Problem 1

Consider a breeding experiment resulting in 10 ‘successes’ out of 100 (independent) trials. The researcher has no real prior opinion about the unknown probability of success $\theta$.

1. Why would a Beta(1,1) prior for $\theta$ be reasonable? Sketch this distribution.

2. Write down the posterior distribution for $\theta$.

3. What is the posterior mean value for $\theta$?

4. Design a Metropolis algorithm to estimate $\theta$. (We wouldn’t do this in practice with this particular example, because we know the answer analytically, but this might be part of a larger problem.)
Solution

1. Beta(\(\alpha=1, \beta=1\)) is equivalent to the Uniform distribution.

2. \(p(\theta|y) \sim Beta(y+\alpha, n-y+\beta) = Beta(11,91)\)
   since \(y=10, n=100\)
   \(Beta(11,91) \propto \theta^{10}(1-\theta)^{90}\)

3. Posterior mean = \((y+\alpha)/(n+\alpha+\beta) = 11/102 = 0.108\)

4. Consider [0,1] as a ‘circle’.
   Initialise: \(\theta^{(1)}=0.5\)
   for (i in 2:1000) {
     sample \(\theta^* \sim Uniform(\theta^{(i-1)}-.1, \theta^{(i-1)}+.1)\)
     adjust: if \(\theta^*>1\) then \(\theta^* = \theta^-1\); if \(\theta^*<1\) then \(\theta^* = 1+\theta^*\)
     calculate \(A = p(\theta^*|y) / p(\theta^{(i-1)}|y) = (\theta^*)^{10}(1-\theta^*)^{90} / \theta^{10}(1-\theta)^{90}\)
     draw \(u \sim Uniform(0,1)\)
     if \(u < A\), take \(\theta^{(i)}=\theta^*\), otherwise take \(\theta^{(i)}= \theta^{(i-1)}\)
   }
Prior Beta(1,1)
# Metropolis algorithm for estimating $p \sim \text{Beta}(a, b)$

# results in matrix p:
# p[,1] = proposed values, p[,2] = accepted values, p[,3] = acceptance ratio

```r
a_10  # parameters of Beta distribution
b_90
N_1000  # number of iterations
p_matrix(0,nrow=N,ncol=3)  # store results
p[1,1]_0.2  # initialise
p[1,2]_0.2
for (i in 2:N) {  # loop
  pold_p[i-1,2]  # current p
  pnew_runif(1,pold-0.1,pold+0.1)  # proposed p
  if (pnew<0) pnew_1+pnew  # adjust if proposed p <0 or >1
  if (pnew>1) pnew_pnew-1
  ratio_(pnew)^a*(1-pnew)^b/((pold)^a*(1-pold)^b)
  u_runif(1,0,1)  # decide if accept proposed value
  if (u<ratio) p[i,2]_pnew
  if (u>=ratio) p[i,2]_pold
  p[i,1]_pnew  # store proposed value and ratio
  p[i,3]_ratio
}
```
<table>
<thead>
<tr>
<th></th>
<th>proposed</th>
<th>accepted</th>
<th>ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.16236505</td>
<td>0.10448786</td>
<td>2.008007e-001</td>
</tr>
<tr>
<td>2</td>
<td>0.11019663</td>
<td>0.11019663</td>
<td>9.573174e-001</td>
</tr>
<tr>
<td>3</td>
<td>0.08624814</td>
<td>0.08624814</td>
<td>9.416614e-001</td>
</tr>
<tr>
<td>4</td>
<td>0.16400424</td>
<td>0.08624814</td>
<td>2.064788e-001</td>
</tr>
<tr>
<td>5</td>
<td>0.15311573</td>
<td>0.08624814</td>
<td>3.329064e-001</td>
</tr>
<tr>
<td>6</td>
<td>0.02256085</td>
<td>0.08624814</td>
<td>6.450435e-004</td>
</tr>
<tr>
<td>7</td>
<td>0.13300017</td>
<td>0.13300017</td>
<td>6.732727e-001</td>
</tr>
<tr>
<td>8</td>
<td>0.05242903</td>
<td>0.13300017</td>
<td>2.694898e-001</td>
</tr>
<tr>
<td>9</td>
<td>0.10375305</td>
<td>0.10375305</td>
<td>1.652981e-001</td>
</tr>
<tr>
<td>10</td>
<td>0.08013098</td>
<td>0.08013098</td>
<td>7.849518e-001</td>
</tr>
<tr>
<td>11</td>
<td>0.03483840</td>
<td>0.08013098</td>
<td>1.824909e-002</td>
</tr>
<tr>
<td>12</td>
<td>0.06088227</td>
<td>0.08013098</td>
<td>4.133978e-001</td>
</tr>
</tbody>
</table>
Problem 2

We are interested in the average milk yield (in litres/day) of a new line of dairy cattle.

