Bayesian Hierarchical Modelling

Likelihood and posterior distributions

- Prior distributions and likelihood provide two sources of information about any problem.
- The likelihood informs about the parameter via the data
- the prior distributions inform via prior beliefs or assumptions.
- When there are large amounts of data, ie the sample size is large, the likelihood will contribute more to the estimation. When the example is data poor then the prior distributions will dominate the analysis.

Definitions

- The likelihood of data $\{y_i\}$ given the parameters $\{\theta_i\}$, is $L(\mathbf{y}|\boldsymbol{\theta})$
- The log likelihood is $l(\mathbf{y}|\boldsymbol{\theta})$.
- Note that $\boldsymbol{\theta}$ does not have to be the same dimension as $\mathbf{y}$.
- The product of the likelihood and the prior distributions is called the posterior distribution.
- This distribution describes the behaviour of the parameters after the data are observed and prior assumptions are made.
- The posterior distribution is defined as :

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto L(\mathbf{y}|\boldsymbol{\theta})\mathbf{g}(\boldsymbol{\theta})$$

where $\mathbf{g}(\boldsymbol{\theta})$ is the joint distribution of the $\boldsymbol{\theta}$ vector.

Poisson-Gamma example in Epidemiology

A simple example of this type of model in is the where the data likelihood is Poisson and there is a common relative risk parameter with a single gamma prior distribution:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto L(\mathbf{y}|\theta)g(\theta)$$

where $g(\theta)$ is a gamma distribution with parameters α, β i.e. $G(\alpha, \beta)$, and $L(\mathbf{y}|\theta) = \prod_{i=1}^{m}\{(e_i\theta)^{y_i}\exp(e_i\theta)\}$ bar a constant only dependent in the data. A compact notation for this model is :

$$y_i|\theta \sim Pois(e_i\theta)$$

$$\theta \sim G(\alpha, \beta).$$

Hierarchical Models

- In the previous section a simple example of a likelihood and prior distribution was given. In that example the prior distribution for the parameter also had parameters controlling its form. These parameters (α, β) can have assumed values, but more usually an investigator will not have a strong belief in the prior parameters values. The investigator may want to estimate these parameters from the data.
- Alternatively and more formally, as parameters within models are regarded as stochastic (and thereby have probability distributions governing their behaviour), then these parameters must also have distributions. These distributions are known as hyperprior distributions, and the parameters are known as hyperparameters.

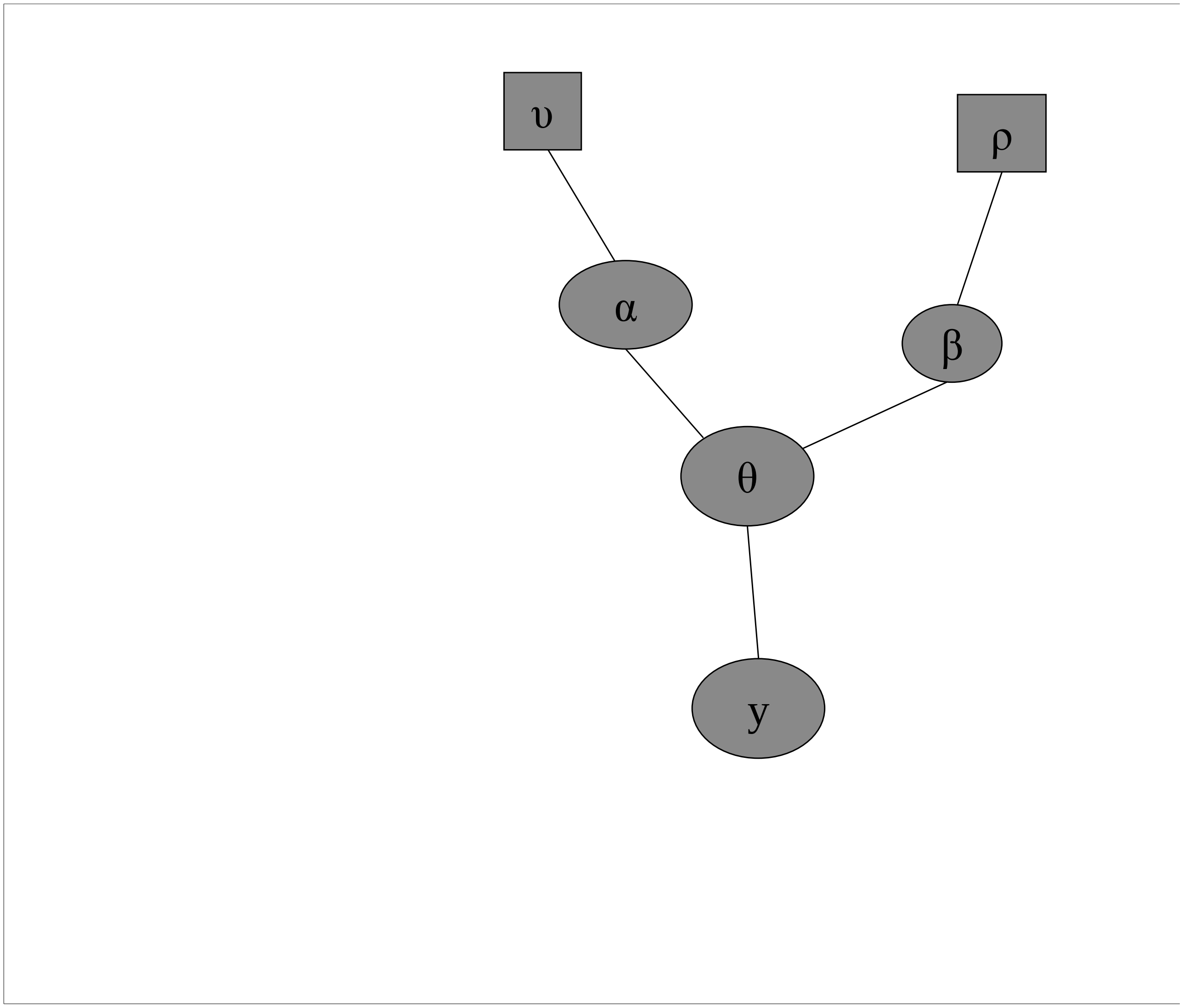
- The idea that the values of parameters could arise from distributions is a fundamental feature of Bayesian methodology and leads naturally to the use of models where parameters arise within *hierarchies*.
- In the Poisson-gamma example there is a two level hierarchy:
- θ has a $G(\alpha, \beta)$ distribution at the first level of the hierarchy and α will have a hyperprior distribution (h_α) as will β (h_β), at the second level of the hierarchy.

