Question 1

$y_i^* \sim N(x_i'\beta, 1)$

$y_i = I(y_i^* > 0)$

$$
\begin{aligned}
P(y_i = 1 \mid x_i) &= P(y_i^* > 0 \mid x_i) \\
&= P\left(\frac{y_i^* - x_i'\beta}{1} > \frac{0 - x_i'\beta}{1} \mid x_i\right) \\
&= P(\underbrace{Z}_{\text{standard normal}} > -x_i'\beta \mid x_i) \\
&= 1 - \Phi(-x_i'\beta) \\
&= \Phi(x_i'\beta)
\end{aligned}
$$

by symmetry of the normal distribution

$y_i^* = x_i'\beta + \varepsilon_i$, $\varepsilon_i \sim \text{Logistic}(0,1)$

$y_i = I(y_i^* > 0)$

$$
\begin{aligned}
P(y_i = 1 \mid x_i) &= P(y_i^* > 0 \mid x_i) \\
&= P(y_i^* - x_i'\beta > 0 - x_i'\beta \mid x_i) \\
&= P(\underbrace{\varepsilon_i}_{\text{standard logistic}} > -x_i'\beta \mid x_i) \\
&= 1 - \text{logit}^{-1}(-x_i'\beta) \\
&= \text{logit}^{-1}(x_i'\beta)
\end{aligned}
$$

by symmetry of the logistic distribution

Question 2

$y_i^* | x_i, \beta \sim N(x_i'\beta, 1)$

$y_i = I(y_i^* > 0)$

$\beta \sim N_4(b_0, B_0)$

$$\pi(y^*, \beta | y, x) \propto L(y|y^*, \beta, x)\, \pi(y^*)\, \pi(\beta)$$

$$\propto \prod_{i=1}^{n} \left[\mathbb{1}(y_i^* > 0) y_i + \mathbb{1}(y_i^* < 0)(1-y_i) \right] \prod_{i=1}^{n} \left[\exp\left(-\tfrac{1}{2}(y_i^* - x_i'\beta)^2\right) \right] |B_0|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(\beta - b_0)' B_0^{-1} (\beta - b_0)\right]$$

$$\pi(y_i^* | \beta, y_i = 1, x) \propto \exp\left[-\tfrac{1}{2}(y_i^* - x_i'\beta)^2\right] \mathbb{1}(y_i^* > 0)$$

$$\pi(y_i^* | \beta, y_i = 0, x) \propto \exp\left[-\tfrac{1}{2}(y_i^* - x_i'\beta)^2\right] \mathbb{1}(y_i^* < 0)$$

$$\boxed{\begin{aligned} y_i^* | \beta, y_i = 1, x &\sim N_+(x_i'\beta, 1) \\ y_i^* | \beta, y_i = 0, x &\sim N_-(x_i'\beta, 1) \end{aligned}}$$

$$\pi(\beta | y^*, y, x) \propto \exp\left[-\tfrac{1}{2} \sum_{i=1}^{n} (-2\beta' x_i y_i^* + \beta' x_i x_i' \beta)\right] \exp\left[-\tfrac{1}{2}(\beta' B_0^{-1} \beta - 2\beta' B_0^{-1} b_0)\right]$$

$$\propto \exp\left[-\tfrac{1}{2}\left[\beta'\left(\sum_{i=1}^{n} x_i x_i' + B_0^{-1}\right)\beta - 2\beta'\left(\sum_{i=1}^{n} x_i y_i^* + B_0^{-1} b_0\right)\right]\right]$$

$$\boxed{\beta | y^*, y, x \sim N_4\left(\left(\sum_{i=1}^{n} x_i x_i' + B_0^{-1}\right)^{-1} \left(\sum_{i=1}^{n} x_i y_i^* + B_0^{-1} b_0\right), \left(\sum_{i=1}^{n} x_i x_i' + B_0^{-1}\right)^{-1} \right)}$$

