Final, Information Theory, 2016 Spring

BY HEHAO (PRIVATE RECALLING VERSION)

1.

Random variable X has probability distribution $\{p_1, ..., p_n\}$, and $P_{\max} = \max_i p_i$,

Prove that, $H(X) \leqslant (1 - p_{\max}) \log(n-1) + h_b(p_{\max})$,

where $h_b(p) = p \log(p^{-1}) + (1-p) \log((1-p)^{-1})$. jensen's ineq.

2.

Considering a block presentation of a binary sequence. For example, for 0011100000111, since it begins with 0, and have 2 0's, 3 1's, 5 0's, 3 1's sequentially, thus it can be present by $[0, 2, 3, 5, 3]$.

Easy to see, the general form of this presentation is $[X_1, L_1, L_2, L_3, ...]$

1) What's the probability distribution of L_1? 1/2, 1/2^2, 1/2^3, 1/2^4 ...

2) Give the entropy of $H(L_1)$, $H(X_1, L_1, ..., L_{100})$. independent, sum them up

3) Give the optimal compress coding scheme for block presentation. And we call this coding scheme 'Block Code'.

4) What's average length of 'Block Code' for $H(X_1, L_1, ..., L_{100})$, and what's the expectation length of original sequence $E\left(\sum_{i=1}^{100} L_i\right)$?

5) Compare the 'Block Code' and direct presentation of the binary sequence.

3.

Consider a continuous random variable X and a discrete random variable.

1) Give the form of condition entropy $h(X|Y)$ and $H(Y|X)$.

2) What about mutual information $I_1(X;Y) = h(X) - h(X|Y)$ and $I_2(X;Y) = H(Y) - H(Y|X)$, are they the same?

4.

X and Z are two indepent variable, $Z \sim \text{Bernoulli}(p)$, X has pd $\{q_1, ..., q_n\}$, and $Y = \text{XZ}$.

1) Give the relationship between H(Y) and H(X), H(Z). H(Y) = pH(X) + H(Z)

2) Which p and q maximize the H(Y)? 用拉格朗日乘子法 — 有简单的方法吗?

3) Consider the communicate channel with input X and output Y. What's the channel capacity $C(p)$? Which p maximize the $C(p)$? Can you give a direct explain?

p log(n) p = 1时成立，直接用定义

5.

We have a bias coin with $\Pr\{\text{head}\} = p, \Pr\{\text{tail}\} = 1 - p$. And we want to use this coin to generate a result from perfect coin. Assume we toss this bais coin n times and get a sequence $\{X_1, ..., X_n\}$, then we use a mapping function $f(X_1, ..., X_n) = \{z_1, ...z_k\}$, the length of mapping output Z, K is also a random variable. And we can define the code rate $R = \frac{E[K]}{n}$.

1) John Von Neumann had given a code scheme. n=2, $f_{01} = 0$, $f_{10} = 1$, $f_{00} = \lambda$, $f_{11} = \lambda$. (λ is null, means outputing nothing.), what's the code rate of this 'Von Neumann Code'.

2) Can you generalize John Von Neumann code from n=2, to any arbitrary n.

3) prove that code rate has upperbound, $R \leqslant h_b(p)$ 这个题目作者的证明方法我觉得值得一看

Brief notes for my solution,

1.

Exploit the concavity of $\log(\bullet)$ and Jense Inequality. Upperbound is reached on $P_0 = p_{\max}$, $p_1 = p_2 = \ldots = p_{n-1} = \frac{1 - p_{\max}}{n-1}$.

2.

1) Geometric distribution with $p = \frac{1}{2}$.

2) 2, 201

3) Using huffman code. Code L_1 into 00...01 ($L_1 - 1$ 0's, plus a 1).

4) Average block code length of $[X_1, L_1, \ldots, L_{100}] = 1 + 2 \times 100 = 201$. $E\left(\sum_{i=1}^{100} L_i\right) = 100E(L_1) = 200$

5) I believe they are the same in performance.

3.

Very similar to the traditional form of conditional entropy.

We believe, if the summation and integration are commutative, $I_1(X;Y) = I_2(X;Y)$.

4.

1) $H(Y) = H(Z) + p\,H(X)$

2) $q_1 = q_2 = \ldots = q_n = \frac{1}{n}$, $p = \frac{n}{n+1}$ can maximize H(Y)

3) $C(p) = p\log(n)$. $p = 1$ maximize the C. Z=0 just like a error happen in the communication, Z=1 means the channel deliver the information correctly.

5.

1) $p(1-p)$

2) [unsure]

Here we propose a function, we group all 2^n by the number of 1's in $x_1, \ldots x_n$.

if $\binom{n}{k} = 2^{a_1} + 2^{a_2} + \ldots + 2^{a_m}$, we assign all $x_1, \ldots x_n$ in this group a ID $\in \left\{0, 1, \ldots, \binom{n}{k}\right\}$, and the range of ID can be divided into following subset $\{0, 1, \ldots, 2^{a_1} - 1\}, \{2^{a_1}, \ldots, 2^{a_1} + 2^{a_2} - 1\}, \ldots$, if the ID of input sequence is in the subset have size 2^{a_i}, the output is the binary presentation of rank of ID in the subset with has length of a_i.

3)

$H(Z,K) = H(Z) + H(K|Z) = H(K) + H(Z|K)$

Because Z decide K, H(K|Z)=0. Consider pd of K $\{p_1, \ldots p_k\}$

$H(Z|K) = \sum p_k H(Z|K=k) = \sum_k p_k \sum \frac{1}{2^k}\log(2^k) = \sum_k p_k\,k = E[K]$

Thus, $H(Z) = H(K) + H(Z|K) \geqslant H(K) = E[K]$

And X decide Z, thus $E[K] \leqslant H(Z) \leqslant H(X) = n\,h_b(p)$, then we get

$R = \frac{E[K]}{n} \leqslant h_b(p)$.