

Q1.

i. $y_t \sim \text{poisson}(\theta)$; prior for θ : $\theta \sim G(\underline{\alpha}, \underline{\beta})$

$$P(y_t = k) = \frac{\theta^k \exp(-\theta)}{k!} \quad f(\theta) = \frac{\underline{\beta}^{\underline{\alpha}}}{\Gamma(\underline{\alpha})} \theta^{\underline{\alpha}-1} \exp(-\underline{\beta} \theta)$$

$$\text{If } y_t \text{ i.i.d., then } P(y_1 = k_1, \dots, y_T = k_T) \propto \prod_{t=1}^T \theta^{k_t} \exp(-\theta) = \exp(-\theta) \theta^{\sum k_t}$$

$$\text{So, } p(\theta | y_t) \propto p(y_t | \theta) p(\theta) \quad (\text{Bayes' Rule})$$

$$\propto \exp(-\theta) \theta^{\sum k_t} \theta^{\underline{\alpha}-1} \exp(-\underline{\beta} \theta)$$

$$= \exp(-\theta(1+\underline{\beta})) \theta^{\underline{\alpha}-1+\sum k_t}$$

$$= \text{Gamma}(\underline{\alpha} + \sum_{t=1}^T k_t, \underline{\beta} + 1)$$

$$= \text{Gamma}(\underline{\alpha} + T \bar{k}, \underline{\beta} + 1). \quad \bar{k} \equiv \frac{1}{T} \sum_{t=1}^T k_t$$

$$\text{ii. } P(y_t | \theta) = \begin{cases} \theta & y_t = 1 \\ 1 - \theta & y_t = 0 \end{cases} \quad \text{prior for } \theta: p(\theta) = \left[\frac{\Gamma(\underline{\alpha}) \Gamma(\underline{\delta})}{\Gamma(\underline{\alpha} + \underline{\delta})} \right]^{-1} \theta^{\underline{\alpha}-1} (1-\theta)^{\underline{\delta}-1}$$

$$= \theta^{y_t} (1-\theta)^{1-y_t}$$

The likelihood of the data, therefore, is

$$P(\{y_t\}_{t=1}^T | \theta) = \prod_{t=1}^T \theta^{y_t} (1-\theta)^{1-y_t}$$

$$= \theta^{\sum y_t} (1-\theta)^{\sum (1-y_t)}$$

$$= \theta^{T \bar{y}} (1-\theta)^{T(1-\bar{y})}$$

$$\text{So, } p(\theta | y_t) \propto p(y_t | \theta) p(\theta)$$

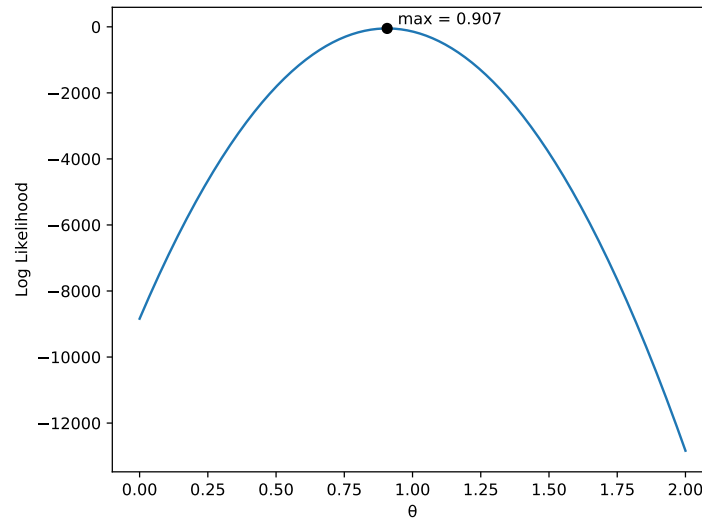
$$= \theta^{T \bar{y}} (1-\theta)^{T(1-\bar{y})} \theta^{\underline{\alpha}-1} (1-\theta)^{\underline{\delta}-1}$$

$$= \theta^{T \bar{y} + \underline{\alpha} - 1} (1-\theta)^{T(1-\bar{y}) + \underline{\delta} - 1}$$

