Information Theory

Problem Sheet 1

(Most questions are from Cover & Thomas, the corresponding question numbers (as in 1st ed.) are given in brackets at the start of the question)

Notation: x, $\mathbf{x}$, $\mathbf{X}$ are scalar, vector and matrix random variables respectively.

The following expressions may be useful: $\sum_{n=1}^{\infty} r^n = \frac{r}{1-r}$ $\quad \sum_{n=1}^{\infty} nr^n = \frac{r}{(1-r)^2}$

1. [2.1] A fair coin is flipped until the first head occurs. Let x denote the number of flips required.
   (a) Find the entropy $H(x)$ in bits.
   (b) A random variable x is drawn according to this distribution. Find an "efficient" sequence of yes-no questions of the form "Is x contained in the set S?". Compare $H(x)$ to the expected number of questions required to determine x.

2. [~2.2] x is a random variable taking integer values. What can you say about the relationship between $H(x)$ and $H(y)$ if
   (a) $y = x^2$
   (b) $y = x^3$

3. [2.3] If $\mathbf{p}$ is an n-dimensional probability vector, what is the maximum and the minimum value of $H(\mathbf{p})$. Find all vectors $\mathbf{p}$ for which $H(\mathbf{p})$ achieves its maximum or minimum value.

4. We write $H(p)$ (with a scalar p) to denote the entropy of the Bernoulli random variable with probability mass vector $\mathbf{p} = [1-p \quad p]$. Prove the following properties of this function:
   (a) $H'(p) = \log(1-p) - \log p$
   (b) $H''(p) = \frac{-\log e}{p(1-p)}$
   (c) $H(p) \geq 2\min(p, 1-p)$
   (d) $H(p) \geq 1 - 4(p - ½)^2$
   (e) $H(p) \leq 1 - 2\log e(p - ½)^2$

5. [2.5] Let x be a discrete random variable and $g(x)$ a deterministic function of it. Show that $H(g(x)) \leq H(x)$ by justifying the following steps:

$$H(x, g(x)) \overset{(a)}{=} H(x) + H(g(x) \mid x) \overset{(b)}{=} H(x)$$
$$H(x, g(x)) \overset{(c)}{=} H(g(x)) + H(x \mid g(x)) \overset{(d)}{\geq} H(g(x))$$

6. [2.6] Show that if $H(y \mid x) = 0$, then y is a function of x, that is for all x with $p(x)>0$, there is only one possible value of y with $p(x,y) > 0$.

7. [~2.7] x_i is a sequence of i.i.d. Bernoulli random variables with $p(x_i=1) = p$ where p is unknown. We want to find a function f that converts n samples of x into a smaller number, K, of i.i.d. Bernoulli random variables, z_i, with $p(z_i=1)=½$. Thus $z_{1:K}=f(x_{1:n})$ where K can depend on the values x_i.

(a) Show that the following mapping for n=4 satisfies the requirements and find the expected value of K, $E(K)$.

0000,1111→ignore;	1010→0;	0101→1;	0001,0011,0111→00;
0010,0110,1110→01;	0100,1100,1101→10;		1000,1001,1011→11

(b) Justify the steps in the following bound on $E(K)$

$$\begin{aligned} nH(p) &\overset{(a)}{=} H(x_{1:n}) \overset{(b)}{\geq} H(z_{1:K}, K) \overset{(c)}{=} H(K) + H(z_{1:K} \mid K) \\ &\overset{(d)}{=} H(K) + E\,K \overset{(e)}{\geq} E\,K \end{aligned}$$

8. [2.10] Give examples of joint random variables x, y and z such that:

(a) $I(x;y \mid z) < I(x;y)$

(b) $I(x;y \mid z) > I(x;y)$

9. [2.12] We can define the "mutual information" between three variables as

$$I(x;y;z) = I(x;y) - I(x;y \mid z)$$

(a) Prove that

$$\begin{aligned} I(x;y;z) = {} & H(x,y,z) - H(x,y) - H(y,z) - H(z,x) \\ & + H(x) + H(y) + H(z) \end{aligned}$$

(b) Give an example where $I(x;y;z)$ is negative. This lack of positivity means that it does not have the intuitive properties of an "information" measure which is why I put "mutual information" in quotes above.

