Lecture Summary 17

The predictive density for a Bayesian model $K(y|\theta)$ and posterior $f(\theta|\text{data})$ is defined as

$$k_p(y) = \int k(y|\theta)\, f(\theta|\text{data})\, d\theta.$$

It is the estimate of the density generating the data.

In some cases, such as a simple exponential model, the predictive can be calculated exactly. For example, if

$$k(y|\theta) = \theta\, e^{-y\theta} \quad \text{and} \quad f(\theta) = \text{Ga}(a, b)$$

then

$$f(\theta|\text{data}) = \text{Ga}(a + n, b + n\bar{y}).$$

Hence

$$\begin{aligned} k_p(y) &= \int \theta\, e^{-y\theta}\, \tfrac{(b+n\bar{y})^{a+n}}{\Gamma(a+n)}\, y^{a+n-1} e^{-y(b+n\bar{y})}\, dy \\ &= (a+n)\,(b+n\bar{y})^{a+n}(b+n\bar{y}+y)^{-(a+n+1)}. \end{aligned}$$

If it is easy to compute $k(y|\theta)$ and $(\theta^{(m)}$ for $m = 1, \ldots, M$ are the output of

a Markov chain sampling the posterior, then we can estimate k_p via Monte Carlo methods:

$$\hat{k}_p(y) = \frac{1}{M} \sum_{m=1}^{M} k(y|\theta^{(m)}.$$

If, on the other hand, we can not compute $k(y|\theta)$, as in the infinite mixture model, we can instead sample from the predictive; i.e., for each $\theta^{(m)}$ we take $y^{(m)}$ from $k(y|\theta^{(m)})$. This gives us

$$y^{(1)}, \ldots, y^{(M)}$$

which can be used to estimate k_p.

The two approaches, when both can be done, should match with each other.