

# Final Exam for 206B

Qi Wang

2022/3/17

## Question a: Descriptive Statistics

First of all, this data is a longitudinal data, it describes the weight gained by pregnant women during a survey. So I reshaped the data into a wide form instead of a long form for further convenience. And also, I made a plot about the personal trend and overall trend for the women's weight gain in our survey.

For each woman, the weight gain can be plotted as in the figure 1. We can see that as time goes by, the weight gained by women are having a increasing trend for almost all the people in the survey. Also, I provided one overall trend between the average weight and time trend.. Since this data includes longitudinal properties, the mean and the summary of the data will be less informative, but the summary are provided as follows, the average or properties of time is not useful since they are similar to each other. Therefore, only information about weights are provided.

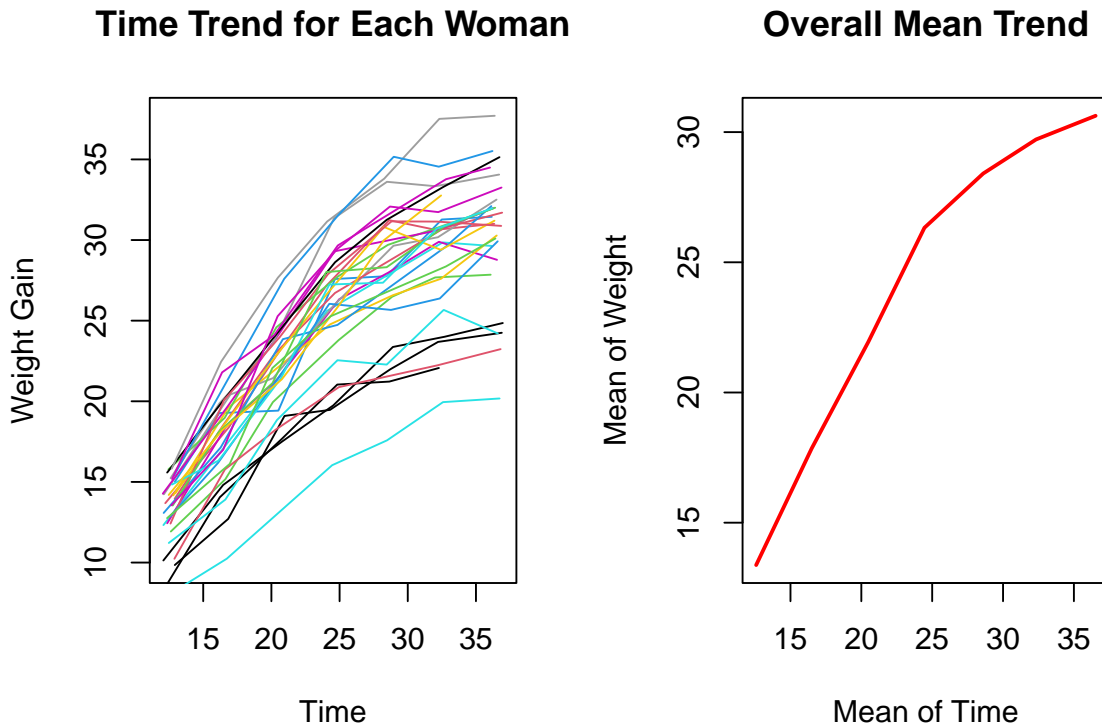


Figure 1: Time Trend of Weight Gain for Each Woman

|    |                |               |               |               |
|----|----------------|---------------|---------------|---------------|
| ## | weight.1       | weight.2      | weight.3      | weight.4      |
| ## | Min. : 9.846   | Min. :12.71   | Min. :17.49   | Min. :19.47   |
| ## | 1st Qu.:12.420 | 1st Qu.:16.22 | 1st Qu.:20.24 | 1st Qu.:25.05 |
| ## | Median :13.613 | Median :18.45 | Median :21.67 | Median :26.98 |
| ## | Mean :13.374   | Mean :17.85   | Mean :21.94   | Mean :26.33   |
| ## | 3rd Qu.:14.700 | 3rd Qu.:19.80 | 3rd Qu.:23.36 | 3rd Qu.:27.98 |
| ## | Max. :15.820   | Max. :22.45   | Max. :27.66   | Max. :31.35   |
| ## | weight.5       | weight.6      | weight.7      |               |
| ## | Min. :21.98    | Min. :23.68   | Min. :24.21   |               |
| ## | 1st Qu.:26.51  | 1st Qu.:27.62 | 1st Qu.:29.06 |               |
| ## | Median :28.15  | Median :30.40 | Median :31.10 |               |
| ## | Mean :28.41    | Mean :29.72   | Mean :30.63   |               |
| ## | 3rd Qu.:31.21  | 3rd Qu.:31.24 | 3rd Qu.:32.38 |               |
| ## | Max. :33.81    | Max. :37.51   | Max. :37.70   |               |

## Question b: Joint Posterior Distribution

I used the handwritten form for this question, and the notes has been attached at the end of the file.

