# MDL exercises, eighth handout

(due April 25th)

1. The prequential plug-in code based on the ML estimator is a universal code. Below we only consider one-dimensional models such that the data are i.i.d. according to all distributions in the model. For such models, the ML estimator is usually defined after 1 observation. Thus the prequential codelength for the second outcome given the first is defined, but the prequential outcome for the first outcome is not defined. Below you may ignore the contribution of that outcome to any codelengths you compute. As last week, just pretend that the contribution of the first outcome is zero. Now we consider possible extensions ($> 1$ outcome) of a sequence of one outcome, restricting ourselves to *i.i.d.* models. Prove or give counterexamples to the following claims.

   a) There is no model for which the prequential ML codelength depends on the order of the extension.

   b) The prequential ML codelength depends on the order of the extension for all 1-dimensional models.

   c) There is no model for which the Bayesian marginal codelength depends on the order of the extension.

   d) The Bayesian marginal codelength depends on the order of the extension for all models.

2. In the lecture, a *source* was defined as a probability distribution on sequences of infinite length. But to make this formally precise requires measure theory, which we want to avoid here. An easier formal definition defines a source to be a sequence of probability distributions, $P^1, P^2, \ldots$: one for each sample size $1, 2, \ldots$. These distributions must satisfy the property of *compatibility*: for all $n$, it must be the case that $P^n(x^n) = \sum_{y \in \Sigma} P^{n+1}(x^n y)$.

   a) Show that a sequence of probability distributions is compatible if and only if we have $P^{n+1}(x_{n+1} \mid x^n) = P^{n+1}(x^{n+1})/P^n(x^n)$.

   b) Why is it not necessary to require compatibility if we directly define a source as a distribution over infinitely many outcomes?

   c) Suppose that we have a parametric source $P^1(x^1 \mid \theta), \ldots$. Show that, using any prior on $\theta$, the sequence of Bayesian marginal distributions is also a source.

   d) Show, by giving an explicit example, that, at least for some i.i.d. models $\{P_\theta \mid \theta \in \Theta\}$ for which NML is well-defined, that the sequence of

NML distributions $P_{\text{nml}}^{(1)}, P_{\text{nml}}^{(2)}, \ldots$ is not a source (hint: you already implicitly did this in a previous homework exercise!)

3. Consider the family of exponential distributions: for $x > 0$, the density of $x$ according to the distribution with parameter $\lambda > 0$ is given by $\lambda e^{-\lambda x}$. We take a uniform prior density over the parameter space. While this prior is improper, the corresponding posterior after observing one or more observations is proper. Calculate the posterior of $\lambda$ after observing a sequence $x_1, x_2, \ldots, x_n$.