Assignment 2

2017-05-11

Question 1: Inference for the Poisson parameter λ:

(a) Given a general Gamma(a,b) prior, derive the posterior distribution of λ, given data $(x_1, \ldots, x_n)$, followed by the posterior predictive distribution of $(z_1, \ldots, z_m)$.

The Likelihood is given by:

$$L(\lambda|x) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod_{i=1}^{n}(x_i!)}$$

The Prior is a *Gamma*(a, b):

$$p(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}, \lambda > 0$$

So the posterior is given by:

$$p(\lambda|x) \propto \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\sum x_i+\alpha-1}e^{-(n+\beta)\lambda}, \lambda > 0,$$

which is equivalent to a $Gamma(\sum x_i + \alpha, n + \beta)$, or $Gamma(n\bar{x} + \alpha, n + \beta)$.

The *Posterior Predictive Distribution* for a Poisson parameter λ with data $(z_1, \ldots, z_m)$ is given by:

$$\begin{aligned}
p(z|X) &= \int_0^\infty p(z|\lambda)p(\lambda|X)d\lambda \\
&= \int_0^\infty Poisson(z|\lambda) \cdot Gamma(\sum x_i + \alpha, n + \beta)d\lambda \\
&= \int_0^\infty \frac{e^{-\lambda}\lambda^z}{z!} \cdot \frac{(n+\beta)^{\sum x_i+\alpha}}{\Gamma(\sum x_i + \alpha)}\lambda^{\sum x_i+\alpha-1}e^{-(n+\beta)\lambda}d\lambda \\
&= \frac{(n+\beta)^{\sum x_i+\alpha}}{\Gamma(\sum x_i + \alpha)\Gamma(z+1)}\int_0^\infty \lambda^{z+\sum x_i+\alpha-1}e^{-(n+\beta+1)\lambda}d\lambda \\
&= \frac{(n+\beta)^{\sum x_i+\alpha}}{\Gamma(\sum x_i + \alpha)\Gamma(z+1)} \cdot \frac{\Gamma(z + \sum x_i + \alpha)}{(n+\beta+1)^{z\sum x_i+\alpha}} \\
&= \frac{\Gamma(z + \sum x_i + \alpha)}{\Gamma(\sum x_i + \alpha)\Gamma(z+1)}\left(\frac{n+\beta}{n+\beta+1}\right)^{\sum x_i+\alpha}\left(\frac{1}{n+\beta+1}\right)^z
\end{aligned}$$

(b) Assuming Gamma(a,b) priors for two Poisson parameters λ_1 and λ_2, derive the posterior for $\phi = \lambda_1/\lambda_2$. (Hint: use nuisance parameter $\mu = \lambda_2$).

Firstly, the likelihood function for two Poissons is:

$$p(x_1, x_2|\lambda_1, \lambda_2) = p(x_1|\lambda_1)p(x_2|\lambda_2) = \frac{e^{-\lambda_1}\lambda_1^{x_1}}{x_1!} \cdot \frac{e^{-\lambda_2}\lambda_2^{x_2}}{x_2!}.$$

Reparameterising by $\phi = \frac{\lambda_1}{\lambda_2}$, we get:

$$p(x_1, x_2|\phi, \lambda_2) = \frac{e^{-\phi\lambda_2}(\phi\lambda_2)^{x_1}}{x_1!} \cdot \frac{e^{-\lambda_2}\lambda_2^{x_2}}{x_2!}.$$

We then compute the Fisher Information Matrix:

$$I(\theta)_{ij} = E\left(-\frac{\partial^2 l}{\partial\theta_i\partial\theta_j}\right)$$
$$\implies F(\phi, \lambda_2) = \begin{bmatrix} \frac{\lambda_2}{\phi} & 1 \\ 1 & \frac{1+\phi}{\lambda_2} \end{bmatrix}$$
$$\implies S(\phi, \lambda_2) = F^{-1}(\phi, \lambda_2) = \begin{bmatrix} \frac{\phi(1+\phi)}{\lambda_2} & -\phi \\ -\phi & \lambda_2 \end{bmatrix}$$

Following the algorithm laid out by Bernardo we can define the marginal and conditional asympototic posteriors for ϕ.

$$d_0(\phi, \lambda_2) = \left[\frac{\phi(1+\phi)}{\lambda_2}\right]^{1/2}$$
$$d_1(\phi, \lambda_2) = \left(\frac{\lambda_2}{1+\phi}\right)^{1/2}$$

According to Corollary 1 of Proposition 2 in Barnardo's paper, because the nuisance parameter space $\Lambda(\phi) = \Lambda$ is independent of ϕ, we can factorise the above equations as

$$d_0^{-1}(\phi, \lambda_2) = \frac{1}{\sqrt{\phi(1+\phi)}} \cdot \sqrt{\lambda_2} = a_0(\phi)b_0(\lambda_2)$$
$$d_1^{-1}(\phi, \lambda_2) = \sqrt{\phi(1+\phi)} \cdot \frac{1}{\sqrt{\lambda_2}} = a_1(\phi)b_1(\lambda_2)$$

which implies that the marginal and conditional reference priors are

$$\pi(\phi) \propto a_0(\phi) = \frac{1}{\sqrt{\phi(1+\phi)}}$$
$$\pi(\lambda|\phi) \propto b_1(\lambda_2) = \frac{1}{\sqrt{\lambda_2}}$$