For the algorithm part, it is a Gibbs sampler algorithm, but we need MCMC when sampling several parameters. I find the full conditional posterior distribution of  $\beta_{0i}$  and  $\bar{\beta}_0$  are a normal distribution, and the full conditional posterior distribution of  $\sigma^2$  and  $\tau^2$  are inverse gamma distribution, which are easy to generate directly from the program. However, the distribution of  $\beta_1$  and  $\beta_2$  are not in closed form, indicating that we have to use the MCMC method when we get samples from this distribution.

### Algorithm Discription:

#### Step 1: Specify an initializing value for all parameters

Give values for  $\beta_{0i}^{(1)}$ ,  $\beta_1^{(1)}$ ,  $\beta_2^{(1)}$ ,  $\bar{\beta}_0^{(1)}$ ,  $\sigma^{2(1)}$ ,  $\tau^{2(1)}$

#### Step 2: Use Metroplis Hasting with Gibbs and loop for i from 1 to 12000

Generate  $\beta_{0i}^{(i+1)}$ ,  $\bar{\beta}_0^{(i+1)}$ ,  $\sigma^{2(i+1)}$ ,  $\tau^{2(i+1)}$  directly from the full conditional posterior distribution. But for  $\beta_1^{(i+1)}$ ,  $\beta_2^{(i+1)}$ , I will use Metropolis Hasting algorithm here. My proposal distribution is a **Random Walk**:

$$q(\theta^{(i+1)}|\theta^{(i)}) \propto \exp\left(-\frac{(\theta^{(i+1)} - \theta^{(i)})^2}{2d^2}\right)$$

First, for  $\beta_1^{(i+1)}$  generation, use the proposal distribution to propose a new value and denote to  $\xi$ , then calculate the transit probability:

$$p = \min\left\{\frac{\pi(\xi)q(\theta^{(i)}|\xi)}{\pi(\theta^{(i)})q(\xi|\theta^{(i)})}, 1\right\}$$

Since this is a random walk,  $q(\theta^{(i)}|\xi) = q(\xi|\theta^{(i)})$ , therefore:

$$p = \min\left\{\frac{\pi(\xi)}{\pi(\theta^{(i)})}, 1\right\}$$

$\pi$  here is just the full conditional probability function of  $\theta$

Generate an index number, for example,  $k$  which follows a uniform distribution between 0 and 1. Compare the value of  $p$  between  $k$ :

$$\begin{aligned} p > k : \theta^{(i+1)} &= \xi \\ p \leq k : \theta^{(i+1)} &= \theta^{(i)} \end{aligned}$$

Then, same method for generate  $\beta_2^{(i+1)}$ . Until now, we have finished one update for all the parameters in a Gibbs sampler.

### Step 3: Burn-in and thinning

Since we have strong auto correlation after generating the raw samples from the posterior distribution, we have to do the burn-in and thinning.

**Burn-in: Discard the front 12000 samples**

**Thinning: For the rest 88000 samples, we get samples every 88 sample, then we get 1000 better samples in total.**

### Step 4: Make inference

Since we have the 1000 samples after the burn-in and thinning, we can make inference based on these samples, for example, the credible interval, posterior mode, posterior mean, median and so on.

## Question c: Hyperparameter Specification

I used this set of hyper parameters:  $\bar{\beta}_1 = 10$ ,  $\bar{\beta}_2 = -0.15$ ,  $u_1^2 = 100$ ,  $u_2^2 = 100$ ,  $\mu_0 = 30$ ,  $v^2 = 1000$ ,  $a_\sigma = 1$ ,  $a_\tau = 1$ ,  $b_\sigma = 1$ ,  $b_\tau = 1$ . We have many ways to specify the hyper parameter, in my method, I just used the empirical Bayesian. I used a simple random parameter for the mean, but I must set the variance to be large to make sure that the prior information is very weak. Then, I run my algorithm, then it did not work well, but I can approximately get a “proper” range of the hyper parameters by definition of the parameter. Then, I used the ones seems perform well to be my final parameter.

