Georgia Tech — School of Mathematics

Math 4280

Information Theory

Final Exam

April 30 2014

First Name : ____

Name : ____

DO NOT WRITE IN THE TABLE BELOW

1.a		3.a		6.a	
1.b		3.b		6.b	
1.c		3.c		7.a	
1.d		3.d		7.b	
2.a		3.e		8	
2.b		4		9.a	
		5		9.b	
				9.c	
				9.d	

WARNING :

Read carefully, read the comments in italic, take your time, do not panic and double check what you write.

Write CLEARLY your answer where it is asked to.

The final exam will last 2 hours and 50 minutes.

1. Let X, Y be two random variables on a finite probability space, which will be denoted by $\Omega_X \times \Omega_Y$.

   (a) Give the definition of the entropies $H(X), H(Y), H(X,Y), H(Y|X)$.

   (b) Write a chain rule formula expressing $H(X,Y)$.

   (c) Prove that $H(Y|X) = 0$ if and only if Y is a function of X.
   (Hint : (i) show that if $Y = g(X)$ then $H(Y|X) = 0$; (ii) conversely if $H(Y|X) = 0$ and if $x \in \Omega_X$ is such that $p(x) > 0$ then there is a unique $y \in \Omega_Y$ such that $p(x,y) > 0$)

   (d) If $Y = g(X)$ prove that $H(Y) \leq H(X)$.
   (Hint : use the chain rule in two different ways)

(Use this page to finish your proof if necessary)

2. Let $X_1 \to X_2 \to \cdots \to X_{n-1} \to X_n$ be a Markov chain made of random variables on finite probability spaces $\Omega_1 \times \cdots \times \Omega_n$

   (a) Give the expression of the joint probability
   (Hint : make sure to express the Markov property)

   (b) Compute the mutual information $I(X_1; X_2, \cdots, X_n)$ and express it in its simplest form.

3. Let $X_1, X_2, \cdots, X_n$ be *i.i.d.*'s drawn from a finite probability space (Ω, p) and with common entropy $H(X)$. For $t \in \mathbb{R}$ let then $C(t)$ be the set of n-sequences of elements in Ω with probability at least 2^{-nt}, namely

$$C_n(t) = \left\{x^n = (x_1, x_2, \cdots, x_n) \in \Omega^n \, ; \, p(x_1, x_2, \cdots, x_n) \geq 2^{-nt}\right\}$$

   (a) Give the expression of the joint probability distribution $p(x_1, x_2, \cdots, x_n)$.

   (b) Give the precise definition of the set A_ϵ^n of ϵ-typical sequences with respect to $p(x)$.

   (c) Give the argument leading to the *asymptotic equipartition property* (AEP), namely showing that $\lim_{n\to\infty} \text{Prob}\{A_\epsilon^n\} = 1$.

   (d) Show that the number of sequence in $C_n(t)$ is less that or equal to 2^{nt}.

   (e) Use the AEP to compute the value t_c of t such that if $t < t_c$ then $\text{Prob}\{C_n(t)\} \stackrel{n\uparrow\infty}{\rightarrow} 0$ while if $t > t_c$ then $\text{Prob}\{C_n(t)\} \stackrel{n\uparrow\infty}{\rightarrow} 1$.

(Use this page to finish your proof if necessary)

(Use this page to finish your proof if necessary)

4. Let

$$X = \begin{cases} 1, & \text{with probability } \ 1/2 \\ 2, & \text{with probability } \ 1/4 \\ 3, & \text{with probability } \ 1/4 \end{cases}$$

Let $X_1, X_2, \cdots, X_n$ be drawn *i.i.d* according to this distribution. Find the limiting behavior of the product

$$Z_n = (X_1 X_2 \cdots X_n)^{1/n}$$

5. Let Ω denote a finite set of cards. Let $\mathcal{P}$ denotes the set of permutations of these cards. An element of $\mathcal{P}$ is nothing but a shuffle (*Note that $\mathcal{P}$ is a group for the composition)*. Let X be a random card variable and let T be a random shuffle variable. Give an argument proving that $H(TX) \geq H(X)$ whenever X and T are stochastically independent.
(Hint : (i) first condition $H(TX)$ given T, (ii) evaluate the conditional entropy of X given T, by remarking that $T^{-1}TX = X$, (iii) use the independence of X and T)

