

**Problem 1 (2.1 Coin flip.)** Score: \_\_\_\_\_. A fair coin is flipped until the first head occurs. Let  $X$  denote the number of flips required.

(a) Find the entropy  $H(X)$  in bits. The following expression may be useful:

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}, \quad \sum_{n=0}^{\infty} nr^n = \frac{r}{(1-r)^2}.$$

(b) A random variable  $X$  is drawn according to this distribution. Find an "efficient" sequence of yes-no questions of the form, "Is  $X$  contained in the set  $S$ ?" Compare  $H(X)$  to the expected number of questions required to determine  $X$ .

**Solution:** (a) The PMF of  $X$  is

$$P(X = n) = \left(\frac{1}{2}\right)^{n-1} \frac{1}{2} = \frac{1}{2^n}, \quad n = 1, 2, 3, \dots \quad (1)$$

The entropy of  $X$  in bits is

$$H(X) = -\sum_{n=1}^{\infty} P(X = n) \log_2 P(X = n) = -\sum_{n=1}^{\infty} \frac{1}{2^n} \log_2 \frac{1}{2^n} = \sum_{n=1}^{\infty} \frac{n}{2^n} = \frac{\frac{1}{2}}{(1 - \frac{1}{2})^2} = 2 \text{ (bits)}. \quad (2)$$

(b) We can use Huffman code to represent the possible  $X$ 's according to the PMF of  $X$ . That is,

$$\begin{aligned} X = 1 &\rightarrow 1, \\ X = 2 &\rightarrow 01, \\ X = 3 &\rightarrow 001, \\ &\dots \\ X = n &\rightarrow \overbrace{00 \dots 0}^{n-1 \text{ zeros in total}} 1, \\ &\dots \end{aligned}$$

Then, we can ask in such way: "Is the first symbol of the code 1?" If the answer is "Yes", we know that  $X = 1$ . If not, we continue to ask: "Is the second symbol of the code 1". If the answer is "Yes", then we know that  $X = 2$ . If not, we continue such asking method until the answer is "Yes" and thus we can determine the  $X$  represented by the code.

This asking method is equivalent to ask: "Is  $X = 1$ ?" "If not, is  $X = 2$ ?" "If not, is  $X = 3$ ?" .....

In such way, the expected number of questions required to determine  $X$  is

$$\frac{1}{2} \times 1 + \frac{1}{2} \times \frac{1}{2} \times 2 + \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times 3 \dots = \sum_{n=1}^{\infty} \frac{n}{2^n} = 2, \quad (3)$$

which equals  $H(X)$  in bits.

□

**Problem 2 (2.2 Entropy of functions.)** Score: \_\_\_\_\_. Let  $X$  be a random variable taking on a finite number of values. What is the (general) inequality relation of  $H(X)$  and  $H(Y)$  if

(a)  $Y = 2^X$ ?

(b)  $Y = \cos X$ ?

**Solution:** (a) Suppose the set of possible values that  $X$  can take on to be  $\{x_1, x_2, \dots, x_N\}$ . The entropy of  $X$  is

$$H(X) = - \sum_{n=1}^N P(X = x_n) \log_2 P(X = x_n). \quad (4)$$

Since  $Y = 2^X$  is a one-to-one mapping, the set of the possible values that  $Y$  can take on is  $\{y_1 = 2^{x_1}, y_2 = 2^{x_2}, \dots, y_n = 2^{x_n}\}$  and the PMF of  $Y$  is

$$P(Y = 2^{x_n}) = P(X = x_n). \quad (5)$$

The entropy of  $Y$  is equal to  $X$ 's:

$$H(Y) = - \sum_{n=1}^N P(Y = 2^{x_n}) \log_2 P(Y = 2^{x_n}) = - \sum_{n=1}^N P(X = x_n) \log_2 P(X = x_n) = H(X). \quad (6)$$

