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?

General Details:

1) we have $i = 1, 2, \ldots, 67$ indexing counties

2) within each county we have two racial groups

$$y_{ir} \sim Bin(n_{ir}, \pi_{ir})$$

$$logit(\pi_{ir}) = \log\left(\frac{\pi_{ir}}{1-\pi_{ir}}\right) = \theta_{ir} \sim Norm(\beta_{0r}, \theta_r^2)$$

$$\pi_{ir} = \frac{\exp(\theta_{ir})}{1+\exp(\theta_{ir})}$$

3) we have the same mean and variance for all θ_{ir} within a fixed r.

4) So, we need to estimate:

2 total β_{0r}

2 total θ^2_r

67×2 total θ_{ir}

Problem 1:

full model:

$$p(\theta_{ir}, \beta_{0r}, \sigma^2_r \mid y_{ir}, n_{ir}) \propto$$

$$\prod_{i=1}^{N=67} \prod_{j=1}^{r=2} \text{Bin}(n_{ir}, \pi_{ir}) \times$$

$$\text{Norm}(\beta_{0r}, \sigma^2_r) \times$$

$$\text{Norm}(0, \tau^2) \times$$

$$IG(0.001, 0.001)$$

$$\tau^2 = 10.000$$

Problem 2

1) Full conditional β_{0r}.

we have a total of 2 β_{0r} for $r=1$ and $r=2$

- β_{0r} appears in a distribution of θ_{ir}.

- for each β_{0r} we have 67 θ_{ir}'s within each level of r.

$$P(\beta_{0r} \mid \cdot) \propto \left[\prod_{i=1}^{67} \text{Norm}(\theta_{ir} \mid \beta_{0r}, \sigma_r^2)\right] \times$$

$$\text{Norm}(\beta_{0r} \mid 0, \tau^2) =$$

$$(2\pi\sigma_r^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_r^2} \sum(\theta_{ir} - \beta_{0r})^2\right) \times$$

$$\times (-2\pi\tau^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\tau^2}\beta_{0r}^2\right) \propto$$

$$\exp\left(-\frac{1}{2\theta^2}\sum(\theta_{ir}-\beta_{0r})^2-\frac{1}{2\tau^2}\beta_{0r}^2\right)=$$

$$\exp\left(-\frac{1}{2\theta^2}\cancel{\sum\theta_{ir}^2}+\frac{1}{2\theta^2}\cdot 2\beta_{0r}\sum\theta_{ir}-\frac{1}{2\theta^2}\sum\beta_{0r}^2-\frac{1}{2\tau^2}\beta_{0r}^2\right)\propto$$

$$\exp\left(\frac{\beta_{0r}}{\theta^2}\sum\theta_{ir}-\frac{1}{2}\beta_{0r}^2\left(\frac{n}{\theta^2}+\frac{1}{\tau^2}\right)\right)$$

After considering Metropolis,

I worked out that this update can be done w/ normal distribution.

Details Below ↓

Alternatively for β or consider a general normal distribution:

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \propto$$

$$\exp\left(-\frac{1}{2\sigma^2}\left(x^2 - 2x\mu + \mu^2\right)\right) \propto$$

$$\exp\left(-\frac{1}{2\sigma^2}x^2 + \frac{1}{\sigma^2}x \cdot \mu\right)$$

This is exactly what we have above:

Change notation of original distribution:

$\sigma \longrightarrow s$, Then, by matching the terms we get

$$\begin{cases} \frac{1}{2\sigma^2} = \frac{1}{2}\left(\frac{n}{s^2} + \frac{1}{\tau^2}\right), \\ \frac{\mu}{\sigma^2} = \frac{\sum \theta_{ir}}{s^2}. \end{cases}$$

$$\frac{1}{\sigma^2} = \left(\frac{n}{s^2} + \frac{1}{\tau^2}\right) \Rightarrow \sigma^2 = \frac{1}{\left(\frac{n}{s^2} + \frac{1}{\tau^2}\right)}$$

