Biostatistics 682: Applied Bayesian Inference
Lecture 5: More on single parameter models

Jian Kang

Department of Biostatistics
University of Michigan, Ann Arbor

Noninformative pirors

- Let the data speak for themselves
- How to guarantee to prior distributions to play a minimal role in the posterior distribution?
- Flat or diffuse improper priors that lead to proper posterior distributions
  - $y \sim \mathrm{N}(\mu, 1)$ with $\pi(\mu) \propto 1$
  - $y \sim \mathrm{N}(0, \sigma^2)$ with $\pi(\sigma^2) \propto 1/\sigma^2$
- Other approaches?
  - Jeffrey's prior
  - Prior for location/scale parameters
  - Weakly informative priors

Jeffrey's invariance principle

- One approach is based on Jeffrey's invariance principle
- Consider one-to-one transformations of the parameter $\phi = h(\theta)$.
- By transformation of variables, the prior density $\pi(\theta)$ is equivalent, in terms of expressing the same beliefs, to the following prior density on ϕ:

$$\pi(\phi) = \pi(\theta)|h'(\theta)|^{-1}. \tag{1}$$

- The Jeffrey's general principle:
  Any rule for determining the prior density $\pi(\theta)$ should yield an equivalent result if applied to the transformed parameter $\phi = h(\theta)$ for any one-to-one transformation h. That is,
  - According the rule, determine $\pi(\theta)$ using $\pi(y \mid \theta)$.
  - According the rule, determine $\pi(\phi)$ using $\pi(y \mid \phi)$.
  - (1) should be satisfied.

The Jeffrey's Prior

- The Jeffrey's prior is given by

$$\pi(\theta) \propto \{I(\theta)\}^{1/2},$$

where $I(\theta)$ is the Fisher information for θ:

$$I(\theta) = E\left[\left\{\frac{d\log\pi(y\mid\theta)}{d\theta}\right\}^2\middle|\theta\right] = -E\left\{\frac{d^2\log\pi(y\mid\theta)}{d\theta^2}\middle|\theta\right\}$$

- Can we verify the Jeffrey's principle?

Since Jeffrey's prior for ϕ is $\pi(\phi) \propto \{I(\phi)\}^{1/2}$. To verify the Jeffrey's principle, we need to show $\{I(\phi)\}^{1/2} \propto \{I(\theta)\}^{1/2}|h'(\theta)|^{-1}$. In fact,

$$\begin{aligned}
I(\phi) &= E\left[\left\{\frac{d\log\tilde{\pi}(y\mid\phi)}{d\phi}\right\}^2\middle|\phi = h(\theta)\right] \\
&= E\left[\left\{\frac{d\log\pi(y\mid\theta = h^{-1}(\phi))}{d\theta}\frac{d\theta}{d\phi}\right\}^2\middle|\theta\right] = E\left[\left\{\frac{d\log\pi(y\mid\theta)}{d\theta}\right\}^2\middle|\theta\right]\left(\frac{d\theta}{d\phi}\right)^2
\end{aligned}$$

Note that

$$\frac{d\theta}{d\phi} = \frac{1}{h'(\theta)}.$$

Example: Normal model

Suppose

$$y \sim \mathrm{N}(\mu, \sigma^2)$$

- When σ^2 is known, the Jeffrey's prior for μ is

$$\pi(\mu) \propto \sqrt{I(\mu)} = \sqrt{\mathrm{E}\left\{\left(\frac{x-\mu}{\sigma^2}\right)^2\right\}} \propto 1$$

  It's an improper prior but leads to proper posterior.
- How about the Jeffrey's prior for μ^3? $\pi(\mu^3) \propto \frac{1}{\mu^2}$
- When μ is known, the Jeffrey's prior for σ^2 is

$$\pi(\sigma^2) \propto \sqrt{I(\sigma^2)} \propto \sqrt{\mathrm{E}\left\{\left(\frac{(x-\mu)^2-\sigma^2}{\sigma^4}\right)^2\right\}} \propto \frac{1}{\sigma^2}$$

- How about the Jeffrey's prior for $\log(\sigma^2)$? $\pi\{\log(\sigma^2)\} \propto 1$.

