Bayesian Statistics

Prior Elicitation

Brani Vidakovic

Professor

School of Industrial and Systems Engineering

Before We Begin…

Priors

- Swords and “Achilles heel” of Bayesian inference
- Garthwhite & Dickey: “…expert personal opinion is of great potential value and can be used more efficiently, communicated more accurately, and judged more critically if it is expressed as a probability distribution.”

Elicitation of Priors

Given a family of distributions and some numerical characteristics (mean, variance, higher moments, quantiles, mode, …) specify the prior.

Example:

- Family exponential, $\theta \sim \text{Exp}(\lambda)$ and $\mathrm{E}\theta = 2 \Rightarrow \frac{1}{\lambda} = 2$, that is $\lambda = \frac{1}{2}$.

- Family exponential and median is equal to 4.

$$\theta \sim \text{Exp}(\lambda), \mathrm{F}\left(\xi_{1/2}\right) = \frac{1}{2} \Rightarrow \mathrm{F}(4) = \frac{1}{2}$$

$$\frac{1}{2} = 1 - e^{-4\lambda} \Rightarrow e^{-4\lambda} = \frac{1}{2} \Rightarrow \lambda = \frac{\log 2}{4} = 0.1733.$$

Example: Elicit beta prior on θ if $\mathrm{E}\theta = \frac{1}{2}$ and Var $\theta = \frac{1}{8}$.

$\theta \sim \mathrm{Be}(\alpha, \beta)$, α, β to be specified.

$$\frac{1}{2} = \mathrm{E}\theta = \frac{\alpha}{\alpha+\beta} \Rightarrow \alpha+\beta = 2\alpha \Rightarrow \boxed{\alpha = \beta} \quad (1)$$

$$\frac{1}{8} = \mathrm{Var}\ \theta = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \overset{(1)}{=} \frac{\alpha^2}{4\alpha^2(2\alpha+1)} = \frac{1}{4(2\alpha+1)}$$

$$4(2\alpha+1) = 8 \Rightarrow 8\alpha = 4 \Rightarrow \alpha = \frac{1}{2}\ (=\beta).$$

In general, if $\mathrm{E}\theta = \mu$ and Var $\theta = \sigma^2$

$\alpha = \mu\left(\frac{\mu(1-\mu)}{\sigma^2} - 1\right)$ and $\beta = (1-\mu)\left(\frac{\mu(1-\mu)}{\sigma^2} - 1\right)$.

Non-Informative Priors

- Bayesian methodology was criticized for subjectivity of priors.
- What if the information contained in a prior is incorrect?
- Bayesians answer to this criticism by offering robust, objective, and/or noninformative choices for priors, when the information about the parameter(s) is not strong.

<u>Invariance Principle in Selecting the Prior</u>

Let $X|\theta \sim \mathrm{f}(x-\theta)$; density is a function of $(x-\theta)$; θ is the <u>location</u> parameter.

- Invariant prior with respect to translation $\pi(\theta) = \pi(\theta - \theta_0)$, for any θ_0

  Solution is $\pi(\theta) =$ const.

  Often called <u>flat</u> prior

- If the parameter of interest is <u>scale</u> parameter, $X|\theta \sim \frac{1}{\theta}\mathrm{f}\left(\frac{x}{\theta}\right)$, then the invariance principle suggests $\pi(\theta) \sim \frac{1}{c}\pi\left(\frac{\theta}{c}\right)$

The choice that satisfies scale invariance is $\pi(\theta) = \frac{1}{\theta}, \theta > 0$

- Both priors are improper (that is, not bona-fide densities)

$$\int_{\mathbb{R}} C d\theta = \infty; \quad \int_0^{+\infty} \frac{1}{\theta} d\theta = \infty$$

- The posteriors could be (and most of the time are) proper densities.