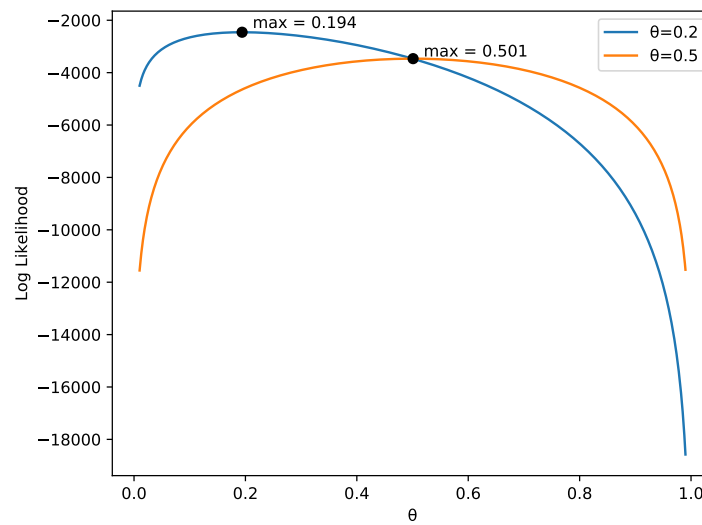
$$\propto \text{Beta}(T \bar{y} + \underline{\alpha}, T(1-\bar{y}) + \underline{\delta})$$

Q2. In all the graphs that follow that are functions of θ , I used a grid for θ with steps of 0.001. So, when I present the “max” or the “mode,” these are just the points on the grid with the highest values.

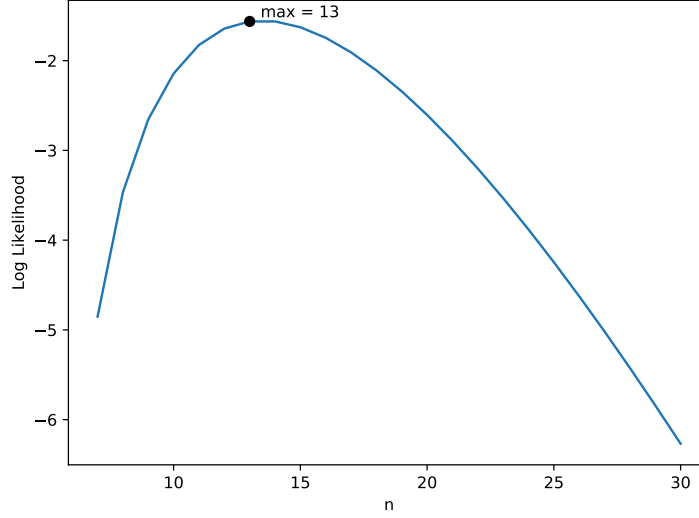
i. The likelihood for β is maximized right around the true value, and doesn't appear to be very flat:



- ii. These likelihoods appear to be a bit more flat, but that could be due to scaling. I don't know of a good way to say that a likelihood is "flat."



The likelihood for n appears relatively well behaved, though it is skewed. I guess this would imply that a maximizer may tend to overestimate n , since the likelihood doesn't fall off as quickly to the right.



iii. The posterior for the conjugate prior is:

$$\begin{aligned}
 p(\theta \mid n, s) &= \text{prior} \times \text{likelihood} \\
 &= \theta^{a-1}(1-\theta)^{b-1}\theta^s(1-\theta)^{n-s} \\
 &= \theta^{a-1+s}(1-\theta)^{b-1+n-s}.
 \end{aligned}$$

