PROBLEM SET 2
STAT 221

WON I. LEE

1. Posterior and Likelihood

The joint posterior of the model is given by the following:

$$p(\mu, \sigma^2, \log\theta|Y, w) \propto p(Y|\mu, \sigma^2, \log\theta, w)p(\mu, \sigma^2, \log\theta|w)$$

where we treat w as constant and fixed (thus conditioned on) throughout the calculation. We have:

$$p(Y|\mu, \sigma^2, \log\theta, w) = \prod_{j=1}^{J}\prod_{n=1}^{N} p(Y_{jn}|\mu, \sigma^2, \log\theta_j, w_j) = \prod_{j=1}^{J}\prod_{n=1}^{N} \mathrm{Pois}(Y_{jn}|w_j\theta_j)$$

and:

$$p(\mu, \sigma^2, \log\theta|w) = p(\log\theta|\mu, \sigma^2, w)p(\mu, \sigma^2|w) \propto \frac{1}{\sigma^2}\prod_{j=1}^{J}\mathcal{N}(\log\theta_j|\mu, \sigma^2)$$

Putting these together, we have:

$$p(\mu, \sigma^2, \log\theta|Y, w) \propto \frac{1}{\sigma^2}\prod_{j=1}^{J}\mathcal{N}(\log\theta_j|\mu, \sigma^2)\prod_{n=1}^{N}\mathrm{Pois}(Y_{jn}|w_j\theta_j)$$

The conditional posterior is given by:

$$p(\log\theta_j|Y, w, \mu, \sigma^2) \propto p(Y_{j\cdot}|\log\theta_j, Y_{-j,\cdot}, w, \mu, \sigma^2)p(\log\theta_j|Y_{-j,\cdot}, w, \mu, \sigma^2)$$

where $Y_{j\cdot}$ indicates $Y_{j1}, \ldots, Y_{jN}$ and $Y_{-j,\cdot}$ indicates all other values of Y (i.e. all other columns). We note that by the presumed i.i.d. assumptions:

$$p(Y_{j\cdot}|\log\theta_j, Y_{-j,\cdot}, w, \mu, \sigma^2) = p(Y_{j\cdot}|\log\theta_j, w) = \prod_{n=1}^{N}\mathrm{Pois}(Y_{jn}|w_j\theta_j)$$

and:

$$p(\log\theta_j|Y_{-j,\cdot}, w, \mu, \sigma^2) = p(\log\theta_j|\mu, \sigma^2) = \mathcal{N}(\log\theta_j|\mu, \sigma^2)$$

so putting this together:

$$p(\log\theta_j|Y, w, \mu, \sigma^2) \propto \mathcal{N}(\log\theta_j|\mu, \sigma^2)\prod_{n=1}^{N}\mathrm{Pois}(Y_{jn}|w_j\theta_j)$$

The conditional posterior is unimodal, as verified by both analytical observation and numerical visualization. That is, we substituted reasonable values for the parameters and computed the posterior numerically at grid-points of $\log\theta_j$ to verify that the posterior is unimodal. It is also log-concave, since taking derivatives twice of the log-posterior according to $\log \theta_j$ yields:

$$\frac{\partial^2}{\partial \theta_j^2} \log p(\log \theta_j | Y_{-j,\cdot}, w, \mu, \sigma^2) = -\sigma^{-2} - n w_j \theta_j < 0$$

2. Simulate Data

Please see `wonlee_ps2_functions.R` to see the desired function.

3. Evaluate Coverage (Simple)

We first started by designing the simulation as requested, with 50 values of Y for each θ value to start. After some preliminary testing, we found that with the requested parameters, i.e. $J = 1000$, $N = 2$, and using 50 data points, we required an average system time of approximately 1200 seconds to perform a simulation for a fixed parameter setting μ, σ^2 and for one draw of $\log \theta$. Thus, we estimated that we could run ≈ 3 such simulations in 1 hour of clock time. Thus, in each job we do the following.

(1) Draw 3 values of θ (for $J = 1000$).
(2) Draw 50 values of Y (for $J = 1000, N = 2$).
(3) Run MCMC to obtain posterior samples.
(4) Compute the lower/upper bounds for our confidence interval.
(5) Compute the number of times each θ value is in the interval.

We do this for 5 times for each set of parameter values (4), resulting in 20 jobs that yield 15 draws of θ for each parameter setting and 50 draws of Y for each θ.

4. Evaluate Coverage with Exposure Weights

5. Evaluate Coverage with Misspecification

6. Interpretation and Results