Name:

Problem	Points	Score
1a	10	
1b	10	
1c	10	
1d	10	
2a	10	
2b	10	
2c	10	
3a	10	
3b	10	
3c	10	
Total	100	

Notes:

1. The exam is closed books/closed notes - except for one page (double-sided) of notes.

2. Please show ALL work. Answers with no supporting explanations or work will be given no credit.

3. Please indicate clearly your answer to the problem. If I can’t read it (and I am the judge of legibility), it is wrong. If I can’t follow your solution (and I get lost easily), it is wrong. All things being equal, neat and legible work will get the higher grade:)

Problem No. 1: General Concepts

(a) Explain the relationship between mutual information and relative entropy. Do not simply write the equations or restate them in words — show some insight. (Hint: recall the discussion in Chap. 1.)

Answer: The relative entropy is a measure of the distance between two distributions

The mutual information I(X;Y) is a measure of the dependence between the two

random variable. $I(X;Y) = \sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)(y)} = D(p(x,y)||p(x)p(y))$

So, mutual information show the distance between p(x,y) and p(x)p(y).

If P(x,y)=p(x)p(y), which means X and Y and independent, D(p(x,y)||p(x)p(y)) = 0,

the distance between them is 0 and thus the mutual information of X and Y is

also 0.

(b) Explain the impact of information theory on the data compression and data transmission problems (Hint: recall the discussion in Chap. 1.)

Answer: The objective of data compression is to get rid of redundancy in data, using information theory, we can get the entropy H of a random variable, which is a lower boundary on the average length of the shortest description of the random variable.

In data transmission, we wish to find codewords that are mutually far apart in the sense that their noisy versions are distinguishable. Such method like huffman coding in information theory can be used to get codewords with high efficiency and low error rate.

(c) Prove the chain rule for relative entropy.

Chain rule for relative entropy: $D(p(x,y)||q(x,y)) = D(p(x)||q(x)) + D(p(y|x)||q(y|x))$

Proof:

$$D(p(x,y)||q(x,y)) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{q(x,y)}$$
$$= \sum_x \sum_y p(x,y) \log \frac{p(x)p(y|x)}{q(x)q(y|x)}$$
$$= \sum_x \sum_y p(x,y) \log \frac{p(x)}{q(x)} + \sum_x \sum_y p(x,y) \log \frac{p(y|x)}{q(y|x)}$$
$$= D(p(x)||q(x)) + D(p(y|x)||q(y|x))$$

(d) Prove the independence bound on entropy:

$$H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i | X_{i-1}, X_{i-2}, \ldots, X_1)$$

Here, we use the inductive proof method,

1. Base: $H(x_1, x_2) = H(x_1) + H(x_2|x_1)$

2. Assume that $H(x_1, x_2, \ldots x_{n-1}) = \sum_{i=1}^{n-1} H(x_i | x_{i-1}, x_{i-2}, \ldots x_1)$ holds,

then

$$H(x_1, x_2, \ldots x_n) = H(x_1, x_2, \ldots x_{n-1}) + H(x_n | (x_{n-1}, x_{n-2}, \ldots x_1))$$
$$= \sum_{i=1}^{n} H(x_i | (x_{i-1}, x_{i-2}, \ldots x_1))$$

So, the formula is true for all the n.

Problem No. 2: Calculations

(a) What is the relationship between $H(Y)$ and $H(X)$ when $Y = \cos X$ and X is a random variable that takes on a finite number of values.

Let $y = g(x)$, then $p(y) = \sum_{x:y = g(x)} p(x)$

Consider any set of x's that map onto a single y. For this set

$$\sum_{x:y = g(x)} p(x)\log p(x) \le \sum_{x:y = g(x)} p(x)\log p(y) = p(y)\log p(y)$$

since log is a monotone increasing function and $p(x) \le \sum_{x:y = g(x)} P(x) = P(y)$

$$\begin{aligned} H(X) &= -\sum_x p(x)\log p(x) \\ &= -\sum_y \sum_{x:y = g(x)} p(x)\log p(x) \\ &\ge -\sum_y p(y)\log p(y) \\ &= H(Y) \end{aligned}$$

for Y=cosX, since it's not necessary one to one, hence $H(X) \ge H(Y)$, and the equality hold only when Y and X are one to one.

(b) 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 $p(x, y) > 0$.

Assume that there exists an x, say x_0 and two different value of y, say y_1 and y_2 such that $p(x_0, y_1) > 0$ and $p(x_0, y_2) > 0$. Then $p(x_0) \ge p(x_0, y_1) + p(x_0, y_2) > 0$, also $p(y_1|x_0)$ and $p(y_2|x_0)$ are not equal to 1 or 0. Thus:

$$\begin{aligned} H(Y|X) &= -\sum_x p(x) \sum_y p(y|x)\log p(y|x) \\ &\ge p(x_0)(-p(y_1|x_0)\log p(y_1|x_0) - p(y_2|x_0)\log p(y_2|x_0)) \\ &> 0 \end{aligned}$$

so, the assumption result in a contradiction to $H(Y|X) = 0$

Therefore, for all x with $p(x) > 0$, there is only one possible value of y with $p(x, y) > 0$.

