Ex. 1
If X and Y are discrete random variables taking only a finite number of values, show that

$$H(X+Y \mid X) = H(Y \mid X).$$

Also show that

$$H(g(X,Y) \mid X) = H(Y \mid X)$$

does not hold generally for $g : \mathbb{R}^2 \to \mathbb{R}$.

Ex. 2
A statistical survey of married couples shows that 70% of men have dark hair, that 25% of girls are blonde, and that 80% of blonde girls marry dark-haired men. How much information about the colour of a man's hair is conveyed by the colour of his wife's hair?

Ex. 3
A random variable X has the binomial distribution with parameters n and p. That is, for $0 \le k \le n$

$$P(X = k) = \binom{n}{k} p^k q^{n-k},$$

where $0 < p < 1$ and $q = 1 - p$. Show that

$$H(X) \le -n(p \log p + q \log q).$$

Ex. 4
Show that, for any positive integer m there exists an instantaneous code over $\Sigma = \{0, 1\}$ that has words of all lengths in the set $\{1, \dots, m\}$.

Ex. 5
What is the maximum number of words in a binary instantaneous code in which the maximum word length is 7?

Ex. 6 (Not to be handed in.)
Work through the detailed example in section 1.6.

Ex. 7 (Not to be handed in.)
Welsh's argument on McMillan's inequality is a bit short. Read it carefully and try to understand the details. Below is the argument (almost a word by word quotation).

Suppose we have a uniquely decipherable code C with word lengths $l_1, \dots, l_N$. If $l = \max_i l_i$, then, for any positive integer r, we have

$$(D^{-l_1} + \ldots + D^{-l_N})^r = \sum_{i=1}^{rl} b_i D^{-i},$$

where b_i is a non-negative integer.
Here b_i counts the number of ways in which a string of length i of symbols from the alphabet Σ can be made up by stringing together r words of lengths chosen from the

set $\{l_1, \ldots, l_N\}$. (This is a partition problem, also studied in the theory of generating functions. Construct some small examples to see what happens here!)
But if the code C is uniquely decipherable, it must be the case that any string of length i formed from codewords must correspond to at most one sequence of code words. Hence we must have

$$b_i \leq D^i, \quad (1 \leq i \leq rl).$$

Hence we obtain

$$(D^{-l_1} + \ldots + D^{-l_N})^r \leq rl.$$

Therefore

$$\sum_{k=1}^{N} D^{-l_k} \leq l^{\frac{1}{r}} r^{\frac{1}{r}},$$

for all positive integers r. We can let $r \to \infty$, so that $l^{\frac{1}{r}} r^{\frac{1}{r}} \to 1$, and McMillan's inequality follows.

Remark:
At the end of the year you will have to write a Thesis. Experience from last years show that students learn the mathematical word processing language LaTeX just in the final steps of their Thesis.
You are encouraged to learn this during the year. good exercise would be to type some of your weekly homework, and to learn LaTeX in small chunks. Essentially one can learn LaTeX by looking at a brief manual, looking just at some examples. LaTeX is installed on the computers in the Lab 357. It's free software. So you can also dowload the package from the web and use on your computer.
A very brief guide: ftp://ftp.ams.org/pub/tex/doc/amsmath/short-math-guide.pdf

To be returned in one week, before the lecture.
My web page contains a collection of related material.
http://www.ma.rhul.ac.uk/∼elsholtz/WWW/lectures/0506mt441/lecture.html