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.)

Mutual information $I(X;Y)$ is a measure of the dependence between two random variables. It is the reduction in the uncertainty of X due to the knowledge of Y. It's the relative entropy of joint entropy of $p(x, y)$ and $p(x) \cdot p(y)$:

$$I(X;Y) = D(p(x, y)||p(x) \cdot p(y))$$

Relative entropy is a measure of the distance between two distributions $p(x)$ and $q(x)$.

Mutual information is a special case of relative entropy. $D(p(x)||q(x)) \geq 0$ with equality iff $p(x) = q(x)$. So, $I(X;Y) = 0$ iff $p(x, y) = p(x) \cdot p(y)$, which means X, Y are independent.

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

Data compression: The minimum codeword length to express a set of data is the entropy $H(X)$ of this set of data.

Data transmission: The channel capacity can be computed from the noise characteristics of the channel. When the speed of transmission below the value of channel capacity, the signal will not be distorted by the channel.

(c) Prove the chain rule for relative entropy.

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

Proof: For (X,Y)~p(x,y), the relative entropy is:

$$D(p(x, y)||q(x, y)) = \sum_{x \in \chi}\sum_{y \in \gamma} p(x, y) \cdot \log\frac{p(x, y)}{q(x, y)}$$

$$= \sum_{x \in \chi}\sum_{y \in \gamma} p(x) \cdot p(y|x) \cdot \log\frac{p(x)p(y|x)}{q(x)q(y|x)}$$

$$= \sum_{x \in \chi}\sum_{y \in \gamma} p(x) \cdot \log\frac{p(x)}{q(x)} + \sum_{x \in \chi}\sum_{y \in \gamma} p(y|x) \cdot \log\frac{p(y|x)}{q(y|x)}$$

$$= D(p(x)||q(x)) + D(p(y|x)||q(y|x))$$

(d) Prove the independence bound on entropy:

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

Proof: 1. For $n = 2$, $H(X_1,X_2) = H(x_1) + H(x_2|x_1)$

2. Suppose for $n = n-1$, the equation is held, i.e.:

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

While $n = n$,

$$H(X_1,X_2,\dots, X_{n-1}, X_n) = H(X_1, X_2, \dots, X_{n-1}) + H(X_n|X_1, X_2, \dots, X_{n-1})$$

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

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

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.

$Y = H(X)$ means Y is a function of X, and more than one values of X may correspond to the same value of Y. So, once we know the value of X, we get the value of Y. But when we know the value of Y, we just know that the value of X is from a specific subset of X. Which means: $H(Y|X) = 0, H(X|Y) \geq 0$.

And, while we decompose $H(X, Y)$ in two ways, we get:

$$\begin{aligned} H(X, Y) &= H(X) + H(Y|X) \\ &= H(Y) + H(X|Y) \end{aligned}$$

where $H(Y|X) = 0$ and $H(X|Y) \geq 0$,

so, $H(X) \geq H(Y)$

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

Since
$$H(Y|X) = -\sum_{x \in \chi} p(x) \sum_{y \in \gamma} p(y|x) \cdot \log p(y|x)$$

$$= \sum_{x \in \chi} \sum_{y \in \gamma} [-p(x) \cdot p(y|x) \cdot \log p(y|x)] = 0,$$

for each component of this summation: $-p(x) \cdot p(y|x) \cdot \log p(y|x) = -p(x, y) \cdot \log p(y|x)$,

since $p(x, y) \geq 0, 1 \geq p(y|x) \geq 0$, $-p(x, y) \cdot \log p(y|x) \geq 0$.

To get the sum of a set of nonnegative numbers come to zero, each of them must be zero. i.e.,

$$p(x, y) \cdot \log p(y|x) = 0 \qquad \forall x,y.$$

so, $p(y|x) = 1$, which means, for each value of x, there is only one value of y corresponds to this x.

So, if $H(Y|X) = 0$, then 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 $H(Z|X) = H(Y|X)$.