(b) Since  $Y = \cos X$  is not necessarily one-to-one, the PMF of  $Y$  is

$$P(Y = y) = \sum_{x \in \{x|y=\cos x\}} P(X = x) \quad (7)$$

The entropy of  $Y$  is equal to or less than  $X$ 's:

$$\begin{aligned} H(Y) &= - \sum_y P(Y = y) \log_2 P(Y = y) \\ &= - \sum_y \left[ \sum_{x \in \{x|y=\cos x\}} P(X = x) \right] \log_2 \left[ \sum_{x \in \{x|y=\cos x\}} P(X = x) \right] \\ &\leq - \sum_y \sum_{x \in \{x|y=\cos x\}} P(X = x) \log_2 P(X = x) \\ &= - \sum_x P(X = x) \log_2 P(X = x) = H(X). \end{aligned} \quad (8)$$

□

**Problem 3 (2.4 Entropy of functions of a random variable.)** Score: \_\_\_\_\_. Let  $X$  be a discrete random variable. Show that the entropy of a function of  $X$  is less than or equal to the entropy of  $X$  by justifying the following steps:

$$\begin{aligned} H(X, g(X)) &\stackrel{(a)}{=} H(X) + H(g(X)|X) \\ &\stackrel{(b)}{=} H(X); \\ H(X, g(X)) &\stackrel{(c)}{=} H(g(X)) + H(X|g(X)) \\ &\stackrel{(d)}{\geq} H(g(X)). \end{aligned}$$

Thus,  $H(g(X)) \leq H(X)$ .

**Proof:** (a)

$$\begin{aligned} H(X, g(X)) &= - \sum_{x,y} P(X = x, g(X) = y) \log_2 P(X = x, g(X) = y) \\ &= - \sum_{x,y} P(X = x) P(g(X) = y|X = x) \log_2 [P(X = x) P(g(X) = y|X = x)] \end{aligned}$$

$$\begin{aligned}
&= - \sum_{x,y} P(X=x)P(g(X)=y|X=x) \log_2 P(X=x) \\
&\quad - \sum_{x,y} P(X=x)P(g(X)=y|X=x) \log_2 P(g(X)=y|X=x) \\
&= - \sum_x P(X=x) \log_2 P(X=x) - \sum_x P(X=x) H(g(X)|X=x) \\
&= H(X) + H(g(X)|X).
\end{aligned}$$

(b) Once the value of  $X$  is given,  $g(X)$  is determined, so

$$H(g(X)|X=x) = 0, \quad (9)$$

$$\Rightarrow H(g(X)|X) = \sum_x P(X=x) H(g(X)|X=x) = 0, \quad (10)$$

and

$$H(X, g(X)) = H(X) + H(g(X)|X) = H(X). \quad (11)$$

(c)

$$\begin{aligned}
H(X, g(X)) &= - \sum_{x,y} P(X=x, g(X)=y) \log_2 P(X=x, g(X)=y) \\
&= - \sum_{x,y} P(g(X)=y) P(X=x|g(X)=y) \log_2 [P(g(X)=y) P(X=x|g(X)=y)] \\
&= - \sum_{x,y} P(g(X)=y) P(X=x|g(X)=y) \log_2 P(g(X)=y) \\
&\quad - \sum_{x,y} P(g(X)=y) P(X=x|g(X)=y) \log_2 P(X=x|g(X)=y) \\
&= - \sum_y P(g(X)=y) \log_2 P(g(X)=y) + \sum_y P(g(X)=y) H(X|g(X)=y) \\
&= H(g(X)) + H(X|g(X)).
\end{aligned}$$

(d) Since entropy of any discrete random variable is no less than 0,

$$H(X|g(X)) \geq 0, \quad (12)$$

we have

$$H(X, g(X)) = H(g(X)) + H(X|g(X)) \geq H(g(X)). \quad (13)$$

Using (a)-(d), we can obtain  $H(g(X)) \leq H(X)$ . □

**Problem 4 (2.5 Zero conditional entropy.) Score: \_\_\_\_\_.** Show that if  $H(Y|X) = 0$ , then  $Y$  is a function of  $X$  [i.e., for all  $x$  with  $p(x) > 0$ , there is only one possible value of  $y$  with  $p(x, y) > 0$ ].

