STA 360/602L: Module 3.6

Noninformative and improper priors

Dr. Olanrewaju Michael Akande

Noninformative and improper priors

- Generally, we must specify both μ_0 and τ_0 to do inference.
- When prior distributions have no population basis, that is, there is no justification of the prior as "prior data", prior distributions can be difficult to construct.
- To that end, there is often the desire to construct noninformative priors, with the rationale being *"to let the data speak for themselves"*.
- For example, we could instead assume a uniform prior on μ that is constant over the real line, i.e., $\pi(\mu) \propto 1 \Rightarrow$ all values on the real line are equally likely apriori.
- Clearly, this is not a valid pdf since it will not integrate to 1 over the real line. Such priors are known as improper priors.
- An improper prior can still be very useful, we just need to ensure it results in a proper posterior.

Jeffreys' prior

- Question: is there a prior pdf (for a given model) that would be universally accepted as a noninformative prior?

- Laplace proposed the uniform distribution. This proposal lacks invariance under monotone transformations of the parameter.

- For example, a uniform prior on the binomial proportion parameter θ is not the same as a uniform prior on the odds parameter $\phi = \dfrac{\theta}{1-\theta}$.

- A more acceptable approach was introduced by Jeffreys. For single parameter models, the Jeffreys' prior defines a noninformative prior density of a parameter θ as

$$\pi(\theta) \propto \sqrt{\mathcal{I}(\theta)}$$

where $\mathcal{I}(\theta)$ is the Fisher information for θ.

Jeffreys' prior

- The Fisher information gives a way to measure the amount of information a random variable Y carries about an unknown parameter θ of a distribution that describes Y.

- Formally, $\mathcal{I}(\theta)$ is defined as

$$\mathcal{I}(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log p(y|\theta)\right)^2 \Bigg| \theta\right] = \int_{\mathcal{Y}} \left(\frac{\partial}{\partial\theta}\log p(y|\theta)\right)^2 p(y|\theta)dy.$$

- Alternatively,

$$\mathcal{I}(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial^2\theta}\log p(y|\theta)\Bigg|\theta\right] = -\int_{\mathcal{Y}} \left(\frac{\partial^2}{\partial^2\theta}\log p(y|\theta)\right) p(y|\theta)dy.$$

- Turns out that the Jeffreys' prior for μ under the normal model, when σ^2 is known, is

$$\pi(\mu) \propto 1,$$

the uniform prior over the real line. Let's derive this on the board.

Inference for mean, conditional on variance using Jeffreys' prior

- Recall that for σ^2 known, the normal likelihood simplifies to

$$\propto \exp\left\{-\frac{1}{2}\tau n(\mu - \bar{y})^2\right\},$$

  ignoring everything else that does not depend on μ.

- With the Jeffreys' prior $\pi(\mu) \propto 1$, can we derive the posterior distribution?

Inference for mean, conditional on variance using Jeffreys' prior

- Posterior:

$$\begin{aligned}\pi(\mu|Y,\tau) &\propto \exp\left\{-\frac{1}{2}\tau n(\mu-\bar{y})^2\right\}\pi(\mu)\\ &\propto \exp\left\{-\frac{1}{2}\tau n(\mu-\bar{y})^2\right\}.\end{aligned}$$

- This is the kernel of a normal distribution with
  - mean $\bar{y}$, and
  - precision $n\tau$ or variance $\dfrac{1}{n\tau} = \dfrac{\sigma^2}{n}$.
- Written differently, we have $\mu|Y,\sigma^2 \sim \mathcal{N}(\bar{y}, \dfrac{\sigma^2}{n})$
- This should look familiar to you. Does it?

