MDL exercises, ninth handout

(due April 27th, 14:00)

Consider MDL model selection between

$$\mathcal{M}_0 = \{P_{0,\sigma} : \sigma > 0\} \text{ and } \mathcal{M}_1 = \{P_{\delta,\sigma} : \sigma > 0, \delta \in \mathbb{R}\}$$

where $P_{\delta,\sigma}$ is the distribution under which $X_1, X_2, \ldots, X_n$ are i.i.d., each with density given by

$$p_{\delta,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x}{\sigma}-\delta\right)^2}.$$

1. Show that $\mathcal{M}_1$ is identical to the family of normal distributions with mean in $\mathbb{R}$ and variance in $\sigma^2 > 0$. That is, if $Q_{\mu,\sigma}$ represents a normal distribution with mean μ and variance σ, show that (i) for every $\sigma > 0, \delta \in \mathbb{R}$, there is a $\mu \in \mathbb{R}$ such that $P_{\delta,\sigma} = Q_{\mu,\sigma}$ and (ii), conversely, for every $\sigma > 0, \mu \in \mathbb{R}$, there is a $\delta \in \mathbb{R}$ such that $P_{\delta,\sigma} = Q_{\mu,\sigma}$.

We associate Bayesian universal measures $\bar{p}_0$ with $\mathcal{M}_0$ and $\bar{p}_1$ with $\mathcal{M}_1$. In both cases, we put the *right Haar prior* $\pi(\sigma) = 1/\sigma$ on the variance σ. For $\bar{p}_1$, we equip δ with some (arbitrary) proper prior density w. Thus, we measure the evidence against $\mathcal{M}_0$ by

$$M(x^n) := \log \frac{\bar{p}_1(x^n)}{\bar{p}_0(x^n)} \tag{1}$$

with $\bar{p}_0(x^n) = \int \sigma^{-1} p_{0,\sigma}(x^n) d\sigma$ and $\bar{p}_1(x^n) = \int_{\sigma>0,\delta\in\mathbb{R}} \sigma^{-1} w(\delta) p_{\delta,\sigma}(x^n) d\sigma d\delta$.

2. Show that $\pi(\sigma) = 1/\sigma$ is improper.

3. (i) Show that $M(x^n)$ is *scale-invariant*. That is, show that for every sequence $x_1, \ldots, x_n$, every $c > 0$,

$$M(x_1, \ldots, x_n) = M(x_1/c, \ldots, x_n/c) \tag{2}$$

   (HINT: re-express the integral over σ in $\bar{p}_0$ and $\bar{p}_1$ as an integral over $\sigma' = c\sigma$).

   (ii) Define $Z^n = (X_1/|X_1|, X_2/|X_1|, \ldots, X_n/|X_1|)$. Use (2) to show that, for arbitrary $X_1 \neq 0$, $X_2, \ldots, X_n$,

$$M(X_1, \ldots, X_n) = M(Z_1, \ldots, Z_n).$$

4. Fix $\sigma > 0$. Let $X_1, X_2, \ldots, X_n \sim$ i.i.d. $P_{\delta,\sigma}$. Let $X'_i = X_i/\sigma$. (i) Show that, for all $\delta \in \mathbb{R}$, the distribution of $X'_1, \ldots, X'_n$ is now i.i.d. $N(\delta, 1)$. (ii) Use (i) to show that, for each fixed δ, the distribution of Z^n is the same under $P_{\delta,\sigma}$, for all $\sigma > 0$ [for question 5. see back side!].

As a consequence of (4)(ii), we can refer to the distribution P'_δ on Z^n without specifying the variance (the distribution does not depend on the variance of the X^n). Let p'_δ be the density of P'_δ. We can now write

$$M(X_1, \ldots, X_n) = \frac{\int_\delta w(\delta) p'_\delta(Z_1, \ldots, Z_n) d\delta}{p'_0(Z_1, \ldots, Z_n)}$$

where Z^n corresponds to X^n as above.

5. Explain the following statement: even though the Bayesian universal measures in (1) are based on improper priors, and therefore do not really define probability distributions, $-\log \bar{p}_0(X^n) - [-\log \bar{p}_1(X^n)]$ *can* be interpreted as a real codelength difference between two codes.