For Jeffrey's prior, we first need the information:

$$\begin{aligned}
 I_\theta &= -\mathbb{E} \left[\frac{\partial^2 \log \ell(\theta; n, s)}{\partial \theta^2} \right] \\
 &= -\mathbb{E} \left[\frac{\partial}{\partial \theta} \left(\frac{s}{\theta} - \frac{n-s}{1-\theta} \right) \right] \\
 &= -\mathbb{E} \left[-\frac{s}{\theta^2} - \frac{n-s}{(1-\theta)^2} \right] \\
 &= \frac{n\theta}{\theta^2} + \frac{n(1-\theta)}{(1-\theta)^2} \\
 &= \frac{n}{\theta} + \frac{n}{1-\theta}
 \end{aligned}$$

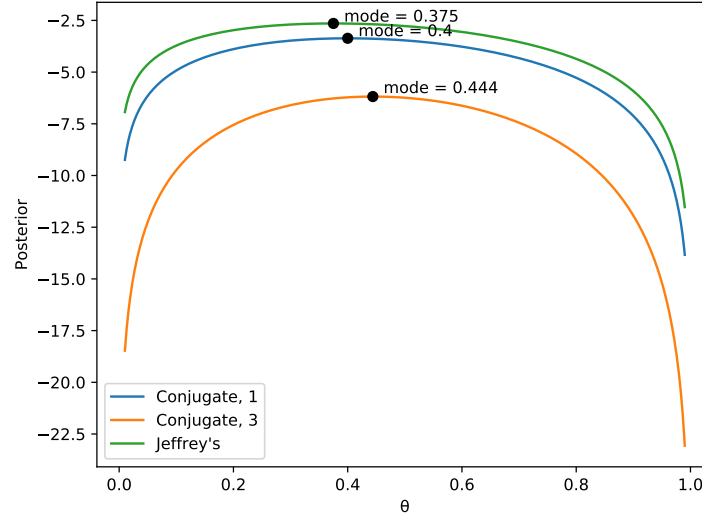
So the prior is proportional to the square root of this:

$$p(\theta) \propto \sqrt{\frac{1}{\theta} + \frac{1}{1-\theta}} = \sqrt{\frac{1}{\theta(1-\theta)}} = \theta^{-1/2}(1-\theta)^{-1/2}$$

So the posterior is:

$$p(\theta \mid n, s) \propto \theta^{s-0.5}(1-\theta)^{n-s-0.5}$$

The three posteriors are plotted below. When $a = b$, this conjugate prior collapses to $(\theta(1-\theta))^{a-1}$. In general, this is maximized at $\theta = 1/2$, so it is like saying that our prior is that $\theta = 1/2$. Setting $a = b = 1$ is just equivalent to setting a uniform prior; thuse, the posterior is just the likelihood of the data—and is therefore maximized at the true parameters. Setting $a = b = 3$ means that we are putting more weight on this $\theta = 1/2$ prior—and we see this in the posterior: the mode is farther to the right. The Jeffrey's prior pulls us in the other direction.



- Q3.** i. I set $\alpha = 1$. Practically, in order to be able to draw from the double exponential, I need the inverse CDF. But before that, I need the CDF. For $x < 0$:

$$F(x) = \int_{-\infty}^x \frac{1}{2} \exp(y) dy = \frac{1}{2} \exp(x)$$

For $x > 0$:

$$\begin{aligned} F(x) &= \int_{-\infty}^x \frac{1}{2} \exp(-|y|) dy \\ &= \int_{-\infty}^0 \frac{1}{2} \exp(y) dy + \int_0^x \frac{1}{2} \exp(-y) dy \\ &= \frac{1}{2} - \frac{1}{2} \exp(-x) \Big|_0^x \\ &= \frac{1}{2} - \left[\frac{1}{2} \exp(-x) - \frac{1}{2} \right] \\ &= 1 - \frac{1}{2} \exp(-x) \end{aligned}$$