Jeffreys' Priors

- Sir Harold Jeffreys (1891-1989)
- Likelihood $\mathrm{f}(x|\theta) \rightarrow$ Fisher Information

$$\mathrm{I}(\theta) = -\mathrm{E}^{x/\theta}\left(\frac{\partial^2 \log \mathrm{f}(x|\theta)}{\partial \theta^2}\right)$$

Jeffreys' suggestion for non-informative prior is

$$\pi(\theta) \propto \det\big(\mathrm{I}(\theta)\big)^{1/2}$$

- Invariance $\phi = \mathrm{h}(\theta), \theta = \mathrm{g}(\phi)$

$$\mathrm{I}^{1/2}(\phi) = \mathrm{I}^{1/2}(\theta) \times \left|\frac{d\theta}{d\phi}\right|.$$

Some important Jeffreys' priors

- $x|\theta \sim \mathrm{N}(\theta, \sigma^2)$, σ^2 known $\qquad \pi(\theta) \propto 1$
- $x|\theta \sim \mathrm{N}(\mu, \theta)$, μ known $\qquad \pi(\theta) = \frac{1}{\theta}$

  θ variance
- $x|\theta \sim \mathrm{Poi}(\theta)$, θ rate $\qquad \pi(\theta) = \frac{1}{\sqrt{\theta}}$
- $x|\theta \sim \mathrm{Bin}(n, \theta)$, θ probability $\qquad \pi(\theta) \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}} \sim \mathrm{Be}\left(\frac{1}{2}, \frac{1}{2}\right)$
- $x|\theta \sim \mathrm{N}(\mu, \theta^2)$, θ standard deviation $\qquad \pi(\theta) = \frac{1}{\theta}$

  $\log\theta \sim$ is uniform on real line

  $\log\theta^2 = 2\log\theta$ also uniform on real line.
- $x|\theta \sim \mathrm{Bin}(n, \theta)$, $\mathrm{logit}(\theta) = \log\frac{\theta}{1-\theta} \sim$ flat prior $\qquad$ Zellner's prior $\propto \theta^{-1}(1-\theta)^{-1}$

Objective Priors

Reference priors (Bernardo, Berger, Pericchi, …)

- Maximizing the divergence (measure of distance) between prior and posterior.
- KL-divergence $\int \pi(\theta|t)\log\frac{\pi(\theta|t)}{\pi(\theta)}d\theta, \quad t = t(x_1, \ldots, x_n)$

sufficient statistic

$$\mathrm{I} = \int \mathrm{m}(t)\left(\int \pi(\theta|t)\log\frac{\pi(\theta|t)}{\pi(\theta)}d\theta\right)dt$$

$$= \int\int \mathrm{h}(t,\theta)\log\frac{\mathrm{h}(t,\theta)}{\mathrm{m}(t)\pi(\theta)}d\theta dt$$

$$\pi^*(\theta) = \underset{\pi(\theta)}{\arg\max}\,\mathrm{I}$$

- For one-dimensional parameters Reference priors and Jeffreys' priors coincide

“Prior Sample Size”

Effective "Sample Size"

- Non-informative prior – a vague attribute.

  For example: p in Bin(n, p)
  - Uniform $\pi(p) = 1(0 \le p \le 1) \equiv \mathrm{Be}(1,1)$
  - Jeffreys $\pi(p) = \mathrm{Be}\left(\frac{1}{2}, \frac{1}{2}\right)$
  - Zellner $\pi(p) \propto \frac{1}{p(1-p)} \sim \text{"}\mathrm{Be}(0,0)\text{"}$

  are all referred as non-informative.

- How to calibrate amount of information carried by the prior?

- Informally

  Information in the prior $\equiv$ information in a sample of size m

m = ESS (Effective Sample Size)

- For $X|\theta \sim \mathrm{Bin}(n, \theta)$ and $\theta \sim \mathrm{Be}(\alpha, \beta)$

$$\frac{\alpha}{\alpha+\beta} \rightarrow \frac{\alpha+x}{\alpha+\beta+n} \Rightarrow \mathrm{ESS} = \alpha+\beta$$

- For $X|\theta \sim \mathrm{Poi}(\theta), \theta \sim \mathrm{Ga}(\alpha, \beta)$

$$\frac{\alpha}{\beta} \rightarrow \frac{\sum X_i + \alpha}{\beta + n} \Rightarrow \mathrm{ESS} = \beta$$

