Work for Information Theory Supervision III

Questions are drawn from a variety of sources: past Tripos questions, MacKay, and some of my own.

Please submit work using the KuDoS Work Submission feature, at least 24 hours before the supervision.

Questions

1. Consider two discrete probability distributions $p(x)$ and $q(x)$ over the same set of four values $\{x\}$ of a random variable:

$p(x)$	1/8	1/8	1/4	1/2
$q(x)$	1/4	1/4	1/4	1/4

   (a) Calculate the cross-entropy $H(p,q)$ (unfortunate notation!) between $p(x)$ and $q(x)$.

   (b) Calculate their Kullback-Leibler divergence $D_{\mathrm{KL}}(p||q)$.

   (c) Comment on the use of the metrics $H(p,q)$ and $D_{\mathrm{KL}}(p||q)$ in machine learning and for calculating the efficiency of codes.

2. Continuous random variables X and Y both have uniform probability density distributions on some interval. For X, $p(x) = 1/2$ if $x \in [0,2]$ otherwise 0, while for Y, $p(y) = 1/8$ if $y \in [0,8]$ otherwise 0. Calculate the differential entropies $h(X)$ and $h(Y)$.

3. (a) Find the optimal input distribution for the Binary Symmetric Channel and show that its capacity is given by $C = 1 - H_2(f)$.

   (b) Consider the Z channel:

$$
\begin{aligned}
P(y=0|x=0) &= 1 \\
P(y=1|x=0) &= 0 \\
P(y=0|x=1) &= f \\
P(y=1|x=1) &= 1-f
\end{aligned}
$$

   Show that the mutual information is given by (p_1 is $P(X=1)$ for source X):

$$I(X;\,Y) = H_2(p_1(1-f)) - p_1 H_2(f)$$

   Hence show that the optimal input distribution is given by

$$p_1^\star = \frac{1/(1-f)}{1 + 2^{H_2(f)/(1-f)}}$$

   (You may use the standard result that $\mathrm{d}/\mathrm{d}f\ H_2(f) = -\log_2(f/(1-f))$.)

   (c) Plot a graph of $p_1^\star$ and investigate what happens as the noise level f becomes close to 1. Observe that $p_1^\star < 0.5$, for all values of f.

   (d) Sketch graphs of the capacity of the binary symmetric channel and the Z channel as a function of f.

4. A continuous communication channel adds Gaussian white noise to signals transmitted through it. The ratio of signal power to noise power is 30 decibels, and the frequency bandwidth of this channel is 10 MHz. Roughly what is the information capacity C of this channel, in bits/second?

5. (a) In lectures, you have seen correlation. A related concept, the autocorrelation of a signal (correlation of a signal with itself after some time lag) is defined as

$$R_{yy}(\ell) = \sum_{n \in Z} y(n)\overline{y(n-\ell)}$$

The following CSV file contains a sine wave ($\sin(2\pi f_0 t_i)$) embedded in additive white Gaussian Noise: `http://ireland.cx/teaching/2122/infoth/autocorrelation.txt`
The sampling frequency is 2 MHz.
Plot the autocorrelation of the signal in the CSV file ($R_{yy}(\ell)$ against ℓ). Hence find the frequency f_0 of the sine wave.

(b) Research and explain how GPS works at the physical and data-link layers. Explain how GPS is implements the spreading of the spectrum of the transmitted signal over a very large bandwidth, and how the wide-bandwidth signal is decoded in a GPS receiver.

6. Explain the concept of entropy in:

(a) Classical thermodynamics ("entropy", developed by Clausius and others)

(b) Statistical mechanics ("statistical entropy", developed by Boltzmann and others – presented in Lecture 12)

(c) Information theory ("information entropy", developed by Shannon)

To what extent do these different notions of entropy relate to each other?