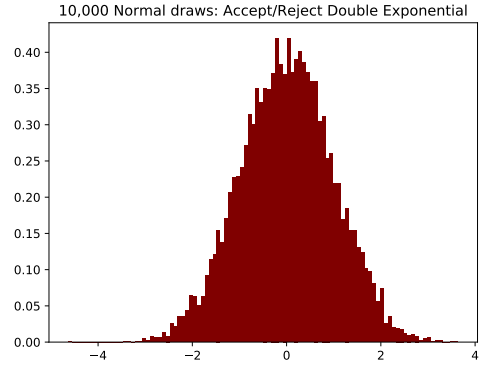
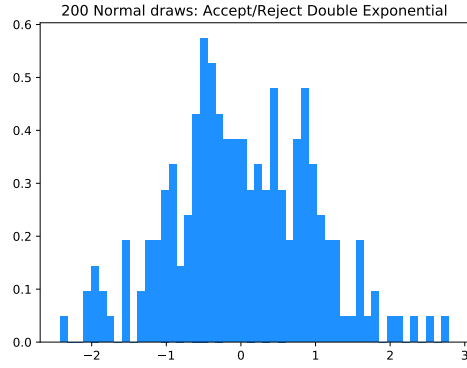
Now we can invert. If $x < 0$, then:

$$u = \frac{\exp(x)}{2} \implies x = \log(2u) \implies F^{-1}(u) = \log(2u) \text{ if } u < \frac{1}{2}.$$

On the other hand, if $x \geq 0$ then

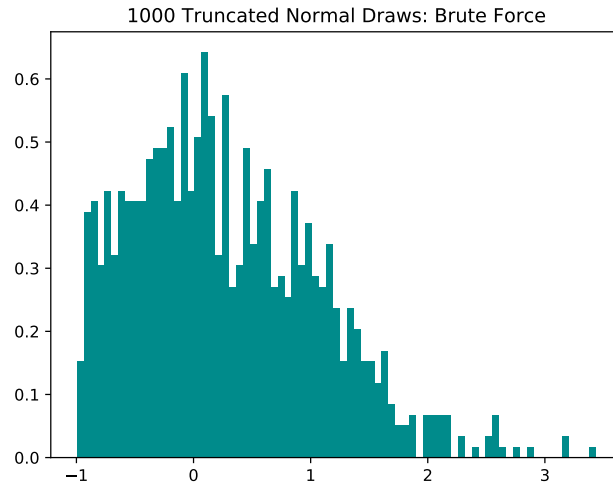
$$u = 1 - \frac{\exp(-x)}{2} \implies x = -\log(2(1-u)) \implies F^{-1}(u) = -\log(2(1-u)) \text{ if } u > \frac{1}{2}.$$

In order to draw 200 observations, it took 263 draws. I also drew 10,000 observations, which took 13142 draws. The (normalized) histograms of these draws look like so:



ii. Set $\mu = 0$, $\sigma = 1$, $\alpha = 1$ and $\underline{\mu} = -1$. I am drawing 1000 truncated normals.

(a) It took 1212 draws from a normal to get 1000 draws from the truncated normal. Here is a histogram of the draws:



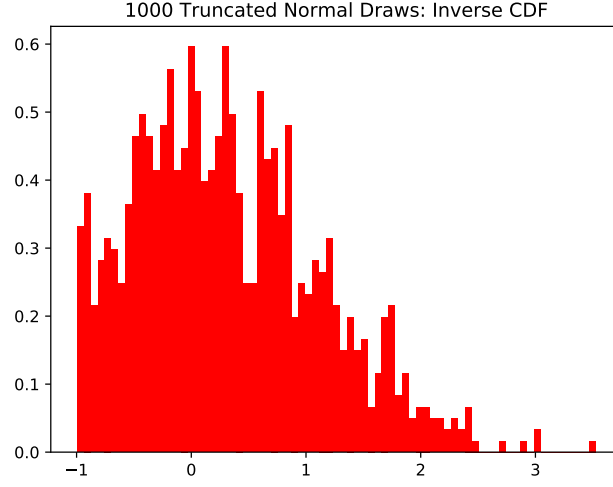
(b) According to Wikipedia, the CDF of the right-truncated normal is given by:

$$F(x) = \frac{\Phi\left(\frac{x-\mu}{\sigma}\right) - \Phi\left(\frac{\underline{\mu}-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{\underline{\mu}-\mu}{\sigma}\right)}$$