Furthermore, the length of a step for a proposal distribution should also be selected, i.e., the standard deviation of the random walk kernel. I get the standard deviation by the approximate posterior sampling by a random selected hyper parameter and then fixed it little by little to fix it with the greatest effective sample size. This is really hard in my project, I used almost 60% of time to get the standard deviation of the random walk kernel. I think I can use adaptive MH but I am not sure whether it is cheating since selecting the variance is also a core part of the MH algorithm. Finally, I chose the standard deviation of `beta_1` proposal to be 0.4 and 0.003 for `beta_2`.

## Question d: MCMC within Gibbs Coding

First, I will check for convergence, from the trace plot in figure 2, we can see that after the burn-in and thinning process, the chains mixed well.

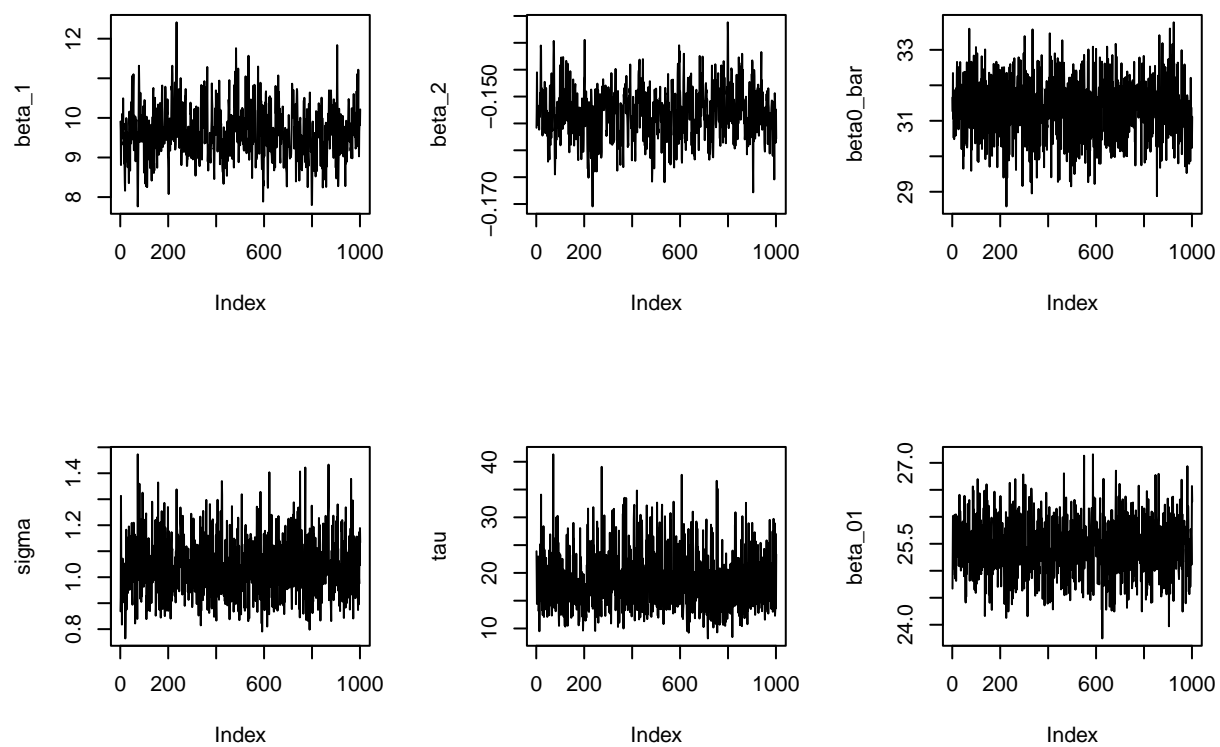


Figure 2: Trace Plot

## Question e: Interpretation

To interpret the parameters of the other parameters except  $\beta_{0i}$ , I simply focused on three concepts, one is posterior mean, the other two is 2.5% quantile and 97.5% quantile, for  $\beta_1, \beta_2, \beta_0, \sigma, \tau$ , I reported a posterior mean and a lower bound and an upper bound. They are got from the Monte Carlo simulation, since we got 1000 samples after burn-in and thinning the chain, the mean of them is just the mean of these samples. By LLN, it converges to the true posterior mean as sample size goes to infinity. And the 2.5% quantile and the 97.5% quantile can construct an interval with a lower bound and an upper bound, this is a 95% credible interval since they are uni-modal distribution (I checked the histogram for  $\beta_1$  and  $\beta_2$ ), so the center is the HPD. The credible interval means, the probability of the true parameter being in this interval is 0.95.

