Chapter 4

Prior distributions

- ➢ Reference priors
  - Jeffreys' rule
  - Improper priors

- ➢ Vague priors

- ➢ Informative priors

- ➢ Hierarchical models

- ➢ Sensitivity to prior

Reference priors

- Historically, considerable research effort has focused at obtaining "non-informative" priors (Note: flat priors are not non-informative in general). However, like the holy grail, this much sought after prize has proved extremely elusive. It is now more common to use the terminology "reference prior" to denote a prior that is considered a default prior (for the particular model in question).

- The most widely used method for obtaining a (potential) reference prior is *Jeffreys's rule*.

- Reference priors are frequently *improper*.

Jeffreys' rule

Jeffreys' rule is motivated by the desire that inference should not depend on how a model is parameterized.

Example: if instantaneous mortality is *m*, then the annual survival rate is $s=e^{-m}$. Some modelers might use *m,* while others might use *s.* Inference should not depend on this arbitrary choice of parameterization.

Jeffreys' rule: The reference prior is obtained as the square root of the determinant of the information matrix for the model.

Jeffreys' rule

Jeffreys' rule is widely accepted for single parameter models, but its use is somewhat more controversial, and often subject to modification, in multi-parameter models….and can also be a chore to calculate.

Example:

- The Jeffreys' prior for the mean of normally distributed data is the flat prior, $\pi(\mu)=1$, and for the standard deviation is the inverse prior $\pi(\sigma)=1/\sigma$ (this is equivalent to $log(\sigma)$ being flat).

Jeffreys' rule

In general, the flat prior is the Jeffreys' prior for "location" parameters and the inverse prior is the Jeffreys' prior for "scale" parameters.

The above priors make intuitive sense:

- If one is totally ignorant of a location parameter, then it could take any value on the real line with equal prior probability.

- If totally ignorant about the scale of a parameter, then it is as likely to lie in the interval 1-10 as it is to lie in the interval 10-100. This implies a flat prior on the log scale.

Improper priors

Priors such as $\pi(\mu)$=1, $\pi(\sigma)$=1/σ are *improper* because they do not integrate to 1. That is, the area under the prior density is not unity (and, in fact, is infinity).

In most cases, improper priors can be used in Bayesian analyses without major problems. However, things to watch out for are:

- In a few models, the use of improper priors can result in improper posteriors.

- Use of improper priors makes model selection and hypothesis testing difficult.

- WinBUGS does not allow the use of improper priors.

Example: Exchange paradox

This seemingly innocent example demonstrates that inconsistencies can arise from using an improper prior.

Exchange paradox:

Amount $\$\theta$ is placed in an envelope, and amount $\$2\theta$ is placed in a second identical envelope. The envelopes are shuffled, and John and Mary are randomly given one envelope each.

John and Mary open their respective envelopes and check its contents (making sure that the other person does not see) and then are given the opportunity to exchange – is it to their advantage to do so?

Exchange paradox continued…

Vague priors

Essentially, these are densities with high spread, such as a normal density with extremely large variance. These give similar prior value over a large range of parameter values.

- In WinBUGS, the flat prior can be approximated by a vague normal density prior, with mean=0 and variance=1,000,000, say.

- The inverse prior, $\pi(\sigma)=1/\sigma$, can be approximated by a Gamma density (with very small shape parameter and rate parameters).

Gamma approximating prior
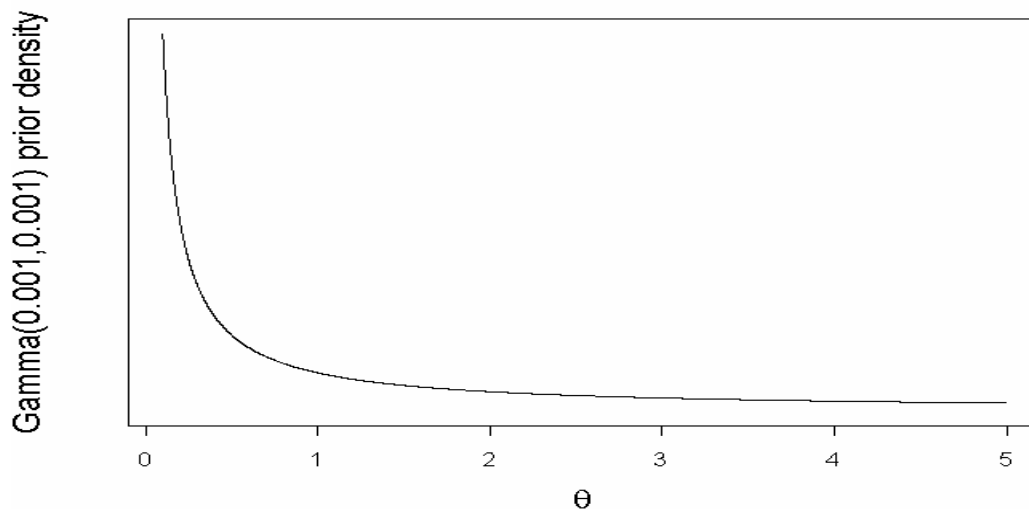
Informative priors

As the name suggests, informative priors convey information concerning prior preference for certain values of the parameters.

