EL606: Homework 8 – Solutions

Due date: Nov. 23, 2004

1. *Problem 8, Ch. 14, Cover&Thomas*
   The quantities defining the Slepian Wolf rate region are $H(X,Y) = H(X)$, $H(Y|X) = 0$ and $H(X|Y) \geq 0$. Hence the rate region is as shown in the Figure 1.

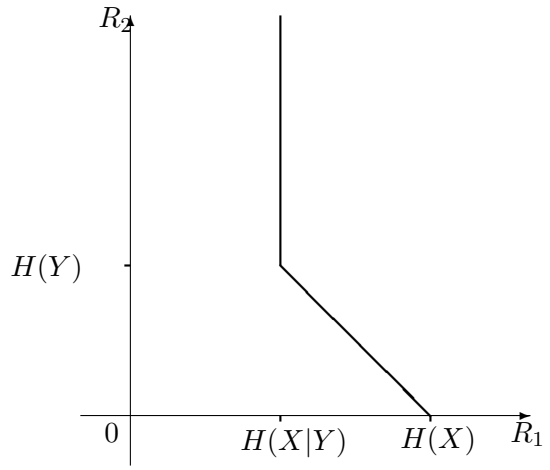


Figure 1: Slepian Wolf rate region for $Y = f(X)$.

2. *Problem 9, Ch. 14, Cover&Thomas*
   $X \sim \text{Bern}(p)$. $Y = X \oplus Z$, $Z \sim \text{Bern}(r)$. Then $Y \sim \text{Bern}(p*r)$, where $p*r = p(1-r)+r(1-p)$. $H(X) = H(p)$. $H(Y) = H(p*r)$, $H(X,Y) = H(X,Z) = H(X) + H(Z) = H(p) + H(r)$. Hence $H(Y|X) = H(r)$ and $H(X|Y) = H(p) + H(r) - H(p*r)$.

   The Slepian Wolf region in this case is shown in Figure 2.

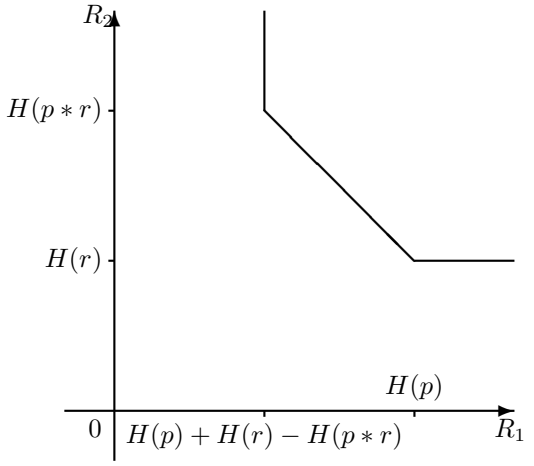


Figure 2: Slepian Wolf region for binary sources

3. *Problem 1, Ch. 9, Cover&Thomas*

   (a) Exponential distribution:

$$\begin{aligned} h(f) &= -\int_0^\infty \lambda e^{-\lambda x}[\ln \lambda - \lambda x]dx \\ &= -\ln \lambda + 1 \text{ nats.} \\ &= \log \frac{e}{\lambda} \text{ bits.} \end{aligned}$$

   (b) Laplace density:

$$\begin{aligned} h(f) &= -\int_{-\infty}^\infty \frac{1}{2}\lambda e^{-\lambda|x|}[\ln \frac{1}{2} + \ln \lambda - \lambda|x|]\, dx \\ &= -\ln \frac{1}{2} - \ln \lambda + 1 \\ &= \ln \frac{2e}{\lambda} \text{ nats.} \\ &= \log \frac{2e}{\lambda} \text{ bits.} \end{aligned}$$

   (c) Sum of two normal distributions:
   The sum of two normal random variables is also normal, so applying the result derived the class for the normal distribution, since $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$,

$$h(f) = \frac{1}{2}\log 2\pi e(\sigma_1^2 + \sigma_2^2) \text{ bits.}$$

4. *Problem 2, Ch. 9, Cover&Thomas*
   Let X_1 and X_2 be normally distributed n-vectors, $\mathbf{X}_i \sim \phi_{K_i}(\mathbf{x})$, $i = 1, 2$. Let the random variable θ have distribution $\Pr\{\theta = 1\} = \lambda$, $\Pr\{\theta = 2\} = 1 - \lambda$, $0 \le \lambda \le 1$. Let θ, $\mathbf{X}_1$, and $\mathbf{X}_2$ be independent and let $\mathbf{Z} = \mathbf{X}_\theta$. Then $\mathbf{Z}$ has covariance $K_Z = \lambda K_1 + (1-\lambda)K_2$. However, $\mathbf{Z}$ will not be multivariate normal. However, since a normal distribution maximizes the entropy for a given variance, we have

$$\frac{1}{2}\ln(2\pi e)^n|\lambda K_1 + (1-\lambda)K_2| \ge h(\mathbf{Z}) \ge h(\mathbf{Z}|\theta) = \lambda\frac{1}{2}\ln(2\pi e)^n|K_1| + (1-\lambda)\frac{1}{2}\ln(2\pi e)^n|K_2| \,.$$

   Thus

$$|\lambda K_1 + (1-\lambda)K_2| \ge |K_1|^\lambda |K_2|^{1-\lambda} \,,$$

   as desired.