So,

$$u = \frac{\Phi\left(\frac{x-\mu}{\sigma}\right) - \Phi\left(\frac{\underline{\mu}-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{\underline{\mu}-\mu}{\sigma}\right)} \implies x = \sigma \Phi^{-1} \left[u \left(1 - \Phi\left(\frac{\underline{\mu}-\mu}{\sigma}\right) \right) + \Phi\left(\frac{\underline{\mu}-\mu}{\sigma}\right) \right] + \mu \equiv F^{-1}(u)$$

Here is a histogram of the draws:



(c) According to Wikipedia, the PDF of a left-truncated normal is:

$$f(x) = \frac{\phi\left(\frac{x-\mu}{\sigma}\right)}{\sigma\left(1 - \Phi\left(\frac{\mu}{\sigma} - \mu\right)\right)} \equiv \frac{\phi\left(\frac{x-\mu}{\sigma}\right)}{\sigma(1-B)}$$

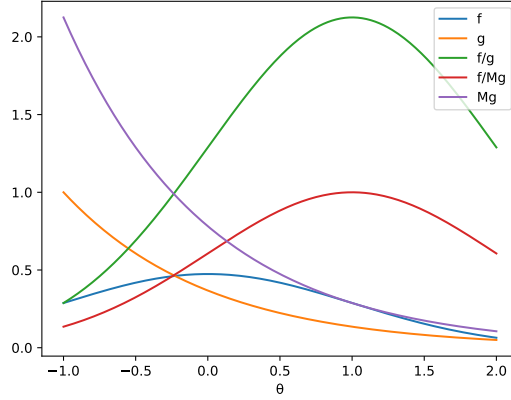
The enveloping pdf is

$$g(x) = \exp\left[-\left(\frac{x-\mu}{\sigma} - \underline{\mu}\right)\right] \mathbb{1}\left\{\frac{x-\mu}{\sigma} \geq \underline{\mu}\right\}$$

So, dividing the target, f , by the candidate, g , (and plugging-in $\mu = 0$ and $\sigma = 1$ for simplicity) produces a bound:

$$\begin{aligned} \frac{f(x)}{g(x)} &= \frac{\frac{1}{\sigma(1-B)\sqrt{2\pi}} \exp(-x^2/2)}{\exp\left[-(x-\underline{\mu})\right] \mathbb{1}\{x \geq \underline{\mu}\}} \\ &= \frac{1}{\mathbb{1}\{x \geq \underline{\mu}\} \sigma(1-B)\sqrt{2\pi}} \exp\left(-\frac{x^2}{2} + (x-\underline{\mu})\right) \\ &\leq \frac{1}{\sigma(1-B)\sqrt{2\pi}} \exp(1.5) \quad \forall x \geq -1, \end{aligned}$$

where the inequality follows from the fact that the exponential term is maximized when $x = 1$. Just so I have a better idea of what's going on, I've plotted a bunch of these quantities.



First, Mg envelopes f , as desired. Next, let's think about what happens when we sample from g . When θ is small, g is above f —but, g is relatively large for small θ , so we will draw from there often. But, we only keep a small fraction of these draws (those smaller than f/Mg). This aligns our draws closer to the distribution of f , where there are relatively few draws that would come from small θ . Similarly, the relatively few draws taken for large θ are kept pretty often, reflecting the fact that $g < f$. Pretty cool!

