Special cases. In the special case $p_? = 0$, the channel is a noiseless binary channel, and $I(X;Y) = H_2(p_0)$.

In the special case $p_0 = p_1$, the term $H_2(p_0 + p_?/2)$ is equal to 1, so $I(X;Y) = 1 - p_?$.

In the special case $p_0 = 0$, the channel is a Z channel with error probability 0.5. We know how to sketch that, from the previous chapter (figure 9.3).

These special cases allow us to construct the skeleton shown in figure 10.9.

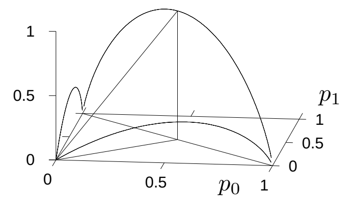

Figure 10.9. Skeleton of the mutual information for the ternary confusion channel.

Solution to exercise 10.5 (p.169). Necessary and sufficient conditions for $\mathbf{p}$ to maximize $I(X;Y)$ are

$$\left.\begin{array}{rcl} \frac{\partial I(X;Y)}{\partial p_i} & = & \lambda \text{ and } p_i > 0 \\ \frac{\partial I(X;Y)}{\partial p_i} & \leq & \lambda \text{ and } p_i = 0 \end{array}\right\} \text{ for all } i, \tag{10.28}$$

where λ is a constant related to the capacity by $C = \lambda + \log_2 e$.

This result can be used in a computer program that evaluates the derivatives, and increments and decrements the probabilities p_i in proportion to the differences between those derivatives.

This result is also useful for lazy human capacity-finders who are good guessers. Having guessed the optimal input distribution, one can simply confirm that equation (10.28) holds.

Solution to exercise 10.11 (p.171). We certainly expect nonsymmetric channels with uniform optimal input distributions to exist, since when inventing a channel we have $I(J-1)$ degrees of freedom whereas the optimal input distribution is just $(I-1)$-dimensional; so in the $I(J-1)$-dimensional space of perturbations around a symmetric channel, we expect there to be a subspace of perturbations of dimension $I(J-1) - (I-1) = I(J-2) + 1$ that leave the optimal input distribution unchanged.

Here is an explicit example, a bit like a Z channel.

$$\mathbf{Q} = \begin{bmatrix} 0.9585 & 0.0415 & 0.35 & 0.0 \\ 0.0415 & 0.9585 & 0.0 & 0.35 \\ 0 & 0 & 0.65 & 0 \\ 0 & 0 & 0 & 0.65 \end{bmatrix} \tag{10.29}$$

Solution to exercise 10.13 (p.173). The labelling problem can be solved for any $N > 2$ with just two trips, one each way across the Atlantic.

The key step in the information-theoretic approach to this problem is to write down the information content of one *partition*, the combinatorial object that is the connecting together of subsets of wires. If N wires are grouped together into g_1 subsets of size 1, g_2 subsets of size 2, ..., then the number of such partitions is

$$\Omega = \frac{N!}{\prod_r (r!)^{g_r} g_r!}, \tag{10.30}$$

and the information content of one such partition is the log of this quantity. In a greedy strategy we choose the first partition to maximize this information content.

One game we can play is to maximize this information content with respect to the quantities g_r, treated as real numbers, subject to the constraint $\sum_r g_r r = N$. Introducing a Lagrange multiplier λ for the constraint, the derivative is

$$\frac{\partial}{\partial g_r}\left(\log \Omega + \lambda \sum_r g_r r\right) = -\log r! - \log g_r + \lambda r, \tag{10.31}$$

which, when set to zero, leads to the rather nice expression

$$g_r = \frac{e^{\lambda r}}{r!}; \tag{10.32}$$

the optimal g_r is proportional to a Poisson distribution! We can solve for the Lagrange multiplier by plugging g_r into the constraint $\sum_r g_r r = N$, which gives the implicit equation

$$N = \mu\, e^{\mu}, \tag{10.33}$$

where $\mu \equiv e^{\lambda}$ is a convenient reparameterization of the Lagrange multiplier. Figure 10.10a shows a graph of $\mu(N)$; figure 10.10b shows the deduced non-integer assignments g_r when $\mu = 2.2$, and nearby integers $g_r = \{1, 2, 2, 1, 1\}$ that motivate setting the first partition to (a)(bc)(de)(fgh)(ijk)(lmno)(pqrst).

This partition produces a random partition at the other end, which has an information content of $\log \Omega = 40.4$ bits, which is a lot more than half the total information content we need to acquire to infer the transatlantic permutation, $\log 20! \simeq 61$ bits. [In contrast, if all the wires are joined together in pairs, the information content generated is only about 29 bits.] How to choose the second partition is left to the reader. A Shannonesque approach is appropriate, picking a random partition at the other end, using the same $\{g_r\}$; you need to ensure the two partitions are as unlike each other as possible.

If $N \neq 2$, 5 or 9, then the labelling problem has solutions that are particularly simple to implement, called Knowlton–Graham partitions: partition $\{1, \ldots, N\}$ into disjoint sets in two ways A and B, subject to the condition that at most one element appears both in an A set of cardinality j and in a B set of cardinality k, for each j and k (Graham, 1966; Graham and Knowlton, 1968).

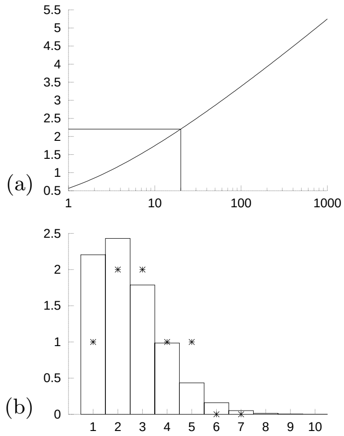

Figure 10.10. Approximate solution of the cable-labelling problem using Lagrange multipliers. (a) The parameter μ as a function of N; the value $\mu(20) = 2.2$ is highlighted. (b) Non-integer values of the function $g_r = {}^{\mu^r}\!/r!$ are shown by lines and integer values of g_r motivated by those non-integer values are shown by crosses.