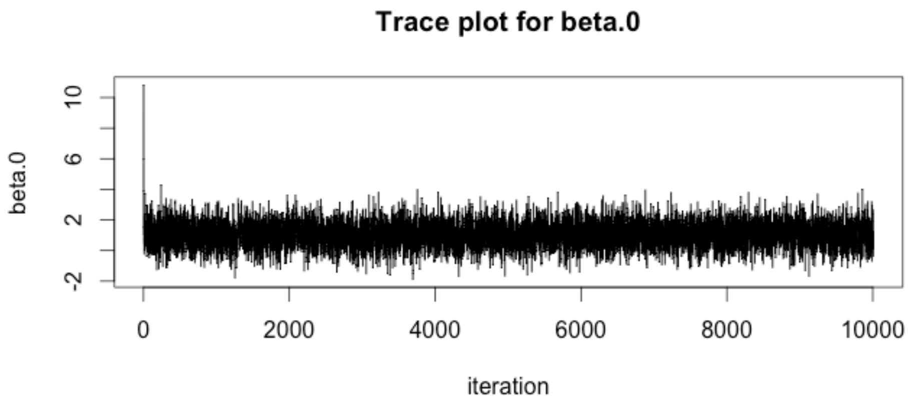
Logistic regression case

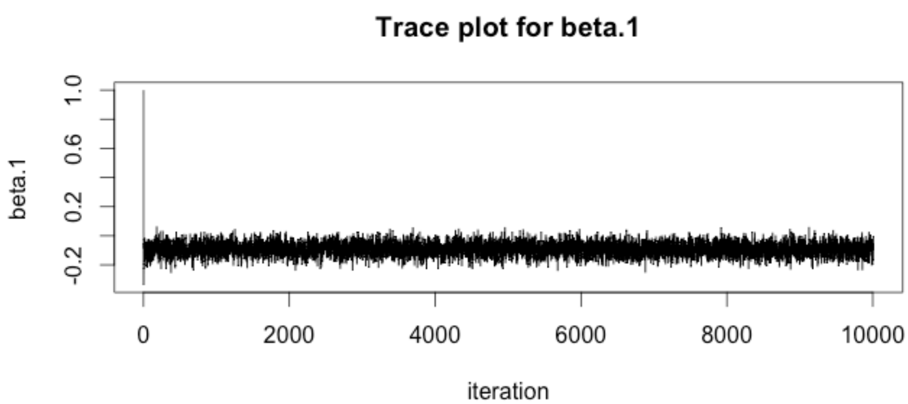
Gibbs sampling in the logistic regression case is more difficult because the normal distribut on β is not a conjugate prior for the logistic regression likelihood.

