ECE 534 Information Theory - MIDTERM

09/27/2017, LH 104.

- You will be given the full 1.25 hours. **Use it wisely!** Many of the problems have short answers; try to find shortcuts. Do questions that you think you can answer correctly first.
- You may bring and use one 8.5x11" double-sided crib sheet.
- No other notes or books are permitted.
- No calculators are permitted.
- Talking, passing notes, copying (and all other forms of cheating) is forbidden.
- Make sure you explain your answers in a way that illustrates your understanding of the problem. Ideas are important, not just the calculation.
- Partial marks will be given.
- Write all answers directly on this exam.

Your name: Solutions

Your UIN: ________________

Your signature: ________________

The exam has 3 questions, for a total of 100 points.

Question:	1	2	3	Total
Points:	30 ~~25~~	~~18~~ 20	~~54~~ 50	100
Score:				

(30

1. *Entropy and the typical set*

Consider a random variable X which takes on the values $0, 2, 4, 8$ with probabilities $p(0) = 1/2, p(2) = 1/4, p(4) = 1/8, p(8) = 1/8$. For parts (c) and (d) and (e) and (f), we consider a sequence of 16 i.i.d. throws, or instances of this random variable.

(a) (6 points) Find the entropy in base 4 of this random variable.

(b) (6 points) Find the entropy in base 2 of this random variable.

(c) (3 points) How many sequences of length ~~8~~ 16 are there in total? 16

(d) (6 points) Give one example of a typical sequence of length $n = \cancel{8}\ 16$, for $\epsilon = 0.1$.

(e) (6 points) Find upper and lower bounds (numbers) on the probability of any typical sequence in $A_{0.1}^{(16)}$.

(f) (6 points) Approximately how many sequences are there in $A_{0.1}^{(16)}$?

(a) $H_4(x) = \frac{1}{2}\log_4(2) + \frac{1}{4}\log_4(4) + \frac{1}{8}\log_4(8) + \frac{1}{8}\log_4(8)$

$= \frac{1}{2}\times\frac{1}{2} + \frac{1}{4}\times 1 + \frac{2}{8}\times\frac{3}{2} = \frac{1}{2} + \frac{3}{8} = \frac{7}{8}$ base 4 bits

(b) $H_2(x) = \frac{1}{2}\log_2(2) + \frac{1}{4}\log_2(4) + \frac{1}{8}\log_2(8) + \frac{1}{8}\log_2(8)$

$= \frac{1}{2} + \frac{1}{4}\times 2 + \frac{1}{8}\times 3 + \frac{1}{8}\times 3 = 1.75$ bits

Alternatively, $H_2(x) = H_4(x)\cdot\log_2(4) = \frac{7}{8}\times 2 = \frac{7}{4} = 1.75$ bits

(c) $4^{16} = (2^2)^{16} = 2^{32}$

(d) ~~The~~ One way to make sure we get a typical sequence is to pick a strongly typical sequence, i.e. with the correct proportions.

e.g. 0000000022224488

(e) For $x \in A_{0.1}^{(16)}$, $2^{-16(1.75+0.1)} \leq p(x) \leq 2^{-16(1.75-0.1)}$

(f) $|A_{0.1}^{(16)}| \approx 2^{16(1.75)}$

Points earned: ____________ out of a possible 46 points

Name:______________________________

Points earned: ______________ out of a possible 0 points

ECE534 Fall 2017 MIDTERM Name:____________________

(20)

2. *Using Fano's inequality.* Consider the joint distribution $p(x, y)$ given below.

X \ Y	a	b	c
1	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$
2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
3	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$

Suppose we use an estimator that estimates X from Y as follows (let $\hat{X}$ denote the estimate):

- If $Y = a$ estimate $\hat{X} = 1$.
- If $Y = b$ estimate $\hat{X} = 2$.
- If $Y = c$ estimate $\hat{X} = 3$.

(a) (6 points) Find the probability of error (an exact number).

(b) (~~6~~ 7 points) State Fano's inequalities (regular and strengthened forms).