- This can be written as :

$$\begin{aligned} y_i|\theta &\sim Pois(e_i\theta) \\ \theta|\alpha,\beta &\sim G(\alpha,\beta) \\ \alpha|\nu &\sim h_\alpha(\nu) \\ \beta|\rho &\sim h_\beta(\rho). \end{aligned}$$

- For thes types of models it is also possible to use a graphical tool to display the linkages in the hierachy. This is known as a *directed acyclic graph* or *DAG* for short.
- On such a graph lines connect the levels of the hierarchy and parameters are nodes at the the ends of the lines.
- Clearly it is important to terminate a hierarchy at an appropriate place, otherwise one could always assume an infinite hierarchy of parameters.
- Usually the cut off point is chosen to lie where further variation in parameters will not affect the lowest level model.
- At this point the parameters are assumed to be fixed.

- For example, in the gamma-Poisson model if you assume α and β were fixed then the Gamma prior would be fixed and the choice of α and β would be uninformed. The data would not inform about the distribution at all. However by allowing a higher level of variation i.e. hyperpriors for α, β, then we can fix the values of ν and ρ without heavily influencing the lower level variation. Below is displayed the DAG for the simple two level gamma-Poisson model just described.



υ
ρ
α
β
θ
y

Posterior Inference

- When a simple likelihood model is employed, often maximum likleihood is used to provide a point estimate and associated variability for parameters.
- This is true for simple epidemiological models.
- For example, in the model $y_i|\theta \sim Pois(e_i\theta)$ the maximum likelihood estimate of θ is the the overall rate for the study region i.e. $\sum y_i / \sum e_i$. On the other hand, the SMR is the maximum likelihood estimate for the model $y_i|\theta_i \sim Pois(e_i\theta_i)$.

- When a Bayesian hierachical model is employed it is no longer possible to provide a simple point estimate for any of the $\theta_i s$.
- This is because the parameter is no longer assumed to be fixed but to arise from a distribution of possible values.
- Given the observed data, the parameter or parameters of interest will be described by the posterior distribution, and hence this distribtuion must be found and examined.
- It is possible to examine the expected value (mean) or the mode of the posterior distribution to give a point estimate for a parameter.

- Just as the maximum likelihood estimate is the mode of the likelihood, then the *maximum aposteriori* estimate is that value of the parameter or parameters at the mode of the posterior distribution.
- More commonly the expected value of the parameter or parameters is used. This is known as the *posterior mean* (or *Bayes estimate*).
- For simple unimodal symetrical distrbutions, the modal and mean estimates coincide.

Finding posterior means and sampling

- For some simple posterior distributions it is possible to find the exact form of the posterior distribution and to find explicit forms for the posterior mean or mode.
- However, it is commonly the case that for reasonably realistic model, it is not possible to obtain a closed form for the posterior distribution.
- Hence it is often not possible to derive simple estimators for parameters such as the relative risk.
- In this situation resort must be made to *posterior sampling* i.e. using simulation methods to obtain samples from the posterior distribution which then can be summarised to yield estimates of relevant quantities.
- In the next section we discuss the use of sampling algorithms for this purpose.

- An exception to this situation where a closed form posterior distribution can be obtained is the gamma-Poisson model where $\alpha,\ \beta$ are fixed.
- In that case, the relative risks have posterior distribution given by:

$$\theta_i \sim G(y_i + \alpha, e_i + \beta)$$

- The posterior expectation of θ_i is $(y_i + \alpha)/(e_i + \beta)$.
- Of course if α and β are not fixed and have hyperprior distributions then the posterior distribution is more complex.

name:	y[i]	type:	stochastic	density:	dpois
mean	mu[i]	upper bound		upper bound