Bayesian computation lab
Astroinformatics Summer School, Penn State University, June 2018.
Instructor: Murali Haran

Suppose observations Y_i, $i = 1, \ldots, n$ are modeled as a linear function of a predictor X_i, with two additive sources of error, one Gaussian with mean 0 and variance τ, and the other exponential with parameter λ. Given a data set $(X_1, Y_1), \ldots, (X_n, Y_n)$, we are interested in estimating β (regression coefficient relating Y's to X's), and the error parameters τ, λ. The Y_is are called exponentially modified Gaussian random variables (EMG). An EMG(μ, τ, λ) random variable is continuous with density

$$f(x; \mu, \tau, \lambda) = \frac{\lambda}{2} \exp\left(\frac{\lambda}{2}(2\mu + \lambda\tau^2 - 2x)\right) \operatorname{erfc}\left(\frac{\mu + \lambda\tau^2 - x}{\sqrt{2}\tau}\right),$$

and erfc is the complementary error function defined as

$$\operatorname{erfc}(x) = \frac{2}{\pi} \int_x^\infty e^{-t^2} dt.$$

Suppose we have independent observations Y_i with fixed (nonrandom) X_i that satisfy $Y_i \sim EMG(\beta X, \tau, \lambda)$, $i = 1, .., n$. These data may be obtained from
http://personal.psu.edu/muh10/expregdat.txt

1. Assume that τ is known to be 2. We are interested in Bayesian inference for β, λ. Let the independent priors for β, λ be $p(\beta), p(\lambda)$ respectively where $p(\beta) \propto 1$ ("flat prior") and $p(\lambda)$ is Gamma$(a = 1, b = 10)$ so $p(\lambda) = \frac{1}{\Gamma(a)b^a}\lambda^{a-1}e^{-\lambda/b}$. The resulting posterior distribution,

$$\begin{aligned}
\pi(\beta, \lambda \mid \mathbf{Y}) &\propto \prod_{i=1}^n f(Y_i \mid \beta, \lambda) p(\lambda) p(\beta) \\
&\propto \prod_{i=1}^n \frac{\lambda}{2} \exp\left(\frac{\lambda}{2}\left(2X_i\beta + \lambda\tau - 2Y_i\right)\right) \operatorname{erfc}\left(\frac{X_i\beta + \lambda\tau - Y_i}{\sqrt{2\tau}}\right) p(\beta)p(\lambda) \\
&\propto \lambda^n \prod_{i=1}^n \operatorname{erfc}\left(\frac{X_i\beta + \lambda\tau - Y_i}{\sqrt{2\tau}}\right) \times \exp\left(\frac{\lambda}{2}\sum_{i=1}^n (2X_i\beta + \lambda\tau - 2Y_i)\right) p(\beta)p(\lambda)
\end{aligned}$$

Hence, full conditionals are:

$$\begin{aligned}
\pi(\beta \mid \lambda, \mathbf{Y}) &\propto \lambda^n \prod_{i=1}^n \operatorname{erfc}\left(\frac{X_i\beta + \lambda\tau - Y_i}{\sqrt{2\tau}}\right) \times \exp\left(\frac{\lambda}{2}\sum_{i=1}^n (2X_i\beta + \lambda\tau - 2Y_i)\right) p(\beta)p(\lambda) \\
\pi(\lambda \mid \beta, \mathbf{Y}) &\propto \lambda^n \prod_{i=1}^n \operatorname{erfc}\left(\frac{X_i\beta + \lambda\tau - Y_i}{\sqrt{2\tau}}\right) \times \exp\left(\frac{\lambda}{2}\sum_{i=1}^n (2X_i\beta + \lambda\tau - 2Y_i)\right) p(\beta)p(\lambda)
\end{aligned}$$

On log scale,

$$\log(\pi(\beta \mid \lambda, \mathbf{Y})) = \sum_{i=1}^{n} \text{logerfc}\left(\frac{X_i\beta + \lambda\tau - Y_i}{\sqrt{2\tau}}\right) + \left(\frac{\lambda}{2}\sum_{i=1}^{n}(2X_i\beta + \lambda\tau - 2Y_i)\right) + \log(p(\beta))$$

$$\log(\pi(\lambda \mid \beta, \mathbf{Y})) = n\log(\lambda) + \sum_{i=1}^{n} \text{logerfc}\left(\frac{X_i\beta + \lambda\tau - Y_i}{\sqrt{2\tau}}\right) + \left(\frac{\lambda}{2}\sum_{i=1}^{n}(2X_i\beta + \lambda\tau - 2Y_i)\right) + \log(p(\lambda))$$

2. Approximate (i) $E_\pi(\lambda)$, (ii) $E_\pi\left(\frac{1}{\beta+\lambda}\right)$, (iii) the posterior correlation between β and λ.

3. Take a look at a histogram and then a smoothed histogram (plot(density(x)) to get an idea of the marginal posterior distributions.

4. Use the R package `mcmcse` and the command `mcse` to find the standard errors of the Monte Carlo approximations of the parameters, as well as the effective sample size (ESS) for each parameter.

5. Determine an appropriate length of the Markov chain

6. Go through the MCMC checklist from the tutorial today. You may have to modify and re-run the algorithm. In particular, try at least 3 different initial values and run the chains for long enough to see if they produce similar results.

7. Once you are confident you are have reliable answers, you are ready to also estimate τ. Assume the same prior for τ as for λ. For this you will need to augment the MCMC code you have been provided. Repeat the above exercise.

8. Suppose your collaborator changes her mind and says that the prior for λ should be Gamma(2,5) instead. Without re-running your MCMC algorithm, use importance sampling to reweight the Markov chain you already have in order to approximate the new posterior expectations of β, λ, τ.

9. Return to the two-dimensional case (where τ is known). Now use a Laplace approximation to approximate the expectations of each parameter. You may find it useful to use the `optim` command in order to do an optimization for the Laplace approximation. Please note that the default of optim is to minimize, not maximize, so you will need to either change the controls, or use the negative of the posterior distribution.