ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 6 Information Theory and Coding

Solutions to homework 3 October 26, 2007

PROBLEM 1.

(a) Consider X and Y to be independent random variables taking values 0 and 1 with equal probability and let $Z = X \oplus Y$ that is the modulo 2 sum of X and Y.

(b) Let $X = Y = Z$ and each take values 0 and 1 with equal probability.

PROBLEM 2.

(a) We have

(i) $H(X|Y) = H(X)$ since X and Y are independent.

(ii) $H(X|K) = H(X)$ since X and K are independent.

(iii) $H(Y|X, K) = 0$ since X and K determine Y.

(iv) $H(X|Y, K) = 0$ since Y and K determine X by the decryptability condition.

(v) $I(X; Y|K) = H(X|K) - H(X|Y, K) = H(X)$ by (iv) and (ii).

(vi) $H(Y|K) = I(X; Y|K) + H(Y|X, K) = H(X)$ by (v) and (iii).

(b) Suppose k a key common to both $\mathcal{K}(x_1)$ and $\mathcal{K}(x_2)$. Then, the pair y_0, k can be decrypted as either x_1 or x_2, contradicting the decryptability condition.

(c) Since $I(X; Y) = 0$ we know that X and Y are independent and thus, $\Pr(Y = y) = \Pr(Y = y|X = x)$ for all x and y. In particular

$$0 < \Pr(Y = y_0) = \Pr(Y = y_0|X = x).$$

Thus for each x, $\mathcal{K}(x)$ is not empty, for otherwise $\Pr(Y = y_0|X = x)$ would have been zero. If any $\mathcal{K}(x)$ had more than one element, then the total number of keys would exceed the number of source letters; thus each $\mathcal{K}(x)$ must have exactly one element.

(d) Given that $X = x$, the only way $Y = y_0$ is when $K = k(x)$. Since X and K are independent this happens with probability $\Pr(K = k(x))$.

(e) We have $\Pr(Y = y_0) = \Pr(Y = y_0|X = x) = \Pr(K = k(x))$. Since the left hand side does not depend on x, the same must be true for the right hand side. Since $k(x)$ exhausts all the keys as x ranges over the source letters, we see that $\Pr(K = k)$ does not depend on k and hence that K is uniformly distributed.

PROBLEM 3. Let X^i denote $X_1, \ldots, X_i$.

(a) By the chain rule for entropy,

$$\frac{H(X_1, X_2, \ldots, X_n)}{n} = \frac{\sum_{i=1}^{n} H(X_i|X^{i-1})}{n} \tag{1}$$
$$= \frac{H(X_n|X^{n-1}) + \sum_{i=1}^{n-1} H(X_i|X^{i-1})}{n} \tag{2}$$
$$= \frac{H(X_n|X^{n-1}) + H(X_1, X_2, \ldots, X_{n-1})}{n}. \tag{3}$$

where we use the notation $X^{i-1} = \{X_1, X_2, \ldots, X_{i-1}$. Since conditioning reduces the entropy we have

$$H(X_n|X^{n-1}) \leq H(X_n|X_{n-1}, X_{n-2}, \ldots, X_{n-i+1})$$

From stationarity it follows that for all $1 \leq i \leq n$,

$$\begin{aligned} H(X_n|X_{n-1}, X_{n-2}, \ldots, X_{n-i+1}) &= H(X_n, X_{n-1}, X_{n-2}, \ldots, X_{n-i+1}) - H(X_{n-1}, X_{n-2}, \ldots, X_{n-i+1}) \\ &= H(X_i, X_{i-1}, X_{i-2}, \ldots, X_1) - H(X_{i-1}, X_{i-2}, \ldots, X_1) \\ &= H(X_i|X_{i-1}, X_{i-2}, \ldots, X_1) \end{aligned}$$

Thus

$$H(X_n|X^{n-1}) \leq H(X_i|X^{i-1}),$$

which further implies, by summing both sides over $i = 1, \ldots, n-1$ and dividing by $n-1$, that,

$$H(X_n|X^{n-1}) \leq \frac{\sum_{i=1}^{n-1} H(X_i|X^{i-1})}{n-1} \tag{4}$$
$$= \frac{H(X_1, X_2, \ldots, X_{n-1})}{n-1}. \tag{5}$$