The joint posterior can be derived with the likelihood and joint prior $(d_1^{-1}(\phi, \lambda_2))$ previously derived.

$$\pi(\phi, \lambda_2|x_1, x_2) \propto \pi(x_1, x_2|\phi, \lambda_2) \cdot \pi(\phi, \lambda_2)$$
$$\propto e^{-(\phi+1)\lambda_2} \cdot \phi^{x_1-1/2}\,(1+\phi)^{1/2} \cdot \lambda_2^{x_1+x_2-1/2}$$

which, I'm pretty sure, can be factored as

$$\pi(\phi, \lambda_2|x_1, x_2) \propto Gamma(\lambda_2|x_1+1/2, 1) \cdot Gamma\left(\phi|x_1, \frac{1}{\lambda_2}\right) \cdot Beta\left(\frac{\phi}{1+\phi}|3/2, 1\right).$$

(c) Derive the Jeffreys prior for (ϕ, μ), and the corresponding marginal posterior for ϕ.

The Jeffreys principle states that the Jeffreys prior is

$$\pi_\phi(\phi) \propto det(I(\phi))^{1/2} = \left(\frac{\lambda_2(1+\phi)}{\lambda_2\phi} - 1\right)^{1/2} = \phi^{-1/2}$$

The marginal posterior can be derived by multiplyin the likelihood by the Jeffrey's prior and then integrating out the nuissance parameter:

$$
\begin{aligned}
\pi(\phi|x_1,x_2) &\propto \int_0^\infty \frac{e^{-\phi\lambda_2}(\phi\lambda_2)^{x_1}}{x_1!}\cdot\frac{e^{-\lambda_2}\lambda_2^{x_2}}{x_2!}\cdot\frac{1}{\sqrt{\phi}} \quad d\lambda_2 \\
&\propto \phi^{x_1-1/2}\int_0^\infty e^{-(\phi+1)\lambda_2}\cdot\lambda_2^{x_1+x_2} \quad d\lambda_2
\end{aligned}
$$

Here we solve the integral:

$$
\int_0^\infty e^{-(\phi+1)\lambda_2}\cdot\lambda_2^{x_1+x_2} \quad d\lambda_2 = (\phi+1)^{-(x_1+x_2+1)}\cdot\Gamma(x_1+x_2+1)
$$

which leads us to

$$
\begin{aligned}
\pi(\phi|x_1,x_2) &\propto \phi^{x_1-1/2}(\phi+1)^{-(x_1+x_2+1)}\cdot\Gamma(x_1+x_2+1) \\
&\propto \frac{\phi^{x_1-1/2}}{(\phi+1)^{-(x_1+x_2+1)}}
\end{aligned}
$$

(d) Derive the reference prior for (ϕ,μ), and the corresponding marginal posterior for ϕ.

We can combine the marginal and conditional refererence priors derived above to determine the joint reference prior.

$$
\pi(\phi,\lambda_2)=\pi(\phi)\pi(\lambda_2|\phi)=\frac{1}{\sqrt{\phi(1+\phi)}}\cdot\frac{1}{\sqrt{\lambda_2}}
$$

Now that we have the marginal reference prior for ϕ, Bernardo states that we can calculate the marginal posterior of ϕ by integrating out the nuisance parameter λ_2.

$$
\begin{aligned}
\pi(\phi|x_1,x_2) &= \pi(\phi)\int_{\Lambda(\phi)} p(x_1,x_2|\phi,\lambda_2)\pi(\lambda_2|\phi)d\lambda_2 \\
&= \pi(\phi)\int_\Lambda \frac{e^{-\phi\lambda_2}(\phi\lambda_2)^{x_1}}{x_1!}\cdot\frac{e^{-\lambda_2}\lambda_2^{x_2}}{x_2!}\cdot\frac{1}{\sqrt{\lambda_2}}d\lambda_2 \\
&\propto \frac{1}{\sqrt{\phi(1+\phi)}}\int_\Lambda e^{-(\phi+1)\lambda_2}\phi^{x_1}\cdot\frac{\lambda_2^{x_1+x_2}}{\sqrt{\lambda_2}}d\lambda_2 \\
&\propto \frac{\phi^{x_1}}{\sqrt{\phi(1+\phi)}}\int_\Lambda e^{-(\phi+1)\lambda_2}\cdot\frac{\lambda_2^{x_1+x_2}}{\sqrt{\lambda_2}}d\lambda_2
\end{aligned}
$$

Here we solve the integral

$$
\int_\Lambda e^{-(\phi+1)\lambda_2}\cdot\frac{\lambda_2^{x_1+x_2}}{\sqrt{\lambda_2}}d\lambda_2=\frac{\Gamma(x_1+x_2-1/2)}{(1+\phi)^{(x_1+x_2+1/2)}}
$$

which leaves

$$
\pi(\phi|x_1,x_2)\propto\frac{\phi^{x_1-1/2}}{(1+\phi)^{(x_1+x_2+1)}}
$$

which is identical to the marginal posterior derived by using the Jeffrey's prior.