The interpretation for  $\beta_{0i}$  are a little different, because it is a longitudinal model, each person has its own  $\beta_0$ , so I presented the table about each  $\beta_{0i}$ . The number of each row corresponds to a specified person with the certain ID (the original order in the data set) at the beginning of this row. It could be interpreted as the posterior mean, 2.5% and 97.5 quantile of the posterior distribution of  $\beta_{0i}$  of patient  $i$ , and there is a 0.95 probability that the  $\beta_{0i}$  of patient  $i$  is in this interval.

Table 1: Summary of Beta\_0i

| Patient ID | Posterior Mean | 2.5% Quantile | 97.5% Quantile |
|------------|----------------|---------------|----------------|
| 1          | 25.44486       | 24.41281      | 26.43446       |
| 2          | 33.13857       | 32.17397      | 34.18924       |
| 3          | 29.00407       | 27.96091      | 30.01486       |
| 4          | 32.30057       | 31.23780      | 33.33045       |
| 5          | 31.16640       | 30.01770      | 32.30795       |
| 6          | 31.14530       | 30.07335      | 32.28220       |
| 7          | 30.78991       | 29.71451      | 31.86817       |
| 8          | 32.81696       | 31.78994      | 33.91261       |
| 9          | 24.39617       | 23.29745      | 25.57270       |
| 10         | 32.69151       | 31.56908      | 33.84870       |
| 11         | 33.81740       | 32.64108      | 35.04703       |
| 12         | 31.94747       | 30.75654      | 33.11219       |
| 13         | 20.35037       | 19.30680      | 21.37550       |
| 14         | 33.88521       | 32.63920      | 35.02479       |
| 15         | 32.80123       | 31.81788      | 33.86894       |
| 16         | 36.18852       | 35.08034      | 37.29795       |
| 17         | 35.44701       | 34.39122      | 36.50632       |
| 18         | 33.50106       | 32.44930      | 34.52689       |
| 19         | 32.71862       | 31.70113      | 33.72264       |
| 20         | 29.99048       | 28.99245      | 31.04918       |
| 21         | 32.20015       | 31.22083      | 33.23413       |
| 22         | 34.52761       | 33.48759      | 35.57734       |
| 23         | 32.88118       | 31.67867      | 34.14022       |
| 24         | 38.98269       | 37.90681      | 40.11786       |
| 25         | 24.83323       | 23.85656      | 25.90365       |
| 26         | 24.96274       | 23.87889      | 26.10766       |
| 27         | 30.72542       | 29.64242      | 31.86827       |
| 28         | 37.65172       | 36.60312      | 38.78132       |
| 29         | 26.21596       | 25.22183      | 27.23138       |
| 30         | 35.92422       | 34.72867      | 37.11669       |

Table 2: Other Parameters Summary

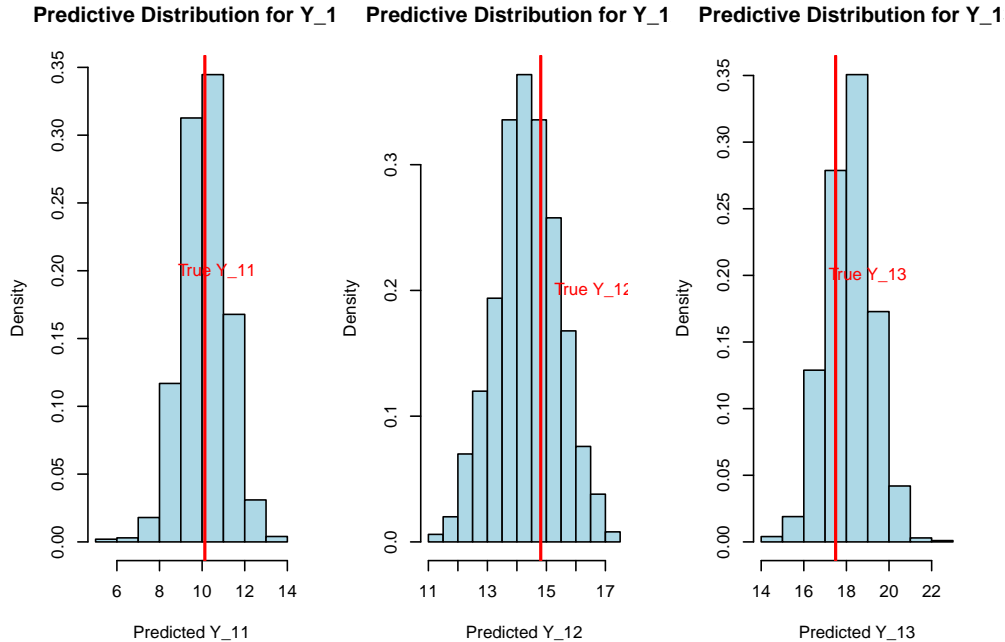
|            | Posterior Mean | 2.5% Lower Bound | 97.5% Upper Bound |
|------------|----------------|------------------|-------------------|
| beta_1     | 9.6289327      | 8.4156206        | 10.896241         |
| beta_2     | -0.1531215     | -0.1619737       | -0.143585         |
| beta_0 bar | 31.3489042     | 29.7064696       | 32.837960         |
| sigma      | 1.0403348      | 0.8477839        | 1.271401          |
| tau        | 18.4801297     | 10.7971255       | 30.768870         |

