Name: Suresh Balakrishnama

Problem	Points	Score
1a	10	
1b	10	
1c	10	
1d	10	
2a	10	
2b	10	
2c	10	
3a	10	
3b	10	
3c	10	
3d	extra 10 points	
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 is a special case of relative entropy, which is a measure of the distance between two probability mass functions. Both, mutual information and relative entropy are good measures only when variables under consideration are dependent variables.

Equation illustrating the relationship: If X and Y are two random variables,

$$I(X,Y) = D\langle p(x,y)||p(x)p(y)\rangle$$

Thus, mutual information is the relative entropy between the joint distribution and the product distribution for the random variables X and Y.

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

Shannon proved that if entropy of a source, a measure of uncertainty of data, is less than the capacity of channel error free communication can be achieved. The data compression should always be minimum whereas data transmission also termed as channel capacity should be maximum. From the references it is possible to construct descriptions providing data compression with average length within one bit of the entropy. They both are two extremes in communication theory. Entropy is the lower bound on the average length of data compression. For a better transmission, codewords (sequences of input symbols to a channel) are used to distinguish the transmitted data from the noise in the channel.

(c) Prove the chain rule for relative entropy.

Statement for chain rule for relative entropy:

$$D((p(x,y)||q(x,y))) = D(p(x)||q(x)) + D(p\langle y|x\rangle||q\langle y|x\rangle)$$

Proof:

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

$$= \sum_x \sum_y p(x,y)\log\frac{p(x)p(y/x)}{q(x)q(y/x)}$$ (based on Bayes's rule)

$$= \sum_x \sum_y p(x,y)\log\frac{p(x)}{q(x)} + \sum_x \sum_y p(x,y)\left(\log\frac{p(y/x)}{q(y/x)}\right)$$ (based on logarithmic addition property)

$$= D(p(x)||q(x)) + D(p\langle y|x\rangle||q\langle y|x\rangle)$$

Hence proved.

(d) Prove the independence bound on entropy:

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

Proof: Let $X_1,X_2,\ldots,X_n$ be drawn according to $p(x_1,x_2,x_3\ldots\ldots x_n)$

then by chain rule for entropies,

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

$$H(X_1,X_2,\ldots,X_n) \le \sum_{i=1}^{n} H(X_i)$$ (since we know conditioning reduces entropy)

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.

Solution:
Here X is a random variable and maps the function: $y = g(x)$ then we have the pmf as:

$$p(y) = \sum_{x,y} p(x)$$

considering a set of x's that map onto a single y, for this set based on Information inequality, we have:

$$\sum_x p(x)\log p(x) \le \left(\sum_{x,y} p(x)\log p(y) = p(y)(\log p(y))\right)$$

since log is a monotony increasing function and $p(x) \le \sum_{x,y} p(x) = p(y)$. Applying this property to the entire range of x and y, we obtain:

$$H(x) = -\sum_x p(x)(\log p(x))$$

$$= \sum_y \sum_x p(x)(\log p(x))$$

$$H(x) \ge -\sum_y p(y)\log p(y) = H(y)$$

with equality iff g has a one-to-one function mapping with probability one.

Considering the function given $p(x) \le \sum_{x,y} p(x) = p(y)$ here cosX is a operator which does not provide one-to-one transformation. The values of y are differ widely based on the values of x and in the range of $-\pi < x < \pi$ and values are different for $x > 2\pi$ and hence we can conclude:

$H(x) \ge H(y)$ for this given function.

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

Solution:

This is a problem involving zero conditional entropy.

Let us consider a value of x, say x_0 and two different values for y say y_1 and y_2 such that $p(x_0, y_1) > 0$ and $p(x_0, y_2) > 0$. Then we can write

$p(x_0) \geq p(x_0, y_1) + p(x_0, y_2) > 0$ and $p\langle y_1|x_0\rangle$ and $p\langle y_1|x_0\rangle$ are definitely not equal to 0 or 1.

