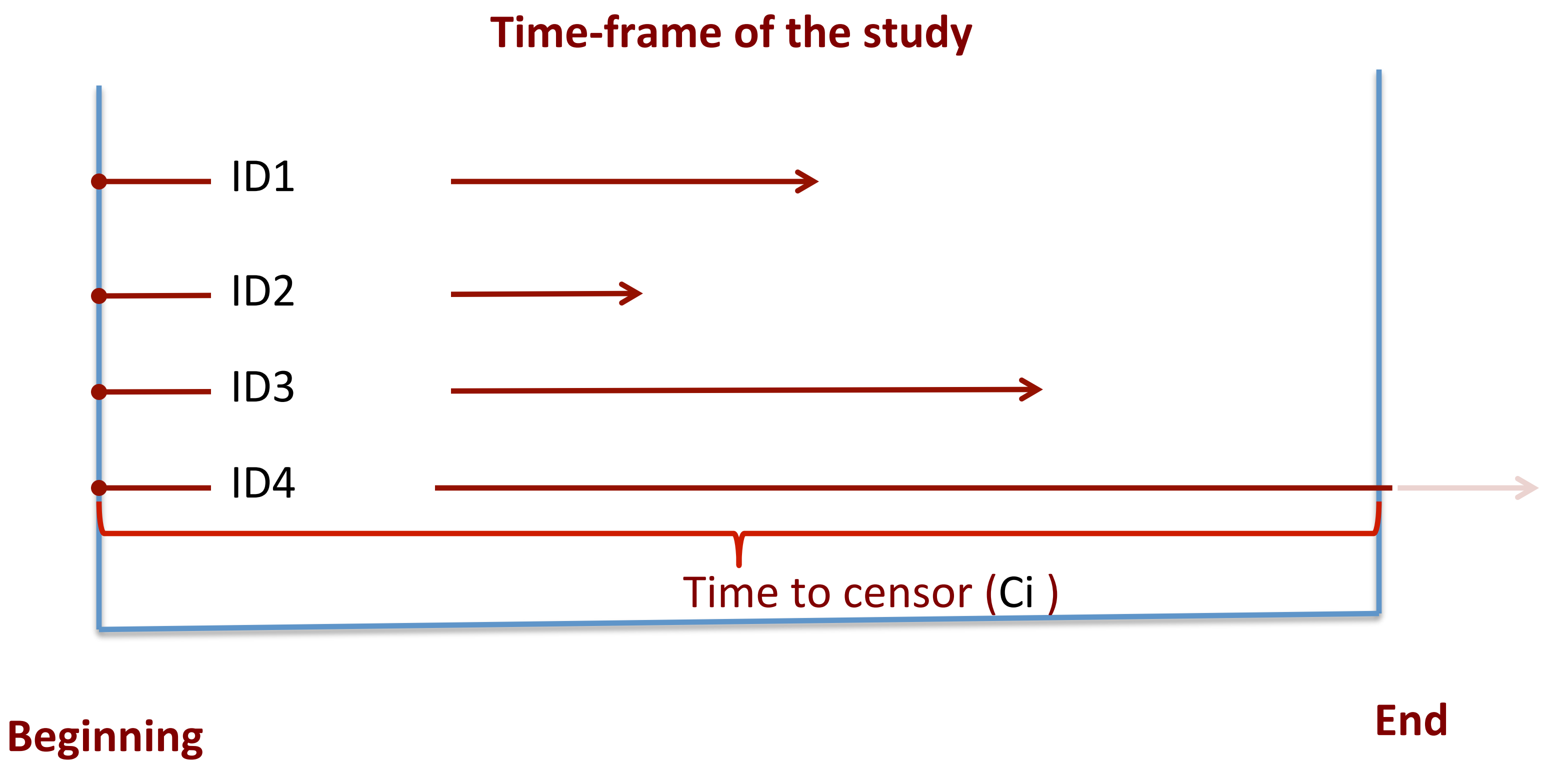
STT 465
Bayesian Multiple Linear Regression:

=> Regression With Censored Data

=> Regression with binary outcomes

Regression with Censored Data

Right-Censoring ($Y_i > C_i$)



IDs 1,2 and 3: time to event is observed.

ID 4: time to event is unknown; however we know the time to censor, and we know that the event will happens after $Y_4 > C4$ (right-censored data)

Notation and datum Likelihood

Data will be defined by a pair of vectors

- y : time to event or time to censoring.
- d : 1 if event, 0 if censored.