5. *Problem 3, Ch. 9, Cover&Thomas*
   For random variables X and Y:

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}_2\left(\mathbf{0}, \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}\right)$$

   we can use the expression for the entropy of a multivariate normal derived in class:

$$h(X,Y) = \frac{1}{2}\log(2\pi e)^2|K| = \frac{1}{2}\log(2\pi e)^2\sigma^4(1-\rho^2).$$

Since X and Y are individually normal with variance σ^2,

$$h(X) = h(Y) = \frac{1}{2}\log 2\pi e\sigma^2.$$

Hence

$$I(X;Y) = h(X) + h(Y) - h(X,Y) = -\frac{1}{2}\log(1-\rho^2).$$

(a) $\rho = 1$. In this case, $X = Y$, and knowing X implies perfect knowledge about Y. Hence the mutual information is infinite, which agrees with the formula.

(b) $\rho = 0$. In this case, X and Y are independent, and hence $I(X;Y) = 0$, which agrees with the formula.

(c) $\rho = -1$. In this case, $X = -Y$, and again the mutual information is infinite as in the case when $\rho = 1$.

6. *Problem 4, Ch. 9, Cover&Thomas*
The probability density function for $Y = X + Z$ is the convolution of the densities of X and Z. Since both X and Z have rectangular densities, the density of Y is a trapezoid. For $a < 1$ the density for Y is

$$p_Y(y) = \begin{cases} (1/2a)(y + (1+a)/2) & -(1+a)/2 \le y \le -(1-a)/2 \\ 1 & -(1-a)/2 \le y \le +(1-a)/2 \\ (1/2a)(-y - (1+a)/2) & +(1-a)/2 \le y \le +(1+a)/2 \end{cases}$$

and for $a > 1$ the density for Y is

$$p_Y(y) = \begin{cases} y + (a+1)/2 & -(a+1)/2 \le y \le -(a-1)/2 \\ 1/a & -(a-1)/2 \le y \le +(a-1)/2 \\ -y - (a+1)/2 & +(a-1)/2 \le y \le +(a+1)/2 \end{cases}$$

(When $a = 1$, the density of Y is triangular over the interval $[-1, +1]$.)

(a) We use the identity $I(X;Y) = h(Y) - h(Y|X)$. It is easy to compute $h(Y)$ directly, but it is even easier to use the grouping property of entropy. First suppose that $a < 1$. With probability $1 - a$, the output Y is conditionally uniformly distributed in the interval $[-(1-a)/2, +(1-a)/2]$; whereas with probability a, Y has a split triangular density where the base of the triangle has width a. As shown in examples in class,

$$\begin{aligned} h(Y) =& H(a) + (1-a)\ln(1-a) + a(\frac{1}{2} + \ln a) \\ =& -a\ln a - (1-a)\ln(1-a) + (1-a)\ln(1-a) + \frac{a}{2} + a\ln a = \frac{a}{2} \text{ nats.} \end{aligned}$$

If $a > 1$ the trapezoidal density of Y can be scaled by a factor a, which yields $h(Y) = \ln a + 1/2a$. Given any value of x, the output Y is conditionally uniformly distributed over an interval of length a, so the conditional differential entropy in nats is $h(Y|X) = h(Z) = \ln a$ for all $a > 0$. Therefore the mutual information in nats is

$$I(X;Y) = \begin{cases} a/2 - \ln a & \text{if } a \le 1 \\ 1/2a & \text{if } a \ge 0\,. \end{cases}$$

As expected, $I(X;Y) \to \infty$ as $a \to 0$ and $I(X;Y) \to 0$ as $a \to \infty$.

(b) As usual with additive noise, we can express $I(X;Y)$ in terms of $h(Y)$ and $h(Z)$:

$$I(X;Y) = h(Y) - h(Y|X) = h(Y) - h(Z)\,.$$

Since both X and Z are limited to the interval $[-1/2, +1/2]$, their sum Y is limited to the interval $[-1, +1]$. The differential entropy of Y is at most that of a random variable uniformly distributed on that interval; that is, $h(Y) \leq 1$. This maximum entropy can be achieved if the input X takes on its extreme values $x = \pm 1$ each with probability $1/2$. In this case, $I(X;Y) = h(Y) - h(Z) = 1 - 0 = 1$. Decoding for this channel is quite simple:

$$\hat{X} = \begin{cases} -1/2 & \text{if } y < 0 \\ +1/2 & \text{if } y \geq 0\,. \end{cases}$$

This coding scheme transmits one bit per channel use with zero error probability. (Only a received value $y = 0$ is ambiguous, and this occurs with probability 0.)

(c) When a is of the form $1/m$ for $m = 2, 3, \ldots$, we can achieve the maximum possible value $I(X;Y) = \log m$ when X is uniformly distributed over the discrete points $\{-1, -1 + 2/(m-1), \ldots, +1 - 2/(m-1), +1\}$. In this case Y has a uniform probability density on the interval $[-1 - 1/(m-1), +1 + 1/(m-1)]$. Other values of a are left as an exercise.