Sample paper 2008: solution to q3.

3 (a) (4 marks) What is meant by a Markov chain $X \to Y \to Z$ and

(b) (3 marks) Show that $X \to Y \to Z$ implies $Z \to Y \to X$.

(c) (8 marks) State and prove the data processing inequality.

(d) (5 marks) Suppose that a Markov chain starts in one of n states, necks down to $k < n$ states and then fans back out to $m > k$ states. Show that the dependence of the first and last variables, X and Z is limited by the bottleneck by showing $I(X, Z) \le \log k$.

*Solution:*So this question is very much book work. Random variable X, Y and Z are said to *form a Markov chain in that order* $X \to Y \to Z$ if the conditional distribution of Z depends only on Y and is conditionally independent of X:

$$p(z|x, y) = p(z|y) \tag{1}$$

for all x, y and z in their respective sets of outcomes. Now

$$p(x, z|y) = \frac{p(x, y, z)}{p(y)} = \frac{p(x, y)p(z|y)}{p(y)} = p(x|y)p(z|y) \tag{2}$$

or, the Markov condition is equivalent to X and Z being conditionally independent given Y, this condition is symmetric in X and Z so

$$X \to Y \to Z \Rightarrow Z \to Y \to X \tag{3}$$

The data processing inequality is Theorem 2.8.1: $X \to Y \to Z$ implies $I(X;Y) \ge I(X;Z)$, it is in the book but is actually pretty simple; basically you expand using the chain rule

$$I(X;Y,Z) = I(X;Z) + I(X;Y|Z) = I(X;Y) + I(X;Z|Y) \tag{4}$$

and then use $I(X;Z|Y) = 0$ which follow from the conditional independence of X and Z, using $I(X;Y|Z) \ge 0$ gives you the proof. Finally, we did the bottleneck before as a problem sheet:

$$I(X, Z) \le I(X;Y) = H(Y) - H(Y|X) \le H(Y) \tag{5}$$

and $H(Y) \le \log k$ because that is the upperbound on the entropy of a variable with k states.