- For $X|\theta \sim \mathrm{N}(\mu, \theta^{-1})$, θ is precision $\equiv \frac{1}{\sigma^2}$

$$\theta \sim \mathrm{Ga}(\alpha, \beta) \Rightarrow \mathrm{ESS} = 2\alpha$$

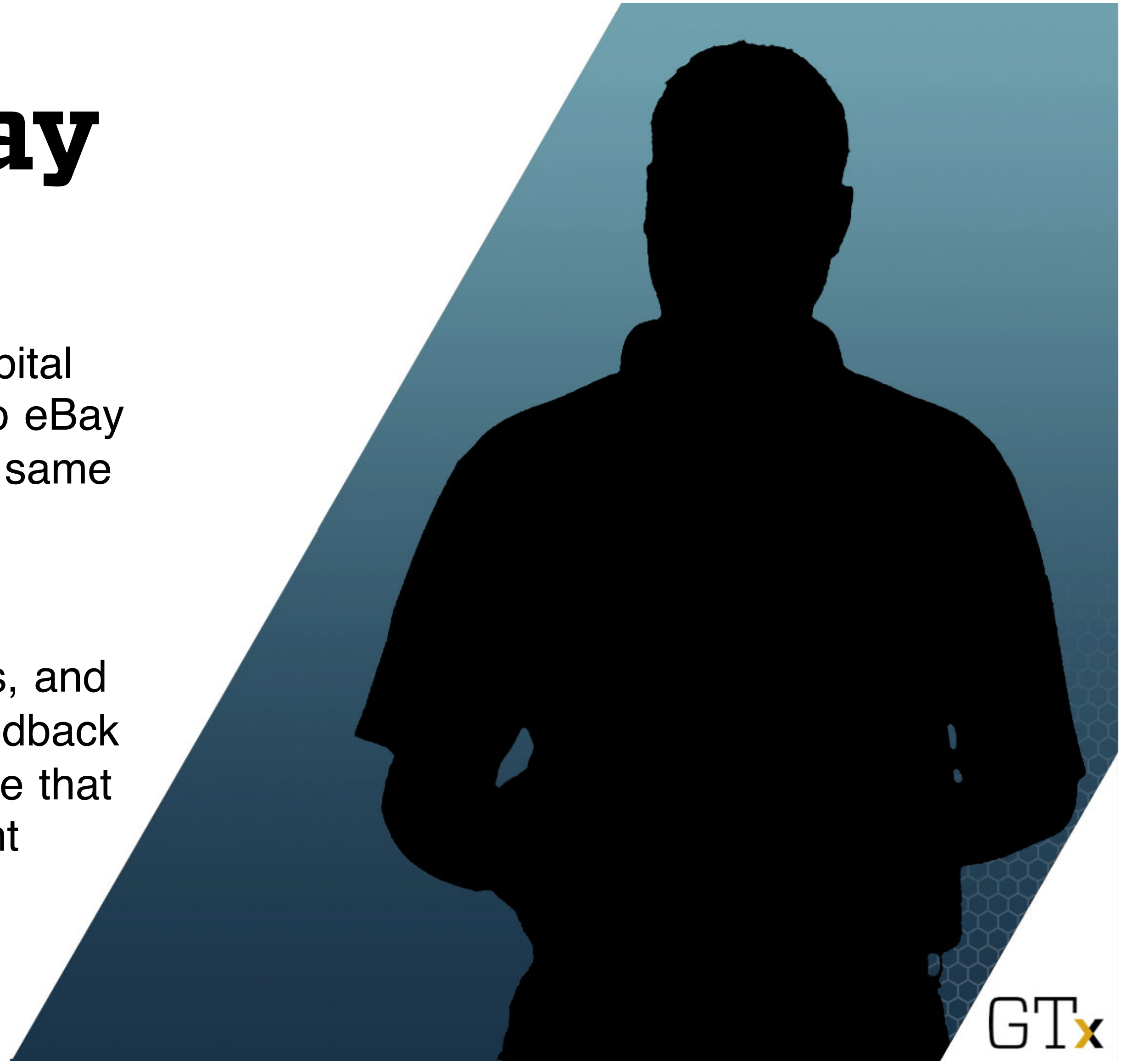
- Spiegelhalter (Community of priors)
  - vague
  - skepticas
  - enthusiastic

Example: eBay Purchase

You decided to purchase a new Orbital Shaking Incubator for your lab. Two eBay sellers are offering this item for the same price, with free shipping.

The seller A has 95% positive feedback from 100 responders, and seller B has 100% positive feedback from 3 responders. We assume that all 103 responders are different unrelated customers.

From which seller to order?



WinBUGS Example

WinBUGS: eBay.odc