10. [2.17] Show that $\log_e(x) \geq 1 - x^{-1}$ for $x>0$.

11. [~2.16] x and y are correlated binary random variables with $p(x=y=0)=0$ and all other joint probabilities equal to 1/3. Calculate $H(x)$, $H(y)$, $H(x|y)$, $H(y|x)$, $H(x,y)$, $I(x;y)$.

12. [~2.22] If $x \to y \to z$ form a markov chain, and for y, the alphabet size $|Y| = k$, show that $I(x;z) \leq \log k$. What does this tell you if $k = 1$?

13. [2.29] Prove the following and find the conditions for equality:

(a) $H(x,y \mid z) \geq H(x \mid z)$

(b) $I(x,y;z) \geq I(x;z)$

(c) $H(x,y,z) - H(x,y) \leq H(x,z) - H(x)$

(d) $I(x;z \mid y) \geq I(z;y \mid x) - I(z;y) + I(x;z)$

Information Theory

Solution Sheet 1

1. (a) $x = n$ means that Tail occurs for the first $n - 1$ flips, while the last flip is Head. Thus, x has distribution $P(x = n) = 2^{-(n-1)}\, 2^{-1} = 2^{-n}$. Thus

$$H(x) = \sum_{n=1}^{\infty} 2^{-n} \log 2^n = \sum_{n=1}^{\infty} 2^{-n} n \log 2$$
$$= \sum_{n=1}^{\infty} n 2^{-n} = \frac{1/2}{(1-1/2)^2} = 2$$

(b) Ask if x = 1, 2, 3, … in turn, i.e., ask the following questions:

Is $x = 1$?
If not, is $x = 1$?
If not, is $x = 2$?
…

Expected number of questions is $\sum_{n=1}^{\infty} n 2^{-n} = 2$.

2. $H(x,y)=H(x)+H(y|x)=H(Y)+H(x|y)$, but $H(y|x)=0$ since Y is a function of x so $H(y)= H(x) - H(x|y) \le H(x)$ with equality iff $H(x|y)=0$ which is true only if x is a function of y, i.e. if y is a one-to-one function of x for every value of x with $p(x)>0$. Hence

(a) $H(y) \le H(x)$ because, for example $1^2 = (-1)^2$

(b) $H(y)=H(x)$

3. Maximum is $\log n$ iff all elements of **p** are equal. Minimum is 0 iff only one element of **p** is non-zero; there are n possible elements that this could be.

4. (a) and (b) are straightforward calculus: easiest to convert logs to base e first. For the others, assume $½ < p < 1$ for covenience (other half follows by symmetry). Since $H''(p) < 0$, $H(p)$ is concave and so lies above the straight line $2 - 2p$ defined in (c).

At $p = ½$ the bound in (e) has the same value and first two derivatives as $H(p)$. For $½ < p < 1$ its second derivative is greater than $H''(p)$ and so the bound follows.

For (d) we consider $D(p) = H(p) - 1 + 4(p - ½)^2$. $D''(p) = 0$ is a quadratic in p and has only two solutions $p = ½ \pm \sqrt{(2 - \log e)/8} = 0.5 \pm 0.26$. Therefore $D'(p)$ increases from 0 at $p = 0.5$ to reach a maximum at $p = 0.76$ and decreases thereafter. This implies that $D'(p) = 0$ has only one solution for $p > ½$ and therefore that $D(p)$ has a single maximum. Since $D(½) = D(1) = 0$ we must have $D(p) > 0$ for $½ < p < 1$.

5. (a) chain rule, (b) $g(x)|x$ has only one possible value and hence zero entropy, (c) chain rule, (d) entropy is positive. We have equality at (d) iff $g(x)$ is a one-to-one function for every x with $p(x)>0$.

6. $H(y \mid x) = \sum_x p(x) H(y \mid x = x)$

All terms are non-negative so the sum is zero only if all terms are zero. For any given term this is true either if $p(x)=0$ or if $H(y|x=x)$ is zero. The second case arises only if $H(y|x=x)$ has only one value, i.e. y is a function of x. The first case is why we needed the qualification about $p(x)>0$ in answers 2 and 4 above.