Plug this into the second equation in the system:

$$\frac{\mu}{\theta^2} = \mu \times \left(\frac{n}{s^2} + \frac{1}{\tau^2}\right) = \frac{\sum \theta_{ir}}{s^2},$$

$$\mu = \frac{\sum \theta_{ir}}{s^2\left(\frac{n}{s^2} + \frac{1}{\tau^2}\right)} = \frac{\sum \theta_{ir}}{n + s^2/\tau^2}$$

Switch notation back to $s \to \theta$

Potentially we can sample β_{0r} from $N\left(\frac{\sum \theta_{ir}}{n + \theta^2/\tau^2}, \frac{1}{\left(\frac{n}{\theta^2} + \frac{1}{\tau^2}\right)}\right)$

2) Full conditional for π_{ir}

$$Y_{ir} \sim \binom{n_{ir}}{y_{ir}} \pi_{ir}^{y_{ir}} (1-\pi_{ir})^{n_{ir}-y_{ir}}$$

$$\log\left(\frac{\pi_{ir}}{1-\pi_{ir}}\right) = \theta_{ir} \sim Norm(\beta_{0r}, \sigma_i^2)$$

$$\pi_{ir} = \frac{\exp(\theta_{ir})}{1+\exp(\theta_{ir})} \implies$$

$$\boxed{Y_{ir} \sim \binom{n_{ir}}{y_{ir}} \left[\frac{\exp(\theta_{ir})}{1+\exp(\theta_{ir})}\right]^{y_{ir}} \left[\frac{1}{1+\exp(\theta_{ir})}\right]^{n_{ir}-y_{ir}}}$$

π_{ir} has a deterministic relationship with θ_{ir}, so we obtain a full conditional for θ_{ir}, and obtain π_{ir} using expit(·) function

$$p(\theta_{ir} \mid \cdot) \propto$$

$$\cancel{\binom{n_{ir}}{y_{ir}}} \left[\frac{\exp(\theta_{ir})}{1+\exp(\theta_{ir})}\right]^{y_{ir}} \left[\frac{1}{1+\exp(\theta_{ir})}\right]^{n_{ir}-y_{ir}} \times$$

$$\exp\left(-\frac{1}{2}(\theta_{ir} - \beta_{0r})^2\right) =$$

$$\exp(\theta_{ir})^{y_{ir}} \times (1+\exp(\theta_{ir}))^{-y_{ir}} \times$$

$$(1+\exp(\theta_{ir}))^{-n_{ir}+y_{ir}} \times \exp\left(-\frac{1}{2}(\theta_{ir}-\beta_{0r})^2\right)$$

$$= \exp(\theta_{ir})^{y_{ir}} \times (1+\exp(\theta_{ir}))^{-n_{ir}} \times$$

$$\exp\left(-\frac{1}{2}(\theta_{ir} - \beta_{0r})^2\right)$$

Will use metropolis updates for this one

3) full conditional for θ^2

As known before, this should be a conjugate prior, so full conditional must be IG as well.

$$p(\theta^2 \mid \cdot) \propto \prod_{i=1}^{n=67} \text{Norm}(\theta_{ir} \mid \beta_{0r}, \theta^2) \times$$

$$\times IG(\theta^2 \mid 0.001, 0.001)$$

$$\propto (2\pi\theta^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\theta^2} \sum (\theta_{ir} - \beta_{0r})^2\right) \times$$

$$(\theta^2)^{-0.001-1} \times \exp(-0.001/\theta^2)$$

Let $\sum_i (\theta_{ir} - \beta_{0r})^2 = A$

$$\propto (\theta^2)^{-\frac{n}{2} - 0.001 - 1} \times$$

$$\exp\left(-\left\{\frac{1}{2} A\right\} \cdot \frac{1}{\theta^2} - 0.001/\theta^2\right)$$

$$= (\theta^2)^{-\left(\frac{n}{2}+0.001\right)-1} \times$$

$$\exp\left(-\left\{\frac{1}{2}A + 0.001\right\}\cdot\frac{1}{\theta^2}\right)$$

$$\Rightarrow p(\theta^2 \mid \cdot) \sim IG\left(\frac{n}{2}+0.001,\ \frac{1}{2}A+0.001\right)$$

Problem 3

Notes for code: we need to get ratios for Metropolis updates.