(e) Based on (c), consider the posterior for ϕ based on uniform priors for λ_1 and λ_2, and comment on when inference based on this posterior could be quite different from that based on the reference posterior from (d). Give a data example, in terms of resulting intervals. (Hint: the above posteriors are related to known pdfs, and by transformation may be simplified even further.).

Before I begin, I should note that you deal with a lot of these issues yourself Frank, in *Consensus priors for multinomial and binomial ratios.*

Firstly, the parameter ϕ can be regarded as a multinomial ratio, a link established by both yourself – in *Section 2.3: Straitforward alternative derivations* – and Bernardo, who uses the transformation $\omega = \frac{phi}{1+phi}$ to arrive at an alternative expression for the marginal posterior density function given above:

$$pi(\phi|x_1, x_2)|x_1, x_2) \left|\frac{d\phi}{d\omega}\right| = \pi(\omega|x_1, x_2) \propto \omega^{x_1 - 1/2}(1-\omega)^{x_2 - 1/2} = Beta(\omega|x_1 + 1/2, x_2 + 1/2)$$

Given this is a multinomial case where the Jeffrey's/Reference posterior is given above, you give the Bayes posterior as

$$\pi_B(\phi|x_1, x_2) \propto \frac{\phi^{x_1}}{(1+\phi)^{(x_1+x_2+2)}}$$

which we can derive a similar alternative expression for:

$$pi(\phi|x_1, x_2)|x_1, x_2) \left|\frac{d\phi}{d\omega}\right| = \pi(\omega|x_1, x_2) \propto \omega^{x_1}(1-\omega)^{x_2} = Beta(\omega|x_1 + 1, x_2 + 1)$$

As usual, the conditions where the difference between the Bayes and Jeffrey's/Reference marginal posteriors becomes apparent is when $x_1 = 0, x_2 \to n$. The Jeffrey's/Reference posterior will have more weight closer to zero than the BL posterior, which we studied more closely in Assignment 1. Also, we know that in terms of mean coverage, BL $\to 1 - \alpha, n \to \infty$, which is not case for J/R.

Question 2: Inference for variance components:

(a) Use e.g. SAS (PROC VARCOMP) to perform a classical analysis of the data in Table 5.1.4 of Box & Tiao (1973), based on finding point estimates only.

(b) Use e.g. SAS (PROC VARCOMP) to perform a classical analysis of the data in Table 5.1.4 of Box & Tiao (1973), based on finding point estimates only.

I'll be using R. First, let's confirm the results reported by Box & Tiao (1973), using frequentist methods.

These results match what was reported by Box & Tiao (1973). Next, we can attempt to fit a linear model using the REML method.

Despite the fact that the result for σ_{btw} is reported as 0, we can observe some slight variation between batches using a dotplot (reordering the Batches to produce a smoother average line).

As a side note, the creator of the lme4 package that contains the Dyestuff2 data set explains the estimate of 0 away by saying that *"indicates that the level of "between-group" variability is not sufficient to warrant incorporating random effects in the model"* – page 25 here: http://lme4.r-forge.r-project.org/lMMwR/lrgprt.pdf.

(b) Use WinBUGS for a Bayesian analysis of (a), and find reasonable point and interval estimates for σ_1^2 and σ_2^2. Include graphs, including one of the joint posterior. [6 marks] Before we begin, I've relabelled the ICC variable as ICC, instead of rho – which I had previously – just to keep things consistent. Also, I found methods for using the ICC method for models with more that two levels but didn't have time to go into them too much: I don't know if they automatically solve the issues we were discussing in class or not, although I assume they would.

I'll be using **OpenBUGS** and an **R** package called **R2OpenBUGS** to run these simulations.

(c) Box & Tiao also studied a 3-component model (Table 5.3.1).
   i. Derive central credible intervals, for the 3 individual components, based on Table 5.3.3.
   ii. Use WinBUGS to do the same, and include graphs.
   iii. While not going as far as Box & Tiao's Figure 5.3.2, produce a graph of the joint posterior of σ_2^2 and σ_3^2, and one of σ_2^2/σ_3^2

Here is where my answers started to diverge somewhat more drastically from those reported by Box and Tiao, which I will discuss further on.

First, let's confirm the results in the text.

Next, the model for simulating in BUGS.

(d) Box, Hunter & Hunter (1976, Chapter 17.3) studied a pigment paste example with three components, focusing on point estimates only. Use WinBUGS again to perform a Bayesian analysis. Include graphs.

Again, any issues that are present in the model above will be reproduced here.