7. (a) The probability of any given value of $x_{1:4}$ depends on the number of 1's and 0's. We create four subsets with equal probabilities to generate a pair of bits and two other subsets to generate one bit only. The expected number of bits generated is

$$E\,K = 8p(1-p)^3 + 10p^2(1-p)^2 + 8p^3(1-p)$$

(b) (a) i.i.d entropies add, (b) functions reduce entropy, (c) chain rule, (d) z_i are i.i.d. with entropy of 1 bit, (e) entropy is positive.

8. (a) This is true for any Markov chain $x \to y \to z$. One possibility is $x=y=z$ all fair Bernoulli variables.

(b) An example of this was given in lectures. A slightly different example is if x and y are fair binary variables and $z=xy$. Knowing z, entangles x and y.

9. (a) $I(x;y;z)=\{H(x)-H(x|y)\}-\{H(x|z)-H(x|y,z)\}=H(x)-\{H(x,y)-H(y)\}-\{H(x,z)-H(z)\}+\{H(x,y,z)-H(y,z)\}$

(b) Use the example from 8(b) above.

10. Define $f(x)=\ln(x)+x^{-1}-1$. This is continuous and differentiable in $(0,\infty)$. Differentiate twice to show that the only extremum occurs at $x=1$ and that it is a minimum. Hence $f(x)\geq f(1)=0$.

11. $H(x)=H(y)=0.918$; $H(x|y)=H(y|x)=0.667$; $H(x,y)=1.58$; $I(x;y)=0.252$.

12. The data processing inequality says that $I(x;z)\leq I(x;y)=H(y)-H(y|x) \leq H(y) \leq \log k$ where the last inequality is the uniform bound on entropy. If $k=1$ then $\log k = 0$ and so x and z must be independent.

13. (a) $H(x,y|z)=H(x|z)+H(y|x,z)\geq H(x|z)$ with equality if y is a function of x and z.

(b) $I(x,y;z)=I(x;z)+I(y;z|x) \geq I(x;z)$ with equality if y and z are conditionally independent given x.

(c) $H(x,y,z)-H(x,y)=H(z|x,y)=H(z|x)-I(y;z|x) \leq H(z|x)=H(x,z)-H(x)$ with equality if y and z are conditionally independent given x.

(d) $I(x,y;z)=I(y;z)+I(x;z|y)= I(x;z)+I(y;z|x)$. Rearrange this to give the inequality which is in fact always an equality (trick question).

Information Theory

Problem Sheet 2

Notation: x, $\mathbf{x}$, $\mathbf{X}$ are scalar, vector and matrix random variables respectively.

1. Use the Kraft inequality to show that it is possible to construct a 4-ary instantaneous code with lengths {1, 1, 2, 2, 2, 2, 2, 2, 2}. Construct such a code for symbols that take the values A, B, …, H, I with probabilities {.15, .15, .1, .1, .1, .1, .1, .1, .1}.

   Calculate the entropy of the input symbols and the expected length of the codewords.

2. The four symbols A, B, C, D are encoded using the following sets of codewords. In each case state whether the code is (i) non-singular, (ii) uniquely decodable and (iii) instantaneous code.

   (a) {1, 01, 000, 001}

   (b) {0, 10, 000, 100}

   (c) {01, 01, 110, 100}

   (d) {0, 01, 011, 0111}

   (e) {10, ~~10~~, 0010, 0111}

   01

3. In a complete D-ary code tree, the end of each branch is either a leaf node of else has D sub-branches. Show that the total number of leaf nodes is one more than a multiple of D–1.

4. For each value of D given below, find a D-ary Huffman code for the probability vector {0.25, 0.2, 0.15, 0.1, 0.1, 0.1, 0.1}. In each case calculate the expected code length. [Note that, to ensure a full code tree, you should add zero-probability symbols so that the total number of symbols is one more than a multiple of D–1]

   (a) D = 3

   (b) D = 4

   (c) D = 5

5. Find a binary instantaneous code with codeword lengths {2, 2, 3, 3, 3, 4, 4}. Find a probability mass vector for which the expected length of this code is equal to the entropy of the source.