**Proof:** The entropy of  $Y|X$  is

$$H(Y|X) = - \sum_{x,y} P(X=x, Y=y) \log_2 P(Y=y|X=x).$$

Since

$$-P(X=x, Y=y) \log_2 P(Y=y|X=x) \geq 0, \quad \forall x, y, \quad (14)$$

for given  $x_0$  and  $y_0$ ,

$$\begin{aligned} H(Y|X) &\geq -P(X = x_0, Y = y_0) \log_2 P(Y = y_0|X = x_0) \\ &= P(X = x_0)P(Y = y_0|X = x_0) \log_2 P(Y = y_0|X = x_0) \geq 0. \end{aligned} \quad (15)$$

If  $H(Y|X) = 0$ , then

$$0 \geq -P(X = x_0)P(Y = y_0|X = x_0) \log_2 P(Y = y_0|X = x_0) \geq 0, \quad (16)$$

so there must be

$$-P(Y = y_0|X = x_0) \log_2 P(Y = y_0|X = x_0) = 0, \quad (17)$$

for  $P(X = x_0) > 0$ . That is,

$$P(Y = y_0|X = x_0) = 0, \text{ or } P(Y = y_0|X = x_0) = 1, \quad (18)$$

$\forall y$  都满足, 且最多  $\exists y_0$  为  $p(\dots) = 1$   
 逻辑不太严密.

for  $P(X = x_0) > 0$ , which means that for all  $x_0$  with  $P(X = x_0) > 0$ , there is only one possible value of  $y_0$  with  $p(X = x_0, Y = y_0) = p(X = x_0)p(Y = y_0|X = x_0) = P(X = x_0) > 0$ .

Therefore, if  $H(Y|X) = 0$ , then  $Y$  is a function of  $X$ . □

**Problem 5 (2.11 Measure of correlation.)** Score: \_\_\_\_\_. Let  $X_1$  and  $X_2$  be identically distributed but not necessarily independent. Let

$$\rho = 1 - \frac{H(X_2|X_1)}{H(X_1)}.$$

(a) Show that  $\rho = \frac{I(X_1; X_2)}{H(X_1)}$ .

(b) Show that  $0 \leq \rho \leq 1$ .

(c) When is  $\rho = 0$ ?

(d) When is  $\rho = 1$ ?

**Solution:** (a)

$$\begin{aligned} \rho &= 1 - \frac{H(X_2|X_1)}{H(X_1)} \\ &= \frac{H(X_1) - H(X_2|X_1)}{H(X_1)} \\ &\quad (\text{since } X_1 \text{ and } X_2 \text{ are identically distributed, } H(X_1) = H(X_2)) \\ &= \frac{H(X_2) - H(X_2|X_1)}{H(X_1)} \\ &= \frac{I(X_1; X_2)}{H(X_1)}. \end{aligned} \quad (19)$$

(b) Since  $H(X_2|X_1) \geq 0$  and  $H(X_1) \geq 0$ , we have

$$\rho = 1 - \frac{H(X_2|X_1)}{H(X_1)} \leq 1. \quad (20)$$

Since  $H(X_2) \geq H(X_2|X_1)$ , we have

$$\rho = \frac{H(X_2) - H(X_2|X_1)}{H(X_1)} \geq 0. \quad (21)$$

(c) When  $X_1$  and  $X_2$  are independent,  $\rho = 0$ .

*Proof.*

$$\rho = \frac{I(X_1; X_2)}{H(X_1)} = 0 \iff I(X_1; X_2) = 0 \iff X_1 \text{ and } X_2 \text{ are independent.} \quad (22)$$

□

(d) When  $X_1$  and  $X_2$  have one-to-one mapping,  $\rho = 1$ .

*Proof.*

$$\rho = 1 - \frac{H(X_2|X_1)}{H(X_1)} = 1 \implies H(X_2|X_1) = 0 \iff X_2 \text{ is a function of } X_1. \quad (23)$$

Accordingly,

$$\rho = \frac{H(X_1) - H(X_2|X_1)}{H(X_2)} = \frac{H(X_2) - H(X_1|X_2)}{H(X_2)} = 0 \implies H(X_2|X_1) = 0 \implies X_1 \text{ is a function of } X_2. \quad (24)$$

Therefore,  $X_1$  and  $X_2$  have one-to-one mapping.

