More on Priors

PUBH 8442: Bayes Decision Theory and Data Analysis

Eric F. Lock
UMN Division of Biostatistics, SPH
elock@umn.edu

01/27/2021

Non-informative priors

- A non-informative prior is intended to convey no prior belief
  - Let the data alone drive inference.
- Example: For $\Theta = \{\theta_1, \ldots, \theta_K\}$

$$P(\theta = \theta_k) = \frac{1}{K}.$$

- Example: For $\Theta = [a, b]$,

$$p(\theta) = \frac{1}{b-a} \quad \text{for } a \leq \theta \leq b.$$

- Note: If $p(\theta) = c \ \forall \ \theta \in \Theta$, then

$$p(\theta \mid \mathbf{y}) \propto p(\mathbf{y} \mid \theta),$$

and so the mode of $p(\theta \mid \mathbf{y})$ is the MLE for θ.

Improper priors

- A "prior" that is not a true probability distribution is *improper*
  - An improper prior can result in an improper posterior

- Example: $p(\theta) = 1$ for $\theta \in (-\infty, \infty)$.
  - Improper because $\int p(\theta)\, d\theta = \infty$
  - If $\int p(\mathbf{y} \mid \theta)\, d\theta < \infty$,

$$p(\theta \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \theta)}{\int p(\mathbf{y} \mid \theta)\, d\theta}$$

is still proper.

- Improper priors should be avoided if possible.
  - But can give reasonable results, especially if posterior is proper.

$\int p(y \mid \theta) p(\theta) < \infty$

Example: Venusian surface

- A spacecraft lands on the planet Venus, and captures a small surface sample.
- The radioactivity of the sample is measured, as the frequency of particle emission.
- $y = 60$ particle emissions are observed in one minute.
- Then the spacecraft dies, due to high surface temperature and pressure.
- Infer θ, the average particle emissions per minute.

$$y \sim \text{Poisson}(\theta).$$

Example: Venusian surface

- Consider the uniform prior

$$p(\theta) = c \text{ for } \theta > 0.$$

  - NOTE: $p(\theta)$ is improper. $\int_0^\infty c = \infty$

$$P(y = k \mid \theta) = \frac{e^{-\theta}\theta^k}{k!}$$

$$\rightarrow p(\theta \mid y = k) \propto p(y = k \mid \theta)\, p(\theta)$$

$$\propto e^{-\theta}\theta^k$$

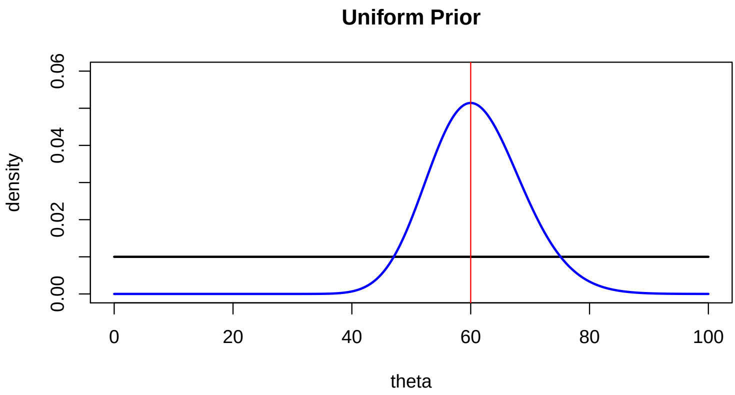
$$\propto \text{Gamma}(k+1, 1)$$

- Posterior IS proper:

$$p(\theta \mid y = 60) = \text{Gamma}(61, 1)$$

Example: Venusian surface

- **Prior** and posterior densities, given $y = 60$:



Code:

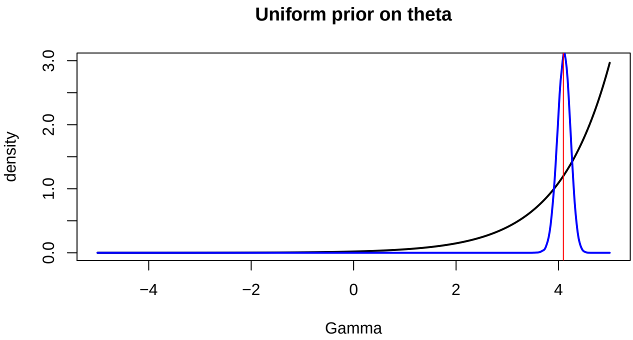
http://www.ericfrazerlock.com/More_on_priors_Rcode1.r

Example: Venusian surface

- Radioactivity by emission frequency is often given on log scale:

$$\gamma = \log(\theta).$$

- A uniform prior for θ implies prior $p^*(\gamma) \propto e^\gamma$ for γ:



-

- Prior is not "non-informative" for γ!