6. [4.2] If $\{x_i\}$ is a stationary stochastic process, show that

$$H(x_i \mid x_{i-1}, x_{i-2}, \cdots, x_{i-n}) = H(x_i \mid x_{i+1}, x_{i+2}, \cdots, x_{i+n}) .$$

   In other words, the conditional entropy given the previous n samples is the same as the conditional entropy given the next n samples.

7. (a) [4.5] Determine the stationary distribution and the entropy rate, $H(\mathsf{X})$, of a Markov process with two states, 0 and 1 and transition matrix

$$T = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} .$$

   (b) Find the values of p and q that maximize $H(\mathsf{X})$.

   (c) If q=1,

   (i) find the maximum value of $H(\mathsf{X})$ and the value of p that attains it.

   (ii) define $s(n)$ to be the number of sequences of length n with non-zero probability. If $s_0(n)$ and $s_1(n)$ are the number of sequences starting with 0 and 1 respectively, show that $s_1(n) = s_0(n-1) = s(n-2)$. Hence show that $s(n) = s(n-1) + s(n-2)$ and derive an explicit expression for $s(n)$.

   (iii) explain why we must have $H(\mathsf{X}) \le \lim_{n\to\infty}\left(n^{-1} \log s(n)\right)$ and calculate the value on the right hand side.

8. If $x_i \in \{A, B\}$ are i.i.d. with probability mass vector $\{0.9, 0.1\}$, determine $H(x_i)$.

   Using binary Huffman codes, determine the coding redundancy (i.e. the difference between the entropy and the number of bits used per symbol) when (a) each x_i is encoded individually, (b) pairs of x_i are coded together (i.e. x_1x_2 followed by x_3x_4 etc) and (c) triplets of x_i are coded together.

Information Theory

Solution Sheet 2

1. Kraft Inequality: $2\times 4^{-1}+7\times 4^{-2}=0.9375\le 1$ so an instantaneous code is possible. A suitable code is {0, 1, 20, 21, 22, 23, 30, 31, 32}.

   Average length of code is 1.7 (equivalent to 3.4 bits since we are using a 4-ary code).

   Entropy of input is 3.1464 bits.

2. Note that Instantaneous ⇒ Uniquely Decodable ⇒ Non-singular

   (a) {1, 01, 000, 001} is an instantaneous code (and therefore uniquely decodable and non-singular).

   (b) {0, 10, 000, 100} is non-singular since no two codewords are the same. However it is not uniquely decodable since AAA and C are indistinguishable (it follows that it is not an instantaneous code either).

   (c) {01, 01, 110, 100} is singular since A and B have the same codeword (hence not uniquely decodable and not an instantaneous code).

   (d) {0, 01, 011, 0111} is uniquely decodable: when you get a 0 it is the start of a new symbol and the previous symbol is given by counting how many 1's since the previous 0. It is not a instantaneous code since the first codeword is a prefix of all the others.

   (e) {10, 01, 0010, 0111} is not an instantaneous code since 01 is a prefix of 0111. Since all codewords have even length we can rewrite it as a 4-ary code {2, 1, 02, 13}. It is easy to see that this is uniquely decodable since 3 only ever appears as the second half of D. Thus if a 1 is followed by anything other than 3, it represents a B.

3. You can create the tree incrementally from an initial tree with only one node. At each stage, you replace an existing leaf node with a set of *D* sub-branches which end in leaf nodes. This process adds *D* new leaf nodes but removes one that existed previously. It therefore adds *D*–1 to the total. Initially, we only had one leaf node and so we will always have one more than a multiple of *D*–1.

4. (a) D = 3. We do not need to append any zero-probability symbols since 7 is one more that a multiple of D–1 already. The codes we get are {2, 00, 01, 02, 10, 11, 12}. The expected length is 1.75.

   (b) D = 4. Again 7 is one more than a multiple of D-1 so we do not need to add any symbols. The codes we get are {1, 2, 3, 00, 01, 02, 03}. The expected length is 1.4.