From $Z = X + Y$, we can see: $p(Z = z|X = x) = p(Y = z - x|X = x)$.

$$H(Z|X) = \sum_{x \in \chi} p(x) \cdot H(Z|X = x)$$

$$= -\sum_{x \in \chi} p(x) \sum_{z \in Z} p(Z=\text{z}|\text{X=x}) \cdot \log p(Z=\text{z}|\text{X=x})$$

$$= -\sum_{x \in \chi} p(x) \sum_{y \in \gamma} p(Y=\text{z-x}|\text{X=x}) \cdot \log p(Y=\text{z-x}|\text{X=x})$$

$$= -\sum_{x \in \chi} p(x) \cdot H(Y|X=\text{x})$$

$$= H(Y|X)$$

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.

In a deck of cards, there are 26 red cards.

1. The probability of drawing two red cards with replacement is $p(x) = \frac{26}{52} \cdot \frac{26}{52} = \frac{1}{4}$, and $p(\bar{x}) = 1 - p(x) = 1 - \frac{1}{4} = \frac{3}{4}$.

So, the entropy of drawing two red cards with replacement is:

$$H(X) = -p(x) \cdot \log p(x) - p(\bar{x}) \cdot \log p(\bar{x})$$
$$= \frac{1}{4} \cdot \log 4 - \frac{3}{4} \cdot \log \frac{4}{3} \approx 0.815 \qquad bits$$

2. The probability of drawing two red cards without replacement is $p(y) = \frac{26}{52} \cdot \frac{25}{51} \approx 0.245$, and $p(\bar{y}) = 1 - p(y) \approx 0.755$.

So, the entropy of drawing two red cards without replacement is:

$$H(Y) = -p(y) \cdot \log p(y) - p(\bar{y}) \cdot \log p(\bar{y})$$
$$= -0.245 \cdot \log 0.245 - 0.755 \cdot \log 0.755 \approx 0.803 \qquad bits$$

Since $0.815 > 0.803$, drawing two red cards with replacement has higher entropy.

(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.

Assume the score of A is 11.

To get a sum no less than 21 points after the first two draws, you need to have one 10-score card from {10, J, Q, K} and one A, or two As. Let's use A to indicate the point of two draws is under 21, and $\bar{A}$ to indicate we get no less than 21 points from two draws.

1. If we don't know the points of the first card, the probability of A and $\bar{A}$ are:

$$p(\bar{A}) = \frac{4}{52} \cdot \frac{16}{51} + \frac{16}{52} \cdot \frac{4}{51} + \frac{4}{52} \cdot \frac{3}{51} \approx 0.053 \text{ and } \quad p(A) = 1 - p(\bar{A}) \approx 0. \quad \cdot$$

So, $H(X) = -p(A) \cdot \log p(A) - p(\bar{A}) \cdot \log p(\bar{A}) \approx 0.278 \qquad bits$.

2. If we know the first card drawn is a king, then the second card must be A for A. The probability of A and $\bar{A}$ are: $p(\bar{A}) = \frac{4}{51} \approx 0.078$ and $p(A) = 1 - p(\bar{A}) \approx 0.922 \cdot$

So, $H(Y) = -p(A) \cdot \log p(A) - p(\bar{A}) \cdot \log p(\bar{A}) \approx 0.395 \qquad bits$.

Since $H(X) < H(Y)$, we can see from the analysis above that knowing the value of the first drawn is a king, we can better predict that if the next card drawn will result in a sum under 21 or not.

(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?

Suppose I have not seen my cards, and each of us just have two cards in hand. Since the dealer has a king, and has reached 20, he must hold two 10-score cards.

1. We are using one deck of cards. There are 14 10-score cards and four As left.

Since the dealer has reached 20, the only chance not to lose is to get a score no less than 20, which means the two cards I need to get must come from {10, J, Q, K, A}, but not two As. The probability to reach it is: $p(x) = \frac{14}{50} \cdot \frac{13}{49} + \frac{14}{50} \cdot \frac{4}{49} + \frac{4}{50} \cdot \frac{14}{49} = 0.12$, and the chance to lose is: $1 - p(x) = 0.88$. Entropy of this case is:

$$H(X) = -p(x) \cdot \log p(x) - (1 - p(x)) \cdot \log(1 - p(x)) = 0.529 \quad bits.$$

