MDL exercises, fourth handout

(due March 17th)

1. a) Let $H(p) = -p \log p - (1-p)\log(1-p)$ denote the binary entropy of a Bernoulli[p] distribution when the probability of observing a zero is p. (The logarithm base is two.) Use Stirling's approximation $\ln(n!) = (n+\frac{1}{2})\ln n - n + \frac{1}{2}\ln 2\pi + O(1/n)$ to show that $\log\binom{n}{\gamma n} = nH(\gamma) - \frac{1}{2}\log n + O(1)$.

   b) More generally, consider a sample space $\mathcal{X} = \{1, \ldots, k\}$ and probability mass functions p on $\mathcal{X}$, given in the form of a vector $p = (p_1, \ldots, p_k)$. Let $H(p) = \sum_{i=1}^{k} -p_i \log p_i$ denote the binary entropy of the distribution with mass function p. Use Stirling's approximation to express $\log\binom{n}{p_1 n \; \ldots \; p_k n} = n!/((p_1 n)! \ldots (p_k n)!)$ up to an $O(1)$ term.

2. Consider two codes for coding sequences of 0s and 1s. One is the Bayesian code with lengths $-\log P_M(x^n)$, where P_M is the Bayesian probability based on a uniform prior over the Bernoulli model. The other is the two-stage code where you first code the number of 1s n_1 in x^n using a uniform code, and then you code the actual sequence with that number of 1's, using again a uniform code over all sequences of length n with n_1 1s.

   Which code is better and why?

3. Markov Chains.

   a) Compute the maximum likelihood estimator $\hat{\theta} = (p_{0\to 1}, p_{1\to 1})$ for a binary first order Markov chain.

   b) Draw X_1, X_2, X_3 from an order 1 Markov chain. Are X_1 and X_3 dependent? What if you know the value of X_2?