Spectral Graph Theory (Summer 2015)
Assignment 2

Instructor: Masoud Khalkhali
Mathematics Department, University of Western Ontario
London, ON, Canada

1. (*Information loss and data processing inequality*) Let $f : X \to Y$ be a map between finite sets, and P a probability distribution on X. Let f_*P be the corresponding *pushforward* probability distribution on Y. Show that
$$H(f_*P) \leq H(P).$$
(Hint: Use Jensen's inequality. H denotes the entropy, and f_*P is defines by
$$f_*P(y) = \sum_{f(x)=y} P(x).$$
The quantity $IL(f) := H(P) - H(f_*P)$ is the information loss under a deterministic map f.)

2. (*Additivity of entropy*) Let $A = \{A_i\}_{1 \leq i \leq n}$ be a disjoint family of events such that $\mathbb{P}(A_i) = p_i$ with $\sum_{i=1}^{n} p_i = 1$. Take another, possibly dependent, sequence of disjoint events $B = \{B_j\}_{1 \leq j \leq m}$ such that $\mathbb{P}(B_j) = q_j$ with $\sum_{j=1}^{n} q_j = 1$. We define
$$q_{kl} = \mathbb{P}(B_l \text{ occurs given that } A_k \text{ occured}).$$
Define, respectively, the entropy of A, B and the conditional entropy

of B with respect to A_k as

$$H(A) := -\sum_{i=1}^{n} p_i \log p_i$$
$$H(B) := -\sum_{j=1}^{m} q_j \log q_j$$
$$H_k(B) := -\sum_{l=1}^{m} q_{kl} \log q_{kl}.$$

Observe that now we have another disjoint family $AB := \{A_k \cap B_l\}_{1\le k\le n, 1\le l\le m}$. Show that

$$H(AB) = H(A) + \sum_{k=1}^{n} p_k H_k(B).$$

What the formula boils down to in the case A_i and B_j are independent i.e. $\mathbb{P}(A_i \cap B_j) = p_i q_j$.

3. Using the notation of the previous question, we define the conditional entropy of B with respect to A as $H_A(B) := \sum_{k=1}^{n} p_k H_k(B)$. Show that $H_A(B) \le H(B)$ i.e. the knowledge of "A occurred" can only decrease the uncertainty in B.

   (Hint: use Jensen's inequality.)

4. The entropy of a probability distribution function $f(x)$ on $\mathbb{R}$ is defined by

$$H(f) = -\int_{-\infty}^{\infty} f(x) \log f(x) dx.$$

   ($\log = \log_e$). Compute the entropy of the normal distribution (with mean μ and variance σ^2)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

   Show that the normal distribution has the maximun entropy among all distributions with mean μ and variance σ^2 on $\mathbb{R}$.

5. (*Maximun entropy principle*) Let X be a finite set. Gven a real number μ and a non-negative number σ^2, show that there is a unique probability

distribution with mean μ, variance σ^2, and with maximum entropy. Compute this maximum entropy. (Hint: Use Lagrange multipliers).