Where does this prior information come from?

- "Expert" opinion.

- Previous experiments of a similar nature. E.g., in fisheries, prior information about population parameters of a fish stock can be obtained from previous work done on other stocks of the same species.
  - This can often be done formally using meta-analysis or hierarchical Bayesian modeling of the existing data.

Specifying informative priors

Rather than trying to directly specify values for the parameters of a prior density, it is often easier to express probability bounds, from which the parameters can then be obtained.

Example: The "expert" may specify that the growth rate of a population is between 0.2 and 0.5 with 90% prior probability, and be happy with a lognormal shaped prior density. A quick bit of math shows that these quantiles correspond to a lognormal with mean of -1.15 and standard deviation of 0.28 on the log scale.

Specifying informative priors

Informative lognormal prior on growth rate
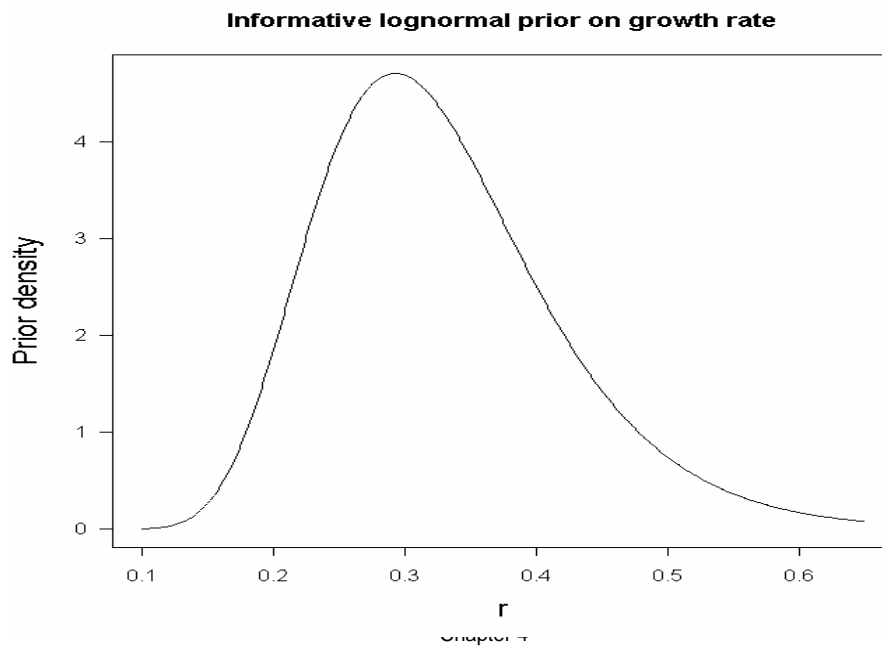
Hierarchical priors

These are two-stage priors, in the sense that a prior is placed on a prior.

Example: The IID Normal(μ,σ^2) example (with known σ^2) used a N(ν,τ^2) prior on μ. The values of ν and τ^2 are specified after due consideration of the prior information (if any) known about μ.

A hierarchical prior for this example would place priors on the values of ν and τ^2. This prior is known as a hyper-prior, and its parameters are known as hyper-parameters.

Hierarchical priors are more flexible than non-hierarchical priors, and make the posterior less sensitivity to the prior.

Hierarchical models

Hierarchical models use hierarchical priors to perform meta-analyses, whereby a number of related experiments are performed and it is desired to combine information.

The relative ease of implementation, and interpretability, of Bayesian hierarchical models (c.f. frequentist "empirical-Bayes" mixture models) is a major strength of the Bayesian approach.

More on this later.

Sensitivity to prior

After doing a Bayesian analysis, it is never long before the inevitable question is asked – "How would the posterior change if you used a different prior?"

- The most common and straightforward approach is to repeat the analysis using a handful of alternative priors.

- Substantial high-powered theoretical research has investigated the sensitivity of the posterior to "arbitrary" changes to the prior, but this work is currently of little help to the practitioner.

- Simpler results exist for specific changes to the prior.

Simple sensitivity example

Recall the IID Normal example where we saw that the posterior mean was given by

$$\nu^* = \frac{\dfrac{n}{\sigma^2}\,\overline{y} + \dfrac{1}{\tau^2}\,\nu}{\dfrac{n}{\sigma^2} + \dfrac{1}{\tau^2}}$$

From this we can see that the change (derivative) in the posterior mean with respect to the prior mean is

$$\frac{d\nu^*}{d\nu} = \frac{\dfrac{1}{\tau^2}}{\dfrac{n}{\sigma^2} + \dfrac{1}{\tau^2}} = \frac{\tau^{*2}}{\tau^2}$$

which is just the ratio of posterior variance to prior variance.

Sensitivity more generally

The sensitivity result from the previous slide can be generalized.

For example, it can be shown that, if parameter μ has a N(ν, τ^2) prior then the derivative of the posterior mean of any parameter θ with respect to ν is the posterior covariance of μ and θ, divided by τ^2.

South Atlantic albacore tuna example