STAT 457 - FINAL

Martha Eichlersmith

2019-12-12

Problem 1

Problem 1a

For the data $Y = (125, 18, 20, 34)$, implement the Gibbs sampler algorithm. Use a flat prior on θ.
Plot θ^i versus iteration i. $Y = (y_1, y_2, y_3, y_4) \propto (2 + \theta, 1 - \theta, 1 - \theta, \theta)$
1. Draw a starting value, $t \sim$Uniform(0,1)
2. Draw a latent value, $Z \sim \text{Binomial}\left(y_1,\ \frac{\theta}{2+\theta}\right)$
3. Draw a parameter, $\theta \sim \text{Beta}(Z + y_4 + 1, y_2 + y_3 + 1)$

Problem 1b

Repeat 1a for $Y = (14, 0, 1, 5)$.

There is a lack of fit for the data in 1b, where the fit appears to be better for data in 1a. Convergence was assessed when values were had a difference less than 10^{-7}.

Problem 1c

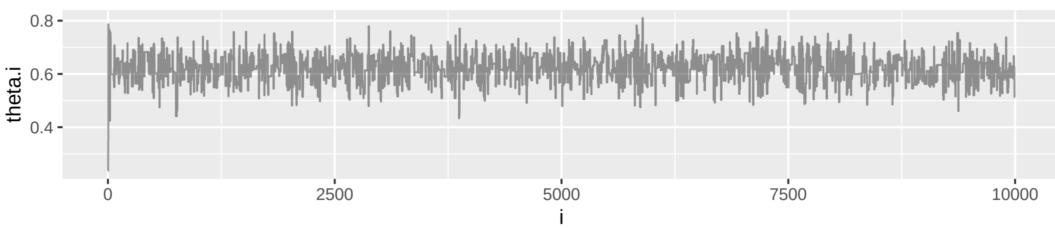
You would expected these values to be similar but it appears that our standard error average is under-estimating the variation in the chain means of θ in this case.

Problem 2

Problem 2a

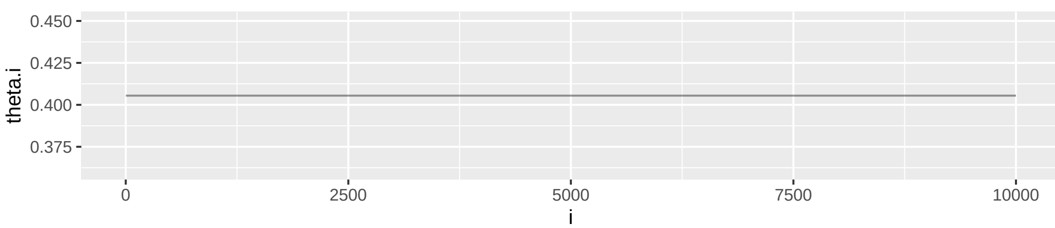
For the genetic linkage model applied to $Y = (125, 18, 20, 34)$, implement the Metropolis algorithm. (use a flat prior on θ). Use one long chain and plot θ^i versus i.
Try several driver functions:

Metropolis for data Y=(125,18,20,34) using the driver Uniform(0,1)



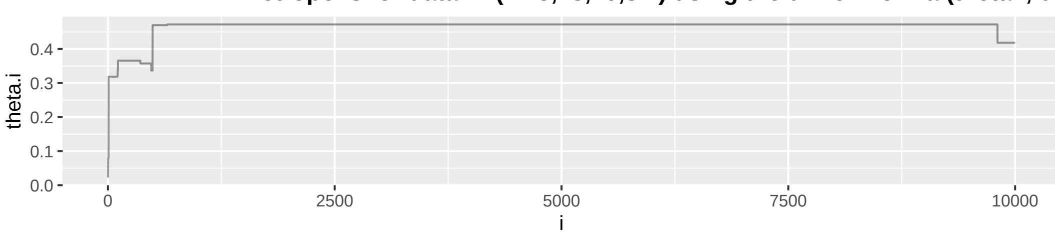
	Name	Value
1	Mean	0.62281
2	SD	0.05181
3	it	10000
4	Start	0.23480

Metropolis for data Y=(125,18,20,34) using the driver Normal(theta.i, 0.01)



