Teoria informacji i kodowania: Lista 11

Zadanie 1. Rozważmy dwa rozkłady: $p_1 = (1 - \alpha, 0, \alpha)$ i $p_2 = (0, 1 - \alpha, \alpha)$. Pokaż bezpośrednio z definicji (nie używając przepustowości kanałów), że optymalnym rozkładem q minimalizującym

$$\min_q \max_{p_\theta} D(p_\theta || q)$$

jest $q = \frac{p_1 + p_2}{2}$.

Zadanie 2. Kod δ Eliasa używa kodu γ do zapisania $\lfloor \log i \rfloor + 1$ i następnie zapisuje i naiwnie. Możemy oczywiście zapisać $\lceil \log i \rceil$ (mniej więcej) kodem δ i potem i naiwnie itd. Poprowadź to rozumowanie do $\log^*$ i zdefiniuj odpowiedni kod.

Zadanie 3 (Kodowanie Tunstalla *)**.** Zwykle kodowania przekształca źródło symbol po symbolu na ciąg o zmiennej długości.

Rozważmy dualny problem kodowanie *zmiennej liczby symboli* ze źródła na kod o stałej długości: Kod o zmiennej długości dla i.i.d. ciągu zmiennych losowych $X_1, \ldots, X_n$, $X \sim p(x)$ ze zbioru $\mathcal{X} = \{0, 1, \ldots, m-1\}$, jest zdefiniowany przez wolny od prefiksów zbiór $A_D \subseteq \mathcal{X}^*$, gdzie $|A_D| = D$. Dowolny ciąg $X_1, \ldots, X_n$ jest reprezentowany jako konkatenacja słów z A_D i reprezentowany jako ciąg elementów z $\{0, \ldots, D-1\}^*$.

Wydajność takiego kodu to

$$R(A_D) = \frac{\log D}{\mathbb{E}[L(A_D)]}$$

gdzie $\mathbb{E}[L(A_D)]$ to oczekiwaną długość wyrażenia z A_D.

1. Udowodnij, że reprezentacja jako konkatenacja elementów z A_D jest najwyżej jedna.

2. Udowodnij, że $R(A_D) \geq H(X)$.

3. Konstruowania A_D można traktować jako konstruowanie m-arnego drzewa, którego liśćmi są ciągi z A_D. Załóżmy, że $D = 1 + k(m-1)$ dla pewnej liczby całkowitej $k \geq 1$. Rozważmy następujący algorytm (Tunstall)

   $A \leftarrow \{0, \ldots, m-1\}$ z z prawdopodobieństwami $p_0, \ldots, p_{m-1}$. ▷ Odpowiada to pełnemu drzewu m-arnemu o głębokości 1.
   while Liczba liści $< D$ **do**
       Rozwiń węzeł o największym prawdopodobieństwie o jeden poziom i dostosuj prawdopodobieństwa. ▷ Na przykład, jeśli 0 jest węzłem o największym prawdopodobieństwie p', to nowym zbiorem jest $A = \{00, 01, \ldots, 0(m-1), 1, \ldots, (m-1)\}$ i prawdopodobieństwach $p'p_0, \ldots, p'p_{m-1}$.

   Pokaż, że algorytm Tunstall jest optymalny w tym sensie, że konstruuje on zbiór o najlepszym $R(A_D)$ dla danego D

4. Pokaż, że istnieje D i A_D takie, że $R(A_D) < H(X) + 1$.

Zadanie 4. Kod Fibonacciego dla liczby n jest konstruowany w następujący sposób:

znajdź najmniejszą liczbę Fibonacciego F_N takie że $n \leq F_N$, reprezentacja będzie miała $N + 1$ bitów.
Ustaw $d(N+1) \leftarrow 1$.
while $N \geq 0$ **do**
    if $F_N \leq i$ **then**
        $d(N) \leftarrow 1, i \leftarrow i - F_N$
    else
        $d(N) \leftarrow 0$
    $N \leftarrow N - 1$

Pokaż, że:

1. Kod Fibonacciego jest kodem jednoznacznie dekodowalnym (wystarczy pokazać, że 11 występuje tylko na końcu kodu)

2. Kod Fibonacciego jest α uniwersalny dla procesów $\mathcal{X} = \{X_i\}_{i=0}^{\infty}$, gdzie X_i są i.i.d. i przyjmują wartości ze zbioru liczb całkowitych dodatnich i rozkład dla X_i jest monotoniczny, tj. $i < j \implies p(i) \geq p(j)$. (wystarczy porównać długość kodu Fibonacciego z długością kodu γ Eliasa, który jest 2 uniwersalny).

Zadanie 5. Rozważmy zmienną X_N o rozkładzie jednostajnym na $\{1, \ldots, N\}$. Niech C_δ oznacza kodowanie δ Eliasa. Pokaż, że:

$$\lim_{n \to \infty} \frac{|C_\delta(X_N)|}{H(X_N)} = 1$$

Zadanie 6 (*). Niech Z będzie dyskretną zmienną losową przyjmująca wartości dodatnie której wartość oczekiwana wynosi $\mu = \mathbb{E}[Z]$. Pokaż, że:

$$H(Z) \leq (\mu + 1)\log(\mu + 1) - \mu \log \mu$$

Sprawdź, że nierówność jest prawdziwa dla rozkładu geometrycznego, tj.

$$\mathbb{P}[Z = k] = (1 - p)^{k-1} p$$

Pokaż, że rozkład geometryczny maksymalizuje entropię przy ustalonej wartości oczekiwanej (na całkowitych dodatnich).

Zadanie 7. Niech $\ell_1, \ldots, \ell_m$ będą dodatnimi liczbami całkowitymi o sumie $\sum_{i=1}^{m} \ell_i = n$. Pokaż, że taki ciąg można zakodować znacznie lepiej niż trywialnie, tzn. istnieje kodowanie używające

$$m \log \frac{n}{m} + \alpha m$$

bitów dla pewnej stałej α.

W tym celu zdefiniuj zmienną losową Z, taką że

$$\mathbb{P}[Z = \ell] = \frac{|\{i \,:\, \ell = \ell_i\}|}{m} \quad ,$$

Niech C: kod Shannona dla zmiennej Z. Pokaż, że

$$\sum_{i=1}^{m} |C(\ell_i)| \leq mH[Z] + m$$

Skorzystaj z poprzedniego zadania.