(c) (~~6~~ 7 points) Evaluate Fano's inequality for the problem and estimator above. Is this a good estimator?

(a) $P_e = \frac{1}{12} + \frac{1}{12} + \frac{1}{12} + \frac{1}{12} + \frac{1}{12} + \frac{1}{12} = \frac{6}{12} = \frac{1}{2}$

$= P\big((X=2 \cap Y=a) \cup (X=3 \cap Y=a) \cup (X=1 \cap Y=b) \cup (X=3 \cap Y=b) \cup (X=1 \cap Y=c) \cup (X=2 \cap Y=c)\big)$

(b) Regular: $P_e \geq \dfrac{H(X|Y) - 1}{\log|\mathcal{X}|}$ for any estimator $\hat{X}$ such that $X \to Y \to \hat{X}$ with $P_e = P(X \neq \hat{X})$ we have $H(P_e) + P_e \log|\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y)$

Strengthened since $\hat{X}: \mathcal{Y} \to \mathcal{X}$ in this case, $H(P_e) + P_e \log(|\mathcal{X}| - 1) \geq H(X|Y)$

or $P_e \geq \dfrac{H(X|Y) - 1}{\log(|\mathcal{X}| - 1)}$

(c) $\log|\mathcal{X}| = \log 3$, $P_e = 1/2$, $H(P_e) = 1$, $H(X|Y) = H(1/4, 1/4, 1/2)$

$= \frac{1}{4}\log 4 + \frac{1}{4}\log 4 + \frac{1}{2}\log 2 = 1.5$ bits

Regular: $P_e \geq \dfrac{1.5 - 1}{\log 3} = \dfrac{0.5}{\log 3}$

Strengthen: $P_e \geq \dfrac{1.5 - 1}{\log 2} = 0.5$ → Yes, good estimator! Achieves Fano!

Points earned: ____________ out of a possible 10 points

 Name:________________________

Points earned: ____________ out of a possible 6 points

150

3. *Short answers.* Be short!

(a) (~~5~~ 6 points) Rigorously (mathematically, not with words or intuitively) show that for a discrete random variable X and *any* function of that random variable $g(X)$, $H(g(X)) \leq H(X)$.

$$H(X, g(X)) = H(X) + \underbrace{H(g(X)|X)}_{0 \text{ since } X \text{ determines } g(X)} \quad \text{chain rule}$$

$$= H(g(X)) + H(X|g(X)) \quad \text{chain rule other way}$$

Thus, since $H(g(X)|X) = 0$, and $H(X|g(X)) \geq 0$,

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

(b) (~~5~~ 6 points) What are the main ingredients needed (sketch the proof roughly) to show that $I(X;Y) \geq 0$?

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

So $I(X;Y) \geq 0$ follows from $D(p(x,y) \| p(x) \cdot p(y)) \geq 0$.

That Divergence between 2 distributions is positive follows almost directly from Jensen's inequality

$$E[f(x)] \geq f E[x] \quad \text{for convex } f.$$

and the fact that $\log(\cdot)$ is concave.

Points earned: ______________ out of a possible 10 points

(c) (6 points) Find the entropies below, when $p(x, y)$ is given by

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

- ~~$H(X)$~~
- ~~$H(X|Y)$~~
- ~~$H(X, Y)$~~
- ~~$H(Y) - H(Y|X)$~~
- ~~$I(X;Y)$~~
- ~~$H(5X)$~~

$\boxed{I(X;Y)}$ only.

- $H(X) = H(1/3) = \frac{1}{3}\log 3 + \frac{2}{3}\log\frac{3}{2}$
- $H(X|Y) = \underbrace{H(X|Y=0)}_{0} \cdot p(y=0) + \underbrace{H(X|Y=1)}_{1}\underbrace{p(y=1)}_{2/3} = \frac{2}{3}$
- $H(X,Y) = 3 \times \frac{1}{3}\log 3 = \log 3$
- $H(Y) - H(Y|X) = H(X) - H(X|Y)$ by symmetry
  $= \cancel{\log 3} \quad \frac{1}{3}\log 3 + \frac{2}{3}\log\frac{3}{2} - \frac{2}{3}$
- $I(X;Y)$ = same as above
- $H(5X) = H(X) = \frac{1}{3}\log 3 + \frac{2}{3}\log\frac{3}{2}$.