Back to the problem at hand: we will accept X if

$$U \leq \exp\left(-\frac{X^2}{2} + (X - \underline{\mu}) - 1.5\right)$$

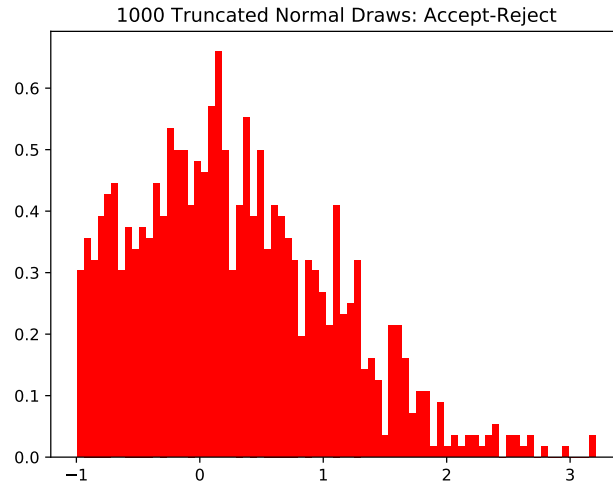
But before we do any of this, we need to calculate the inverse CDF associated with g . If $x \leq \underline{\mu}$, then $G(x) = 0$. If $x \geq \underline{\mu}$, then

$$\begin{aligned} G(x) &\equiv \int_{-\infty}^x \exp[-(y - \underline{\mu})] \mathbb{1}\{y \geq \underline{\mu}\} dy \\ &= 0 + \int_{\underline{\mu}}^x \exp[-(y - \underline{\mu})] dy \\ &= -\exp(-(y - \underline{\mu})) \Big|_{\underline{\mu}}^x \\ &= 1 - \exp(-(x - \underline{\mu})) \end{aligned}$$

The inverse CDF is therefore

$$u = 1 - \exp(-(x - \underline{\mu})) \implies x = \underline{\mu} - \log(1 - u) \equiv G^{-1}(u).$$

In order to draw the 1000 draws, it took me 2103 draws. Here they are:



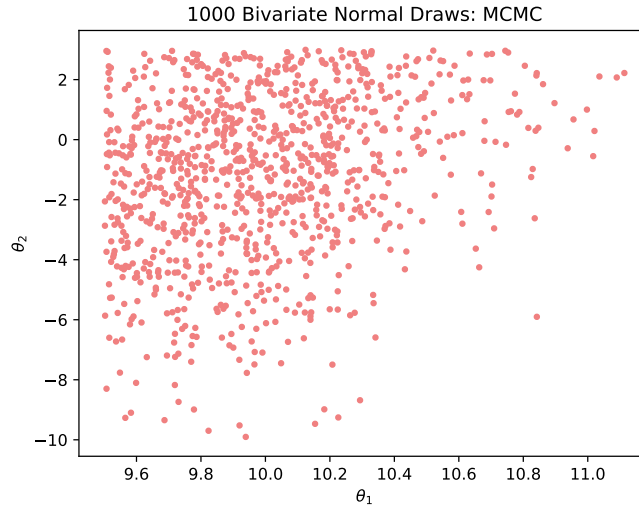
iii. Define

$$\Sigma \equiv \begin{bmatrix} \sigma_1^2 & \gamma \\ \gamma & \sigma_2^2 \end{bmatrix} \quad \mu \equiv \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \rho \equiv \frac{\gamma}{\sigma_1 \sigma_2}$$

By Wikipedia,

$$\theta_i \mid \theta_j \sim N\left(\mu_i + \frac{\sigma_i}{\sigma_j} \rho (\theta_j - \mu_j), (1 - \rho^2) \sigma_i^2\right).$$

I kept all draws in which both θ_1 and θ_2 were within the bounds I set—just as in the “brute force” approach, above. I set $\sigma_1 = 0.5$, $\sigma_2 = 4$, $\gamma = 1$, $\mu_1 = 10$, $\mu_2 = 0$, and required $\theta_1 \in [9.5, 100.0]$ and $\theta_2 \in [-10, 3]$. It took 1960 draws to get the 250 burn-in draws and 1000 kept-draws. Here is a scatter plot of the retained draws:



Q4. The probit model specifies that¹

$$y_i^* = \beta_0 + \beta_1 x_i + e_i \quad e_i \sim N(0, \sigma^2)$$

¹Following the lecture notes, the Handbook of Econometrics chapter quoted therein, and Joe.

