Solutions

1. Compression / inference question

Mark allocation: **2** for using the Huffman algorithm correctly to construct a binary tree. **2** for correctly assigning binary strings to the tree branches for example:

a_i	p_i	$\log_2 \frac{1}{p_i}$	l_i	$c(a_i)$
a	0.09091	3.5	4	0000
b	0.09091	3.5	4	0001
c	0.09091	3.5	4	0100
d	0.09091	3.5	4	0101
e	0.09091	3.5	4	0110
f	0.09091	3.5	4	0111
g	0.09091	3.5	3	100
h	0.09091	3.5	3	101
i	0.09091	3.5	3	110
j	0.09091	3.5	3	111
k	0.09091	3.5	3	001

The entropy is $\log_2 11 = 3.4594$ [1]

and the expected length is $L = 3 \times \frac{5}{11} + 4 \times \frac{6}{11}$ which is $3\frac{6}{11} = 3.54545$. [1]

Length - entropy $= 0.086$. [1]

(b) Marks are earned for any of the following steps to a sensible answer. Not all steps are required.

1 mark for saying f_A roughly 1/3.

1 for giving fractional precision of this estimate, roughly same as the fractional precision of the predicted number of heads, $\sqrt{Npq}/Np = \sqrt{600 \times 2/3}/600 = 20/600 = 1/30$, or a graph showing that this fact was appreciated.

1 for saying the posterior probability of each effectiveness is likelihood function times prior, normalized.

1 for saying the posterior probability density of f_B is $\propto f_B$.

1 for a sketch of the posterior of f_B which looks like /|.

1 for marking the value 1/3 on this graph somehow.

1 for stating that the integral of $P(f_B)$ up to 1/3 is important.

1 for correct evaluation of (integral / total integral) $\simeq 1/9$. And for stating that $P(f_B > f_A) = 8/9$.

If the answer given is outrageously off compared with the correct ballpark (eg smaller than 33% or bigger than 99%) the maximum attainable mark will be one less than the original maximum.

(c) Marks up to the maximum will be given for intelligent discussion of the following issues. Marking scheme will be adjusted in the light of the candidates' answers.

(i) Squaring the probability assumes independence. But there could be natural causes common to both deaths, for example genetic or environmental causes. This makes the probability of two deaths bigger than the $\frac{1}{72{,}000{,}000}$ figure. Genetic conditions exist such that the probability of death is $\frac{1}{4}$. The probability of death by this route is [Up to 3 marks or 4 if flawless and detailed.]

$$P(\text{the parents have the relevant genes}) \times \frac{1}{4} \times \frac{1}{4}.$$

[2]

(ii) The fact that the data are unlikely is not enough to imply that she's guilty. In order to turn the probability of the data given a hypothesis into the probability of guilt given the data, Bayes's theorem should be used. This requires prior probabilities to be assigned to the hypotheses, and the probability of the data under the alternative hypotheses to be evaluated. [3]

(iii) She claimed the deaths were natural, not that they were "sudden infant death syndrome", which is a subset of natural deaths. So his statistical data are not the relevant ones to the case. [2]

(iv) Empirical data on the actual rate of double deaths in affluent families could be mentioned. For example, 'I know a couple who had a double death'. These data kick into touch the absurd $\frac{1}{72,000,000}$ figure, which would expect only one double death in an affluent family every few centuries. [2]

2. Question on channels and encoding

The mutual information is

We can evaluate $H(Y)$ by saying 'is it b; and if not, is it a or c?' –

$$H(Y) = H_2(p/3) + (1 - p/3).$$

[2]

$H(Y|X) = p \log_2 3$. [1]

So $I = H(Y) - H(Y|X)$. [1]

$$I = H_2(p/3) + (1 - p/3) - p \log_2 3.$$

Differentiate I. [1]

Doing it correctly [1]

$$\frac{\partial I}{\partial p} = \frac{1}{3} \log \frac{1 - p/3}{p/3} - \frac{1}{3} - \log 3.$$

Setting to zero and rearranging [1]

$$\log \frac{1 - p/3}{p/3} = 1 + 3 \log 3.$$

So [1]

$$p/3 = \frac{1}{1 + 2^{1+3 \log_2 3}}.$$

Incidentally, $2^{1+3 \log_2 3} = 54$ so

$$p/3 = \frac{1}{1 + 54} = \frac{1}{55}$$

and

$$p = 3/55 = 0.05454.$$

(b) Spies.

The capacity of the channel is exactly $\log_2 52!$ [1]

which is roughly $52 \log_2 52$ [1 for Stirling formula]

$\simeq 52 \times 6 \simeq 300$.

The exact answer is $\log_2 52! = 225.6$. [1 for a reasonable numerical answer. This one is

More accurate approximate answer is $52(\log_e 52 - 1)/\log_e 2 = 52 \times 4.26 = 221$.

The answer $52(\log_2 52 - 1) = 244$, while erroneous, will also be given full credit.

Practical method: ideal answer will describe *arithmetic coding.* The first card has distribution $\frac{1}{52}$ over all available cards. The second $\frac{1}{51}$ over the remaining cards, etc. Apart from termination details, which may lose 2 bits, this method will encode bits an order of the cards optimally. A suboptimal solution inspired by the arithmetic coding approach is

- Order the 52 cards in a standard way.
- Choose the first card from the leftmost 32 cards using the first 5 bits.
- Choose the second card from the leftmost 32 remaining using 5 bits.
- Choose the third card from the leftmost 32 remaining using 5 bits.
- ⋮
- Choose the twenty-first card from the 32 remaining using 5 bits.

At this point, $21 \times 5 = 105$ bits have been conveyed. Similarly, can convey $16 \times 4 = 64$ in the next 16 cards and $8 \times 3 = 24$ and $4 \times 2 = 8$ and $2 \times 1 = 2$ and 1×0 in the final card. Total is 105+64+24+8+2 = 203.

Suboptimal because not all permutations were realizable.

The noisy channel's capacity is $\log 51!$ which is less by 5.7 bits than the first channel. The answer $\log 52! - \log 51$ is also fine.

Any viable scheme for 52 cards can be preserved by using one card as a placeholder and converting to an analogous 51-card scheme for communicating information.

3. Essay question

Expected topics:

K means clustering, soft K means clustering. Definition of the algorithms. Their interpretation as maximum likelihood mixture density modelling, using a mixture of Gaussians. Problems with soft K means. Maximum likelihood can blow up.

Labelled data: make a Bayesian model of each cluster and do classification using Bayes theorem. Or make a parameterized classification model (single neuron, for example) and train it on the data. Objective function for single neuron training. Motivation for single neuron. Overfitting problem. Regularization. Backpropagation algorithm for more general classification models.

Monte carlo approach to clustering. Mixture of gaussians fitted by Gibbs sampling.