(c) X **and** Y **are random variables, and let** $Z = X + Y$**. Show that** $H(Z|X) = H(Y|X)$**.**

$Z = X + Y$, so $p(Z = z|X = x) = p(Y = z - x|X = x)$

$$
\begin{aligned}
H(Z|X) &= \sum p(x) H(Z|X{=}\mathrm{x}) \\
&= -\sum_x p(x) \sum_z p(Z{=}\mathrm{z}|\mathrm{X}{=}\mathrm{x}) \log p(Z{=}\mathrm{z}|\mathrm{X}{=}\mathrm{x}) \\
&= \sum_x p(x) \sum_y p(Y{=}\mathrm{z\text{-}x}|\mathrm{X}{=}\mathrm{x}) \log p(Y{=}\mathrm{z\text{-}x}|\mathrm{X}{=}\mathrm{x}) \\
&= \sum_x p(x) H(Y|X{=}\mathrm{x}) \\
&= H(Y|X)
\end{aligned}
$$

Problem No. 3: A Deck of Cards

(a) A deck of cards contains 52 cards, evenly distributed between hearts, diamonds, spades, and clubs. Which has higher entropy: drawing two red cards with or without replacement.

For with replacement case, probability of getting two red cards is: $P(x) = \frac{1}{4}$,

H= p(x)log(p(x)) + (1-p(x))log(1-(p(x)) = 0.8113

for without replacement case, probability of getting two red cards is: $p(x) = \frac{1}{2} \times \frac{25}{51}$

so, entropy H = 0.8034

We can see that entropy is higher for with replacement case, which means more uncertainty.

(b) Suppose the first card drawn is a king. What are the chances that the next card drawn will result in a sum under 21. Prove that you are better off knowing that the first card drawn was a king.

If we don't know the first card is king, there are two way in getting sum of 21, 1) first card is card of 10 points and second is Ace; 2) first card is Ace and second is card with 10 points. so, the chances of getting a sum under 21 is:

$1 - \frac{16}{52} \times \frac{4}{51} - \frac{4}{52} \times \frac{15}{51} = 0.953$, the entropy for this distribution is 0.274

if we know the first card is king, to get a sum of 21 points, the second care must be Ace, so the chances of getting a sum under 21 is:

$1 - \frac{4}{51} = 0.922$, the entropy for this distribution is 0.395, this is higher than the entropy for not knowing the first card is king, a higher entropy means the uncertainty that u can't reach 21 is high, which means a better chance to reach 21 points.

so, we can find that knowing that first card is a king, we have a better chance to reach a score of 21 points, which means we are better off.

(c) You are playing blackjack with a dealer who is using one or two decks of cards. The player to your right has a king exposed on the table, and tells you he has reached a score of 20. Can you use information learned in this class to determine what your safest betting strategy might be? Or is this simply a straightforward probability calculation?

If I believe what the player said, from 3 b), I learn that there are very few chance for me to win, so I will give up this game, on this point, one deck or two will make little difference.

However, since it's a game, I have reason to suspect the truth of that. I can still guess what his another card is. Using one deck, the distribution for another card is (4/51 (A), 4/51 (2), 4/51(3),...4/51 (9), 15/51(10)), the entropy for such a distribution is 3.112. The distribution while using two decks is (8/103(A),8/103(2),... 8/103(9), 31/103(10)), entropy is 3.098.

We can see that there are still much uncertainty on the results but one or two decks make only little difference. So I will need to look at my card, suppose I have 18 points (10+8), using one deck, the probability that he can reach a score higher than 18 or equal to 18 is 44.9%, so I still stand a better chance to win.

(d) Suppose that two cards have been drawn by yourself and another player. The two cards you are holding sum to 17. Prove that if I tell you the other player's first card was a king, you are better able to make an intelligent bet (meaning you can better predict whether your sum will be greater than the other player's sum). Try to do this using entropy or other such concepts.

Answer: Suppose the card of mine is 10+7, if the other player's first card is a king, if his overal score is below 17, then other card must be from (2, 3, 4, 5, 6), the probability is 40.8%. The entropy for the distribution of sum lower than 17 points or not are 0.976.

If I don't know the first card of the other player, the possible cases that his score can be less then 17 points are the combanition of the following:

Table 1:

10	9	8	7	6	5	4	3	2
2, 3, 4, 5 ,6	2, 3, 4, 5, 6, 7	2, 3, 4,5, 6, 7, 8	2, 3,4, 5, 6, 7	6, 5, 4, 3, 2	5, 4, 3, 2, A	4, 3, 2 A	3, 2, A	2, A

The overal probability is 42.7%, the entropy for the distribution of sum lower than 17 points or not are 0.985.

We can see that knowing the first card is a king, the entropy will be lower than without knowing the first card is king, this confirms to the idea that conditional intropy H(X|Y) should be lower than H(X), so we can better predict the reselt.