One way to model this is as follows.

Let milk yield \( y \sim N(\mu, \sigma^2) \), with \( \mu \) and \( \sigma^2 \) unknown.

Based on previous experiments, set the following priors:

\[ \mu \sim N(\mu_0 = 10, \sigma^2 / \kappa_0) \]
\[ \kappa_0 \text{ represents ‘the equivalent number of prior measurements’, so here we set } \kappa_0 = 20. \]

\[ \sigma^2 \sim \text{Inv-} \chi^2 (\nu_0 = 2, \sigma_0^2 = 1). \]
\[ \sigma_0^2 \text{ is the ‘best guess’ at } \sigma^2. \]
\[ \nu_0 \text{ represents the ‘degrees of freedom’ (how much we believe our estimate of } \sigma_0^2); \text{ the larger the value, the stronger the belief.} \]
Problem 2

Priors: $\mu \sim N(\mu_0=10, \sigma^2/\kappa_0), \kappa_0=20.$  
$\sigma^2 \sim \text{Inv-}\chi^2 (\nu_0=2, \sigma_0^2=1).$

A sample of 100 animals from the new line gives a sample mean milk yield of 10L/day with a sample standard deviation of 2L, so $\bar{y} = 10, s^2 = 4$

- Assuming that milk yield is normally distributed, write down an appropriate likelihood and prior for this problem.
- Develop a Gibbs algorithm to estimate the unknown mean and variance of the distribution.
Solution

1. Sample

\[ \sigma^2 \mid y \sim \text{Inv-Chi}^2(v_n, \sigma^2_{n}) \]

2. Sample

\[ \mu \mid \sigma^2, y \sim \mathcal{N}(\mu_n, \sigma^2 / \kappa_n) \]

\[
\mu_n = \frac{\kappa_0}{\kappa_0 + n} \mu_0 + \frac{n}{\kappa_0 + n} \bar{y} \\
\kappa_n = \kappa_0 + n; \quad v_n = v_0 + n \\
v_n \sigma_n^2 = v_0 \sigma_0^2 + (n - 1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n} (\bar{y} - \mu_0)^2
\]
\[ \mu_n = \frac{20}{20+100} \times 10 + \frac{100}{20+100} \times 10 = 10 \]
\[ \kappa_n = 20+100=120 \]
\[ \nu_n = 2+100 = 102 \]
\[ \sigma_n^2 = 2 \times 1 + 99 \times 4 + \frac{20(100)}{(20+100)}(10-10)^2 / 102 \]
\[ = \frac{398}{102} = 3.9 \]

1. Sample \( \sigma^2 \sim \text{Inv-\( \chi^2 \)}(102,3.9) \)
2. Sample \( \mu \sim \text{N}(10,\sigma^2/120) \)
Inverse chi-square (102,3.9)
ISSUES IN MODELLING

- Choosing a prior
- Initial values
- Reparametrisation
- Model checking
- Model averaging
- Other applications
Interpretations of Prior Distributions

1. Based on previous experiments, physical properties etc

2. Objective representations of what is rational to believe about a parameter

3. As a subjective measure of what a particular individual, “you,” actually believes
Care with ‘noninformative’ priors

- Central problem: specifying a prior distribution for a parameter about which nothing is known
- If $\theta$ can only have a finite set of values, it seems natural to assume all values equally likely \textit{a priori}
- This can have odd consequences. For example specifying a uniform prior on regression models:


assigns prior probability $6/16$ to 3-variable models and prior probability only $4/16$ to 2-variable models
Uniform prior = ignorance?

• Natural to use a uniform prior, but if $\theta$ is uniform, an arbitrary function of $\theta$ is not.

• Eg, earlier we saw that a uniform distribution on a probability translates to a strong assumption about the odds. Do we really mean this?

