MDL exercises, eighth handout

(due April 20, 14:15)

1. A *probabilistic source* may be 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

   $$P^{n+1}(x_{n+1} \mid x^n) = P^{n+1}(x^{n+1}) / P^n(x^n). \tag{1}$$

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

   e) 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 λ after observing a sequence $x_1, x_2, \ldots, x_n$.