Thus, using the definition of conditional entropy

$$H\langle Y|X\rangle = -\sum_x p(x) \sum_y p\langle y|x\rangle \log p\langle y|x\rangle$$

$$H\langle Y|X\rangle \geq p(x_0)(-p\langle y_1|x_0\rangle \log p\langle y_1|x_0\rangle) - p\langle y_2|x_0\rangle \log p\langle y_2|x_0\rangle > 0$$

since $t \log t \geq 0$ for $0 \leq t \leq 1$ and is strictly positive for not equal to 0 or 1.
However $H\langle Y|X\rangle \geq 0$ truly contradicts the statement in given in the problem. Therefore the conditional entropy is zero $H\langle Y|X\rangle = 0$ if and only if Y is a function of X.

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

Solution:

Given that, $Z = X + Y$,
then $p\langle Z = z|X = x\rangle = p\langle Y = z - x|X = x\rangle$

$H(Z|X) = \sum p(x)H\langle Z|X = x\rangle$ based on the definition of conditional entropy

$= -\sum_x p(x)\sum_z p\langle Z = z|X = x\rangle \log p\langle Z = z|X = x\rangle$

$= \sum_x p(x)\sum_y p\langle Y = z - x|X = x\rangle \log p\langle Y = z - x|X = x\rangle$

$= \sum p(x)H(Y|X = x)$

$= H(Y|X)$

Hence proved.

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.

Solution:

$$Prob(withreplacement) = \frac{26}{52} \times \frac{26}{52} = 0.25$$

$$Prob(withoutreplacement) = \frac{26}{52} \times \frac{25}{51} = 0.245$$

Entropy values are:
with replacement:

$$H(x) = -\sum_x p(x) \log p(x) = 0.25 \log 0.25 = 0.50 bits$$

without replacement:

$$H(x) = -\sum_x p(x) \log p(x) = 0.245 \log 0.245 = 0.497 bits$$

Hence, from the calculations Entropy of drawing two cards with replacement contains 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.

Solution:

Let us consider two separate cases in this problem (here the value of A is 11):

Case I: This is the case where the first card drawn is any card from the deck of cards.
Case II: This is the case where the first card drawn is a King.

Computation for Case I:

Instead of finding the probability of getting a sum of two cards under 21 we will find out the probability of obtaining over 21 and then use the probability rule of $\sum_x p(x) = 1$.

So, let us compute the probability of obtaining a sum over 21.
Now, to get a sum of over 21 the events possible are: $\{(A,10), (10,A)\}$

Probability of getting this event $= \left(\frac{4}{52} \times \frac{19}{51} + \frac{16}{52} \times \frac{4}{51}\right) = 0.052$

Hence, probability of getting an event with sum of two cards under 21
$= (1 - 0.052) = 0.948$

Let's calculate the entropy with this probability:
$$H(x) = -\sum_x p(x) \log p(x) = 0.948 \log 0.948 = 0.073 bits$$

Computation for Case II:
Here also let us first compute the probability of obtaining a sum over 21.
Now, to get a sum of over 21 the events possible are: $\{(K,10), (10,K)\}$

Probability of getting this event $= \left(\frac{4}{52} \times \frac{4}{51}\right) = 0.006$

Hence, probability of getting an event with sum of two cards under 21 known that the first card is a King $= (1 - 0.006) = 0.994$
Let's calculate the entropy with this probability:
$$H(x) = -\sum_x p(x) \log p(x) = 0.994 \log 0.994 = 0.0086 bits$$

Hence, more entropy indicates more information and less entropy indicates less information so it would be wiser to invest on the second case where it is less entropy and betting is safer. Hence, it is proved that we are better off when we know that first card is a king.

(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?

Solution:

We have to compute, here, for two different cases, and we can provide conclusion in each case separately (value of A is chosen as 10).

Case I: for one deck of cards
We will find the what is the probability of getting a score of 20, that indicates that the second card of the other player is any one card of J/Q/K/A/10.

Probability of getting second card from this combination is:

P(X=J/Q/K/10/A) = $\frac{4}{51} + \frac{4}{51} + \frac{3}{51} + \frac{4}{51} + \frac{4}{51} = \frac{19}{51}$

hence, let's find the probability of not getting sum of 20:

Prob of not getting sum of 20 = $1 - \frac{19}{51} = 0.6274$

Entropy: $H(X) = -0.6274 \log 0.6274 = 0.4219 bits$

Case II: for two deck of cards
We will find in this case what is the probability of getting a score of 20, that indicates that the second card of the other player is any one card of J/Q/K/A/10 and then compare the entropy obtained here to the case I to draw conclusion.

