PUBH 7440: Intro to Bayesian Analysis
Midterm (Take-Home Portion) — Due March 14

Incidence of low weight births in PA: [Insert text saying why looking at the incidence of low birth weight is important]. Here, we let y_{ir} denote the number of low weight births from mothers of race r ($r = 1$ white, $r = 2$ black) in county i out of a total of n_{ir} births. To model these data, we will assume:

$$y_{ir} \sim \text{Bin}\left(n_{ir}, \pi_{ir}\right), \text{ where logit}\left(\pi_{ir}\right) = \theta_{ir} \sim \text{Norm}\left(\beta_{0r}, \sigma_r^2\right),$$

and where π_{ir} represents the incidence rate. Assuming standard priors for $\beta_{0r} \sim \text{Norm}\left(0, \tau^2\right)$ and $\sigma_r^2 \sim \text{IG}\left(0.001, 0.001\right)$, with $\tau^2 = 10{,}000$, answer the following questions:

1. Write the full hierarchical model.

2. Derive the full-conditional distributions for β_{0r}, π_{ir}, and σ_r^2. Which parameters have full-conditional distributions we can sample from directly, and which parameters require Metropolis steps to sample?

3. Write code to fit the model, and use $\beta_{0r} = 0$ and $\sigma_r^2 = 1$ as initial values.

   - Make history plots of β_{0r} and σ_r^2 for both races and assess model convergence. Is burn-in required? If so, how much?

4. Suppose we're interested in investigating racial disparities in the incidence of low weight births. Using the β_{0r} terms, make a histogram of the posterior distribution of the log odds ratio. Does this indicate evidence of a "significant" racial disparity? (Hint: The log odds ratio is represented by γ_1 in the conventional regression model parameterization, $E\left[\theta_{ir} \,|\, \boldsymbol{\gamma}, \sigma_r^2\right] = \gamma_0 + \gamma_1 \times (r-1)$ where $r = 1, 2$, so you'll need to first write γ_1 as a function of the β_{0r} parameters.)

5. Now suppose we're interested in *geographic* trends in the incidence of low weight births by race and in their racial disparities. Using the mapping code from HW3/HW4, make the following maps:

   - The incidence of low weight births for white mothers.
   - The incidence of low weight births for black mothers.
   - The black/white ratio of the incidence of low weight births.

6. Finally, make histograms of posterior distribution of the black/white ratio of the incidence of low weight births in Philadelphia County ($i = 51$) and Sullivan County ($i = 57$) and compare these to their respective crude estimates (i.e., the ratio of the crude incidence rates, y_{ir}/n_{ir}, for black and white mothers in both counties) and the statewide averages (i.e, the ratio of $\sum_i y_{ir} / \sum_i n_{ir}$ for black and white mothers). Are the posterior distributions consistent with either/both of these estimates based on the data? From a statistical perspective, would you have any reservations about presenting these results?