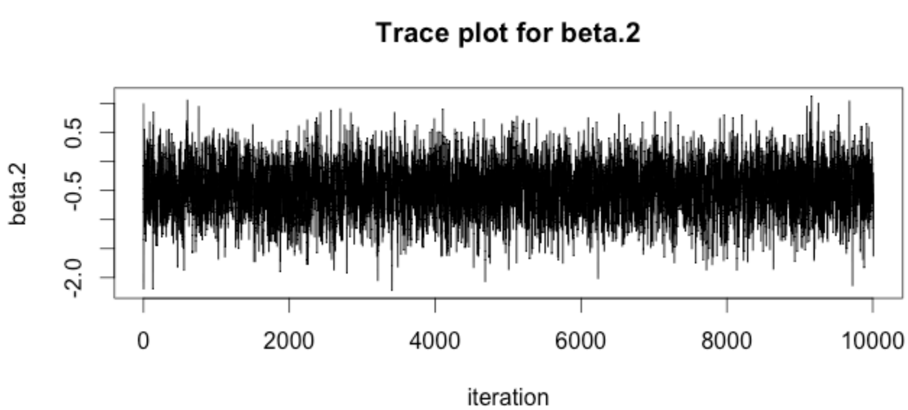
Kelvin Niu

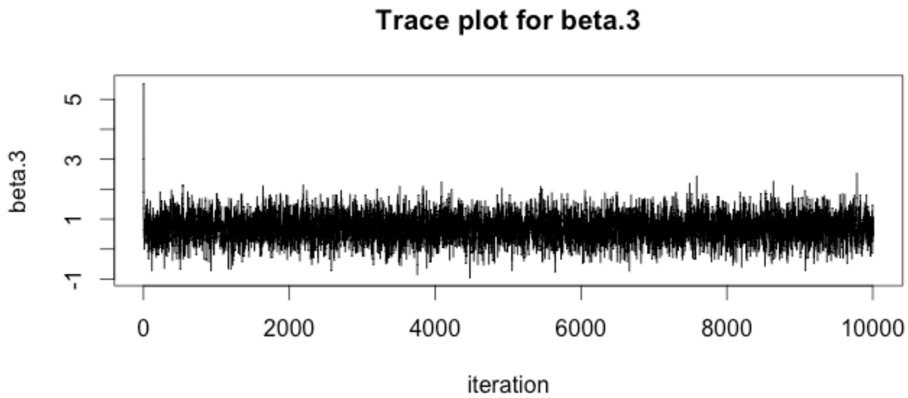
Lab 11

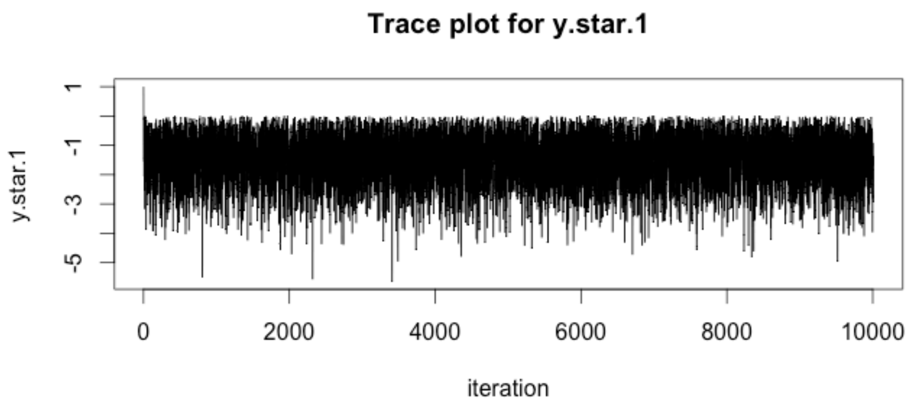
Using the full conditionals above and prior $b_0 = (0, 0, 0, 0)'$ and $B_0 = I$, I performed 10,000 iterations of Gibbs sampling. Below are the trace plots for the elements of β and one y_i^*:





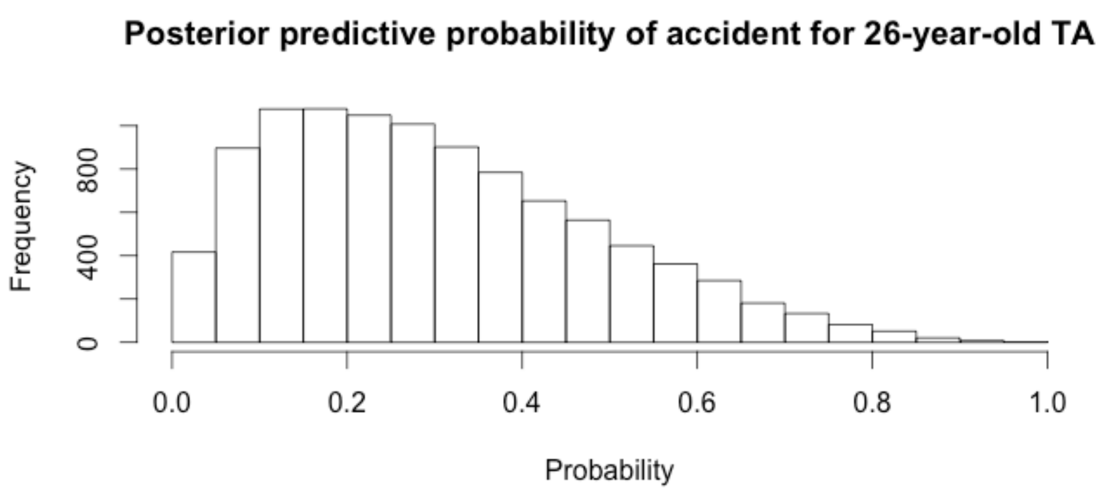






10 burn-ins were thrown out for each parameter for the analysis to follow.

<u>26-year old TA</u>
For each iteration, I calculated the probability of an accident [probability that a draw from $N(x'\beta, 1) > 0$] for the TA $x = (1, 26, 0, 1)'$. Below is a histogram of these probabilities:



The mean probability that the TA was involved in an accident in the last six months is **0.2950**.

Mercedes parking

For each iteration, I calculated the probability of an accident [probability that a draw from $N(x_i'\beta, 1) > 0$] for $x_1' = (1, 17, 1, 0)$ and $x_2' = (1, 18, 1, 1)$. The mean probability for x_1 is 0.1837 and the probability for x_1 is 0.3659. **Thus, I would rather have the 17-year-old who has done a driver's education course and hates statistics park my Mercedes**.