Prior Information and Prior Distributions

Donald E. Brown

School of Data Science
University of Virginia
Charlottesville, VA 22904

Bayes Theorem

- For variables $\mathbf{X}$ and model parameters $\boldsymbol{\theta}$, Bayes theorem tells us how to obtain posterior beliefs about $\boldsymbol{\theta}$ after observing $\mathbf{X}$:

$$\underbrace{p(\boldsymbol{\theta}|\mathbf{X})}_{\text{posterior}} = \frac{\underbrace{p(\mathbf{X}|\boldsymbol{\theta})}_{\text{likelihood}}\underbrace{p(\boldsymbol{\theta})}_{\text{prior}}}{\underbrace{p(\mathbf{X})}_{\text{evidence}}}$$

- How do we obtain the prior?

Characteristics of Prior Information

- Impact of the choice of prior - small, medium, large
- Prior information does not precisely determine the prior distribution
- Common approaches
  - Base case or rates in existing populations
  - Subjective probabilities
  - Entropy maximization
  - Conjugate priors
  - Noninformative priors

Definition of Conjugate Prior

Posterior has same distribution with different parameters as the prior, so with:

$$\underbrace{p(\boldsymbol{\theta}|\mathbf{X})}_{\text{posterior}} = \frac{\underbrace{p(\mathbf{X}|\boldsymbol{\theta})}_{\text{likelihood}}\underbrace{p(\boldsymbol{\theta})}_{\text{prior}}}{\underbrace{p(\mathbf{X})}_{\text{evidence}}}$$

Then:

$$p(\boldsymbol{\theta}|\mathbf{X}) \sim p(\boldsymbol{\theta})$$

Advantages to Conjugate Priors

- Convenient choice for a prior
- Analytical solution
- Fast computation
- Fast prediction

Disadvantages to Conjugate Priors

- Minimizes the importance of the data
- Restrictive
- Right solution to the wrong problem
- Needs sensitivity analysis to determine the impacts