1) Ratios for β_{0r}

$r(\beta_{0r}') =$

$$\frac{\exp\left(\frac{\beta_{0r}^*}{\sigma^2}\sum\theta_{ir} - \frac{1}{2}\beta_{0r}^{*2}\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)\right)}{\exp\left(\frac{\beta_{0r}'}{\sigma^2}\sum\theta_{ir} - \frac{1}{2}\beta_{0r}'^2\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)\right)} =$$

$$\exp\left(\frac{\sum\theta_{ir}}{\sigma^2}\left(\beta_{0r}^* - \beta_{0r}'\right)\right) \times$$

$$\exp\left(-\frac{1}{2}\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)\left(\beta_{0r}^{*2} - \beta_{0r}'^2\right)\right)$$

On the log scale

$$\frac{\sum \theta_{ir}}{\sigma^2}\left(\beta_{0r}^{*} - \beta_{0r}'\right) - \frac{1}{2}\left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right)\left(\beta_{0r}^{*2} - \beta_{0r}'^{2}\right)$$

2) Ratios for θ_{ir}

$$r(\theta_{ir}^{*}) = \frac{\exp(\theta_{ir}^{*})^{y_{ir}} \times \left(1+\exp(\theta_{ir}^{*})\right)^{-u_{ir}} \times \exp\left(-\frac{1}{2}(\theta_{ir}^{*}-\beta_{or})^2\right)}{\exp(\theta_{ir}')^{y_{ir}} \times \left(1+\exp(\theta_{ir}')\right)^{-u_{ir}} \times \exp\left(-\frac{1}{2}(\theta_{ir}'-\beta_{or})^2\right)} =$$

$$= \left[\frac{\exp(\theta_{ir}^{*})}{\exp(\theta_{ir}')}\right]^{y_{ir}} \times \left[\frac{1+\exp(\theta_{ir}^{*})}{1+\exp(\theta_{ir}')}\right]^{-u_{ir}} \times$$

$$\times \exp\left(-\frac{1}{2}\left((\theta_{ir}^{*}-\beta_{or})^2 - (\theta_{ir}'-\beta_{or})^2\right)\right)$$

On the log scale:

$$y_{ir}(\theta_{ir}^{*} - \theta_{ir}^{i}) - n_{ir}\left[\log(1+\exp(\theta_{ir}^{*})) - \log(1+\exp(\theta_{ir}^{i}))\right] -$$

$$-\frac{1}{2}\left((\theta_{ir}^{*} - \beta_{0r})^2 - (\theta_{ir}^{i} - \beta_{0r})^2\right)$$

Problem 5

$$\text{logit}(\pi_{ir}) = \log\left(\frac{\pi_{ir}}{1-\pi_{ir}}\right) = \theta_{ir}.$$

$$E[\theta_{ir} \mid \beta_{0r}, \sigma^2_{0r}] = \beta_{0r}$$

$$E[\log(\pi_{ir}) - \log(1-\pi_{ir})] =$$

$$E[\log(\pi_{ir})] - E[\log(1-\pi_{ir})] = E[\theta_{ir}] = \beta_{0r}$$

$$E\left[\log\left(\frac{\pi_{ir}}{1-\pi_{ir}}\right)\right] = \beta_{0r}$$

let p_{0r} be the rate of events within group r.

Then, log odds ratio is given by

$$\log\left[\frac{\frac{p_{01}}{1-p_{01}}}{\frac{p_{02}}{1-p_{02}}}\right] = \log\left[\frac{p_{01}}{1-p_{01}}\right] - \log\left[\frac{p_{02}}{1-p_{02}}\right]$$

So this basically boils down to

$\underline{\beta_{01} - \beta_{02}}$. |