

# STAT 238 - Bayesian Statistics

## Lecture Twenty Seven

Spring 2026, UC Berkeley

Aditya Guntuboyina

06 April 2026

### 1 Gibbs Sampler for Probit Regression

Our next application of the Gibbs sampler is to Probit Regression which was introduced in Homework Four. The model is given by the following. We observe data  $(y_1, x_1), \dots, (y_n, x_n)$  with  $y_1, \dots, y_n$  binary and  $x_1, \dots, x_n \in \mathbb{R}^p$ .  $x_1, \dots, x_n$  are treated non-random and  $y_1, \dots, y_n$  random. The parameter is  $\beta \in \mathbb{R}^p$ . The probit regression model assumes that  $y_1, \dots, y_n$  are independent given  $\beta$  with

$$y_i | \beta \sim \text{Bernoulli}(\Phi(x_i' \beta))$$

where  $\Phi(\cdot)$  is the standard normal cdf. Let  $\pi(\cdot)$  be our prior for  $\beta$ . The posterior is then given by:

$$\pi(\beta | \text{data}) \propto \pi(\beta) \prod_{i=1}^n (\Phi(x_i' \beta))^{y_i} (1 - \Phi(x_i' \beta))^{1-y_i}.$$

Suppose we take the (improper) uniform prior for  $\beta$  so that the posterior becomes

$$\pi(\beta | \text{data}) \propto \prod_{i=1}^n (\Phi(x_i' \beta))^{y_i} (1 - \Phi(x_i' \beta))^{1-y_i}.$$

It is not actually clear how to use the Gibbs sampler on the above posterior. We can try decomposing  $\beta$  as  $(\beta_1, \dots, \beta_p)$  and then attempt to obtain samples from the conditional:

$$\pi(\beta_j | \beta_{-j}, \text{data}) \propto \prod_{i=1}^n (\Phi(x_i' \beta))^{y_i} (1 - \Phi(x_i' \beta))^{1-y_i}.$$

I am not sure how to obtain samples from the above conditional.

There is a trick that people use to run the Gibbs sampler for probit regression. Let us introduce random variables  $w_1, \dots, w_n$  which, conditional on  $\beta$ , are independent with

$$w_i | \beta \sim N(x_i' \beta, 1).$$

We then let

$$y_i = I\{w_i > 0\}.$$

It is then clear that  $y_1, \dots, y_n$  defined thus have the same likelihood as in the probit regression model. Let  $w = (w_1, \dots, w_n)$ . The posterior of  $(w, \beta)$  given the data (data =  $y$  =

$(y_1, \dots, y_n)$  is

$$\begin{aligned}\pi(w, \beta | \text{data}) &\propto \pi(\beta) \pi(w | \beta) \pi(y | w, \beta) \\ &\propto \prod_{i=1}^n \exp\left(\frac{-1}{2}(w_i - x'_i \beta)^2\right) \prod_{i=1}^n (I\{w_i > 0, y_i = 1\} + I\{w_i \leq 0, y_i = 0\}) \\ &\propto \exp\left(\frac{-1}{2}\|w - X\beta\|^2\right) \prod_{i=1}^n (I\{w_i > 0, y_i = 1\} + I\{w_i \leq 0, y_i = 0\}).\end{aligned}$$

where  $X$  is, as usual, the design matrix with rows  $x'_1, \dots, x'_n$ .

We shall now run the Gibbs sampler with the decomposition  $\theta = (w, \beta)$ . For this, we need to sample from the conditionals  $\beta | w, \text{data}$  and  $w | \beta, \text{data}$ .

$$\begin{aligned}\pi(\beta | w, \text{data}) &\propto \pi(w, \beta | \text{data}) \\ &\propto \exp\left(\frac{-1}{2}\|w - X\beta\|^2\right) \prod_{i=1}^n (I\{w_i > 0, y_i = 1\} + I\{w_i \leq 0, y_i = 0\}) \\ &\propto \exp\left(\frac{-1}{2}\|w - X\beta\|^2\right)\end{aligned}$$

For a fixed  $w \in \mathbb{R}^n$ , let

$$\hat{\beta}_w := \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \|w - X\beta\|^2 = (X'X)^{-1}X'w$$

so that

$$\|w - X\beta\|^2 = \|w - X\hat{\beta}_w\|^2 + \|X\hat{\beta}_w - X\beta\|^2 = \|w - X\hat{\beta}_w\|^2 + (\beta - \hat{\beta}_w)'(X'X)(\beta - \hat{\beta}_w).$$

Thus

$$\pi(\beta | w, \text{data}) \propto \exp\left(\frac{-1}{2}(\beta - \hat{\beta}_w)'(X'X)(\beta - \hat{\beta}_w)\right).$$