   (c) D = 5. We need to two zero-probability symbols to make the number of symbols one more than a multiple of 4. The codes we get are then {1, 2, 3, 4, 00, 01, 02, [03, 04]} where the last two codes are in brackets because they are not used. Not that adding the zero-probability symbols ensures that the unused codes are the longest ones. The expected length is now 1.3.

5. A suitable code is {00, 01, 100, 101, 110, 1110, 1111}.

   To attain the source entropy, all the symbol probabilities must equal 2^{-l}. This gives us {0.25, 0.25, 0.125, 0.125, 0.125, 0.0625, 0.0625} as the probabilities.

6. We can write the following:

$$\begin{aligned}H(x_i \mid x_{i-1},x_{i-2},\cdots,x_{i-n}) &\overset{(a)}{=} H(x_i,x_{i-1},x_{i-2},\cdots,x_{i-n})-H(x_{i-1},x_{i-2},\cdots,x_{i-n})\\ &\overset{(b)}{=} H(x_{i+n},x_{i+n-i},x_{i+n-2},\cdots,x_i)-H(x_{i+n},x_{i+n-1},\cdots,x_{i+1})\\ &\overset{(c)}{=} H(x_i \mid x_{i+1},x_{i+2},\cdots,x_{i+n})\end{aligned}$$

   Where (a) and (c) follow from the definition of conditional entropy and (b) follows from stationarity: we add n onto the time index of the first term and $n+1$ onto the time index of the second.

7. (a) To find the stationary distribution, we write

$$T^T\begin{pmatrix}a\\1-a\end{pmatrix}=\begin{pmatrix}1-p & q\\ p & 1-q\end{pmatrix}\begin{pmatrix}a\\1-a\end{pmatrix}=\begin{pmatrix}a\\1-a\end{pmatrix}$$

$$\Rightarrow\quad (1-p)a+q(1-a)=(1-p-q)a+q=a$$

$$\Rightarrow\quad a=q(p+q)^{-1}$$

   It follows that the entropy rate is given by

$$H(\mathsf{X})=aH(p)+(1-a)H(q)=(qH(p)+pH(q))(p+q)^{-1}$$

This is a weighted average of $H(p)$ and $H(q)$ with the weights equal to their frequencies in the stationary distribution.

(b) We know that $H(\mathsf{X}) = H(x_n \mid x_{n-1}) \le H(x_n) \le 1$

If $p = q = 0.5$ then $H(\mathsf{X}) = H(p) = H(q) = 1$ and this is the maximum attainable.

(c-i) Substituting $q = 1$ into the previous answer gives $H(\mathsf{X}) = (p+1)^{-1}H(p)$. Differentiating and setting this to zero gives $H(p) = (p+1)H'(p)$ from which we get $p = (1-p)^2$ and $p = 1.5 - \sqrt{1.25} = 0.382$. This gives $H(\mathsf{X}) = 0.694$ bits.

(c-ii) If the first bit is 0 the next can be either 0 or 1 which means that $s_0(n) = s(n-1)$. However $q = 1$, if the first bit of a sequence is 1, then the next must be 0. It follows that $s_1(n) = s_0(n-1) = s(n-2)$. Hence $s(n) = s_0(n) + s_1(n) = s(n-1) + s(n-2)$ with the initial conditions $s(1) = 2$ and $s(2) = 3$. The solutions to the recurrence relation (which is that of the Fibonacci series for those interested in such things) have the form $s(n) = a^n$ where a satisfies $a^2 = a + 1 \quad \Rightarrow \quad a = 0.5 \pm \sqrt{1.25} = \{-0.618, 1.618\}$. Imposing the initial conditions gives $s(n) = -0.17 \times -0.618^n + 1.17 \times 1.618^n$. For large n, the first term is insignificant and $s(n) \approx 1.17 \times 1.618^n$.

(c-iii) By definition $H(\mathsf{X}) = \lim_{n\to\infty}\left(n^{-1}H(X_{1:n})\right)$. An upper bound on $H(x_1, x_2, \ldots x_n)$ is given by $\log s(n)$ which is attained if all $s(n)$ possible sequences have equal probability. Thus $\lim_{n\to\infty}\left(n^{-1}\log s(n)\right) = \log 1.618 = 0.694$ is an upper bound for $H(\mathsf{X})$.

