MIDTERM

Tuesday, 3rd November, 2009, 13:15-17:15

This exam has 5 problems and 80 points in total.

Instructions

- You are allowed to use 1 sheet of paper for reference. No mobile phones or calculators are allowed in the exam.
- You can attempt the problems in any order as long as it is clear which problem is being attempted and which solution to the problem you want us to grade.
- If you are stuck in any part of a problem do not dwell on it, try to move on and attempt it later.
- Please solve every problem on **separate paper sheets**.
- It is your responsibility to **number the pages** of your solutions and write on the first sheet the **total number of pages** submitted.

Some Preliminaries

- A sequence of random variables $\{X_n\}$ **converges** toward X **in probability** if

$$\lim_{n\to\infty} \Pr[|X_n - X| \geq \varepsilon] = 0,$$

  for any $\varepsilon > 0$. For example the Weak Law of Large Numbers implies that if $X_1, X_2, \ldots$ is a sequence of i.i.d. random variables, and $S_n = \frac{1}{n}\sum_{i=1}^{n} X_n$, then

$$\lim_{n\to\infty} \Pr[|S_n - \mathbb{E}[X]| \geq \varepsilon] = 0.$$

  In other words, S_n converges to $\mathbb{E}[X]$ in probability.

- The following approximations might be useful.

$$0\log_2 0 = 0 \qquad \log_2 3 = 1.58 \qquad \log_2 5 = 2.32 \qquad \log_2 6 = 2.58$$

Good Luck!

Problem 1 (12 pts)

Let the three discrete random variables X, Y, Z be related by $Z = X - Y$, where $X, Y \in \{0, \ldots, m-1\}$.

(a) Compare $H(X|Y)$ and $H(Z)$. *[4pts]*

(b) When is $H(X|Y)$ equal to $H(Z)$? *[2pts]*

(c) Let $U \leftrightarrow V \leftrightarrow (W, T)$ form a Markov chain. Prove that *[6pts]*

$$I(U; W) + I(U; T) \leq I(U; V) + I(W; T).$$

Hint: For example, add $I(U; T|W)$ to both sides of the inequality and simplify using chain rule. Also use data processing inequality on Markov chain $U \leftrightarrow V \leftrightarrow (W, T)$.

Problem 2 (13 pts)

Let $X_1, X_2, \cdots$ be independent identically distributed random variables drawn according to the probability distribution $p(x)$, $x \in \mathcal{X}$, *i.e.*, $p(x_1, \cdots, x_n) = \prod_{i=1}^n p(x_i)$.

(a) What does $[p(x_1, \cdots, x_n)]^{\frac{1}{n}}$ converge in probability to? *[4pts]*

Let $f(x)$ be a function from $\mathcal{X}$ to the interval $(0, 1]$.

(b) What does $[\prod_{i=1}^n f(x_i)]^{\frac{1}{n}}$ converge in probability to? *[3pts]*

(c) How does $\mathbb{E}\left(\prod_{i=1}^n f(x_i)\right)^{\frac{1}{n}}$ compare to $\mathbb{E}f(X_1)$? Next, what implication does this have on the relationship between the result in (b) and $\mathbb{E}f(X_1)$? *[6pts]*
Hint: Use Jensen's inequality on the function $g(u) = u^{\frac{1}{n}}$, for $u \in (0, 1]$. That is, determine whether the function $g(u)$ is convex or concave in the interval $u \in (0, 1]$.

Problem 3 (15 pts)

Two fair dice are thrown together. Each dice has an outcome in the set of numbers $\{1, \ldots, 6\}$, and hence there are 36 possible outcomes of the two dice. Each dice is fair, and has a uniform probability of yielding any of the outputs $\{1, \ldots, 6\}$, *i.e.*, each outcome for a single dice occurs with probability $\frac{1}{6}$ and the dice take values independent of one another. Let X denote the sum of the two numbers that show up. A random variable Y which takes its values in $\{A, B, C\}$ can be constructed from X. A, B, and C can be any objects.

