Solutions to extra credit problems in lab 4.

4.

$$p(\alpha,\beta|\underline{X}) \propto \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right)^{10}\left[\prod^{10}\frac{\Gamma(x_j+\alpha)\Gamma(n_j-x_j+\beta)}{\Gamma(\alpha+\beta+n_j)}\right]\alpha^{c-1}\exp(-\alpha d)\,\beta^{e-1}\exp(-\beta f)$$

In the R-code, I re-used (+modified) the same code that was given in *bikes.R* (I guess this was ok). However I encountered some problems, which were due to the updated proposal distribution and undefined logarithm values (log(<0)). I'm not sure that was it even purpose to use *bikes.R* with slight modifications, so my code might be wrong from the beginning. Anyway, I didn't have time to fix it. The only changes I made were to update the *ab.lprior*, *aprime* and *bprime*-functions.

5.

a)

$y_i \sim Bern(p_i)$
$logit(p_i) = \beta_0 + \beta_1 * logincome_i + \beta_2 * distance_i + \beta_3 * dropout_i + \beta_4 * college_i +$
$\beta_5 * group_i$
$\forall i \in [1, \cdots, 5] \, : \, \beta_i \sim N(0, 1000)$

b)

```
1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

        Mean      SD  Naive SE Time-series SE
b0 -1.835201 1.02483 0.0187107      0.0505936
b1  0.339143 0.14434 0.0026352      0.0064973
b2 -0.139797 0.08273 0.0015104      0.0018760
b3 -0.696892 0.45036 0.0082225      0.0085419
b4  1.735697 0.70319 0.0128385      0.0133433
b5 -0.002299 0.01181 0.0002157      0.0002477

2. Quantiles for each variable:

       2.5%      25%       50%       75%   97.5%
b0 -3.79408 -2.53117 -1.826000 -1.129390 0.18370
b1  0.05770  0.24109  0.339154  0.434290 0.61478
b2 -0.31125 -0.19472 -0.136056 -0.082585 0.01596
b3 -1.57676 -1.01448 -0.692069 -0.385753 0.15906
b4  0.42150  1.26107  1.699027  2.193413 3.18430
b5 -0.02643 -0.01013 -0.002183  0.005716 0.02033
```

Due to the output of the Gibbs sampler, especially the factors β_1, β_2 and β_5 are relatively small (in terms of mean and quantiles). Therefore the corresponding explanatory variables (*income*, *distance*, *group*) might not be very informative in this model.

6.

(6.) $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad , i = 1, \dots, n \quad , \varepsilon_i \sim N(0, \tau)$

$$P(\beta_0, \beta_1, \tau) \propto \frac{1}{\tau} \qquad n = 15$$

$$\therefore y_i \mid \beta_0, \beta_1, \tau, \underline{x} \sim N\left(\beta_0 + \beta_1 x_i , \frac{1}{\tau}\right)$$

Lkhd:

$$P(y \mid \beta_0, \beta_1, \tau, \underline{X}) = \prod^n P(y_i \mid \beta_0, \beta_1, \tau, x_i) = \prod^n \left(\sqrt{2\pi\tau^{-1}}\right)^{-\frac{1}{2}} \exp\left(\frac{1}{2\tau^{-1}} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2\right)$$

$$\propto \left(\sqrt{\tau^{-1}}\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\tau^{-1}} \sum (y_i - \beta_0 - \beta_1 x_i)^2\right)$$

$$= \tau^{\frac{n}{4}} \exp\left(-\frac{\tau}{2} \sum (y_i - \beta_0 - \beta_1 x_i)^2\right)$$

Joint

$$P(y, \beta_0, \beta_1, \tau, \underline{X}) = P(y \mid \beta_0, \beta_1, \tau, \underline{X})\, P(\beta_0, \beta_1, \tau) = P(\beta_0, \beta_1, \tau \mid y, \underline{x})$$

$$\propto \tau^{\frac{n}{4}} \tau^{-1} \exp\left(-\tau \tfrac{1}{2} \sum (y_i - \beta_0 - \beta_1 x_i)^2\right) = \tau^{\frac{n}{4}-1} \exp\left(-\tau \tfrac{1}{2} \sum (y_i - \beta_0 - \beta_1 x_i)^2\right)$$

