

$H(Y | X) = \sum_x P(x)H(Y | x) = P(x=1)H(Y | x=1) + P(x=0)H(Y | x=0)$
 so the mutual information is:

$$I(X; Y) = H(Y) - H(Y | X) \quad (9.33)$$

$$= H_2(0.575) - [0.5 \times H_2(0.15) + 0.5 \times 0] \quad (9.34)$$

$$= 0.98 - 0.30 = 0.679 \text{ bits.} \quad (9.35)$$

Solution to exercise 9.12 (p.151). By symmetry, the optimal input distribution is $\{0.5, 0.5\}$. Then the capacity is

$$C = I(X; Y) = H(Y) - H(Y | X) \quad (9.36)$$

$$= H_2(0.5) - H_2(f) \quad (9.37)$$

$$= 1 - H_2(f). \quad (9.38)$$

Would you like to find the optimal input distribution without invoking symmetry? We can do this by computing the mutual information in the general case where the input ensemble is $\{p_0, p_1\}$:

$$I(X; Y) = H(Y) - H(Y | X) \quad (9.39)$$

$$= H_2(p_0f + p_1(1 - f)) - H_2(f). \quad (9.40)$$

The only p -dependence is in the first term $H_2(p_0f + p_1(1 - f))$, which is maximized by setting the argument to 0.5. This value is given by setting $p_0 = 1/2$.

Solution to exercise 9.13 (p.151). **Answer 1.** By symmetry, the optimal input distribution is $\{0.5, 0.5\}$. The capacity is most easily evaluated by writing the mutual information as $I(X; Y) = H(X) - H(X | Y)$. The conditional entropy $H(X | Y)$ is $\sum_y P(y)H(X | y)$; when y is known, x is uncertain only if $y = ?$, which occurs with probability $f/2 + f/2$, so the conditional entropy $H(X | Y)$ is $fH_2(0.5)$.

$$C = I(X; Y) = H(X) - H(X | Y) \quad (9.41)$$

$$= H_2(0.5) - fH_2(0.5) \quad (9.42)$$

$$= 1 - f. \quad (9.43)$$

The binary erasure channel fails a fraction f of the time. Its capacity is precisely $1 - f$, which is the fraction of the time that the channel is reliable. This result seems very reasonable, but it is far from obvious how to encode information so as to communicate *reliably* over this channel.

Answer 2. Alternatively, without invoking the symmetry assumed above, we can start from the input ensemble $\{p_0, p_1\}$. The probability that $y = ?$ is $p_0f + p_1f = f$, and when we receive $y = ?$, the posterior probability of x is the same as the prior probability, so:

$$I(X; Y) = H(X) - H(X | Y) \quad (9.44)$$

$$= H_2(p_1) - fH_2(p_1) \quad (9.45)$$

$$= (1 - f)H_2(p_1). \quad (9.46)$$

This mutual information achieves its maximum value of $(1 - f)$ when $p_1 = 1/2$.

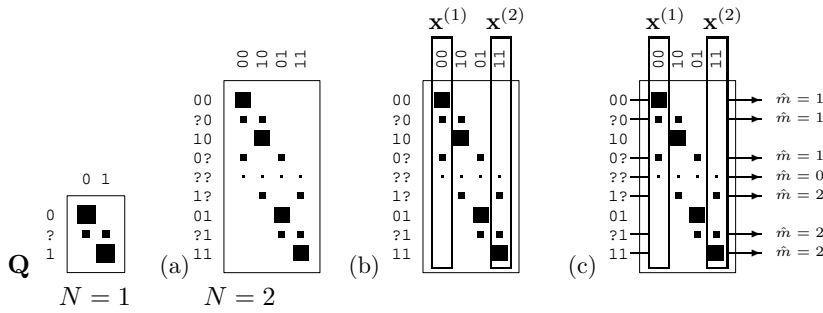


Figure 9.10. (a) The extended channel (\$N = 2\$) obtained from a binary erasure channel with erasure probability 0.15. (b) A block code consisting of the two codewords 00 and 11. (c) The optimal decoder for this code.

Solution to exercise 9.14 (p.153). The extended channel is shown in figure 9.10. The best code for this channel with \$N = 2\$ is obtained by choosing two columns that have minimal overlap, for example, columns 00 and 11. The decoding algorithm returns ‘00’ if the extended channel output is among the top four and ‘11’ if it’s among the bottom four, and gives up if the output is ‘??’.