2. We are using two decks of cards. There are 30 10-score cards and 8 As left.

As analyzed above, the probability for me to get a score no less than 20 is: $p(y) = \frac{30}{102} \cdot \frac{29}{101} + \frac{30}{102} \cdot \frac{8}{101} + \frac{8}{102} \cdot \frac{30}{101} = 0.131$, and the chance to lose is: $1 - p(x) = 0.869$. Entropy of this case is:

$$H(Y) = -p(y) \cdot \log p(y) - (1 - p(y)) \cdot \log(1 - p(y)) = 0.560 \quad bits.$$

From the analysis above we can see that, after the dealer reaching 20, no matter we are using one deck or two decks, the chance to beat him is so small that I would rather give up this time.

(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.

Assume that we are using one deck of cards, and the two cards that I'm holding is one 10-score card and one 7. If somebody reach 21 points, game is over.

1. Case X.

Suppose I don't know the cards that other player is holding.

The chance for him to get a score higher than me is that he gets:

1). one 8 and another card from {10, J, Q, K, A},

2). two cards from {9, 10, J, Q, K},

3). one from {7, 8, 9} and one A.

The probability for him to get a score higher than 17, but not 21 is:

$$p(x_2) = \frac{4}{50} \cdot \frac{19}{49} + \frac{19}{50} \cdot \frac{4}{49} + \frac{19}{50} \cdot \frac{18}{49} + \frac{11}{50} \cdot \frac{4}{49} + \frac{4}{50} \cdot \frac{11}{49} = 0.2376,$$

The probability for me to win now is: $p(x_1) = 0.7624$.

Entropy for this condition is:

$$H(X_1) = -p(x_1) \cdot \log p(x_1) - p(x_2) \cdot \log p(x_2) = 0.791 \qquad bits.$$

Under this situation, if I assume that the other player scored higher than I did, and ask for the third card, the third card must come from {2, 3, 4}, otherwise, I'll lose. The probability of not sum over 21, i.e., I still hold the chance to win, is: $p(x_{21}) = \frac{12}{48} \cdot p(x_2) = 0.0594$, and the probability for sum over 21 is: $p(x_{22}) = 0.1782$.

Entropy for this condition is:

$$H(X_2) = -p(x_1) \cdot \log p(x_1) - p(x_{21}) \cdot \log p(x_{21}) - p(x_{22}) \cdot \log p(x_{22}) = 0.984 \qquad bits.$$

$H(X_1) < H(X_2)$ means the uncertainty of asking for the third card is greater than just betting now. And the probability for me to win after two draws is greater than 50%. So I can bet now, but will not ask for the third card.

2. Case Y.

Suppose I know that the first card of the other player is a king. Since the game is still going, his second card is not A.

The chance for him to get a score higher than mine is that his second card is from

{8, 9, 10, J, Q, K}. The probability of this is: $p(y_2) = \frac{22}{49} = 0.449$, and the probability for me to win now is: $p(y_1) = 0.551$.

Entropy for this condition is:

$$H(Y_1) = -p(y_1) \cdot \log p(y_1) - p(y_2) \cdot \log p(y_2) = 0.992 \qquad bits.$$

Under this situation, if I assume that he scored higher than I did, and ask for the third card, the probability of not sum over 21 is:

$p(y_{21}) = \frac{12}{48} \cdot p(y_2) = 0.11225$, and the chance to sum over 21 is: $p(y_{22}) = 0.33675$.

Entropy for this condition is:

$$H(Y_2) = -p(y_1) \cdot \log p(y_1) - p(y_{21}) \cdot \log p(y_{21}) - p(y_{22}) \cdot \log p(y_{22}) = 1.402 \qquad bits.$$

For this case, we get $H(Y_1) < H(Y_2)$, and the probability for me to win after two draws is greater than 50%, which means I'd better bet now, but not ask for the third card too.

And we can see that the difference between $H(X_1)$ and $H(X_2)$ is smaller than the difference between $H(Y_1)$ and $H(Y_2)$, so for case Y, we are more confident to decide just bet now.