For notation purposes, we will use

y_i : to denote an observed time to event or time at censoring.

$\tilde{y}_i$: to denote un-observed time to event, only applies to censored points.

Likelihood Function (single-data-point)

Observed Data-points

$$p\left(y_i \mid X,\beta,\sigma_\varepsilon^2,d_i=1\right)=\frac{Exp\left\{-\frac{\left(y_i-x_i'\beta\right)^2}{2\sigma_\varepsilon^2}\right\}}{\sqrt{2\pi\sigma_\varepsilon^2}}$$

Censored Data-points

$$p\left(\tilde{y}_i > y_i \mid X,\beta,\sigma_\varepsilon^2,d_i=0\right)=\int_{-\infty}^{y_i}\frac{Exp\left\{-\frac{\left(\tilde{y}_i-x_i'\beta\right)^2}{2\sigma_\varepsilon^2}\right\}}{\sqrt{2\pi\sigma_\varepsilon^2}}d\tilde{y}_i$$

Likelihood Function

$$p\left(y \mid X,\beta,\sigma_\varepsilon^2,d\right)=\prod_{i:d_i=1}\frac{e^{\left\{-\frac{(y_i-x_i'\beta)^2}{2\sigma_\varepsilon^2}\right\}}}{\sqrt{2\pi\sigma_\varepsilon^2}}\times\prod_{i:d_i=0}\int_{-\infty}^{y_i}\frac{Exp\left\{-\frac{(\tilde{y}_i-x_i'\beta)^2}{2\sigma_\varepsilon^2}\right\}}{\sqrt{2\pi\sigma_\varepsilon^2}}d\tilde{y}_{\boldsymbol{i}}$$

- The above function can be used to derive Max. Likelihood Estimates (see survreg of survival package)

```
## Example: Maximum Likelihood with Censored Data ##
  library(survival)
  y=rnorm(1000,sd=4)
  yCensored=y
  threshold=2
  d=ifelse(y<threshold,1,0)
  yCensored[d==0]=threshold
  mean(y); sd(y)
  mean(yCensored); sd(yCensored)
  fm=survreg(Surv(time=y,event=d)~1,dist="gaussian")
  summary(fm)
```

Likelihood Function

$$p\left(y \mid X,\beta,\sigma_\varepsilon^2,d\right)=\prod_{i:d_i=1}\frac{e^{\left\{-\frac{(y_i-x_i'\beta)^2}{2\sigma_\varepsilon^2}\right\}}}{\sqrt{2\pi\sigma_\varepsilon^2}}\times\prod_{i:d_i=0}\int_{-\infty}^{y_i}\frac{Exp\left\{-\frac{(\tilde{y}_i-x_i'\beta)^2}{2\sigma_\varepsilon^2}\right\}}{\sqrt{2\pi\sigma_\varepsilon^2}}d\tilde{y}_{\boldsymbol{i}}$$

- Bayesian Analysis with this likelihood is difficult because the integrals in the 2nd term do not have closed form and all the fully conditionals will not have closed form.

- Therefore, instead, we will use Data Augmentation. With data augmentation we will perform the integrals using Monte Carlo Methods.

- In data augmentation we exploit the following equality:

$$\int_{-\infty}^{y_i}\frac{Exp\left\{-\frac{(\tilde{y}_i-x_i'\beta)^2}{2\sigma_\varepsilon^2}\right\}}{\sqrt{2\pi\sigma_\varepsilon^2}}d\tilde{y}_{\boldsymbol{i}}=\int_{-\infty}^{\infty}\frac{Exp\left\{-\frac{(\tilde{y}_i-x_i'\beta)^2}{2\sigma_\varepsilon^2}\right\}}{\sqrt{2\pi\sigma_\varepsilon^2}}\times 1\left(y_i<\tilde{y}_i\right)d\tilde{y}_{\boldsymbol{i}}$$

Data Augmentation

Bayesian Likelihood (with data augmentation)

$$p\left(y \mid X, \beta, \sigma_\varepsilon^2, d\right) = \prod_{i:d_i=1} \frac{e^{\left\{-\frac{(y_i - x_i'\beta)^2}{2\sigma_\varepsilon^2}\right\}}}{\sqrt{2\pi\sigma_\varepsilon^2}} \times \prod_{i:d_i=0} \frac{Exp\left\{-\frac{(\tilde{y}_i - x_i'\beta)^2}{2\sigma_\varepsilon^2}\right\}}{\sqrt{2\pi\sigma_\varepsilon^2}} 1\left(\tilde{y}_i > y_i\right)$$

$\tilde{y}_i$: Unobserved time to event.