6. **Reminder :** *a source code $\mathcal{C}$, for a random variable X defined on the finite probability space (Ω, p), is a map $\mathcal{C}$ defined on Ω with values into the set of finite words $\mathcal{D}^*$ of a D-ary alphabet. If $\mathcal{C}(x)$ denotes the codeword associated with x then $\ell(x)$ denotes the length of $\mathcal{C}(x)$. The expected length of the code is $L = \sum_{x\in\Omega} p(x)\ell(x)$. Then $\mathcal{C}$ is called* nonsingular *if the map $\mathcal{C}$ is one-to-one. The extension $\mathcal{C}^*$ of $\mathcal{C}$ is the map defined on the set Ω^* of finite strings of elements of Ω into $\mathcal{D}^*$ defined by $\mathcal{C}^*(x_1, x_2, \cdots, x_n) = \mathcal{C}(x_1)\mathcal{C}(x_2)\cdots\mathcal{C}(x_n)$. Then $\mathcal{C}$ is called* uniquely decodable *if $\mathcal{C}^*$ is nonsingular as well. It is called* prefix free *or* instantaneous *if no codeword is the prefix of another one.* □

   (a) Re-prove the inequality $L(\mathcal{C}) \geq H_D(X)$ where H_D denotes the entropy expressed in the $\log_D$ basis.
   (Hint : express the difference as a Kullback-Leibler distance between p and the Kraft distribution)

   (b) Let $L_k(\mathcal{C}) = \sum_{x\in\Omega} p(x)\ell(x)^k$ and let L_d be the minimum of the $L_k(\mathcal{C})$ over all uniquely decodable codes, while L_p denote the minimum of the $L_k(\mathcal{C})$ over all prefix free codes. Then what is the inequality relating L_p to L_d? Justify your answer by a complete argument.

(Use this page to finish your proof if necessary)

7. Let $\Omega = \{0, 1, 2, 3\}$ be endowed with the addition modulo 4. Let X be a random variable over Ω. Let Z be a random variable over Ω taking on only the values $\{1, 2, 3\}$ with equal probabilities. A discrete memoryless channel is obtained from using $Y = X + Z \pmod 4$ and assuming that X, Z are stochastically independent.
   (Hint : the Lagrange multiplier method might be useful in this exercise)

   (a) Compute the channel capacity.

   (b) Gives the maximizing probability p^* for X.

(Use this page to finish your proof if necessary)

8. Let $f(\vec{x})$ be a probability density on $\mathbb{R}^n$ and let $h(\vec{X}) = -\int_{\mathbb{R}^n} f(\vec{x}) \log f(\vec{x})\, d\vec{x}$ denote the corresponding differential entropy. If A is an $n \times n$ invertible matrix show that $h(A\vec{X}) = \log|\det(A)| + h(\vec{X})$.

9. Let X be a binary random variable distributed uniformly over $\Omega = \{0, 1\}$. Let $\widehat{X}$ be the random variable over Ω obtained from an encoding-decoding scheme with $p(\widehat{x}|x)$ fixed. Let d be the distorsion function given by $d(0, 1) = a$, $d(1, 0) = b$.

   (a) Give the joint distribution of $(X, \widehat{X})$.

   (b) Compute explicitly $I(X; \widehat{X})$ in terms of the previous data.

   (c) What are the values of $d(0, 0)$ and $d(1, 1)$?

   (d) Give the explicit expression of the average distorsion D in terms of the previous data.
   (Hint : give the definition of the average distorsion)
   (Remark : in this example, the rate distorsion function cannot be expressed in closed form)

(Use this page to finish your proof if necessary)

(Use this page to finish your proof if necessary)