Является ли кодирование разделимым?
а) $L = \{1, 2, 4, 9, 50\}$;
б) $L = \{1, 2, 4, 17, 98\}$.

10.2 Построить двоичный префиксный код с заданной последовательностью длин кодовых слов:
а) $L = \{1, 2, 3, 3\}$;
б) $L = \{1, 2, 4, 4, 4\}$;
в) $L = \{1, 2, 3, 3, 4, 4, 4, 4\}$.

10.3 Может ли набор чисел L быть набором длин кодовых слов разделимого кода в q-значном алфавите?
а) $L = \{1, 2, 2, 3\}$;
б) $L = \{2, 2, 2, 4, 4, 4\}$.

10.4 Пусть в алфавитном двоичном коде C таком, что $|C| > 2^n$ каждое слово имеет длину, не превышающую n. Может ли код C быть разделимым?

10.5 Построить коды Фано и Хаффмена, найти стоимости кодирований для источников Бернулли с вероятностями букв:
а) $P = \{0,5; 0,2; 0,1; 0,09; 0,08; 0,03\}$;

б) $P = \{0,4; 0,2; 0,1; 0,1; 0,1; 0,1\}$;
в) $P = \{0,4; 0,3; 0,1; 0,07; 0,06; 0,04; 0,03\}$.

10.6 Доказать, что кодирование Фано префиксное.

10.7 Построить оптимальный код для q-значного источника Бернулли с данными вероятностями букв:
а) $P = \{0,3; 0,2; 0,2; 0,2; 0,05; 0,05\}, q = 3$;
б) $P = \{0,4; 0,2; 0,1; 0,1; 0,1; 0,05; 0,05\}, q = 3$.

10.8 Для заданного q указать набор вероятностей P, при котором существует q-значный префиксный код с заданным набором длин кодовых слов L, являющийся оптимальным. Построить этот код.
а) $q = 2, L = \{1, 2, 4, 4, 5, 5\}$;
б) $q = 2, L = \{2, 2, 2, 3, 3\}$.

10.9 Доказать, что энтропия $\mathcal{H}(A)$ источника Бернулли A неотрицательна и равна 0 $\Longleftrightarrow$ когда одна из вероятностей букв равна 1, а остальные вероятности равны 0.

10.10 Дан источник Бернулли A с вероятностями букв $\{p_1, \ldots, p_k\}$, где $\sum_{i=1}^{k} p_i = 1$. Доказать, что справедливо неравенство $\mathcal{H}(A) \leq \log k$.

10.11 Дан источник Бернулли с вероятностями букв $\{p_1, \ldots, p_k\}$, где $\sum_{i=1}^{k} p_i = 1$. Доказать, что для любых неотрицательных чисел q_i, $i = 1, \ldots, k$, таких что $\sum_{i=1}^{k} q_i = 1$, выполняется неравенство $-\sum_{i=1}^{k} p_i \log q_i \geq -\sum_{i=1}^{k} p_i \log p_i$.

10.12 Префиксное q-значное кодирование Σ называется *полным*, если для любого слова v в кодирующем алфавите справедливо одно из следующих условий:
1) слово v является префиксом (необязательно собственным) некоторого слова из Σ;
2) некоторое слово из Σ является собственным префиксом слова v.
Доказать, что префиксный код с q-значным кодирующим алфавитом полный $\Longleftrightarrow$ выполняется равенство $\sum_{i=1}^{r} q^{-l_i} = 1$, где $l_1, \ldots, l_r$ — длины кодовых слов.

Семинар 11 "Код Шеннона, графы Маркова".
Коды Шеннона. Теорема Шеннона. Теоремы Маркова. Граф Маркова.