8. $H(x_i) = 0.469$ bits .

(a) $p\{A, B\} = \{0.9, 0.1\}$ giving Huffman lengths of $\{1, 1\}$ with $E(L) = 1$. Redundancy = 0.531 bits.

(b) $p\{AA, AB, BA, BB\} = \{0.81, 0.09, 0.09, 0.01\}$ giving Huffman lengths of $\{1, 2, 3, 3\}$ with $E(L)/2 = 1.29/2 = 0.645$ bits per input symbol. Redundancy = 0.176 bits.

(c) $p\{AAA, AAB, ABA, BAA, ABB, BAB, BBA, BBB\} =$ $\{0.729, 0.081*3, 0.009*3, 0.001\}$ giving Huffman lengths of $\{1, 3, 3, 3, 5, 5, 5, 5\}$ with $E(L)/3 = 1.598/3 = 0.533$ bits per input symbol . Redundancy = 0.064 bits.

Information Theory

Problem Sheet 3

Notation: x, $\mathbf{x}$, $\mathbf{X}$ are scalar, vector and matrix random variables respectively.

1. The Z channel. The Z-channel has binary input and output alphabets and transition probabilities $p(y|x)$ given by the following matrix:

$$Q = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix} \quad x, y \in \{0,1\}$$

Find the capacity of the Z-channel and the maximizing input probability distribution.

2. Calculate the capacity of the following channel with probability transition matrix

$$Q = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{bmatrix} \quad x, y \in \{0,1,2\}$$

3. Differential entropy. Evaluate the differential entropy $h(X) = -\int f \ln f$ for the following:
(a) The exponential density $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$.
(b) The sum of X_1 and X_2; where X_1 and X_2 are independent normal random variables with means μ_i and variances σ_i^2 for $i = 1, 2$.

4. Parallel channels and waterfilling. Consider a pair of parallel Gaussian channels, i.e.,

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$$

where

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N\left(0, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\right)$$

and there is a power constraint $E(X_1^2 + X_2^2) \leq 2P$. Assume that $\sigma_1^2 \geq \sigma_2^2$. At what power does the channel stop behaving like a single channel with noise variance σ_2^2 and begin behaving like a pair of channels?

5. Rate-distortion function. Let

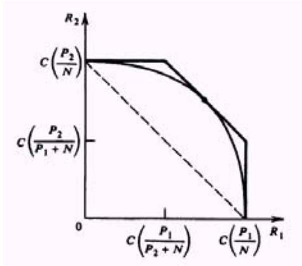
$$D(R) = \min_{p(\hat{x}|x): I(X;\hat{X}) \leq R} Ed(X, \hat{X})$$

be the rate-distortion function.
(a) Is $D(R)$ a decreasing or increasing function of R?
(b) Is $D(R)$ convex or concave in R?

6. Describe the capacity region of a two-user multiple access channel. Interpret the corner points (i.e., why can one of the users achieve the capacity as if the other user were absent?) Verify the following equality for the corner point:

$$C\left(\frac{P_1}{N}\right) + C\left(\frac{P_2}{P_1 + N}\right) = C\left(\frac{P_1 + P_2}{N}\right)$$

where $C(x) = (\log(1+x))/2$ is the capacity function.



7. Slepian-Wolf region for binary sources. Let X_i be i.i.d. Bernoulli(p). Let Z_i be i.i.d. Bernoulli(r), and let Z be independent of X. Finally, let Y = X⊕Z (mod 2 addition). Let X be described at rate R_1 and Y be described at rate R_2: What region of rates allows recovery of X, Y with probability of error tending to zero?