□

□

**Problem 6 (2.12 Example of entropy.)** Score: \_\_\_\_\_. Let  $p(x, y)$  be given by

X \ Y	0	1
	$\frac{1}{3}$	$\frac{1}{3}$
0	$\frac{1}{3}$	$\frac{1}{3}$
1	0	$\frac{1}{3}$

Find:

(a)  $H(X)$ ,  $H(Y)$ .

(b)  $H(X|Y)$ ,  $H(Y|X)$ .

(c)  $H(X, Y)$ .

(d)  $H(Y) - H(Y|X)$ .

(e)  $I(X; Y)$ .

(f) Draw a Venn diagram for the quantities in parts (a) through e.

**Solution:** (a) The PMF of  $X$  is

$$P(X = 0) = P(X = 0, Y = 0) + P(X = 0, Y = 1) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}, \quad (25)$$

$$P(X = 1) = P(X = 1, Y = 0) + P(X = 1, Y = 1) = 0 + \frac{1}{3} = \frac{1}{3}. \quad (26)$$

The PMF of  $Y$  is

$$P(Y = 0) = P(X = 0, Y = 0) + P(X = 1, Y = 0) = \frac{1}{3} + 0 = \frac{1}{3}, \quad (27)$$

$$P(Y = 1) = P(X = 0, Y = 1) + P(X = 1, Y = 1) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}. \quad (28)$$

The entropy of  $X$  is

$$H(X) = -P(X = 0) \log_2 P(X = 0) - P(X = 1) \log_2 P(X = 1) = -\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} = 0.918 \text{ (bits)}. \quad (29)$$

The entropy of  $Y$  is

$$H(Y) = -P(Y = 0) \log_2 P(Y = 0) - P(Y = 1) \log_2 P(Y = 1) = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} = 0.918 \text{ (bits)}. \quad (30)$$

(b) The conditional PMF of  $X|Y$  is

$$P(X = 0|Y = 0) = 1, \quad (31)$$

$$P(X = 1|Y = 0) = 0, \quad (32)$$

$$P(X = 0|Y = 1) = \frac{1}{2}, \quad (33)$$

$$P(X = 1|Y = 1) = \frac{1}{2}. \quad (34)$$

The conditional PMF of  $X|Y$  is

$$P(Y = 0|X = 0) = \frac{1}{2}, \quad (35)$$

$$P(Y = 1|X = 0) = \frac{1}{2}, \quad (36)$$

$$P(Y = 0|X = 1) = 0, \quad (37)$$

$$P(Y = 1|X = 1) = 1. \quad (38)$$

The conditional entropy of  $X|Y$  is

$$\begin{aligned} H(X|Y) &= -P(X = 0, Y = 0) \log_2 P(X = 0|Y = 0) - P(X = 1, Y = 0) \log_2 P(X = 1|Y = 0) \\ &\quad - P(X = 0, Y = 1) \log_2 P(X = 0|Y = 1) - P(X = 1, Y = 1) \log_2 P(X = 1|Y = 1) \\ &= -\frac{1}{3} \log_2 1 - 0 - \frac{1}{3} \log_2 \frac{1}{2} - \frac{1}{3} \log_2 \frac{1}{2} \\ &= \frac{2}{3} \text{ (bits)}. \end{aligned} \quad (39)$$

The conditional entropy of  $Y|X$  is

$$\begin{aligned} H(Y|X) &= -P(X = 0, Y = 0) \log_2 P(Y = 0|X = 0) - P(X = 0, Y = 1) \log_2 P(Y = 1|X = 0) \\ &\quad - P(X = 1, Y = 0) \log_2 P(Y = 0|X = 1) - P(X = 1, Y = 1) \log_2 P(Y = 1|X = 1) \\ &= -\frac{1}{3} \log_2 \frac{1}{2} - \frac{1}{3} \log_2 \frac{1}{2} - 0 - \frac{1}{3} \log_2 1 \\ &= \frac{2}{3} \text{ (bits)}. \end{aligned} \quad (40)$$