(a) Find $H(X|Y)$ if Y is constructed from X as follows: *[6pts]*

X	2	3	4	5	6	7	8	9	10	11	12
Y	A	B	C	C	C	C	C	C	C	B	A

This says that

$$Y = \begin{cases} A & \text{if } X \in \{2, 12\} \\ B & \text{if } X \in \{3, 11\} \\ C & \text{if } X \in \{4, \cdots, 10\} \end{cases}$$

(b) Construct Y from X such that knowing Y gives the maximum information about X, *i.e.,* construct Y such that that $I(X;Y) = H(X) - H(X|Y)$ is maximized, note that this also means $H(X|Y)$ is minimized. *[6pts]*

(c) Suppose now that the dice are faulty and only $\{1, 1\}$ or $\{2, 6\}$ can occur. That is, out of the 36 outcomes of the dice, only the two possibilities $\{1, 1\}$ or $\{2, 6\}$ can occur. Now, repeat (b) so that Y gives the maximum information about X, *i.e.,* $I(X;Y)$ is maximized. *[3pts]*

Problem 4 (20 pts)

A loaded dice with outcome X in the set of numbers $\{1, \ldots, 6\}$ has a non-uniform probability, $p_1 = \frac{1}{12}$, $p_2 = \frac{1}{9}$, $p_3 = \frac{1}{18}$, $p_4 = \frac{1}{6}$, $p_5 = \frac{1}{12}$, $p_6 = \frac{1}{2}$ where $p_i = \Pr\{X = i\}$.

(a) Find the entropy $H(X)$ in bits. *[4pts]*

(b) You are allowed to ask yes-no (binary) questions of the form "Is X contained in the set S?" What is the sequence of questions to ask to guess X with the minimum number of questions on average? *[5pts]*

The same dice is tossed until the first 6 occurs. Let Y denote the number of tosses required. For example if the outcome of the tossing is $2, 6$, then $Y = 2$; or if the outcome of the tossing is $1, 4, 2, 4, 6$, then $Y = 5$.

(c) Find $\Pr\{Y = k\}$. *[2pts]*

(d) Find the entropy $H(Y)$ in bits. *[3pts]*

(e) You are again allowed to ask yes-no (binary) questions of the form "Is Y contained in the set S?" What is the sequence of questions to ask to guess Y with the minimum number of questions on average? *[4pts]*

(f) Compare $H(Y)$ to the expected number of questions you need to ask in part (e) to determine Y. *[2pts]*

Hint: The following expressions might be useful:

$$\sum_{k=0}^{\infty} r^k = \frac{1}{1-r} \qquad \sum_{k=0}^{\infty} k r^k = \frac{r}{(1-r)^2}.$$

Problem 5 (20pts)

A source produces a sequence of bits through a finite state machine (FSM) as follows: The source has two states S_1 and S_2 as described in Fig 1, where the machine is in state A_i at time i. The machine starts from state $A_1 = S_1$, at time $i = 1$. At each state, the source flips a fair coin ($\Pr(H) = \frac{1}{2} = \Pr(T)$) and decides what to output and whether to change its state or not. The FSM determining the output as well as the state-transition is depicted in Figure 1. So at the end, after n coin tosses, a sequence $X_1, \cdots, X_{2n}$ of 0's and 1's is produced that satisfies certain constraints.

(a) Compute $\Pr(x_{2i}, x_{2i-1} | x_{2i-2}, \cdots, x_1)$ for $i \geq 2$. *[6pts]*

(b) Model the stochastic process $X_1, \cdots, X_{2n}$ with a Markov process. *[5pts]*

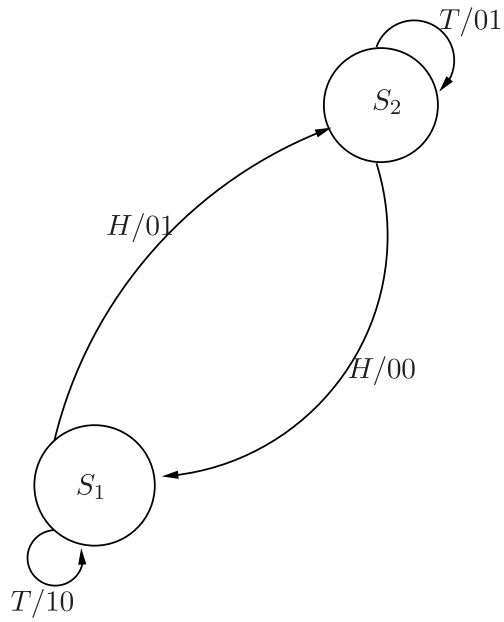

Figure 1: The finite state machine of the source explained in problem 5. Note that the labels (for example, $H/01$) on the arrows show the outcome of the coin toss and the corresponding output of the FSM. For example, if current state $A_i = S_1$, and the coin toss yields H, then the FSM outputs $X_{2i-1} = 0, X_{2i} = 1$ and makes a state transition to S_2, *i.e.*, $A_{i+1} = S_2$.

(c) Relate entropy rate of the source to entropy rate of the Markov process you suggest in part (b). *[3pts]*

(d) Calculate entropy rate of the source. *[6pts]*