

Lecture 2: Prior Distributions

Instructor: Fei Tan

 @econdojo  @BusinessSchool101  Saint Louis University

Course: Introduction to Bayesian Statistics

Date: January 31, 2026

The Road Ahead

1. Preliminary
2. Specifying Prior Distributions

Normal Linear Regression

Model

$$y_i = \beta_1 x_{i1} + \cdots + \beta_K x_{iK} + u_i, \quad u_i | x_i \sim_{i.i.d.} \mathcal{N}(0, \sigma^2), \quad i = 1, \dots, n$$

- Compact notation: $y = X\beta + u$, where

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nK} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

- Likelihood function

$$f(y|\beta, \sigma^2) \propto \left(\frac{1}{\sigma^2} \right)^{n/2} \exp \left[-\frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta) \right]$$

Conjugate Priors

- Normal-inverse-gamma (type-2) prior

$$\pi(\beta, \sigma^2) = \underbrace{\mathcal{N}(\beta | \beta_0, \sigma^2 B_0)}_{\pi(\beta | \sigma^2)} \underbrace{\text{IG-2}(\sigma^2 | \alpha_0/2, \delta_0/2)}_{\pi(\sigma^2)}$$

- Exercise: posterior is of same family

$$\pi(\beta, \sigma^2 | y) = \underbrace{\mathcal{N}(\beta | \beta_1, \sigma^2 B_1)}_{\pi(\beta | \sigma^2, y)} \underbrace{\text{IG-2}(\sigma^2 | \alpha_1/2, \delta_1/2)}_{\pi(\sigma^2 | y)}$$

where

$$B_1 = (X'X + B_0^{-1})^{-1}$$

$$\beta_1 = B_1(X'y + B_0^{-1}\beta_0)$$

$$\alpha_1 = \alpha_0 + n$$

$$\delta_1 = \delta_0 + y'y + \beta_0'B_0^{-1}\beta_0 - \beta_1'B_1^{-1}\beta_1$$

- Exercise: $\pi(\beta | y) = t_{\alpha_1}(\beta_1, \frac{\delta_1}{\alpha_1} B_1)$, $\pi(\sigma^2 | y) = \text{IG-2}(\frac{\alpha_1}{2}, \frac{\delta_1}{2})$

(Im)proper Priors

- Proper priors integrate to unity, e.g.
 - $\pi(\sigma^2) = \text{IG-2}(\alpha_0/2, \delta_0/2)$
 - equivalently, $h = 1/\sigma^2$ (precision), $\pi(h) = \mathcal{G}(\alpha_0/2, \delta_0/2)$
- Improper priors are not integrable
 - improper vs. uninformative/diffuse prior
 - e.g. $\pi(\beta) \propto c > 0, \pi(\sigma) \propto 1/\sigma$ (Jeffreys prior)
 - posterior may still be proper
- We work with proper prior
 - available information/methods to avoid improper prior
 - $m(y)$ based on improper prior can be manipulated

Hierarchical Models

Model

Hyperparameter prior: $\alpha \sim \pi(\alpha|\delta)$

Parameter prior: $\theta \sim \pi(\theta|\alpha)$

Likelihood: $y \sim f(y|\theta)$

- Remarks
 - $f(y|\theta, \alpha) = f(y|\theta) \Rightarrow \alpha$ not identified
 - introduce α to facilitate computation/modeling
- Examples
 - Student- t error: $u_i|x_i \sim_{i.i.d.} t_\nu(0, \sigma^2), \nu \sim \pi(\nu|\nu_0)$
 - DSGE prior for VAR as will be covered later

Training Sample Priors

Bayesian updating

$$\text{Training sample } y_1 : \quad \pi(\theta|y_1) \propto f(y_1|\theta)\pi(\theta) \Rightarrow \underbrace{\pi(\theta|\alpha(y_1))}_{\text{posterior}}$$

$$\text{Remaining sample } y_2 : \quad \pi(\theta|y_2, \alpha(y_1)) \propto f(y_2|\theta) \underbrace{\pi(\theta|\alpha(y_1))}_{\text{prior}}$$

- Consider linear regression
 - improper prior: $\pi(\beta) \propto c, \pi(\sigma) \propto 1/\sigma$
 - proper joint posterior: $\beta|\sigma^2, y \sim \mathcal{N}(\hat{\beta}, \sigma^2(X'X)^{-1})$, where
 $\hat{\beta} = (X'X)^{-1}X'y$, and $\sigma^2|y \sim \text{IG-2}((n - K)/2, S^2/2)$, where
 $S^2 = (y - X\hat{\beta})'(y - X\hat{\beta})$

Conditionally Conjugate Priors

- Independent priors

$$\pi(\beta, \sigma^2) = \underbrace{\mathcal{N}(\beta | \beta_0, B_0)}_{\pi(\beta)} \underbrace{\text{IG-2}(\sigma^2 | \alpha_0/2, \delta_0/2)}_{\pi(\sigma^2)}$$

- Exercise: *conditional* posteriors are of same family

$$\pi(\beta | \sigma^2, y) = \mathcal{N}(\beta | \beta_1, B_1), \quad \pi(\sigma^2 | \beta, y) \propto \text{IG-2}(\sigma^2 | \alpha_1/2, \delta_1/2)$$

where

$$\begin{aligned} B_1 &= (\sigma^{-2} X' X + B_0^{-1})^{-1} \\ \beta_1 &= B_1 (\sigma^{-2} X' y + B_0^{-1} \beta_0) \\ \alpha_1 &= \alpha_0 + n \\ \delta_1 &= \delta_0 + (y - X\beta)'(y - X\beta) \end{aligned}$$

Readings

- Garthwaite, Kadane & O'Hagan (2005), "Statistical Methods for Eliciting Probability Distributions," *Journal of the American Statistical Association*
- O'Hagan et al. (2006), "Uncertain Judgements: Eliciting Experts' Probabilities," John Wiley & Sons