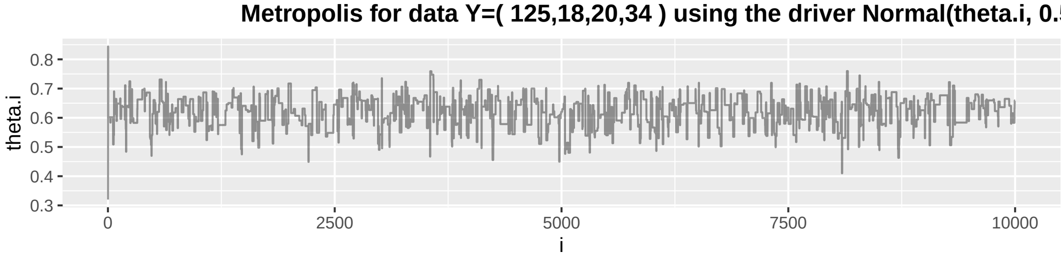
	Name	Value
1	Mean	0.40549
2	SD	0.00000
3	it	10000
4	Start	0.40549

Metropolis for data Y=(125,18,20,34) using the driver Normal(theta.i, 0.10)



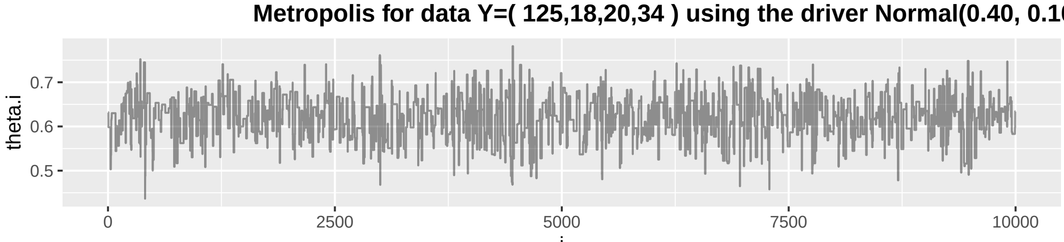
	Name	Value
1	Mean	0.46506
2	SD	0.02872
3	it	10000
4	Start	0.02234

Metropolis for data Y=(125,18,20,34) using the driver Normal(theta.i, 0.50)



	Name	Value
1	Mean	0.62252
2	SD	0.05229
3	it	10000
4	Start	0.31998

Metropolis for data Y=(125,18,20,34) using the driver Normal(0.40, 0.10)

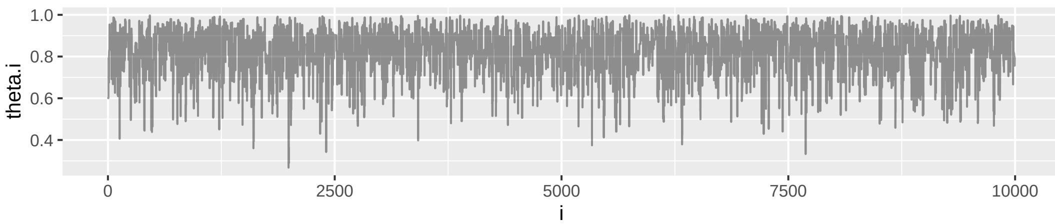


	Name	Value
1	Mean	0.62122
2	SD	0.04950
3	it	10000
4	Start	0.63208

Problem 2b

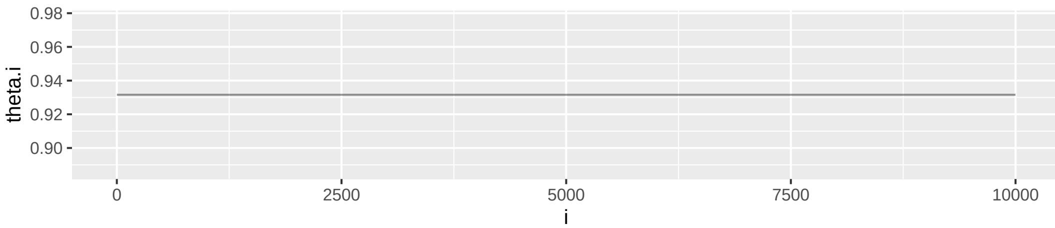
Repeat 2a for $Y = (14, 0, 1, 5)$

Metropolis for data Y=(14,0,1,5) using the driver Uniform(0,1)



