Homework 2

Paul Hewson

December 14, 2009

1 A computer exercise with the Binomial distribution

Consider the estimation of the sex ratio within a population of human births. There is a medical condition called *placenta previa* where the placenta can be too low. In an early study of babies born under such conditions, we find that of $n = 980$ births, $x = 437$ were female. We **know** that the population of female births in the general population is 0.485. We wish to determine whether the proportion of female babies born following *placenta previa* is lower than the proportion of babies born in the population.

In order to do this, we wish to have a posterior estimate of π, the proportion of female babies born following *placenta previa*.

In particular, we wish to see whether the 95% credible interval we calculate from any posterior distribution contains the value 0.485.

1. For these data, where $x = 437$ and $n = 980$, give the maximum likelihood estimates for π. Give a 95% Wald confidence interval based on the formula $\hat{\pi} \pm 1.96 \times I(\hat{\pi})^{-1/2}$.

2. Estimate and summarise the posterior distribution for π given a Jeffreys' prior. You should summarise your results using the mean, mode as well as giving the 95% credible interval.

3. Estimate and summarise the posterior distribution for π using four informative Beta priors. You should choose two priors that suggest a mean of 0.500 and two priors that suggest a mean of 0.485. Each of these two priors should suggest a different variance. You should report your prior carefully, including information on the equivalent sample size. As before, summarise your results as a posterior mean, mode and estimate the 95% credible interval for your posterior distribution.

4. Briefly comment on how sensitive the results are to changes in the prior distribution. In other words, briefly comment on how much the posterior distributions are altered by changes in your prior assumptions.

5. Briefly comment on whether the proportion of female births is lower among *placenta previa* mothers than in the general population.

2 An algebraic exercise with the Poisson distribution

1. For a sample of data $x_1, x_2, \ldots, x_n$, recall that the Poisson likelihood is given by:

$$L(\lambda) \propto \lambda^{\sum x_i} e^{-n\lambda} \text{ for } 0 < \lambda < \infty$$

and the corresponding log-likelihood:

$$\log L(\lambda) = \sum x_i \log \lambda - n\lambda$$

Show that the Jeffreys prior is given by:

$$p(\lambda) \propto \frac{1}{\sqrt{\lambda}}$$

2. Give the expression for the posterior density (ignoring any constants not involving λ) when using the Jeffreys prior.

3. Recall that the Gamma distribution has density:

$$f(\lambda) = \frac{r^a}{\Gamma(a)} \lambda^{(a-1)} e^{-\lambda r}$$

   for $x > 0$.

   Give the general expression for the posterior for λ when using a gamma prior (ignoring constants not involving λ).

4. We have data on accidents for $n = 8$ weeks, namely $\boldsymbol{x} = (3, 2, 0, 8, 2, 4, 6, 1)$. We believe that the accident rate should be around 2.5, and wish to use this value ($\lambda_{Prior} = 2.5$) as our prior mean when modelling the posterior of the Poisson rate parameter λ. However, we wish to ensure that the equivalent sample size of our prior is small relative to the data. We have $n = 8$ datapoints, and wish n_{eq} to be 2.5.

   Use the formula:

$$\frac{\lambda_{Prior}}{n_{eq}} = \frac{a}{r^2}$$

   to provide suitable prior values for a and r that give a Prior Mean of 2.5and an equivalent sample size of 2.5. Give an formula for the resulting posterior distribution.

3 Optional exercises

For the accident data in section 2, you have given an expression for the Posterior distribution. There are two *optional* exercises you may like to try:

- Summarise the posterior for these data with your chosen prior (i.e. report the posterior mean, median and 95% credible interval.
- Sketch the posterior function