Probability of getting second card from this combination is:

P(X=J/Q/K/10/A) = $\frac{8}{103} + \frac{8}{103} + \frac{7}{103} + \frac{8}{103} + \frac{8}{103} = \frac{39}{103}$

hence, let's find the probability of not getting sum of 20:

Prob of not getting sum of 20 = $1 - \frac{39}{103} = 0.6213$

Entropy: $H(X) = -0.6213 \log 0.6213 = 0.4266 bits$

Conclusion:
The entropy with one deck of cards is less than the entropy obtained using two deck of cards. Hence, it is safer to bet when a single deck of cards is used, there's a higher chance that the other player's card sum is less than 20 and I can win in this situation.

(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.

Solution:

Here again, we will consider two separate cases (value of A is chosen as 10):

Case I:
To find the entropy when the first card of other player is revealed as K and to
to find out if his sum is greater than 17.

For easy computation, let us compute the probability of getting sum less than 17.
If the first card is a King, then the second card should be one from {2,3,4,5,6}.

P(X=2/3/4/5/6) = $\frac{4}{50}+\frac{4}{50}+\frac{4}{50}+\frac{4}{50}+\frac{4}{50}=\frac{20}{50}$

Hence, probability of getting sum greater than 17 is:

P(X>17) = $1-\frac{20}{50}=0.6$

Entropy: $H(X) = -0.6\log 0.6 = 0.4421 bits$

Case II:
There are two events possible for us to get a sum of 17 -- $\{(7, 10), and (8, 9)\}$
To find the entropy when the first card of other player is not revealed and to
to find out if his sum is greater than 17. So to find the other person's entropy we will have to calculate the probability of getting all the events if our card combination are any one of the above mentioned event.

For easy computation, let us compute the probability of getting sum greater than 17 and then apply the probability rule to find the probability of sum less than 17.
The combination can be anything from this sample space:
for $(7, 10)$ the sample space is:

$$S = \{(7,10), (8,9), (9,8), (10,7), (9,10)(9,9), (10,8), (8,10), (10,9), (10,10)\}$$

P(X=x,y)= $\frac{3}{50}\cdot\frac{19}{49}+\frac{14}{50}\cdot\frac{4}{49}+\frac{4}{50}\cdot\frac{19}{49}+\frac{4}{50}\cdot\frac{4}{49}+\frac{4}{50}\cdot\frac{4}{49}+$

= $\frac{4}{50}\cdot\frac{19}{49}+\frac{19}{50}\cdot\frac{3}{49}+\frac{19}{50}\cdot\frac{4}{49}+\frac{19}{50}\cdot\frac{4}{49}+\frac{19}{50}\cdot\frac{4}{49}+\frac{19}{50}\cdot\frac{18}{49}$

where $(x, y)\varepsilon S$

The combination can be anything from this sample space:
for $(8, 9)$ the sample space is:

$S = \{(7,10), (8,9), (9,8), (10,7), (9,10)(9,9), (10,8), (8,10), (10,9), (10,10)\}$

P(X=x,y)= $\frac{4}{50}\cdot\frac{20}{49}+\frac{3}{50}\cdot\frac{3}{49}+\frac{3}{50}\cdot\frac{20}{49}+\frac{3}{50}\cdot\frac{20}{49}+\frac{3}{50}\cdot\frac{3}{49}+$

= $\frac{3}{50}\cdot\frac{2}{49}+\frac{3}{50}\cdot\frac{20}{49}+\frac{20}{50}\cdot\frac{4}{49}+\frac{20}{50}\cdot\frac{3}{49}+\frac{20}{50}\cdot\frac{3}{49}+\frac{20}{50}\cdot\frac{19}{49}$

where $(x, y)\varepsilon S$

Hence, total probability of getting sum greater than 17 with both events is:

P(X<17) = $1-\frac{808+804}{2450} = 0.342$

Entropy: $H(X) = -0.342\log 0.342 = 0.5293 bits$

Conclusion:
Comparing the entropies, entropy in Case I is lesser compared to entropy in Case II hence the betting should be done when the first card of other player is revealed as king. So if neither of other player's card is revealed we should not bet as happened in Case II.