To interpret the parameters of the other parameters except  $\beta_{0i}$ , I simply focused on three concepts, one is posterior mean, the other two is 2.5% quantile and 97.5% quantile, for  $\beta_1, \beta_2, \bar{\beta}_0, \sigma, \tau$ , I reported a posterior mean and a lower bound and an upper bound. They are got from the Monte Carlo simulation, since we got 1000 samples after burn-in and thinning the chain, the mean of them is just the mean of these samples. By LLN, it converges to the true posterior mean as sample size goes to infinity. And the 2.5% quantile and the 97.5% quantile can construct an interval with a lower bound and an upper bound, this is a 95% credible interval since they are uni-modal distribution (I checked the histogram for  $\beta_1$  and  $\beta_2$ ), so the center is the HPD. The credible interval means, the probability of the true parameter being in this interval is 0.95.

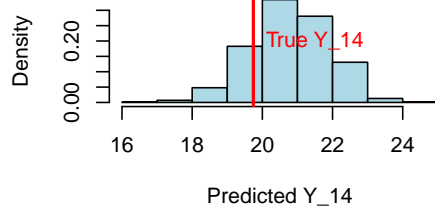
The interpretation for  $\beta_{0i}$  are a little different, because it is a longitudinal model, each person has its own  $\beta_0$ , so I presented the table about each  $\beta_{0i}$ . The number of each row corresponds to a specified person with the certain ID (the original order in the data set) at the beginning of this row. It could be interpreted as the posterior mean, 2.5% and 97.5 quantile of the posterior distribution of  $\beta_{0i}$  of patient  $i$ , and there is a 0.95 probability that the  $\beta_{0i}$  of patient  $i$  is in this interval.

## Question f: Pridictive distribution and model checking

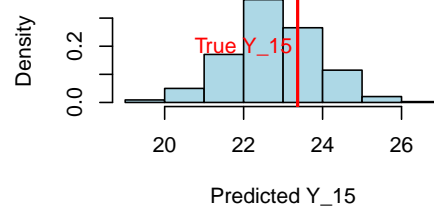
Here are plots supporting my models. My predictive distribution performs well for the time point 1, 2, 5, 6, 7, and a little away from the true value for the time point 4, but it is okay from the figure about the trend between value and time.



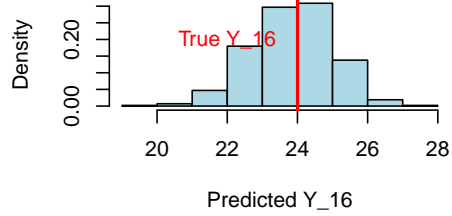
**Predictive Distribution for Y<sub>14</sub>**



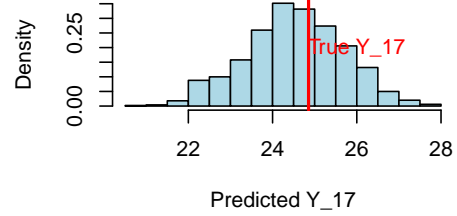
**Predictive Distribution for Y<sub>15</sub>**



**Predictive Distribution for Y<sub>16</sub>**



**Predictive Distribution for Y<sub>17</sub>**



**Comparisons Between True and Predictive Values**

