Hierarchical Models

Brendon J. Brewer

Department of Statistics
The University of Auckland

https://www.stat.auckland.ac.nz/~brewer/

Hierarchical models

Hierarchical models are useful ways of specifying priors in complex situations with lots of unknown parameters.

An example

Suppose you want to measure some properties of the (frequency) distribution of masses of some stars, but your mass measurements contain noise.

Let $\{m_1, m_2, ..., m_N\}$ be the true masses. If you could, you'd infer some parameters from the ms, perhaps with the following assumptions:

$$m_i | m_{\min}, \alpha \sim \text{Pareto}(m_{\min}, \alpha) \tag{1}$$
$$m_{\min} \sim \text{Something} \tag{2}$$
$$\alpha \sim \text{Something} \tag{3}$$

But if we don't have the ms...

An example

Let x_i be a noisy measurement of mass m_i:

$$x_i | m_i \sim \text{Normal}(m_i, \sigma_i^2) \tag{4}$$

Then we have $p(\boldsymbol{x}|\boldsymbol{m})$ and $p(\boldsymbol{m}|m_{\min}, \alpha)$.

Bayes

Posterior distribution for unknowns given knowns:

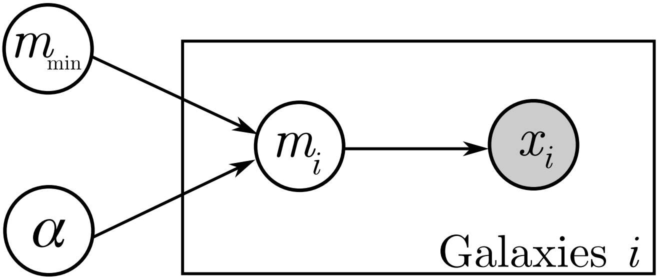
$$
\begin{aligned}
p(m_{\min}, \alpha, \boldsymbol{m}|\boldsymbol{x}) &\propto p(m_{\min}, \alpha, \boldsymbol{m}) p(\boldsymbol{x}|m_{\min}, \alpha, \boldsymbol{m}) \qquad (5) \\
&\propto p(m_{\min}, \alpha) p(\boldsymbol{m}|m_{\min}, \alpha) p(\boldsymbol{x}|\boldsymbol{m}) \qquad (6)
\end{aligned}
$$

The data we wish we had (the xs) are now unknown parameters.

Their prior is defined *conditional on other parameters* called 'hyperparameters'.

Probabilistic Graphical Model

PGM (also known as DAG for Directed Acyclic Graph)



Different parameterisation

The prior $p(m_{\min}, \alpha, \boldsymbol{m})$ is highly correlated.

It's usually better to make the prior more independent. In this case, we can define $u_i \sim \text{Uniform}(0, 1)$, and obtain m_i from

$$m_i := m_{\min} \left(1 - u_i\right)^{-1/\alpha}. \tag{7}$$

That's the 'inverse transform method'.