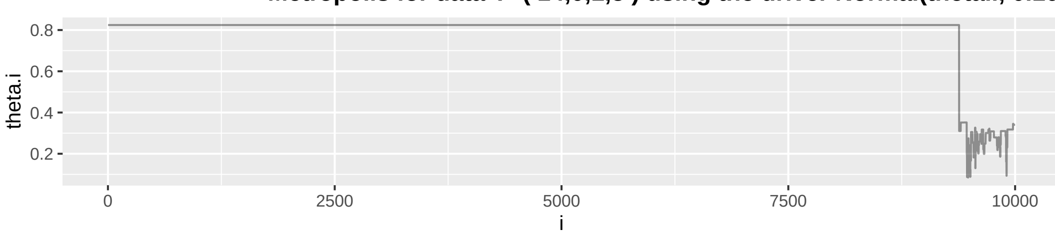
	Name	Value
1	Mean	0.82968
2	SD	0.10753
3	it	10000
4	Start	0.60549

Metropolis for data Y=(14,0,1,5) using the driver Normal(theta.i, 0.01)



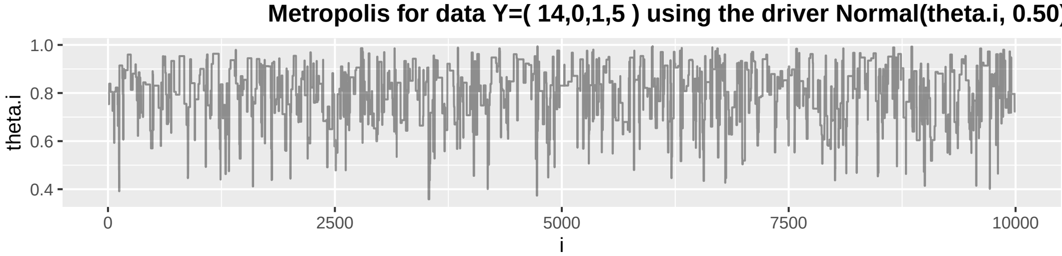
	Name	Value
1	Mean	0.93159
2	SD	0.00000
3	it	10000
4	Start	0.93159

Metropolis for data Y=(14,0,1,5) using the driver Normal(theta.i, 0.10)



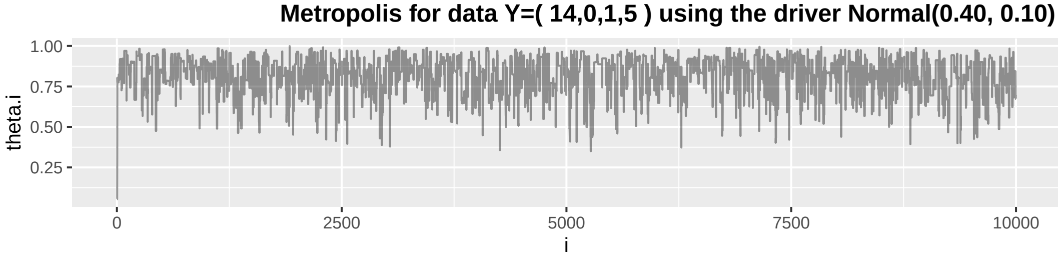
	Name	Value
1	Mean	0.79095
2	SD	0.13086
3	it	10000
4	Start	0.82435

Metropolis for data Y=(14,0,1,5) using the driver Normal(theta.i, 0.50)



	Name	Value
1	Mean	0.80984
2	SD	0.11698
3	it	10000
4	Start	0.75380

Metropolis for data Y=(14,0,1,5) using the driver Normal(0.40, 0.10)



	Name	Value
1	Mean	0.81881
2	SD	0.11488
3	it	10000
4	Start	0.05359

Problem 2c

Compute both the posterior mean and standard deviation for both data sets.
Compare to results from the previous problem.

In problem 1 the means were similar, but in problem 2 the means vary depending on the driver. Some of the means in problem 2 are close to the means in problem 1.

Problem 2d

Problem 2d(1)

For each of the drives in part 2a, run 20 chains with independent starting values.
Compute the averages of the θ's in each chain.

Problem 2d(2)

Calculate the standard deviation of the 20 averages. Interpret this value.

Problem 2d(3)

Compute the standard deviation of the θ's in each chain. Divide each SD by the square root of the number of iterations.
Average these "standard errors".

Problem 2d(4)

Compare the 2d(2) values to 2d(3). Would you expect these number to be similar or different?
Compare to the results of Exercise 1c.

