Nested Sampling and the Evaluation of the 'Evidence' for Bayesian Model Selection

Paul D. Baines, Nicholas Ulle
University of California, Davis

February 27, 2014

1 Introduction

- Explanation of Nested Sampling
- Intuitive explanation of the algorithm
- Basics of computing the evidence and Bayesian model selection

Example I

Let:

$$Y_i \sim N(\mu, \sigma^2), \qquad i = 1, \ldots, n,$$

with prior $p(\mu) \propto N(\mu_0, \tau_0^2)$ and σ^2 known. Letting $C = (2\pi)^{-(n+1)/2}(\tau_0^2)^{-1/2}(\sigma^2)^{-n/2}$, the evidence, or marginal likelihood, is:

$$\begin{aligned}
p(y) &= \int p(y_1, \ldots, y_n|\mu)p(\mu)d\mu = \int p(\mu)\prod_{i=1}^{n} p(y_i|\mu)d\mu \\
&= \int C \times \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2 - \frac{1}{2\tau_0^2}(\mu-\mu_0)^2\right\} d\mu \\
&= C \times \int \exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2}+\frac{1}{\tau_0^2}\right)\mu^2 + \frac{1}{2}\mu\left(\frac{\mu_0}{\tau_0^2}+\frac{\sum_{i=1}^n y_i}{\sigma^2}\right) - \frac{1}{2}\left(\frac{\mu_0^2}{\tau_0^2}+\frac{\sum_{i=1}^n y_i^2}{\sigma^2}\right)\right\} d\mu \\
&= C \times \exp\left\{-\frac{1}{2}\left(\frac{\mu_0^2}{\tau_0^2}+\frac{\sum_{i=1}^n y_i^2}{\sigma^2}\right) + \frac{1}{2}\left(\frac{n}{\sigma^2}+\frac{1}{\tau_0^2}\right)^{-1}\left[\frac{\mu_0}{\tau_0^2}+\frac{\sum_{i=1}^n y_i}{\sigma^2}\right]^2\right\} \\
&\times \int \exp\left\{-\frac{1}{2}\left(\frac{n}{\sigma^2}+\frac{1}{\tau_0^2}\right)\left(\mu - \left(\frac{n}{\sigma^2}+\frac{1}{\tau_0^2}\right)^{-1}\left[\frac{\mu_0}{\tau_0^2}+\frac{\sum_{i=1}^n y_i}{\sigma^2}\right]\right)^2\right\} d\mu \\
&= C \times \exp\left\{-\frac{1}{2}\left(\frac{\mu_0^2}{\tau_0^2}+\frac{\sum_{i=1}^n y_i^2}{\sigma^2}\right) + \frac{1}{2}\left(\frac{n}{\sigma^2}+\frac{1}{\tau_0^2}\right)^{-1}\left[\frac{\mu_0}{\tau_0^2}+\frac{\sum_{i=1}^n y_i}{\sigma^2}\right]^2\right\} \times (2\pi)^{1/2}\left(\frac{n}{\sigma^2}+\frac{1}{\tau_0^2}\right)^{-1/2} \\
&= C \times (2\pi)^{1/2}\left(\frac{n}{\sigma^2}+\frac{1}{\tau_0^2}\right)^{-1/2} \exp\left\{-\frac{1}{2}\left(\frac{\mu_0^2}{\tau_0^2}+\frac{\sum_{i=1}^n y_i^2}{\sigma^2}\right) + \frac{1}{2}\left(\frac{n}{\sigma^2}+\frac{1}{\tau_0^2}\right)^{-1}\left[\frac{\mu_0}{\tau_0^2}+\frac{\sum_{i=1}^n y_i}{\sigma^2}\right]^2\right\}.
\end{aligned}$$

This will allow us to verify the results obtained using nested sampling. In the simple case where $\mu_0 = 0$, $\tau_0^2 = 1$, $n = 1$, $\sigma^2 = 1$ we obtain:

$$Z = (2\pi)^{-1/2} (2)^{-1/2} \exp\left\{-\frac{1}{2}y^2 + \frac{1}{2}(2)^{-1} y^2\right\} = \frac{1}{2\sqrt{\pi}} \exp\left\{-\frac{y^2}{4}\right\}.$$

2 Example II

Here we take a look at the classic mixture of normals:

$$Y_i = \sum_{j=1}^{K} I_{ij} Z_{ij}, \qquad i = 1, \ldots, n,$$

where:

$$I_i = (I_{i1}, \ldots, I_{iK}) \sim \text{Multinomial}(1, p),$$
$$Z_{ij} \overset{iid}{\sim} N(\mu_j, 1).$$

The parameters in the model are the mixture proportions $p = (p_1, \ldots, p_K)$ (with $\sum_j p_j = 1$) and the mixture locations $\mu = (\mu_1, \ldots, \mu_k)$. The number of mixture components K will be fixed for a given model, and we will use the evidence to motivate a model selection procedure to select the appropriate K. For convenience we choose conditionally conjugate priors for μ and p:

$$\mu \sim N\left(\mu_0, \tau_0^2\right), \qquad p \sim \text{Dirichlet}(\alpha),$$

where μ_0, τ_0^2 and α are fixed hyperparameters chosen by the analyst.

2.1 Posterior Distributions

The Y_i are conditionally independent, so the likelihood is

$$\begin{aligned} f(y|\mu, p) &= \prod_{i=1}^{n} f(y_i|\mu, p) = \prod_{i=1}^{n} \left[\sum_{\text{all } I_i} f(y_i|\mu, I_i) f(I_i|p)\right] \\ &= \prod_{i=1}^{n} \left[\sum_{j=1}^{K} (2\pi)^{-1/2} \exp\left(-\frac{1}{2}(y_i - \mu_j)^2\right) p_j\right]. \end{aligned}$$

The posterior is therefore

$$\begin{aligned} f(\mu, p|y) &= \frac{1}{Z} f(y|\mu, p) f(\mu) f(p) \\ &= \frac{1}{Z}\left\{\prod_{i=1}^{n}\left[\sum_{j=1}^{K}(2\pi)^{-1/2}\exp\left(-\frac{1}{2}(y_i - \mu_j)^2\right)p_j\right]\right\} \\ &\times \left\{(2\pi\tau_0^2)^{-1/2}\exp\left(-\frac{1}{2}(\mu - \mu_0\mathbf{1})'(\mu - \mu_0\mathbf{1})\right)\right\} \\ &\times \left\{\frac{\Gamma\left(\sum_{u=1}^{K}\alpha_u\right)}{\prod_{v=1}^{K}\Gamma(\alpha_v)}\prod_{j=1}^{K}p_j^{\alpha_j - 1}\right\}, \end{aligned}$$

where $\mathbf{1}$ denotes the $K \times 1$ vector of ones.
TODO

2.2 Evaluating the Evidence: Analytically

TODO

2.3 Evaluating the Evidence: Nested Sampling

TODO

2.4 Evaluating the Evidence: Other Methods

TODO