• “ignorance about $\theta$” does not imply “ignorance about $\gamma$”. The notion of “prior ignorance” may be untenable.
The Jeffreys Prior
(single parameter)

- Jeffreys prior is arguably an objective prior. It corresponds to the expected Fisher Information. All parametrizations lead to the same prior. (see Box and Tiao, 1973, Section 1.3)
- Jeffrey’s prior for a Binomial likelihood is a Beta density with parameters $\frac{1}{2}$, $\frac{1}{2}$.
- Other Jeffreys priors:
  - Poisson($\lambda$): $\pi(\lambda) \propto \lambda^{-1/2}$
  - Normal($\mu$): $\pi(\mu) = 1, \mu \in \mathbb{R}$
  - Normal($\sigma$): $\pi(\sigma) = 1/\sigma, \sigma > 0$
Non-informative priors

• May not want priors to be influential
• Distinguish
  - *primary* parameters of interest
  - *secondary* structure used for smoothing etc.
• Location parameters (eg regression coefficients): Normal (0, 0.0001)
  - standard deviation of 100
  - effectively a uniform prior
Non-informative priors (cont)

- Careful! An improper prior can give an improper posterior distribution (distribution doesn’t integrate to one, so isn’t a ‘real’ distribution, so estimates can’t be trusted)
- Eg: Scale parameters (eg precision of random effects)
  - at the second level of a hierarchy a uniform prior gives an improper distribution
- Options:
  - “just proper” eg Gamma(1E-3, 1E-3) as on previous slide
  - s.d. ~ Uniform (0, r)
  - proper prior
Subjective priors

- Determination of subjective priors is an area of current research. Subjective priors can be potentially useful but difficult to elicit and use.
- Difficult to assess the usefulness of a subjective posterior. What does it tell us?
- Don’t be misled by the term “subjective”; all data analyses involve appreciable personal elements
Acceptance Rate

The desired acceptance rate of a Metropolis-Hastings algorithm has also been a matter of recent research. Optimal rates for random walk algorithms have been carefully investigated by Roberts et al. [84] and corresponding guidelines have been suggested. As described and illustrated by Robert and Casella ([80], pp. 252-254), high acceptance rates are desirable if the proposal density \( g \) approximates the target \( f \) such that \( f = g \) is bounded for uniform ergodicity. However, low acceptance rates are preferable if a random walk proposal is adopted. These authors also propose the use of the rejected values in a Metropolis-Hastings algorithm through Rao-Blackwellisation and give references to other acceleration methods.
Reparameterisation

• In regression problems
  - *rescale* quantitative covariates where appropriate: improves stability of the parameter estimates
  - *standardise* quantitative covariates about their mean: makes parameters more orthogonal, eg rats example...
Reparametrisation (cont)

• For fixed effects (‘non-informative’ priors)
  - use corner point constraints,
    eg, kidney…
  - or eliminate grand mean and calculate contrasts separately…

• For random effects models, try hierarchical centring, eg:
  - uncentred
  - fully centred

• If prior variance of a random effect is large relative to the error variance, centring reduces posterior correlation between the random effects
Model criticism and selection

- Lack of well-established techniques for Bayesian model choice in software
- Difficulty of implementing some methods (e.g., cross-validation) in a Bayesian framework
- Not interested in “Is model true?” but “Do model deficiencies affect substantive inferences?”
- Compare observed statistics with values predicted under the model
  - if the model is adequate, replicated data generated under the model should look similar to the observed data
Model Averaging

Instead of choosing a single model based on the above methods, an increasingly common practice is model averaging. This is the practice of combining expected values obtained from different models (perhaps describing different dimensions or different combinations of variables) weighted by their corresponding posterior probabilities. Of course, adoption of this approach depends on the aim of the analysis and achieving a balance between improved estimation and easy interpretation.
Other Issues

- Length of burnin
- Total length of run
- Number of chains
- Dependence in chains
- Choice of algorithm
- Choice of proposal distribution
- Subjective and expert priors
- Speeding up convergence
General strategy for complex modelling in BUGS

• Start with simple models which have been used in other software or in examples and for which answers are known
• Develop more complex models incrementally
• Check final answers by starting from different initial values, running for long periods and using different parametrisations
• Perform a few updates before undertaking long runs, to assess timings and examine ballpark results.