In other words,

$$\beta | w, \text{data} \sim N\left(\hat{\beta}_w, (X'X)^{-1}\right).$$

The next task is to figure out the conditional  $w | \beta, \text{data}$ :

$$\begin{aligned}\pi(w | \beta, \text{data}) &\propto \pi(w, \beta | \text{data}) \\ &\propto \exp\left(\frac{-1}{2}\|w - X\beta\|^2\right) \prod_{i=1}^n (I\{w_i > 0, y_i = 1\} + I\{w_i \leq 0, y_i = 0\}) \\ &\propto \prod_{i=1}^n \exp\left(\frac{-1}{2}(w_i - x'_i \beta)^2\right) (I\{w_i > 0, y_i = 1\} + I\{w_i \leq 0, y_i = 0\})\end{aligned}$$

Thus  $w_1, \dots, w_n$  are independent conditional on  $\beta$  and data. Further

$$\pi(w_i | \beta, \text{data}) = \pi(w_i | \beta, y_i) \propto \exp\left(\frac{-1}{2}(w_i - x'_i \beta)^2\right) (I\{w_i > 0, y_i = 1\} + I\{w_i \leq 0, y_i = 0\}).$$

It makes sense now to separate the two cases  $y_i = 1$  and  $y_i = 0$ . For  $y_i = 1$ , we get

$$\pi(w_i | \beta, y_i = 1) \propto \exp\left(\frac{-1}{2}(w_i - x'_i \beta)^2\right) I\{w_i > 0\}.$$

Thus, conditional on  $\beta$  and  $y_i = 1$ , the random variable  $w_i$  has the  $N(x'_i\beta, 1)$  distribution truncated to the positive half-line.

Here is a fact (from the wikipedia article for truncated normal distributions: [https://en.wikipedia.org/wiki/Truncated\\_normal\\_distribution#Generating\\_values\\_from\\_the\\_truncated\\_normal\\_distribution](https://en.wikipedia.org/wiki/Truncated_normal_distribution#Generating_values_from_the_truncated_normal_distribution)) that is useful for simulating from truncated normal distributions.

**Fact 1.1.** Consider a random variable  $X_{trunc}$  having the truncated  $N(\mu, \sigma^2)$  truncated to the interval  $(a, b)$  (here  $a < b$ ;  $a$  is allowed to be  $-\infty$  and  $b$  is allowed to be  $+\infty$ ). The density of  $X_{trunc}$  is equal to the conditional density of  $X$  conditioned on  $a < X < b$  where  $X \sim N(\mu, \sigma^2)$ .

$X_{trunc}$  can be simulated using the equation:

$$X_{trunc} = \mu + \sigma\Phi^{-1}((1 - U)\Phi(a_0) + U\Phi(b_0)) \quad \text{where } a_0 := \frac{a - \mu}{\sigma} \text{ and } b_0 := \frac{b - \mu}{\sigma}$$

and  $U$  has the uniform distribution on  $(0, 1)$ . I will leave as an exercise for you to verify that the right hand side above indeed has the required truncated normal distribution.

Using this fact, it is straightforward to verify that  $w_i|\{\beta, y_i = 1\}$  can be simulated by

$$x'_i\beta + \Phi^{-1}((1 - U)\Phi(-x'_i\beta) + U).$$

This follows directly from the fact above by taking  $\mu = x'_i\beta$ ,  $\sigma = 1$ ,  $a = 0$  and  $b = \infty$  (this implies that  $a_0 = -x'_i\beta$  and  $b_0 = \infty$ ).

One can argue analogously to deduce that  $w_i|\{\beta, y_i = 0\}$  can be simulated by

$$x'_i\beta + \Phi^{-1}(U\Phi(-x'_i\beta)).$$

The Gibbs sampler for the probit model can be therefore implemented by the following model:

1. Initialize at  $\beta^{(0)}$  and  $w^{(0)}$ .
2. Repeat the following for  $t = 1, \dots, N$ 
  - a) Pick  $J$  to be either 1 or 2 with equal probability.
  - b) If  $J = 1$ , take  $\beta^{(t+1)} \sim N((X'X)^{-1}X'w^{(t)}, (X'X)^{-1})$  and  $w^{(t+1)} = w^{(t)}$ .
  - c) If  $J = 2$ , take  $\beta^{(t+1)} = \beta^{(t)}$  and

$$w_i = \begin{cases} x'_i\beta^{(t)} + \Phi^{-1}((1 - U_i)\Phi(-x'_i\beta^{(t)}) + U_i) & \text{if } y_i = 1 \\ x'_i\beta^{(t)} + \Phi^{-1}(U_i\Phi(-x'_i\beta^{(t)})) & \text{if } y_i = 0 \end{cases}$$

where  $U_i \sim \text{Unif}(0, 1)$ .