(c) The joint entropy of  $X$  and  $Y$  is

$$\begin{aligned} H(X, Y) &= -P(X = 0, Y = 0) \log_2 P(X = 0, Y = 0) - P(X = 1, Y = 0) \log_2 P(X = 1, Y = 0) \\ &\quad - P(X = 0, Y = 1) \log_2 P(X = 0, Y = 1) - P(X = 1, Y = 1) \log_2 P(X = 1, Y = 1) \\ &= -\frac{1}{3} \log_2 \frac{1}{3} - 0 - \frac{1}{3} \log_2 \frac{1}{3} - \frac{1}{3} \log_2 \frac{1}{3} \\ &= \log_2 3 = 1.585 \text{ (bits)}. \end{aligned} \quad (41)$$

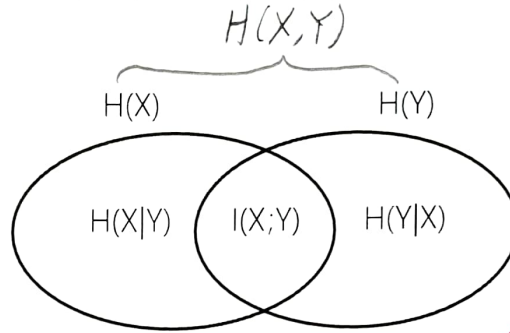
(d)

$$H(Y) - H(Y|X) = 0.251 \text{ (bits)}. \quad (42)$$

(e) The mutual information of  $X$  and  $Y$  is

$$I(X;Y) = H(Y) - H(Y|X) = 0.251 \text{ (bits)}. \quad (43)$$

(f) As shown in figure 1.

Figure 1: Venn diagram of  $H(X)$ ,  $H(Y)$ ,  $H(X|Y)$ ,  $H(Y|X)$ ,  $H(X,Y)$  and  $I(X;Y)$ 

□

**Problem 7 (8.1 Differential entropy.)** Score: \_\_\_\_\_. Evaluate the differential entropy  $h(X) = -\int f \ln f$  for the following:

(a) The exponential density,  $f(x) = \lambda e^{-\lambda x}$ ,  $x \geq 0$ .(b) The Laplace density,  $f(x) = \frac{1}{2} \lambda e^{-\lambda|x|}$ .(c) The sum of  $X_1$  and  $X_2$ , where  $X_1$  and  $X_2$  are independent random variables with means  $\mu_i$  and variances  $\sigma_i^2$ ,  $i = 1, 2$ .**Solution:** (a) The differential entropy of the exponential density is

$$\begin{aligned} h(f) &= - \int_0^{+\infty} \lambda e^{-\lambda x} \ln[\lambda e^{-\lambda x}] dx \\ &= - \int_0^{+\infty} \lambda e^{-\lambda x} \ln \lambda dx - \int_0^{+\infty} \lambda e^{-\lambda x} (-\lambda x) dx \\ &= e^{-\lambda x} \ln \lambda \Big|_0^{+\infty} - \lambda \int_0^{+\infty} x d(e^{-\lambda x}) \\ &= -\ln \lambda - \lambda x e^{-\lambda x} \Big|_0^{+\infty} + \lambda \int_0^{+\infty} e^{-\lambda x} dx \\ &= 1 - \ln \lambda \text{ (nats)} \\ &= \log_2 \frac{e}{\lambda} \text{ (bits)}. \end{aligned} \quad (44)$$

(b) The differential entropy of the Laplace density is

$$h(f) = - \int_{-\infty}^{+\infty} \frac{1}{2} \lambda e^{-\lambda|x|} \ln \left( \frac{1}{2} \lambda e^{-\lambda|x|} \right) dx = - \int_0^{+\infty} \lambda e^{-\lambda x} \ln \left( \frac{1}{2} \lambda e^{-\lambda x} \right) dx = \log_2 \frac{e}{\lambda} \text{ (bits)}. \quad (45)$$

(c)  $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ , thus

$$h(X_1 + X_2) = \frac{1}{2} \log_2 2\pi e(\sigma_1^2 + \sigma_2^2) \text{ (bits)}. \quad (46)$$

□