Exercise Sheet #12

Hendrik Wernecke <wernecke@th.physik.uni-frankfurt.de>
Philip Trapp <trapp@th.physik.uni-frankfurt.de>

Problem 1 (*Kullback-Leibler Divergence*) 4 Pts

Calculate the Kullback-Leibler divergence of the two exponential distributions p, q given by,

$$p(x) \propto \mathrm{e}^{-\lambda_1 x}, \qquad q(x) \propto \mathrm{e}^{-\lambda_2 x},$$

with parameters $\lambda_1, \lambda_2 > 0$. When is the distance minimal?

Problem 2 (*Tsallis entropy*) 8 Pts

(a) Show that

$$\log x = \lim_{c \to 0} \frac{x^c - 1}{c}, \qquad x, c \in \mathbb{R}, \quad x \geq 0.$$

(b) Using the previous expression generalise the Shannon entropy $H[p]$ to get the Tsallis entropy:

$$H_q[p] = \frac{1}{1-q} \sum_k [(p_k)^q - p_k], \qquad 0 < q \leq 1,$$

i. e. show that

$$\lim_{q \to 1} H_q[p] = H[p], \qquad H_q[p] \geq 0.$$

(c) Prove that the Tsallis entropy is a non-extensive quantity:

$$H_q[p] = H_q[p_X] + H_q[p_Y] + (1-q) H_q[p_X] H_q[p_Y], \qquad p = p_X p_Y,$$

where X and Y are two statistically independent systems.

Problem 3 (*Source Coding Theorem*) 8 Pts

(a) Compute the entropy of the alphabet in the English language.

(b) Consider the old telephone writing system (see figure 1), in which the alphabet A-Z is encoded onto sequences of the numbers 2 to 9. Example: A is encoded as 2, B is encoded as 22, C is encoded as 222, D is encoded as 3, and so forth. Compute the average number of bits per letter.

(c) How can you improve the telephone writing system? Justify.

Hint: In the English language, the frequency of letters is:

a= 8.167%	b= 1.492%	c= 2.782%	
d= 4.253%	e=12.702%	f= 2.228%	
g= 2.015%	h= 6.094%	i= 6.966%	
j= 0.153%	k= 0.772%	l= 4.025%	
m= 2.406%	n= 6.749%	o= 7.507%	
p= 1.929%	q= 0.095%	r= 5.987%	s= 6.327%
t= 9.056%	u= 2.758%	v= 0.978%	
w= 2.360%	x= 0.150%	y= 1.974%	z= 0.074%

Figure 1: A typical keyboard of a phone.