Problem 3

Problem 3a

Consider the 1-way variance components model

$$Y_{ij} = \theta_i + \epsilon_{ij}$$

where Y_{ij} is the jth observation from the ith group, θ_i is the effect, ϵ_{ij}=error, $i = 1, ..., K$ and $j = 1, ..., J$. It is assume that $\epsilon_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$ and $\theta_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma_\theta^2)$. Under the prior specification $p(\sigma_\epsilon^2, \sigma_\theta^2, \mu) = p(\sigma_\epsilon^2)p(\sigma_\theta^2)p(\mu)$, with $p(\sigma_\theta^2) = \text{InverseGamma}(a_1, b_1)$, $p(\sigma_\epsilon^2) = \text{InverseGamma}(a_2, b_2)$, and $p(\mu) = \mathcal{N}(\mu_0, \sigma_0^2)$.
Let $\overline{Y}_i = \frac{1}{J}\sum_{j=1}^{J} Y_{ij}$ and $\theta = (theta_1, \cdots, \theta_k)$.
Show the following:

Problem 3a(1)

$$p(\mu \mid \theta, \sigma_\epsilon^2, \sigma_\theta^2, Y) = \mathcal{N}\left(\frac{\sigma_\theta^2\mu_0 + \sigma_0^2\sum\theta_i}{\sigma_\theta^2 + K\sigma_0^2},\ \frac{\sigma_\theta^2\sigma_0^2}{\sigma_\theta^2 + K\sigma_0^2}\right)$$

Problem 3a(2)

$$p(\theta_i \mid \mu, \sigma_\epsilon^2, \sigma_\theta^2, Y) = \mathcal{N}\left(\frac{J\sigma_\theta^2}{J\sigma_\theta^2 + \sigma_\epsilon^2} \cdot \overline{Y}_i + \frac{\sigma_\epsilon^2}{J\sigma_\theta^2 + \sigma_\epsilon^2} \cdot \mu, \quad \frac{\sigma_\theta^2\sigma_\epsilon^2}{J\sigma_\theta^2 + \sigma_\epsilon^2}\right)$$

Problem 3a(3)

$$p(\sigma_\epsilon^2 \mid \mu, \theta, \sigma_\theta^2, Y) = \text{InverseGamma}\left(a_2 + \frac{KJ}{2},\ b_2 + \frac{1}{2}\sum_{i=1}^{K}\sum_{j=1}^{J}(Y_{ij} - \theta_i)^2\right)$$

Problem 3a(4)

$$p(\sigma_\theta^2 \mid \mu, \theta, \sigma_\epsilon^2, Y) = \text{InverseGamma}\left(a_1 + \frac{K}{2},\ b_1 + \frac{1}{2}\sum_{i=1}^{K}(\theta_i - \mu)^2\right)$$

Problem 3b

Run the Gibbs sampler for the data below. Use one chain of length 75,000. Take $p(\mu) = \mathcal{N}(0, 10^{12})$, $p(\sigma_\epsilon^2) = IG(0, 0)$, and $p(\sigma_\theta^2) = IG(1, 1)$. For each θ_i, for σ_ϵ, and for θ_θ, plot the simulated value at iteration j versus j. Summarize each posterior marginal.

Problem 3c

Repeat 3b using the prior specification $p(\mu) = \mathcal{N}(0, 10^{12})$, $p(\sigma_\epsilon^2) = IG(0, 0)$, and $p(\sigma_\theta^2) = IG(0, 0)$. Does this specification violate the Hobart_Casella conditions? Describe what happens to the Gibbs sampler chain in this case.

Problem 5

Suppose that X and Y have exponential conditional distributions restricted over the interval $(0, B)$, i.e. $p(x \mid y) \propto y \exp\{-yx\}$ for $0 < x < B < \infty$ and $p(y \mid x) \propto x \exp\{-xy\}$ for $0 < y < B < \infty$, where B is known constant.

Problem 5a

Take $m = 1$ and $B = 3$. Run the data augmentation algorithm using these conditionals. (Hist: Reject the exponential deviates that lie outside $(0, B)$). How did you assess convergence of this chain? Obtain the marginal for x using the mixture of conditionals $p(x \mid y)$, mixed over the simulated y deviates in your chain.

Problem 5b

Show that the marginal for x is proportional to $(1 - \exp\{-Bx\})/x$. Compare your results in 5a to this curve.

Problem 5c

Repeat 5a and 5b using $B = \infty$. Describe what happens. Is the marginal for x a proper density in this case?