ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

School of Computer and Communication Sciences

Handout 9 | Information Theory and Coding

Solutions to homework 4 | October 16, 2012

PROBLEM 1. Let $\mathcal{H}(p) = -p \log p - (1-p) \log p$ denote the entropy of a binary valued random variable with distribution $p, 1-p$. The entropy per symbol of the source is

$$\mathcal{H}(p_1) = -p_1 \log p_1 - (1-p_1)\log(1-p_1)$$

and the average symbol duration (or time per symbol) is

$$T(p_1) = 1 \cdot p_1 + 2 \cdot p_2 = p_1 + 2(1-p_1) = 2 - p_1 = 1 + p_2 \,.$$

Therefore the source entropy per unit time is

$$f(p_1) = \frac{\mathcal{H}(p_1)}{T(p_1)} = \frac{-p_1 \log p_1 - (1-p_1)\log(1-p_1)}{2-p_1} \,.$$

Since $f(0) = f(1) = 0$, the maximum value of $f(p_1)$ must occur for some point p_1 such that $0 < p_1 < 1$ and $\partial f / \partial p_1 = 0$.

$$\frac{\partial}{\partial p_1} \frac{\mathcal{H}(p_1)}{T(p_1)} = \frac{T(\partial \mathcal{H} / \partial p_1) - \mathcal{H}(\partial T / \partial p_1)}{T^2}$$

After some calculus, we find that the numerator of the above expression (assuming natural logarithms) is

$$T(\partial H / \partial p_1) - H(\partial T / \partial p_1) = \ln(1-p_1) - 2 \ln p_1 \,,$$

which is zero when $1 - p_1 = p_1^2 = p_2$, that is, $p_1 = \frac{1}{2}(\sqrt{5}-1) = 0.61803$, the reciprocal of the golden ratio, $\frac{1}{2}(\sqrt{5}+1) = 1.61803$. The corresponding entropy per unit time is

$$\frac{\mathcal{H}(p_1)}{T(p_1)} = \frac{-p_1 \log p_1 - p_1^2 \log p_1^2}{2 - p_1} = \frac{-(1+p_1^2)\log p_1}{1+p_1^2} = -\log p_1 = 0.69424 \text{ bits.}$$

PROBLEM 2.

(a) The number of 100-bit binary sequences with three or fewer ones is

$$\binom{100}{0} + \binom{100}{1} + \binom{100}{2} + \binom{100}{3} = 1 + 100 + 4950 + 161700 = 166751 \,.$$

The required codeword length is $\lceil \log_2 166751 \rceil = 18$. (Note that the entropy of the source is $-0.005 \log_2(0.005) - 0.995 \log_2(0.995) = 0.0454$ bits, so 18 is quite a bit larger than the 4.5 bits of entropy per 100 source letters.)

(b) The probability that a 100-bit sequence has three or fewer ones is

$$\sum_{i=0}^{3} \binom{100}{i} (0.005)^i (0.995)^{100-i} = 0.60577 + 0.30441 + 0.7572 + 0.01243 = 0.99833$$

Thus the probability that the sequence that is generated cannot be encoded is $1 - 0.99833 = 0.00167$.

(c) In the case of a random variable S_n that is the sum of n i.i.d. random variables $X_1, X_2, \ldots, X_n$, Chebyshev's inequality states that

$$\Pr(|S_n - n\mu| \geq a) \leq \frac{n\sigma^2}{a^2},$$

where μ and σ^2 are the mean and variance of X_i. (Therefore $n\mu$ and $n\sigma^2$ are the mean and variance of S_n.) In this problem, $n = 100$, $\mu = 0.005$, and $\sigma^2 = (0.005)(0.995)$. Note that $S_{100} \geq 4$ if and only if $|S_{100} - 100(0.005)| \geq 3.5$, so we should choose $a = 3.5$. Then

$$\Pr(S_{100} \geq 4) \leq \frac{100(0.005)(0.995)}{(3.5)^2} \approx 0.04061\,.$$

This bound is much larger than the actual probability 0.00167.

PROBLEM 3.

(a) Since the $X_1, \ldots, X_n$ are i.i.d., so are $q(X_1), q(X_2), \ldots, q(X_n)$, and hence we can apply the strong law of large numbers to obtain

$$\begin{aligned}
\lim -\frac{1}{n} \log q(X_1, \ldots, X_n) &= \lim -\frac{1}{n} \sum \log q(X_i) \\
&= -E[\log q(X)] \quad \text{w.p. } 1 \\
&= -\sum p(x) \log q(x) \\
&= \sum p(x) \log \frac{p(x)}{q(x)} - \sum p(x) \log p(x) \\
&= D(p||q) + H(X).
\end{aligned}$$

(b) Again, by the strong law of large numbers,

$$\begin{aligned}
\lim -\frac{1}{n} \log \frac{q(X_1, \ldots, X_n)}{p(X_1, \ldots, X_n)} &= \lim -\frac{1}{n} \sum \log \frac{q(X_i)}{p(X_i)} \\
&= -E\left[\log \frac{q(X)}{p(X)}\right] \quad \text{w.p. } 1 \\
&= -\sum p(x) \log \frac{q(x)}{p(x)} \\
&= \sum p(x) \log \frac{p(x)}{q(x)} \\
&= D(p||q).
\end{aligned}$$

PROBLEM 4.

(a) It is easy to check that W is an i.i.d. process but Z is not. As W is i.i.d. it is also stationary. We want to show that Z is also stationary. To show this, it is sufficient to prove that the distribution of the process does not change by shift in the time domain.

$$
\begin{aligned}
p_Z(Z_m = a_m, Z_{m+1} &= a_{m+1}, \ldots, Z_{m+r} = a_{m+r}) \\
&= \frac{1}{2} p_X(X_m = a_m, X_{m+1} = a_{m+1}, \ldots, X_{m+r} = a_{m+r}) \\
&+ \frac{1}{2} p_Y(Y_m = a_m, Y_{m+1} = a_{m+1}, \ldots, Y_{m+r} = a_{m+r}) \\
&= \frac{1}{2} p_X(X_{m+s} = a_m, X_{m+s+1} = a_{m+1}, \ldots, X_{m+s+r} = a_{m+r}) \\
&+ \frac{1}{2} p_Y(Y_{m+s} = a_m, Y_{m+s+1} = a_{m+1}, \ldots, Y_{m+s+r} = a_{m+r}) \\
&= p_Z(Z_{m+s} = a_m, Z_{m+s+1} = a_{m+1}, \ldots, Z_{m+s+r} = a_{m+r}),
\end{aligned}
$$

where we used the stationarity of the X and Y processes. This shows the invariance of the distribution with respect to the arbitrary shift s in time which implies stationarity.

(b) For the Z process we have

$$
\begin{aligned}
H(Z) &= \lim_{n\to\infty} \frac{1}{n} H(Z_1, \ldots, Z_n) \\
&= \lim_{n\to\infty} \frac{1}{n} H(Z_1, \ldots, Z_n | \Theta) \\
&= \frac{1}{2} H(X_0) + \frac{1}{2} H(Y_0) = 1.
\end{aligned}
$$

W process is an i.i.d process with the distribution $p_W(a) = \frac{1}{2}p_X(a) + \frac{1}{2}p_Y(a)$. From concavity of the entropy, it is easy to see that $H(W) = H(W_0) \geq \frac{1}{2}H(X_0) + \frac{1}{2}H(Y_0) = 1$. Hence, the entropy rate of W is greater than the entropy rate of Z and the equality holds if and only if X_0 and Y_0 have the same probability distribution function.