Full-cond for τ

$$\tau \mid \beta_0, \beta_1, y, \underline{X} \sim \text{Gamma}\left(\frac{n}{4}, \left(\tfrac{1}{2} \sum (y_i - \beta_0 - \beta_1 x_i)^2\right)^{-1}\right)$$

Full-cond for β_0:

$$P(\beta_0 \mid \beta_1, \tau, y, \underline{X}) \propto \exp\left(-\tau \tfrac{1}{2} \sum (y_i - \beta_0 - \beta_1 x_i)^2\right)$$

$$\propto \exp\left(-\tau \tfrac{1}{2} \sum \beta_0^2 - 2y_i\beta_0 + 2\beta_0\beta_1 x_i\right) = \exp\left(-\tau \tfrac{1}{2} \sum \beta_0^2 - 2(y_i - \beta_1 x_i)\beta_0\right)$$

$$= \exp\left(-\tau \tfrac{1}{2} \left[n\beta_0^2 - 2\left[\sum (y_i - \beta_1 x_i)\right]\beta_0\right]\right)$$

$$= \exp\left(-\tfrac{1}{2} \tau n \left(\beta_0 - \tfrac{1}{n} \sum (y_i - \beta_1 x_i)\right)^2\right)$$

$$\therefore \beta_0 \mid \beta_1, \tau, y, \underline{x} \sim N\left(\frac{1}{n} \sum (y_i - \beta_1 x_i), \frac{1}{\tau n}\right)$$

Full-cond for β_1:

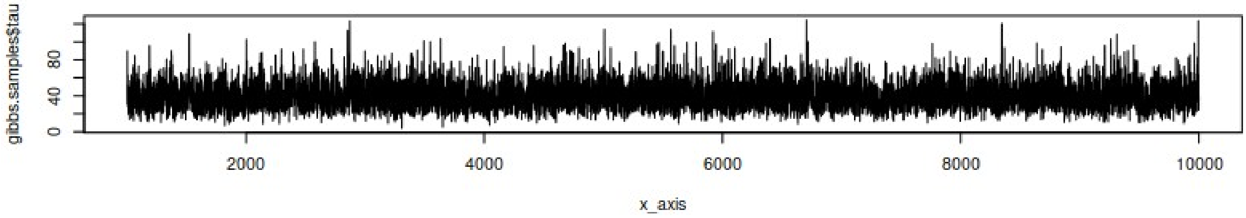
$$P(\beta_1 \mid \beta_0, \tau, y, \underline{x}) \propto \exp\left(-\tau \tfrac{1}{2} \sum (y_i - \beta_0 - \beta_1 x_i)^2\right)$$

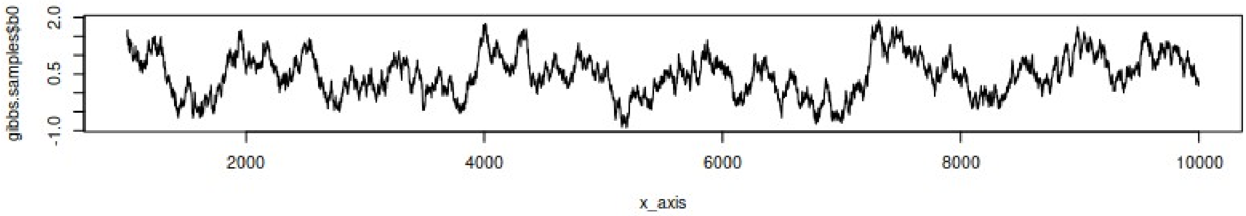
$$\propto \exp\left(-\tau \tfrac{1}{2} \sum \beta_1^2 x_i^2 - 2y_i x_i \beta_1 + 2\beta_0\beta_1 x_i\right) = \exp\left(-\tau \tfrac{1}{2} \sum \beta_1^2 x_i^2 - 2(y_i x_i - \beta_0 x_i)\beta_1\right)$$