$\Rightarrow$ With data augmentation, the model unknowns involve not only the parameters (effects, variances, etc.) but also the un-observed time to events.

$\Rightarrow$ Therefore, in our sampler we need to sample also the un-observed time to event of the censored points.

Fully Conditional Distribution

Likelihood Function

$$p\left(y \mid X,\beta,\sigma_\varepsilon^2,d\right)=\prod_{i:d_i=1}\frac{e^{\left\{-\frac{\left(y_i-x_i'\beta\right)^2}{2\sigma_\varepsilon^2}\right\}}}{\sqrt{2\pi\sigma_\varepsilon^2}}\times\prod_{i:d_i=0}\frac{Exp\left\{-\frac{\left(\tilde{y}_i-x_i'\beta\right)^2}{2\sigma_\varepsilon^2}\right\}}{\sqrt{2\pi\sigma_\varepsilon^2}}1\left(\tilde{y}_i>y_i\right)$$

Prior

$$p\left(\beta,\sigma_\varepsilon^2,\sigma_\beta^2,\ldots\right)$$

Joint Posterior:

$$p\left(\beta,\sigma_\varepsilon^2,\sigma_\beta^2,\ldots \mid y,d\right)\propto\left[\prod_{i:d_i=1}\frac{e^{\left\{-\frac{\left(y_i-x_i'\beta\right)^2}{2\sigma_\varepsilon^2}\right\}}}{\sqrt{2\pi\sigma_\varepsilon^2}}\times\prod_{i:d_i=0}\frac{Exp\left\{-\frac{\left(\tilde{y}_i-x_i'\beta\right)^2}{2\sigma_\varepsilon^2}\right\}}{\sqrt{2\pi\sigma_\varepsilon^2}}1\left(\tilde{y}_i>y_i\right)\right]\times p\left(\beta,\sigma_\varepsilon^2,\sigma_\beta^2,\ldots\right)$$

Fully Conditional (truncated normal):

$$p\left(\tilde{y}_i \mid d_i=0\right)\propto\frac{e^{\left\{-\frac{\left(\tilde{y}_i-x_i'\beta\right)^2}{2\sigma_\varepsilon^2}\right\}}}{\sqrt{2\pi\sigma_\varepsilon^2}}1\left(\tilde{y}_i>y_i\right)$$

Outline of a Gibbs Sampler

$\Rightarrow$ Time to event of censored data points follow have a truncated normal fully conditional distribution.

$\Rightarrow$ The other fully conditional distributions are as those of the standard linear regression without censoring.

$\Rightarrow$ Therefore, relative to a Gibbs sampler for a model without censoring, we just need to add a step where we 'impute' the un-observed time to events with samples drawn from the corresponding fully conditionals (truncated normal, in our case).

Sampler in GitHub:

Regression with Binary Outcomes

Regression with Binary Outcomes

- Binary outcomes follow Bernoulli distributions

$$y_i \in \{0,1\} \qquad\qquad p\left(y_i \middle| \theta\right) = \theta^{y_i}\left(1-\theta\right)^{1-y_i}$$