We observe $y_i = 1$ if $y_i^* > 0$, and $y_i = 0$ otherwise.

$$P(y_i = 1 \mid x_i, \beta) = P(y_i^* > 0 \mid x_i, \beta) = P(\beta_0 + \beta_1 x_i + e_i > 0 \mid x_i, \beta) = \Phi(\beta_0 + \beta_1 x_i) \quad (1)$$

Our prior for β

$$\underline{\beta} \sim N(\underline{\mu}, \underline{g}(X'X)^{-1})$$

where $\underline{\mu}$ and \underline{g} are parameters to be specified—note that this is the g -prior for the variance. Now for the marginal of y^* :

$$y_i^* \mid \beta, x_i \stackrel{d}{=} \beta_0 + \beta_1 x_i + e_i \mid x_i, \beta \sim N(\beta_0 + \beta_1 x_i, \sigma^2).$$

Conditioning on y_i gives:

$$y_i^* \mid y_i, x_i, \beta = \begin{cases} TN_{(0, \infty)}(\beta_0 + \beta_1 x_i, \sigma^2) & \text{if } y_i = 1 \\ TN_{(-\infty, 0]}(\beta_0 + \beta_1 x_i, \sigma^2) & \text{if } y_i = 0. \end{cases}$$

We know from class that, if the variance of e_i is known then

$$\beta \mid y_i, x_i, \sigma^2, g \sim N(\bar{\beta}, \sigma^2 \bar{B}) \quad \bar{B} = (g(X'X)^{-1} + X'X)^{-1} \quad \bar{\beta} = \bar{B}(g^{-1}(X'X)\underline{\mu} + X'y^*).$$

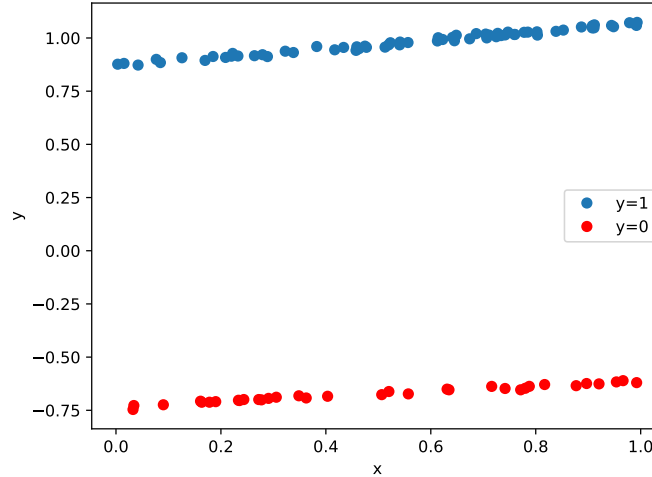
So, using the following parameters:

$$\underline{\mu} = [0, 0] \quad g = 10 \quad \underline{y}_i^* = 0 \quad \sigma^2 = 1$$

And drawing β then y_i^* , I get the following estimates:

| | Prior | Mean | Standard Deviation |
|-----------|-------|------|--------------------|
| β_0 | 0 | 0.19 | 0.24 |
| β_1 | 0 | 0.42 | 0.38 |

And the probability of $y_i = 1$ given $x_i = 2$ is 0.85, by (1). For fun, here are the estimated y_i^* 's and x_i 's in a scatter plot.



Q5.