Jeffreys priors

- For a one-parameter model $p(\mathbf{y} \mid \theta)$, the Fisher information is

$$I(\theta) = -E_{\mathbf{y} \mid \theta} \left[\frac{d^2}{d\theta^2} \log p(\mathbf{y} \mid \theta) \right] .$$

- The Jeffreys prior is proportional to the square root of I:

$$p(\theta) \propto [I(\theta)]^{1/2}$$

- Jeffreys priors are invariant under all one-to-one transformations:
  - If $\gamma = h(\theta)$ and $p(\theta)$ is Jeffreys for θ, then the induced prior for γ is Jeffreys for γ.
- Satisfy other "non-informative" properties (Box & Tiao, Sec 1.3).

Fact ①: For $y = h(x)$ and x has density $p(x)$,

density of y is $\left|\frac{dx}{dy}\right| p(x)$ where $x = h^{-1}(y)$.

$\therefore$ need to show $[I(r)]^{\frac{1}{2}} = [I(\theta)]^{1/2} \cdot \left|\frac{d\theta}{dr}\right|$

Fact ②: $E_{y|\theta} \underbrace{\frac{d \log p(y|\theta)}{d\theta}}_{\text{score statistic}} = 0$

By chain rule and product rule:

$$\frac{d^2}{dr^2} \log p(y|r) = \frac{d \log p(y|\theta)}{d(\theta)} \cdot \frac{d^2\theta}{dr^2} + \frac{d^2 \log p(y|\theta)}{d\theta^2} \cdot \left(\frac{d\theta}{dr}\right)^2$$

By ②, $I(r) = -E_{y|\theta}\left[\frac{d^2 \log p(y|\theta)}{d\theta^2}\right] \cdot \left(\frac{d\theta}{dr}\right)^2 = I(\theta)\left(\frac{d\theta}{dr}\right)^2$

$$\longrightarrow \sqrt{I(r)} = \sqrt{I(\theta)} \cdot \left|\frac{d(\theta)}{dr}\right|$$

Jeffreys priors

- For multi-parameter $\theta = (\theta_1, \ldots, \theta_K)$, the Fisher information is a matrix with i, j element

$$I_{ij}(\theta) = -E_{\mathbf{y} \mid \theta} \left[\frac{d^2}{d\theta_i \theta_j} \log p(\mathbf{y} \mid \theta) \right].$$

- And the Jeffreys prior is

$$p(\theta) \propto [\det I(\theta)]^{1/2}$$

Jeffreys priors

- For a Poisson(θ) model, the Jeffreys prior for θ is

$$p(\theta) \propto \frac{1}{\sqrt{\theta}}$$

- For a Binomial(n, θ) model, the Jeffreys prior for θ is

$$p(\theta) = \text{Beta}(0.5, 0.5)$$

  - Homework

Elicited priors

- Often desire an expert to elicit the prior for a given experiment
- Issues:
  - Human intuition generally approximates uncertainty poorly
  - Non-statistician experts have little knowledge of probability distributions
- Have expert create a probability histogram
- Or specify certain desirable properties (e.g., quantiles) and choose a known distribution to match those properties.
- Prior elicitation software: `http://www.tonyohagan.co.uk/shelf/`.

Poisson-gamma model

- Let $y \sim \text{Poisson}(\theta)$ and $\theta \sim \text{Gamma}(\alpha, \beta)$:

$$p(\theta) \propto \theta^{\alpha-1} e^{-\beta\theta}.$$

- Then, $p(\theta \mid y) = \text{Gamma}(\alpha + y, \beta + 1)$

Example: Venusian surface (cont.)

- An expert in planetary radioactivity is specifies percentiles for radioactivity of Venusian surface
  - 10th percentile: 1, 90th percentile: 80
  - "80 percent sure that between 1 and 80 parts will emit per minute"
- Fit a Gamma prior using this information.
  - Code: http://www.ericfrazerlock.com/More_on_priors_Rcode2.r
  - $\theta \sim \text{Gamma}(0.62, 0.02)$
- Posterior after $y = 60$:

$$\text{Gamma}(60.62, 1.02)$$

Example: Venusian surface (cont.)

- **Prior** and posterior densities, given $y = 60$: