ADVANCED BAYESIAN MODELING

2016 Presidential Polls Data and Model Assumptions

2016 Presidential Polls Example

Seven polls conducted within days of 2016 U.S. presidential election for two-way race (Clinton v. Trump):

$$
\begin{aligned}
y_j &= \text{Clinton lead (percentage points) in poll } j \\
\sigma_j &= \text{half margin of error of } y_j \\
j &= 1, \ldots, 7
\end{aligned}
$$

Regard σ_js as fixed and known.

Model similar to one proposed in BDA3, Sec. 5.4:

$$
\begin{aligned}
y_j \mid \theta_j &\sim \mathrm{N}(\theta_j, \sigma_j^2) \qquad j = 1, \ldots, 7 \\
\theta_j \mid \mu, \tau &\sim \mathrm{N}(\mu, \tau^2) \qquad j = 1, \ldots, 7 \\
\mu &\sim \text{flat on } (-\infty, \infty) \\
\tau &\sim \text{flat on } (0, \infty)
\end{aligned}
$$

The fundamental assumptions:

- *Poll results have normal sampling distribution.*

  Not exactly true, but nearly true because of central limit theorem.

- *Poll means are exchangeable under prior.*

  More like an assertion: We choose to let the data distinguish between the polls, without incorporating any prior knowledge (other than margin of error).

- *Poll means have a normal population (prior) distribution.*

  Conjugacy makes model easier to analyze, but not justified by any actual prior information. Should be checked.

- *Hyperparameters have flat prior.*

  Mathematically convenient, and seems noninformative, but posterior implications are unclear (e.g., too much probability toward large τ values?). Should be checked.