Improper prior

- Let's be very objective with the prior selection. In fact, let's be extreme!
  - If we let the normal variance $\to \infty$ then our prior on μ is $\propto 1$ (recall the Jeffreys' prior on μ for known σ^2).
  - If we let the gamma variance get very large (e.g., $a, b \to 0$), then the prior on σ^2 is $\propto \dfrac{1}{\sigma^2}$.
- $\pi(\mu, \sigma^2) \propto \dfrac{1}{\sigma^2}$ is improper (does not integrate to 1) but does lead to a proper posterior distribution that yields inferences similar to frequentist ones.
- For that choice, we have

$$\mu | Y, \tau \sim \mathcal{N}\left(\bar{y}, \frac{1}{n\tau}\right)$$

$$\tau | Y \sim \text{Gamma}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$$

Analysis with noninformative priors

- Recall the Pygmalion data:
  - Accelerated group (A): 20, 10, 19, 15, 9, 18.
  - No growth group (N): 3, 2, 6, 10, 11, 5.
- Summary statistics:
  - $\bar{y}_A = 15.2$; $s_A = 4.71$.
  - $\bar{y}_N = 6.2$; $s_N = 3.65$.
- So our joint posterior is

$$\mu_A | Y_A, \tau_A \sim \mathcal{N}\left(\bar{y}_A, \frac{1}{n_A \tau_A}\right) = \mathcal{N}\left(15.2, \frac{1}{6\tau_A}\right)$$

$$\tau_A | Y_A \sim \text{Gamma}\left(\frac{n_A - 1}{2}, \frac{(n_A - 1)s_A^2}{2}\right) = \text{Gamma}\left(\frac{6-1}{2}, \frac{(6-1)(22.17)}{2}\right)$$

$$\mu_N | Y_N, \tau_N \sim \mathcal{N}\left(\bar{y}_N, \frac{1}{n_N \tau_N}\right) = \mathcal{N}\left(6.2, \frac{1}{6\tau_N}\right)$$

$$\tau_N | Y_N \sim \text{Gamma}\left(\frac{n_N - 1}{2}, \frac{(n_N - 1)s_A^2}{2}\right) = \text{Gamma}\left(\frac{6-1}{2}, \frac{(6-1)(13.37)}{2}\right)$$

Monte Carlo sampling

It is easy to sample from these posteriors:

```
aA <- (6-1)/2
aN <- (6-1)/2
bA <- (6-1)*22.17/2
bN <- (6-1)*13.37/2
muA <- 15.2
muN <- 6.2
tauA_postsample_impr <- rgamma(10000,aA,bA)
thetaA_postsample_impr <- rnorm(10000,muA,sqrt(1/(6*tauA_postsample_impr)))
tauN_postsample_impr <- rgamma(10000,aN,bN)
thetaN_postsample_impr <- rnorm(10000,muN,sqrt(1/(6*tauN_postsample_impr)))
sigma2A_postsample_impr <- 1/tauA_postsample_impr
sigma2N_postsample_impr <- 1/tauN_postsample_impr
```

Monte Carlo sampling

- Is the average improvement for the accelerated group larger than that for the no growth group?
  - What is $\Pr[\mu_A > \mu_N | Y_A, Y_N)$?

```
mean(thetaA_postsample_impr > thetaN_postsample_impr)
```

```
## [1] 0.9951
```

- Is the variance of improvement scores for the accelerated group larger than that for the no growth group?
  - What is $\Pr[\sigma^2_A > \sigma^2_N | Y_A, Y_N)$?

```
mean(sigma2A_postsample_impr > sigma2N_postsample_impr)
```

```
## [1] 0.7041
```

- How does the new choice of prior affect our conclusions?

Recall the job training data

- Data:
  - No training group (N): sample size $n_N = 429$.
  - Training group (T): sample size $n_A = 185$.
- Summary statistics for change in annual earnings:
  - $\bar{y}_N = 1364.93$; $s_N = 7460.05$
  - $\bar{y}_T = 4253.57$; $s_T = 8926.99$
- Using the same approach we used for the Pygmalion data, answer the questions of interest.

What's next?

Move on to the readings for the next module!