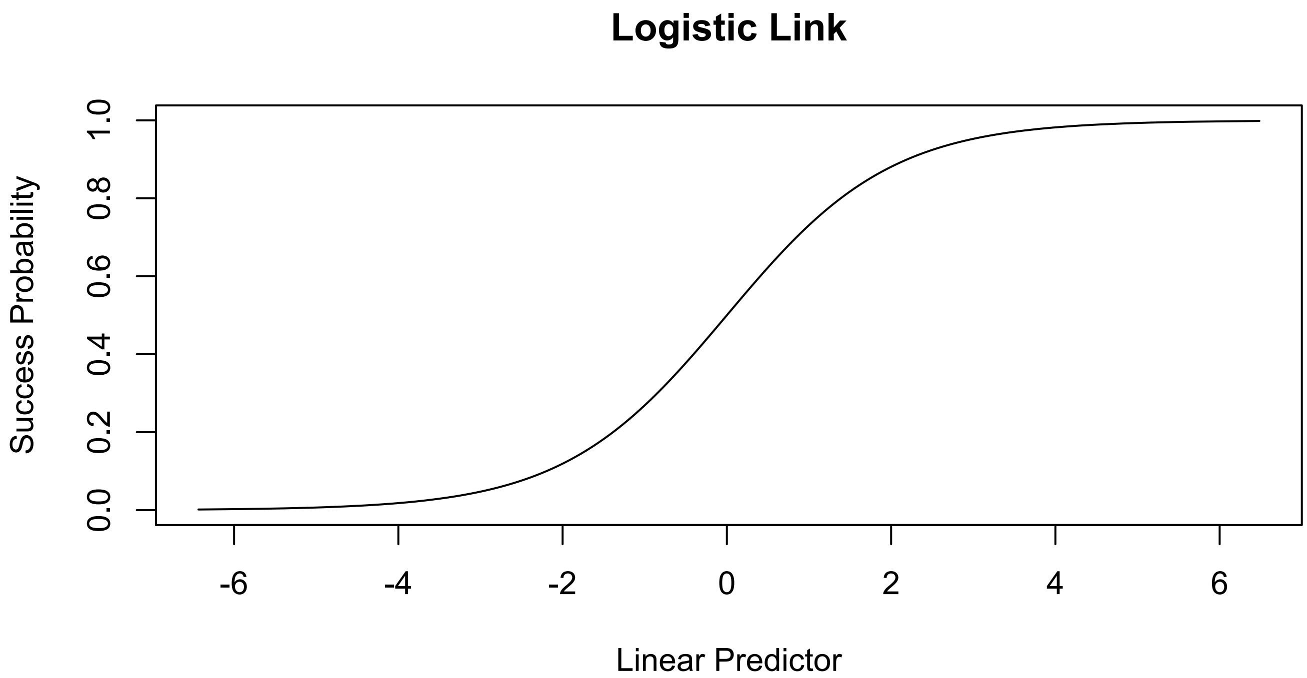
- Regression with Binary outcomes: we want to make θ a function of one or more predictors.
- Problem:
  - θ lives in the 0-1 interval,
  - while a regression function $\sum_{j=1}^{p} x_{ij}\beta_j$ lives in the real line.

- To get around this we need to introduce a link function that maps from the real line to the 0-1 interval.

- The most commonly used links are the logit and probit link.

Logistic Regression

$$p\left(y_i = 1\right) = \theta \qquad \log\left(\frac{\theta}{1-\theta}\right) = \sum_{j=1}^{p} x_{ij}\beta_j \qquad \Rightarrow \theta = \frac{e^{\sum_{j=1}^{p} x_{ij}\beta_j}}{1 + e^{\sum_{j=1}^{p} x_{ij}\beta_j}}$$



Logistic Regression

- Likelihood Function $$p(y)=\prod_{i=1}^{n}\theta^{y_i}(1-\theta)^{1-y_i}$$

$$\log\left(\frac{\theta}{1-\theta}\right)=\eta_i=\sum_{j=1}^{p}x_{ij}\beta_j \qquad \Rightarrow \theta_i=\frac{e^{\sum_{j=1}^{p}x_{ij}\beta_j}}{1+e^{\sum_{j=1}^{p}x_{ij}\beta_j}}$$

$$p(y)=\prod_{i=1}^{n}\left[\frac{e^{\eta_i}}{1+e^{\eta_i}}\right]^{y_i}\left[1-\frac{e^{\eta_i}}{1+e^{\eta_i}}\right]^{1-y_i}$$

- The above likelihood can be maximized with respect to regression coefficient to obtain Max. Likelihood Estimates.

```
glm(y~X,family=binomial)
```

Threshold Model (Probit Link)

