p9 q10
JGD

Model Answer: Information Theory and Coding, Question 2.

2001

1.

Example of one such code (there are others as well):

Letter	*Code*
A	1
B	01
C	001
D	0000
E	0001

This is a uniquely decodable code, and it also has the prefix property that no symbol's code is the beginning of a code for a different symbol.

[4 marks]

The shortest possible average code length per symbol is equal to the entropy of the distribution of symbols, according to Shannon's Source Coding Theorem. The entropy of this symbol alphabet is:

$$H = -\sum_i p_i \log_2(p_i) = 1/2 + 2/4 + 3/8 + 4/16 + 4/16 = 1(7/8)$$

bits, and the average code length per symbol for the above prefix code is also (just weighing the length in bits of each of the above letter codes, by their associated probabilities of appearance): $1/2 + 2/4 + 3/8 + 4/16 + 4/16 = 1(7/8)$ bits. Thus no code can be more efficient than the above code.

[2 marks]

2.

For a string of data of length N bits, the upper bound on its Minimal Description Length is N. The reason is that this would correspond to the worst case in which the shortest program that can generate the data is one that simply lists the string itself.

[2 marks]

It is often impossible to know whether one has truly found the shortest possible description of a string of data. For example, the string:
`0110101000001001111001100110011111110011 10...`
passes most tests for randomness and reveals no simple rule which generates it, but it turns out to be simply the binary expansion for the irrational number $\sqrt{2} - 1$.

[2 marks]

3.

The bandlimiting constraint (either just a highest frequency component in the case of Nyquist sampling, or the bandwidth limitation to one octave in the case of Logan's Theorem), is remarkably severe. It ensures that the signal cannot vary unsmoothly between the sample points (i.e. it must be everywhere a linear combination of shifted sinc functions in the Nyquist case), and it cannot remain away from zero for very long in Logan's case. Doing so would violate the stated frequency bandwidth constraint.

[3 marks]

4.

The autocorrelation integral for a (real-valued) signal $f(x)$ is:

$$g(x) = \int f(y)f(x+y)dy$$

i.e. $f(x)$ is multiplied by a shifted copy of itself, and this product integrated, to generate a new signal as a function of the amount of the shift.

Signals differ from noise by tending to have some coherent, or oscillatory, component whose phase varies regularly; but noise tends to be incoherent, with randomly changing phase. The autocorrelation integral shifts the coherent component systematically from being in-phase with itself to being out-of-phase with itself. But this self-reinforcement does not happen for the noise, because of its randomly changing phase. Therefore the noise tends to cancel out, leaving the signal clean and reinforced. The process works best for purely coherently signals (sinusoids) buried in completely incoherent noise. Sinusoids would be perfectly extracted from the noise.

[5 marks]

Autocorrelation as a noise removal strategy depends on the noise being just added to the signal. It would not work at all for multiplicative noise.

[2 marks]