

Homework January 20, 2015 (to be returned on January 27)

1) A source produces a character  $x$  that is chosen randomly from the alphabet  $\mathcal{A} = \{0, 1, \dots, 9, a, b, \dots, z\}$ ; with probability  $1/3$ ,  $x$  is a numeral  $(0, \dots, 9)$ ; with probability  $1/3$ ,  $x$  is a vowel  $(a, e, i, o, u)$ ; and with probability  $1/3$  it is one of the 21 consonants. All numerals are equiprobable, and the same goes for vowels and consonants. Estimate the entropy of  $X$  regrouping first the events in the three groups of numerals, vowels and consonants and then calculating the entropy within the groups.

2) A random variable  $x \in \{0, 1, 2, 3\}$  is selected by flipping a bent coin with bias  $f$  to determine whether the outcome is in  $\{0, 1\}$  or  $\{2, 3\}$ ; then either flipping a second bent coin with bias  $g$  or a third bent coin with bias  $h$ , respectively. Write down the probability distribution of  $X$ . Use also the regrouping property of the Shannon entropy to find the entropy of  $X$ .

3) An unbiased coin is flipped until one head is thrown. What is the entropy of the random variable  $x \in \{1, 2, 3, \dots\}$ , the number of flips? Repeat the calculation for the case of a biased coin with probability  $f$  of coming up heads. Solve the problem both directly and by using the regrouping property of the Shannon entropy.

4) The frequency  $p_n$  of the  $n$ -th most frequent word in English is roughly approximated by  $p_n = 0.1/n$  for  $n \leq 12367$  and zero otherwise. This remarkable  $1/n$  law is known as Zipf's law, and approximately applies to the word frequencies of many languages (Zipf, 1949). If we assume that English is generated by picking words at random according to this distribution, what is the entropy of English (per word)? [This calculation can be found in Prediction and entropy of printed English, C.E. Shannon, Bell Syst. Tech. J. 30, pp.5064 (1950), though some numerical errors are present so copying is not recommended!]

5) You are given 16 balls, all of which are equal in weight except for one that is either heavier or lighter. You are also given a bizarre two-pan balance that can report only two outcomes: the two sides balance or the two sides do not balance. Design a strategy to determine which is the odd ball in as few uses of the balance as possible.