Lecture 11-12 Examples and Exercises

October 9th and October 11th, 2022

Example (2.10)

Let X_1 and X_2 be discrete random variables drawn according to probability mass functions $p_1(\cdot)$ and $p_2(\cdot)$ over the respective alphabets $\mathcal{X}_1 = \{1, 2, \ldots, m\}$ and $\mathcal{X}_2 = \{m+1, \ldots, n\}$. Let

$$X = \begin{cases} X_1, & \text{with probability } \alpha \\ X_2, & \text{with probability } 1-\alpha. \end{cases}$$

(a) *Find $H(X)$ in terms of $H(X_1)$ and $H(X_2)$ and α.*

(b) *Maximize over α to show that $2^{H(X)} \leq 2^{H(X_1)} + 2^{H(X_2)}$ and interpret using the notion that $2^{H(X)}$ is the effective alphabet size.*

Example (2.30)

Find the probability mass function $p(x)$ that maximizes the entropy $H(X)$ of a nonnegative integer-valued random variable X subject to the constraint

$$EX = \sum_{n=0}^{\infty} np(n) = A$$

for a fixed value $A > 0$. Evaluate this maximum $H(X)$.

Example (2.33)

Let $Pr(X = i) = p_i,\ i = 1, 2, \cdots, m$ *and let* $p_1 \geq p_2 \geq p_3 \geq \cdots p_m$. *The minimal probability of error predictor of* X *is* $\hat{X} = 1$, *with resulting probability of error* $P_e = 1 - p_1$. *Maximize* $H(\boldsymbol{p})$ *subject to the constraint* $1 - p_1 = P_e$ *to find a bound on* P_e *in terms of* H.

Example (3.5)

Let $X_1, X_2, \ldots$ *be an i.i.d. sequence of discrete variables wit entropy* $H(X)$*. Let*

$$C_n(t) = \{x^n \in \mathcal{X}^n : p(x^n) \geq 2^{-nt}\}$$

denote the subset of n*-sequences with probabilities* $\geq 2^{-nt}$*.*

(a) *Show* $|C_n(t)| \leq 2^{nt}$*.*

(b) *For what values of* t *does* $P(\{X^n \in C_n(t)\}) \to 1$*?*

Example (4.30)

$$P = [P_{ij}] = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

Let X_1 be uniformly distributed over the states $\{0, 1, 2\}$. Let $\{X_i\}_1^\infty$ be a Markov chain with transition matrix P, thus $P(X_{n+1} = j | X_n = i) = P_{ij}, i, j \in \{0, 1, 2\}$.

(a) *Is $\{X_n\}$ stationary?*

(b) *Find $\lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n)$.*

Now consider the derived process $Z_1, Z_2, \ldots, Z_n$, where

$$\begin{aligned} Z_1 &= X_1 \\ Z_i &= X_i = -X_{i-1} \pmod 3, \; i = 2, \ldots, n. \end{aligned}$$

Thus Z^n encodes the transitions, not the states.

(c) *Find $H(Z_1, Z_2, \ldots, Z_n)$.*

(d) *Find $H(Z_n)$ and $H(X_n)$, for $n \geq 2$.*

(e) *Find $H(Z - n | Z_{n-1})$ for $n \geq 2$.*

(f) *Are Z_{n-1} and Z_n independent for $n \geq 2$?*

Example (4.31)

Let $\{X_i\} \sim$ *Bernoulli*(p). *Consider the associated Markov chain* $\{Y_i\}_{i=1}^n$ *where* $Y_i =$ *(the number of 1's in the current run of 1's). For example, if* $X^n = 101110\ldots$, *we have* $Y^n = 101230\ldots$.

(a) *Find the entropy rate of* X^n.

(b) *Find the entropy rate of* Y^n.

Example (5.25)

Suppose that the message probabilities are given in decreasing order $p_1 > p_2 \geq \cdots \geq p_m$.

(a) *Prove that for any binary Huffman code, if the most probable message symbol has probability* $p_1 > 2/5$, *then that symbol must be assigned a codeword of length* 1.

(b) *Prove that for any binary Huffman code, if the most probable message symbol has probability* $p_1 < 1/3$, *then that symbol must be assigned a codeword of length* ≥ 2.

Example (5.44)

Find the word lengths of optimal binary encoding of $p = (\frac{1}{100}, \frac{1}{100}, \ldots, \frac{1}{100})$.

Example

Let $X_1, X_2, \dots, X_n$ are pairwise independent random variables, distributed identically as Bern(0.5). Then:

A. *Show that: $H(X_1, X_2, X_3) \leq 3$. When is equality achieved?*

B. *Show that: $H(X_1, X_2, X_3) \geq 2$. When is equality achieved?*

Example

Let $Z_1, Z_2, Z_3, \ldots$ be i.i.d. random variables that take values "0" and "1" with equal probability. Further, let

$$X_i = \sum_{j=1}^{i} Z_j, \text{ for } 1 \leq i \leq n.$$

Find $I(X_1; X_2, X_3, \ldots, x_n)$.

Example

Assume that a sequence of symbols from the random variable X as below using the code C_3. Imagine picking one bit at random from the binary encoded sequence $\boldsymbol{x} = c(x_1)c(x_2)c(x_3)\cdots$. What is the probability that this bit is a 1?

C_3:

a_i	$c(a_i)$	p_i	$h(p_i)$	l_i
a	0	$1/2$	1.0	1
b	10	$1/4$	2.0	2
c	110	$1/8$	3.0	3
d	111	$1/8$	3.0	3