$$\tilde{y}_i = \sum_{j=1}^{p} x_{ij}\beta_j + \varepsilon_i$$

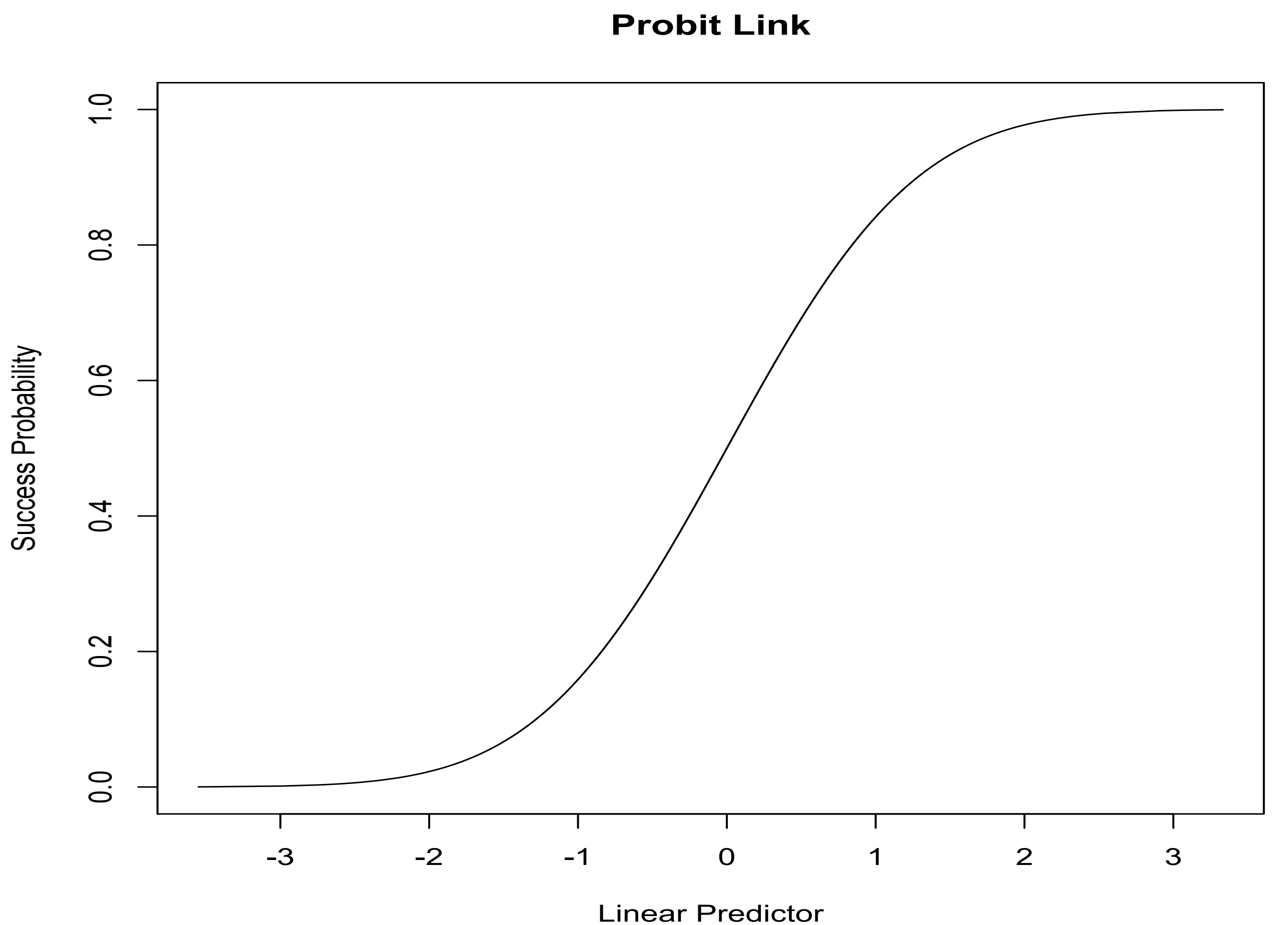
$$y_i = \begin{cases} 0 & \tilde{y}_{\boldsymbol{i}} < 0 \\ 1 & \text{Otherwise} \end{cases}$$

$$\varepsilon_i \overset{iid}{\sim} N(0,1)$$

$$\begin{aligned} p(y_i = 1) = p(\tilde{y}_{\boldsymbol{i}} > 0) &= p\left(\sum_{j=1}^{p} x_{ij}\beta_j + \varepsilon_i > 0\right) \\ &= p\left(\varepsilon_i > -\sum_{j=1}^{p} x_{ij}\beta_j\right) \\ &= p\left(-\varepsilon_i < \sum_{j=1}^{p} x_{ij}\beta_j\right) \\ &= \text{pnorm}\left(q = \sum_{j=1}^{p} x_{ij}\beta_j\right) \end{aligned}$$

Threshold Model (Probit link)

$$p\left(y_i = 1\right) = \theta \qquad \theta = pnorm\left(\sum_{j=1}^{p} x_{ij}\beta_j\right)$$



Threshold Model (Probit Link)