(d) (6 points) If $Y_1 \to Y_2 \to Y_3 \to \cdots \to Y_m$ forms a Markov chain, what is $I(Y_1; Y_2, Y_3, \cdots, Y_m)$? (simplify as much as possible)

chain rule of MI

$$I(Y_1; Y_2, Y_3, \ldots, Y_m) = I(Y_1;Y_2) + \underbrace{I(Y_1;Y_3|Y_2)}_{0} + \cdots + \underbrace{I(Y_1; Y_m|Y_2,\ldots,Y_{m-1})}_{0}$$

by markov chain.

$$= I(Y_1;Y_2)$$

(e) (6 points) State and prove the chain rule for entropy – include all steps and reasons for the equalities, and do not use the definition of mutual information.

WANT: $H(X,Y) = H(X) + H(Y|X)$

$$H(X,Y) = -\sum_x \sum_y p(x,y) \log p(x,y)$$

always $p(x,y) = p(x)\cdot p(y|x)$

$$= -\sum_x \sum_y p(x,y) \log (p(x)\cdot p(y|x))$$

$$= -\sum_x \underbrace{\sum_y p(x,y)}_{=p(x)} \log p(x) - \sum_x\sum_y p(x,y)\log p(y|x)$$

marginal

$$= -\sum_x p(x)\log p(x) - \sum_x\sum_y p(x,y)\log p(y|x)$$

$$= H(X) + H(Y|X)$$

definitions.

(f) (8 points) Let $\cdots, X_{-1}, X_0, X_1, \cdots$ be a stationary (not necessarily Markov) stochastic process. Which of the following statements are true and which are false? Show it/

1. $H(X_n|X_0) = H(X_{-n}|X_0)$
2. $H(X_n|X_1, X_2, \cdots, X_{n-1}, X_{n+1})$ is nonincreasing in n

1. TRUE:

chain rule

$$H(X_n, X_0) = H(X_0) + H(X_n|X_0)$$

by stationary $\parallel$

$$H(X_0, X_{-n}) = H(X_0) + H(X_{-n}|X_0)$$

so $H(X_n|X_0) = H(X_{-n}|X_0)$

2. TRUE:

stationarity

$$H(X_n|X_1, X_2, \ldots, X_{n-1}, X_{n+1}) = H(X_{n+1}|X_2, X_3, \ldots, X_n, X_{n+2})$$

conditioning reduces entropy

$$\geq H(X_{n+1}|X_1, X_2, X_3, \ldots, X_n, X_{n+2})$$

So, non-increasing in n.

(g) (6 points) What is the difference between a uniquely decodable code and an instantaneous code?

Both can be decoded 1:1 (1:1 mapping btw. codewords and symbols). An instantaneous code is a prefix code and is uniquely decodable. No codewords are prefixes of other codewords and codewords can be decoded in real time, as see the letters. Uniquely decodable may have to wait until end of sequence of codewords to decode.

(h) (6 points) Is the mutual information $I(X;Y)$ of discrete random variables (X,Y) with joint probability mass function $p(x,y)$ concave or convex in $p(x|y)$ for fixed $p(y)$? Formally (mathematically) state what it means (no need to prove this, just give the mathematical definition).

It is convex in $p(x|y)$ for given $p(y)$.

This means that for $0 \le \lambda \le 1$; for $p_1(x|y), p_2(x|y)$ and defining $p_1(x,y) \triangleq p(y)p_1(x|y)$, $p_2(x,y) \triangleq p(y)p_2(x|y)$ that

$$I(X;Y)\Big|_{\lambda p_1(x,y)+(1-\lambda)p_2(x,y)} \le \lambda I(X;Y)\Big|_{p_1(x,y)} + (1-\lambda) I(X;Y)\Big|_{p_2(x,y)}$$