Combining (3) and (5) yields,

$$\frac{H(X_1, X_2, \ldots, X_n)}{n} \leq \frac{1}{n}\left[\frac{H(X_1, X_2, \ldots, X_{n-1})}{n-1} + H(X_1, X_2, \ldots, X_{n-1})\right] \tag{6}$$
$$= \frac{H(X_1, X_2, \ldots, X_{n-1})}{n-1}. \tag{7}$$

(b) By stationarity we have for all $1 \leq i \leq n$,

$$H(X_n|X^{n-1}) \leq H(X_i|X^{i-1}),$$

which implies that,

$$H(X_n|X^{n-1}) = \frac{\sum_{i=1}^{n} H(X_n|X^{n-1})}{n} \tag{8}$$
$$\leq \frac{\sum_{i=1}^{n} H(X_i|X^{i-1})}{n} \tag{9}$$
$$= \frac{H(X_1, X_2, \ldots, X_n)}{n}. \tag{10}$$

Problem 4. By the chain rule for entropy,

$$\begin{aligned} H(X_0|X_{-1},\ldots,X_{-n}) &= H(X_0,X_{-1},\ldots,X_{-n}) - H(X_{-1},\ldots,X_{-n}) && (11)\\ &= H(X_0,X_1,\ldots,X_n) - H(X_1,\ldots,X_n) && (12)\\ &= H(X_0|X_1,\ldots,X_n), && (13)\end{aligned}$$

where (12) follows from stationarity.

Problem 5. For a Markov chain, X_0 and X_n are independent given X_{n-1}. Thus

$$H(X_0|X_nX_{n-1}) = H(X_0|X_{n-1})$$

But, since conditioning reduces entropy,

$$H(X_0|X_nX_{n-1}) \leq H(X_0|X_n).$$

Putting the above together we see that $H(X_0|X_{n-1}) \leq H(X_0|X_n)$.

Problem 6.

$X_1, X_2, \ldots$ are i.i.d. with distribution $p(x)$. Hence $f(X_i)$ are also i.i.d. and

$$\begin{aligned} \lim(\Pi_{i=1}^n f(X_i))^{\frac{1}{n}} &= \lim 2^{\log(\Pi_{i=1}^n f(X_i))^{\frac{1}{n}}}\\ &= 2^{\lim \frac{1}{n}\sum \log f(X_i)}\\ &= 2^{E(\log(f(X))} \qquad \text{a.s.}\end{aligned}$$

by the strong law of large numbers. Note: The abbreviation a.s. stands for 'almost surely', which is synonymous with 'with probability 1'.

(a) Let random variable Z_i represent the multiplicative gain of the gambler for toss i. Z_i is i.i.d., taking the value 2 with probability 0.5 and the value $\frac{1}{3}$ with probability 0.5. The gambler fortune S_n at time n can be described by

$$S_n = \Pi_{i=1}^n Z_i$$

Using the result above

$$\begin{aligned} \lim_{n\to\infty} S_n^{\frac{1}{n}} &= 2^{E(\log(Z))} \qquad \text{a.s.}\\ &= 2^{0.5\log(\frac{2}{3})} = \sqrt{\frac{2}{3}} \qquad \text{a.s.}\end{aligned}$$

(b) For any $\epsilon > 0$, we can find n large enough so that

$$S_n^{\frac{1}{n}} < \sqrt{\frac{2}{3}} + \epsilon \qquad \text{a.s.}$$

Raising to the n^{th} power we have that

$$S_n < (\sqrt{\frac{2}{3}} + \epsilon)^n \qquad \text{a.s.}$$

As $S_n \geq 0$ and the upper bound tends to 0 as $n \to \infty$, we have that

$$\lim_{n\to\infty} S_n = 0 \qquad \text{a.s.}$$

(c)

$$E(S_n|S_{n-1}) = \frac{1}{2}(2S_{n-1} + \frac{1}{3}S_{n-1}) = \frac{7}{6}S_{n-1}$$

Taking an expectation over all possible values of S_{n-1}, we get

$$E(S_n) = \frac{7}{6}E(S_{n-1})$$

Using the fact that $S_1 = 1$, we can compute $E(S_n) = (\frac{7}{6})^n$.

(d) Since $\lim_{n\to\infty} S_n = 0$ a.s.,

$$E \lim_{n\to\infty} S_n = 0$$

(Since expectation is an integral we do not bother with measure 0 events to compute the integral) . Whereas

$$\lim_{n\to\infty} E(S_n) = \lim_{n\to\infty} (\frac{7}{6})^n = \infty$$

Therefore

$$E \lim_{n\to\infty} S_n \neq \lim_{n\to\infty} E(S_n)$$