Example: Binomial model

Suppose

$$y \sim \text{Binomial}(n, \theta)$$

Then

$$I(\theta) = \frac{n}{\theta(1-\theta)}$$

- The Jeffrey's prior for θ is

  $$\pi(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}.$$

  This is

  $$\theta \sim \text{beta}(1/2, 1/2)$$

- What is the prior for $\text{logit}(\theta)$? Let $\phi = \text{logit}(\theta)$. Then $\theta = 1/\{1+\exp(-\phi)\}$ and $\frac{d\theta}{d\phi} = \exp(-\phi)/\{1+\exp(-\phi)\}^2$

  $$\pi(\phi) \propto \frac{\exp(-\phi/2)}{1+\exp(-\phi)} \Rightarrow \pi\left\{\text{logit}(\theta)\right\} \propto \frac{\theta}{2}$$

Pivotal Quantity

- A function of observation and unknown parameters whose probability distribution does not depend on the unknown parameters.
- Different principles may give (slightly) different noninformative prior distributions.
- But for two cases – location parameters and scale parameters – all principles seem to agree.

Location Parameters

- If a density of y is such that $\pi(y-\theta \mid \theta)$ is a function that is free of θ and y, say, $f(u)$, where $u = y - \theta$, then $y - \theta$ is a pivotal quantity, and θ is called a location parameter.
- Noninformative prior distribution for θ should ensure that $y - \theta$ also a pivotal quantity (free of the unknown true parameter θ_0) under the posterior distribution of θ.
$$\pi(u \mid \theta) = \pi(u \mid y) = f(u).$$
- Note that
$$\pi(\theta \mid y) \propto \pi(\theta)\pi(y \mid \theta).$$
- Thus
$$\pi(\theta) \propto 1.$$
- Example: normal distribution, t-distribution, Laplace distribution

Scale Parameters

- If a density of y is such that $\pi(y/\theta \mid \theta)$ is a function that is free of θ and y, say, $g(u)$, where $u = y/\theta$, then y/θ is a pivotal quantity, and θ is called a scale parameter.
- Noninformative prior distribution for θ should ensure that y/θ also a pivotal quantity (free of the unknown true parameter θ_0) under the posterior distribution of θ.

$$\pi(u \mid \theta) = \pi(u \mid y) = g(u)$$

- Note that

$$\pi(y \mid \theta) = \frac{1}{\theta}\pi(u \mid \theta)$$

$$\pi(\theta \mid y) = \frac{y}{\theta^2}\pi(u \mid y)$$

$$\pi(\theta \mid y) \propto \pi(y \mid \theta)\pi(\theta).$$

- Thus

$$\pi(\theta) \propto 1/\theta.$$

Difficulties with Noninformative Priors

- Searching for a prior distribution that is always vague seems misguided: If the likelihood is truly dominant in a given problem, then the choice among a range of relatively flat prior densities cannot matter.
- For many problems, there is no clear choice for a vague prior distribution, since a density that is flat or uniform in one parameterization will not be in another. For example, for normal model $y \sim \mathrm{N}(\mu, 1)$, we can assume $\pi(\mu) \propto 1$, i.e., a uniform flat prior. How about $\pi\{\exp(\mu)\}$? still uniform?
- Noninformative priors are often useful when it does not seem to be worth the effort to quantify one's real prior knowledge as a probability distribution, as long as one is willing to perform the mathematical work to check that the posterior density is proper

Weakly Informative Priors

- Prior distribution is weakly informative if it is proper but is set up so that the information it does provide is intentionally weaker than whatever actual prior knowledge.
- Examples: $\mu \sim \mathrm{N}(0, 10^6)$, $\sigma^2 \sim \mathrm{G}^{-1}(0.001, 0.001)$.
- Principles for setting up the weakly informative priors.
- Start with some version of a noninformative prior distribution and then add enough information so that inferences are constrained to be reasonable
- Start with a strong, highly informative prior and broaden it to account for uncertainty in ones' prior beliefs and in the applicability of any historically based prior distribution to new data.