Statistical Modelling: Exercises

A. C. Davison and J. J. Forster

April 11, 2008

The work provided here is intended to take students up to half a week to complete. Students should talk to their supervisors to find out whether or not their department requires this work as part of any formal accreditation process (APTS itself has no resources to assess or certify students). It is anticipated that departments will decide on the appropriate level of assessment locally, and may choose to drop some (or indeed all) of the parts, accordingly.

0. If you have not already done so, complete the three APTS week practical sessions.

1. In a Bayesian context, write the marginal density of the observed data y as

$$\Pr(y) = \int f(y \mid \theta)\pi(\theta)\, d\theta = \int \exp\{-h(\theta)\}\, d\theta,$$

say, and suppose that the log likelihood is $O(n)$, and the log prior is $O(1)$. Apply the Laplace approximation to the integral, and show that if $O(1)$ terms are neglected, then

$$-2\log \Pr(y) \;\doteq\; \mathrm{BIC} = -2\log f(y \mid \widehat{\theta}) + p\log n, \quad n \to \infty,$$

where p is the dimension of θ, assumed fixed.

2. The data frame `bacteria` are discussed in Chapter 10 of *Modern Applied Statistics with S (4th edition)* by Venables and Ripley (Springer, 2002). They are available in R by loading the library `MASS`. The response `y` indicates presence or absence of a particular bacteria when assessed on 50 individuals (`ID`) at each of up to 6 time points (`week`). Each individual has received one of three treatments (`trt`: `placebo`/`drug`/`drug+`).

   Model the dependence of `y` on `trt` and `week` using binary GLMs and GLMMs (to account for intra-subject dependence in the response), fitted by maximum likelihood and associated approximations. Functions which you might wish to investigate for doing this include `glmmPQL` (from the `MASS` library), `glmmML` (from the library of the same name) and `lmer` (from the `lme4` library). Use the library documentation provided to learn about the required arguments of these functions. Compare the inferences obtained by different fitting methods (quadrature, Laplace, PQL).

3. For the `bacteria` data Venables and Ripley (2002, p297) propose the binary GLMM with

$$Y_{ij} \sim \text{Bernoulli}(\mu_{ij}), \quad g(\mu_{ij}) = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_2 x_{3ij} + b_{0i}, \quad b_{0i} \sim N(0, \sigma_b^2)$$

   where X_1, X_2, X_3 are the three binary explanatory variables $I(\texttt{trt} = \texttt{drug})$, $I(\texttt{trt} = \texttt{drug+})$, and $I(\texttt{week} > 2)$) and g is the logit link function.

   If g is the probit link (Φ^{-1}), then a Bayesian analysis of this model, using a Gibbs sampler, is straightforward, utilising the following latent variable formulation (also described briefly on slide 163 of the lecture notes): The GLMM above (with $g = \Phi^{-1}$) is equivalent to

$$Y_{ij} = I(Z_{ij} > 0), \quad Z_{ij} \sim \mathrm{N}(\mu_{ij}, 1), \quad \mu_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_2 x_{3ij} + b_{0i}, \quad b_{0i} \sim N(0, \sigma_b^2)$$

   where the Z_{ij} are latent continuous-valued variables, one for each observed Y_{ij}.

(a) Establish the equivalence above, and draw the DAG for the latent variable model, and the corresponding undirected conditional independence graph for the vertices (Y, Z, β, b, σ^2).

As the Z_{ij} are unobserved, they can also be generated in any Gibbs sampler scheme. It is immediately obvious that, given Z, the conditional distributions for β, b, σ_b^2 are exactly as for a LMM (with known error variance $\sigma^2 = 1$). Hence, a Gibbs sampler for this GLMM can be obtained by a straightforward modification of our LMM Gibbs sampler from Practical 2. We simply need to generate the Z_{ij} at each step.

(b) Show that the conditional distribution for $Z_{ij}|Y, \beta, b, \sigma^2$ is $N(\mu_{ij})$, restricted to the range $(0, \infty)$ when $Y_{ij} = 1$, or the range $(-\infty, 0]$ when $Y_{ij} = 0$.

A (not particularly efficient) way of generating a single $N(\mu, \sigma^2)$ variable restricted to the range (a, b) in R is

```
mu+sigma*qnorm(runif(1,pnorm(a,mu,sigma),pnorm(b,mu,sigma)))
```

(c) Modify the R function you used for an LMM Gibbs sampler in Practical 2, to perform a Bayesian analysis of the model above. Use the initial diffuse priors $\beta_i \overset{\text{ind}}{\sim} N(0, 25)$ and $\sigma^{-2} \sim \text{Gamma}(0.01, 0.01)$. It is reasonable to suppose *a priori* that the probability of bacteria presence decreases over time. Perform an alternative analysis with the more informative prior distribution $\beta_3 \sim N(-2, 4)$. How are your results affected?

(d) Compare your results with the logit model results obtained by maximum likelihood. [Note that, if g_1 and g_2 are logit and probit links respectively, then linear approximation gives $g_1 \approx 4g_2/(2\pi)^{1/2}$.]

4. In the context of the EM algorithm, use the fact that $\int f(u \mid y; \theta)\, du = 1$ for all y and θ to show that

$$\begin{aligned} 0 &= \mathrm{E}\left\{ \frac{\partial \log f(U \mid Y; \theta)}{\partial \theta} \middle| Y = y; \theta \right\}, \\ 0 &= \mathrm{E}\left\{ \frac{\partial^2 \log f(U \mid Y; \theta)}{\partial \theta \partial \theta^{\mathrm{T}}} + \frac{\partial \log f(U \mid Y; \theta)}{\partial \theta} \frac{\partial \log f(U \mid Y; \theta)}{\partial \theta^{\mathrm{T}}} \middle| Y = y; \theta \right\}. \end{aligned}$$

Now establish that

$$\frac{\partial \ell(\theta)}{\partial \theta} = \left. \frac{\partial Q(\theta; \theta')}{\partial \theta} \right|_{\theta' = \theta}, \quad \frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta^{\mathrm{T}}} = \left. \left\{ \frac{\partial^2 Q(\theta; \theta')}{\partial \theta \partial \theta^{\mathrm{T}}} + \frac{\partial^2 Q(\theta; \theta')}{\partial \theta \partial \theta'^{\mathrm{T}}} \right\} \right|_{\theta' = \theta}. \tag{1}$$

Deduce that even if $\ell(\theta)$ is inaccessible, its derivatives may be obtained from those of $Q(\theta; \theta')$ and used in a generic maximization algorithm. The second of these formulae also provides standard errors for the maximum likelihood estimate $\widehat{\theta}$ when $Q(\theta; \theta')$ is known but $\ell(\theta)$ is not.

Check this in the special case of the negative binomial example of the lectures, and hence give the Newton–Raphson step for maximization of the observed-data log likelihood, even though $\ell(\theta)$ itself is unknown. Write a program to compare the convergence of the EM and Newton–Raphson algorithms in that example.