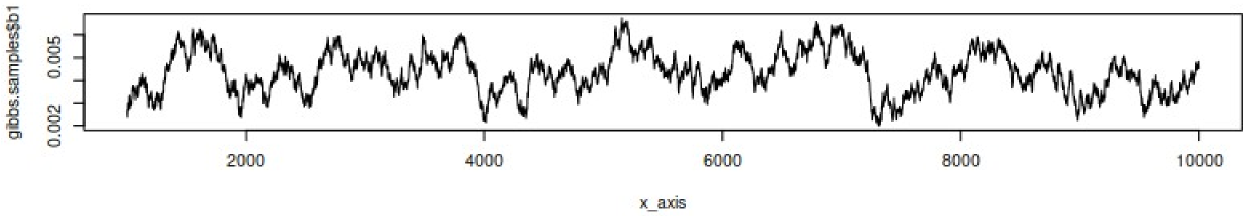
$$= \exp\left(-\tau \tfrac{1}{2} \left[\left(\sum x_i^2\right)\beta_1^2 - 2\left[\sum (y_i x_i - \beta_0 x_i)\right]\beta_1\right]\right)$$

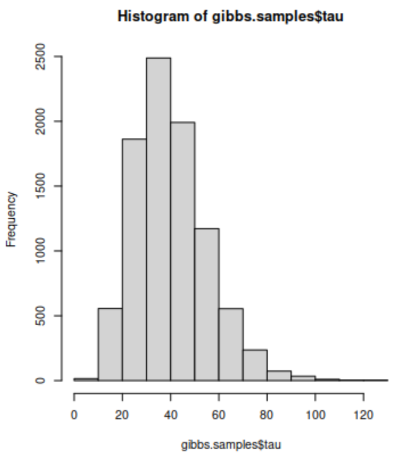
$$= \exp\left(-\tfrac{1}{2} \tau \left(\sum x_i^2\right)\left(\beta_1^2 - \frac{\sum (y_i x_i - \beta_0 x_i)}{\sum x_i^2}\right)^2\right)$$

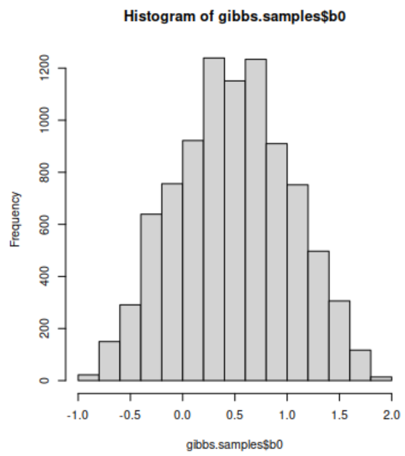
$$\therefore \beta_1 \mid \beta_0, \tau, y, \underline{x} \sim N\left(\frac{\sum (y_i x_i - \beta_0 x_i)}{\sum x_i^2}, \frac{1}{\tau \sum x_i^2}\right)$$

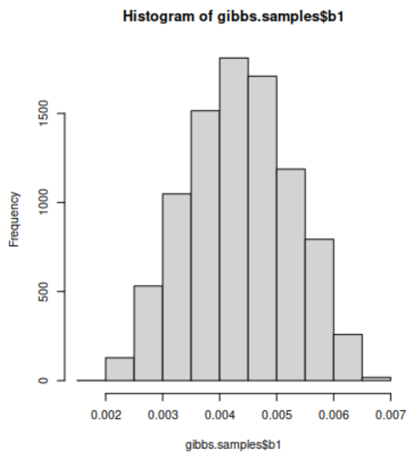

gibbs.samples$tau
x_axis

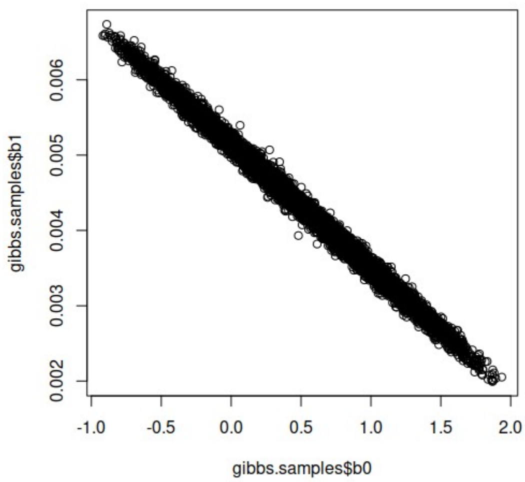

gibbs.samples$b0
x_axis


gibbs.samples$b1
x_axis


Histogram of gibbs.samples$tau
Frequency
gibbs.samples$tau


Histogram of gibbs.samples$b0
Frequency
gibbs.samples$b0


Histogram of gibbs.samples$b1
Frequency
gibbs.samples$b1


gibbs.samples$b1
gibbs.samples$b0