MDL exercises, tenth handout (final obligatory homework exercises)

(due May 21st)

1. Let $f(x)$ be a density function on $[0, \infty)$ with fixed mean $1/\lambda$. Define $g(x) = \lambda e^{-\lambda x}$, the density function of the exponential distribution on the same domain and with the same mean. Show that $H(f)$ is maximized by choosing $f = g$, by evaluating $0 \leq D(f \| g)$.

2. Jensen's inequality states that $E[f(X)] \geq f(E[X])$ for conve^x f. Use this inequality to find (a) a lower bound on the entropy $H(P)$ for a distribution P on a finite sample space, and (b) an upper bound on this entropy (Hint: for the upper bound, rewrite the entropy as $-\sum_x P(x)(f(1/P(x)))$ with $f \equiv -\log$). Compare this upper bound to the entropy for the uniform distribution on that sample space and for a nonuniform distribution on that space. In which case is the bound tighter?

3. For two distributions P_0 and P_1 defined on the same space $\mathcal{X}$ with $P_0 \neq P_1$, let P_α be the α- mixture between P_0 and P_1, i.e. $P_\alpha(x) = (1-\alpha)P_0(x) + \alpha P_1(x)$. Show that the entropy $H(P_\alpha)$ is strictly concave as a function of $\alpha \in [0, 1]$.

4. Consider the following three families of distributions. For each of these models, prove that they are an exponential family. HINT: you can show that a family is an exponential family by rewriting it in the exponential form $\frac{1}{Z(\beta)} e^{\beta\phi(x)} r(x)$ for some function $\phi(x)$.

   a) The set of all distributions on $\{0, 1\}$ with mean $E[X] = \theta$, for all $0 \leq \theta \leq 1$ (How is this set of distributions called?).

   b) The set of all normal distributions with a variance of one, for all means $\mu \in \mathbb{R}$.

   c) The set of power law distributions, also known as the *Pareto family*: $P_\theta(n) = n^{-\theta} / \sum_{n=1}^{\infty} n^{-\theta}$ for $n \in \{1, 2, \ldots\}$ and $\theta > 1$.

5. This question refers back to questions 4(a)-4(b).

   a) Is the distribution corresponding to θ in question 4(a) the maximum entropy distribution among *all* distributions on $\{0, 1\}$ with mean $E[X] = \theta$? Why (not)?

   b) Is the distribution corresponding to mean μ in question 4(b) the maximum entropy distribution among *all* distributions on $\mathbb{R}$ with mean μ? Why (not)?