8. Describe the capacity region of a two-user Gaussian broadcast channel and sketch it.

Information Theory

Solution Sheet 3

1. First we express I(X;Y), the mutual information between the input and output of the Z-channel, as a function of x = Pr(X = 1):

$$H(Y \mid X) = \mathrm{P}(X = 0)\cdot 0 + \mathrm{P}(X = 1)\cdot 1 = x$$
$$H(Y) = H(P(Y=1)) = H(x/2)$$
$$I(X;Y) = H(Y) - H(Y \mid X) = H(x/2) - x$$

Taking the derivative, we have

$$\frac{dI(X;Y)}{dx} = \frac{1}{2}\log_2 \frac{1-x/2}{x/2} - 1$$

which is equal to zero for x = 2/5. So the capacity of the Z-channel in bits is H(1/5) – 2/5 = 0.722 – 0.4 = 0.322.

2. Since the channel is symmetric,

$$C = \log|Y| - H(Q(1,:)) = \log_2 3 - 1 = 0.58 \text{ bits}$$

3. (a) log(e/λ) bits.

$$h(X) = -\int_0^\infty f(x)\log f(x)dx = -\log e\int_0^\infty f(x)\ln f(x)dx$$
$$= -\log e\int_0^\infty \lambda e^{-\lambda x}(\ln\lambda - \lambda x)dx = -\log e(\ln\lambda - \int_0^\infty \lambda x e^{-\lambda x}d\lambda x)$$
$$= -\log e(\ln\lambda - 1) = -\log e\cdot\ln(\lambda/e) = \log(e/\lambda) \text{ bits}$$

(b) The sum of two normal random variables is also normal, so applying the result derived the class for the normal distribution,

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

4. By the result of Lecture 14, it follows that we will put all the signal power into the channel with less noise until the total power of noise + signal in that channel equals the noise power in the other channel. After that, we will split any additional power "evenly" (in the sense that the power of noise + signal in one channel is equal to that in the other channel) between the two channels. Thus the combined channel begins to behave like a pair of parallel channels when the signal power is equal to the difference of the two noise powers, i.e., when $2P = \sigma_1^2 - \sigma_2^2$.

5. (a) Decreasing.

(b) Convex.

6. Capacity region

$$R_1 < C\left(\frac{P_1}{N}\right)$$
$$R_2 < C\left(\frac{P_2}{N}\right)$$
$$R_1 + R_1 < C\left(\frac{P_1+P_2}{N}\right)$$

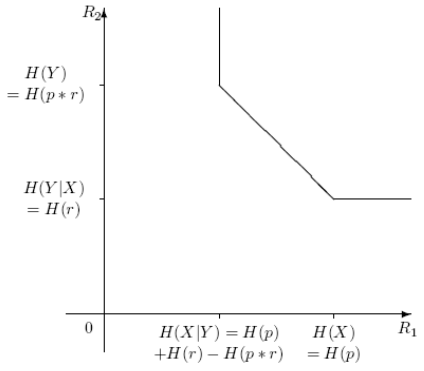
Interpretation: onion-peeling.
Verification:

$$\begin{aligned} C(\frac{P_1+P_2}{N}) &= \frac{1}{2}\log(1+\frac{P_1+P_2}{N}) \\ &= \frac{1}{2}\log(\frac{N+P_1+P_2}{N}) \\ &= \frac{1}{2}\log(\frac{N+P_1+P_2}{N+P_1}\cdot\frac{N+P_1}{N}) \\ &= \frac{1}{2}\log(\frac{N+P_1+P_2}{N+P_1}) + \frac{1}{2}\log(\frac{N+P_1}{N}) \\ &= C(\frac{P_2}{P_1+N}) + C(\frac{P}{N_1}) \end{aligned}$$

7. The Slepian-Wolf region is given by

$$R_1 \ge H(X \mid Y)$$
$$R_2 \ge H(Y \mid X)$$
$$R_1 + R_2 \ge H(X,Y)$$

X = Bernoulli(*p*). *Y* = *X*⊕*Z*, *Z* = Bernoulli(*r*). Then *Y* = Bernoulli(*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*).
This region is shown in the next page.



8. The capacity region is

$$R_1 \le C\left(\frac{\alpha P}{N_1}\right)$$

$$R_2 \le C\left(\frac{(1-\alpha)P}{\alpha P + N_2}\right)$$

where $0 \le \alpha \le 1$. It looks like this: