Diskrete Stochastik und Informationstheorie – 18 Jun 2014

Exercise 41 (Shannon code with respect to wrong distribution)**.** Suppose that we encode X using a Shannon code with lengths $\ell(C(x)) = \lceil -\log q(x) \rceil$ assuming the distribution q of X. Show that if the real distribution of X is p, we have:

$$H(p) + D(p \parallel q) \leq \mathbb{E}_p(\ell(C(X))) < H(p) + D(p \parallel q) + 1\,.$$

Exercise 42 (Relative entropy is the cost of miscoding)**.** Let the random variable X take values in $\{1, 2, 3, 4, 5\}$. Consider two possible distributions p and q of X, as well as two binary encodings C_1 and C_2:

x	$p(x)$	$q(x)$	$C_1(x)$	$C_2(x)$
1	$1/2$	$1/2$	0	0
2	$1/4$	$1/8$	10	100
3	$1/8$	$1/8$	110	101
4	$1/16$	$1/8$	1110	110
5	$1/16$	$1/8$	1111	111

(a) Calculate $H(p)$ and show that C_1 is optimal for p. Same for q and C_2.

(b) Calculate $D(p \parallel q)$ and quantify the loss when encoding with C_2 under p. Same for $D(q \parallel p)$ and encoding with C_1 under q.

Exercise 43 (Optimal compression of a Markov source)**.** Consider the homogeneous Markov chain $X = (X_n)_{n\in\mathbb{N}}$ on the state space $S = \{s_1, s_2, s_3\}$ with transition matrix

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{pmatrix}.$$

(a) Design three binary encodings C_1, C_2 and C_3 of S such that the Markov process X can be sent with maximal compression when for each state, the code is selected depending on the previous state (code C_i if the previous state was s_i).

(b) What is the average message length of the next symbol conditioned on the current symbol $X_n = s_i$ for $i = 1, 2, 3$?

(c) What is the unconditional average number of bits per source symbol? Relate this to the entropy of the Markov chain.

Exercise 44 (Typical sets for Markov chains)**.** A discrete source emits a sequence (X_n) of bits that follow a stationary Markov chain with the transition matrix

$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix}.$$

(a) Find the entropy rate of X_n.

(b) How many “1”s will you typically observe in a sequence of 100 digits?

(c) Give bounds on the size of the typical sets $A_\varepsilon^{(n)}$.

(d) Use Markov's inequality to bound the probability of observing a source sequence that contains more than k “1”s. Use this bound to find the value k such that the probability of observing more than k times a 1 is less than $\varepsilon = 0.001$.