- Likelihood Function

$$p(y)=\prod_{i=1}^{n}\theta^{y_i}\left(1-\theta\right)^{1-y_i} \qquad \theta = pnorm\left(\sum_{j=1}^{p}x_{ij}\beta_j\right)$$

$$p(y)=\prod_{i=1}^{n}\left[pnorm\left(\sum_{j=1}^{p}x_{ij}\beta_j\right)\right]^{y_i}\left[1-pnorm\left(\sum_{j=1}^{p}x_{ij}\beta_j\right)\right]^{1-y_i}$$

- The above function can be used to derive Max. Likelihood Estimates

```
glm(y~X, family=binomial(link=probit))
```

- ML and Bayesian analysis can be difficult because the integrals involved do not have closed forms.
- Instead we will use ‘data augmentation’

Bayesian Model For Binary Outcomes

(Bayesian) Likelihood Function

$\tilde{y}_i$: Unobserved liability.

$$p\left(\tilde{y} \mid X, y, \beta\right) = \prod_{i:y_i=1} N\left(\tilde{y}_i \middle| x_i'\beta, 1\right) 1\left(\tilde{y}_i > 0\right) \times \prod_{i:y_i=0} N\left(\tilde{y}_i \middle| x_i'\beta, 1\right) 1\left(\tilde{y}_i \le 0\right)$$

$$= \prod_{i:y_i=1} \frac{e^{\left\{-\frac{\left(\tilde{y}_i - x_i'\beta\right)^2}{2}\right\}}}{\sqrt{2\pi}} 1\left(\tilde{y}_i > 0\right) \times \prod_{i:y_i=0} \frac{e^{\left\{-\frac{\left(\tilde{y}_i - x_i'\beta\right)^2}{2}\right\}}}{\sqrt{2\pi}} 1\left(\tilde{y}_i \le 0\right)$$

Prior $\quad p\left(\beta, \sigma_\varepsilon^2, \sigma_\beta^2, \ldots\right)$

Joint Posterior

$$p\left(\tilde{y}, \beta, \sigma_\varepsilon^2, \sigma_\beta^2, \ldots \mid X, y,\right) \propto \prod_{i:y_i=1} \frac{e^{\left\{-\frac{\left(\tilde{y}_i - x_i'\beta\right)^2}{2}\right\}}}{\sqrt{2\pi}} 1\left(\tilde{y}_i > 0\right) \times \prod_{i:y_i=0} \frac{e^{\left\{-\frac{\left(\tilde{y}_i - x_i'\beta\right)^2}{2}\right\}}}{\sqrt{2\pi}} 1\left(\tilde{y}_i \le 0\right) \times p\left(\beta, \sigma_\varepsilon^2, \sigma_\beta^2, \ldots\right)$$

Fully Conditional Distributions

$$p\left(\tilde{y}_i \mid y_i = 1, ELSE\right) \propto e^{\left\{-\frac{\left(\tilde{y}_i - x_i'\beta\right)^2}{2}\right\}} 1\left(\tilde{y}_i > 0\right)$$

$$p\left(\tilde{y}_i \mid y_i = 0, ELSE\right) \propto e^{\left\{-\frac{\left(\tilde{y}_i - x_i'\beta\right)^2}{2}\right\}} 1\left(\tilde{y}_i \le 0\right)$$

- These are truncated normal densities with mean $x_i'\beta$ and variance equal to 0. Truncated below or above zero depending on whether y_i *is* equal to 1 or 0, respectively.

Outline of a Gibbs Sampler

- Once we have sampled the un-observed liabilities the likelihood function becomes the standard Gaussian likelihood for data $\tilde{y}$.

$$p(\tilde{y} \mid X, y, \beta) = \prod_{i:y_i=1} N(\tilde{y}_i | x_i'\beta, 1) 1(\tilde{y}_i > 0) \times \prod_{i:y_i=0} N(\tilde{y}_i | x_i'\beta, 1) 1(\tilde{y}_i \leq 0)$$

$$= \prod_i N(\tilde{y}_i | x_i'\beta, 1)$$

- Therefore, the fully conditionals for all the parameters (regression coefficients, variances, etc.) are the same as the ones we discussed for the linear model, with one exception, the error variance is fixed to 1, and we do not need to sample it.

- Consequently, only a few modifications are needed to adapt the sampler we developed for the linear regression to accommodate binary outcomes (see sampler in GitHub).