Solution to exercise 9.15 (p.155). In example 9.11 (p.151) we showed that the mutual information between input and output of the Z channel is

$$I(X; Y) = H(Y) - H(Y | X) = H_2(p_1(1 - f)) - p_1 H_2(f). \quad (9.47)$$

We differentiate this expression with respect to \$p_1\$, taking care not to confuse \$\log_2\$ with \$\log_e\$:

$$\frac{d}{dp_1} I(X; Y) = (1 - f) \log_2 \frac{1 - p_1(1 - f)}{p_1(1 - f)} - H_2(f). \quad (9.48)$$

Setting this derivative to zero and rearranging using skills developed in exercise 2.17 (p.36), we obtain:

$$p_1^*(1 - f) = \frac{1}{1 + 2^{H_2(f)/(1-f)}}, \quad (9.49)$$

so the optimal input distribution is

$$p_1^* = \frac{1/(1 - f)}{1 + 2^{(H_2(f)/(1-f))}}. \quad (9.50)$$

As the noise level \$f\$ tends to 1, this expression tends to \$1/e\$ (as you can prove using L’Hôpital’s rule).

For all values of \$f\$, \$p_1^*\$ is smaller than \$1/2\$. A rough intuition for why input 1 is used less than input 0 is that when input 1 is used, the noisy channel injects entropy into the received string; whereas when input 0 is used, the noise has zero entropy.

Solution to exercise 9.16 (p.155). The capacities of the three channels are shown in figure 9.11. For any \$f < 0.5\$, the BEC is the channel with highest capacity and the BSC the lowest.

Solution to exercise 9.18 (p.155). The logarithm of the posterior probability ratio, given \$y\$, is

$$a(y) = \ln \frac{P(x = 1 | y, \alpha, \sigma)}{P(x = -1 | y, \alpha, \sigma)} = \ln \frac{Q(y | x = 1, \alpha, \sigma)}{Q(y | x = -1, \alpha, \sigma)} = 2 \frac{\alpha y}{\sigma^2}. \quad (9.51)$$

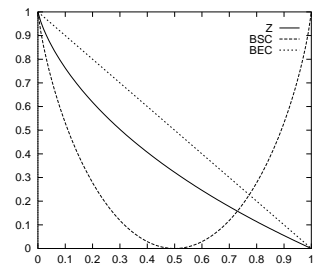


Figure 9.11. Capacities of the Z channel, binary symmetric channel, and binary erasure channel.

Using our skills picked up from exercise 2.17 (p.36), we rewrite this in the form

$$P(x=1 | y, \alpha, \sigma) = \frac{1}{1 + e^{-a(y)}}. \quad (9.52)$$

The optimal decoder selects the most probable hypothesis; this can be done simply by looking at the sign of $a(y)$. If $a(y) > 0$ then decode as $\hat{x} = 1$.

The probability of error is

$$p_b = \int_{-\infty}^0 dy Q(y | x=1, \alpha, \sigma) = \int_{-\infty}^{-x\alpha} dy \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} = \Phi\left(-\frac{x\alpha}{\sigma}\right). \quad (9.53)$$

Random coding

Solution to exercise 9.20 (p.156). The probability that $S = 24$ people whose birthdays are drawn at random from $A = 365$ days all have *distinct* birthdays is

$$\frac{A(A-1)(A-2)\dots(A-S+1)}{A^S}. \quad (9.54)$$

The probability that two (or more) people share a birthday is one minus this quantity, which, for $S = 24$ and $A = 365$, is about 0.5. This exact way of answering the question is not very informative since it is not clear for what value of S the probability changes from being close to 0 to being close to 1.

The number of pairs is $S(S-1)/2$, and the probability that a particular pair shares a birthday is $1/A$, so the *expected number* of collisions is

$$\frac{S(S-1)}{2} \frac{1}{A}. \quad (9.55)$$

This answer is more instructive. The expected number of collisions is tiny if $S \ll \sqrt{A}$ and big if $S \gg \sqrt{A}$.

We can also approximate the probability that all birthdays are distinct, for small S , thus:

$$\begin{aligned} \frac{A(A-1)(A-2)\dots(A-S+1)}{A^S} &= (1)(1-1/A)(1-2/A)\dots(1-(S-1)/A) \\ &\simeq \exp(0) \exp(-1/A) \exp(-2/A) \dots \exp(-(S-1)/A) \end{aligned} \quad (9.56)$$

$$\simeq \exp\left(-\frac{1}{A} \sum_{i=1}^{S-1} i\right) = \exp\left(-\frac{S(S-1)/2}{A}\right). \quad (9.57)$$