STAT 431 — Applied Bayesian Analysis — Course Notes

Jeffreys' Prior

Fall 2022

Objective Priors

Priors that use subjective information are not always desirable:

- subjectivity may draw criticism
- different analysts obtain different results despite using same data and data model
- subjective prior specification may be difficult for high-dimensional or poorly-understood parameters

May prefer an automatic procedure for producing priors that reflect minimal prior information: **objective priors**.

May also call them "noninformative" or "uninformative."

Several schemes for producing objective priors are available — see BSM, Section 2.3.

One of the oldest and most popular:

Jeffreys' prior

Fisher Information

Consider random data $\boldsymbol{Y}$, following a model with continuous scalar parameter θ, and defined by densities

$$f(\boldsymbol{y} \mid \theta)$$

Assume $f(\boldsymbol{y} \mid \theta)$ continuously differentiable in θ for each $\boldsymbol{y}$.

The **(expected) Fisher information** is

$$I(\theta) \;=\; -\,\mathrm{E}\left(\frac{\partial^2}{\partial\theta^2} \ln f(\boldsymbol{Y} \mid \theta) \;\middle|\; \theta\right)$$

(Note: $f(\boldsymbol{Y} \mid \theta)$ is the *random* likelihood function, since $\boldsymbol{Y}$ is random, and the expectation is over $\boldsymbol{Y}$, not θ)

Example: Binomial Proportion

Binomial model ($\theta \in (0,1)$)

$$f(y \mid \theta) \;=\; \binom{n}{y} \theta^y (1-\theta)^{n-y} \qquad y = 0, 1, \ldots n$$

$$\ln f(y \mid \theta) \;=\; \ln \binom{n}{y} \;+\; y \ln \theta \;+\; (n-y) \ln(1-\theta)$$

$$\frac{\partial^2}{\partial \theta^2} \ln f(y \mid \theta) \;=\; -\frac{y}{\theta^2} \;-\; \frac{n-y}{(1-\theta)^2}$$

$$\mathrm{E}\left(\frac{\partial^2}{\partial \theta^2} \ln f(Y \mid \theta) \;\middle|\; \theta \right) \;=\; \mathrm{E}\left(-\frac{Y}{\theta^2} \;-\; \frac{n-Y}{(1-\theta)^2} \;\middle|\; \theta \right)$$

$$
\begin{aligned}
\mathrm{E}\left(\frac{\partial^2}{\partial \theta^2} \ln f(Y \mid \theta) \;\middle|\; \theta \right) &= \mathrm{E}\left(-\frac{Y}{\theta^2} - \frac{n-Y}{(1-\theta)^2} \;\middle|\; \theta \right) \\
&= -\frac{\mathrm{E}(Y \mid \theta)}{\theta^2} - \frac{n - \mathrm{E}(Y \mid \theta)}{(1-\theta)^2}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}\left(\frac{\partial^2}{\partial\theta^2} \ln f(Y \mid \theta) \;\middle|\; \theta \right) &= \mathrm{E}\left(-\frac{Y}{\theta^2} - \frac{n-Y}{(1-\theta)^2} \;\middle|\; \theta \right) \\
&= -\frac{\mathrm{E}(Y \mid \theta)}{\theta^2} - \frac{n - \mathrm{E}(Y \mid \theta)}{(1-\theta)^2} \\
&= -\frac{n\theta}{\theta^2} - \frac{n - n\theta}{(1-\theta)^2}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}\left(\frac{\partial^2}{\partial \theta^2} \ln f(Y \mid \theta) \;\middle|\; \theta \right) &= \mathrm{E}\left(-\frac{Y}{\theta^2} - \frac{n-Y}{(1-\theta)^2} \;\middle|\; \theta \right) \\
&= -\frac{\mathrm{E}(Y \mid \theta)}{\theta^2} - \frac{n - \mathrm{E}(Y \mid \theta)}{(1-\theta)^2} \\
&= -\frac{n\theta}{\theta^2} - \frac{n - n\theta}{(1-\theta)^2} \\
&= -\frac{n}{\theta} - \frac{n}{1-\theta} = -\frac{n}{\theta(1-\theta)}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}\left(\frac{\partial^2}{\partial\theta^2} \ln f(Y \mid \theta) \;\middle|\; \theta \right) &= \mathrm{E}\left(-\frac{Y}{\theta^2} - \frac{n-Y}{(1-\theta)^2} \;\middle|\; \theta \right) \\
&= -\frac{\mathrm{E}(Y \mid \theta)}{\theta^2} - \frac{n - \mathrm{E}(Y \mid \theta)}{(1-\theta)^2} \\
&= -\frac{n\theta}{\theta^2} - \frac{n - n\theta}{(1-\theta)^2} \\
&= -\frac{n}{\theta} - \frac{n}{1-\theta} = -\frac{n}{\theta(1-\theta)}
\end{aligned}
$$

So

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

Jeffreys' Prior

The **Jeffreys' prior (JP)** density for a data model with scalar parameter θ is

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

Note:

- Defined only if Fisher information exists for all θ.
- Not necessarily proper.

Example: Binomial Proportion

$$\pi(\theta) \;\propto\; \sqrt{\frac{n}{\theta(1-\theta)}} \;\propto\; \theta^{-1/2}(1-\theta)^{-1/2}$$

[Draw ...]

Recognize as kernel of a $\mathrm{Beta}(1/2, 1/2)$.

So the JP for the binomial model is $\mathrm{Beta}(1/2, 1/2)$.

Important Property:

A JP is **invariant to reparameterization**:

If $\gamma = g(\theta)$ is a (smooth) **reparameterization**, then the JPs for γ and θ give equivalent posterior distributions (under change of variables).

(See BSM, Section 2.3.1, for more information.)

Eg: A common reparameterization of the binomial model is

$$\gamma \;=\; \mathrm{logit}(\theta) \;=\; \ln\!\left(\frac{\theta}{1-\theta}\right) \qquad (0 < \theta < 1)$$

For multiple parameters, the generalization of JP is

$$\pi(\boldsymbol{\theta}) \quad \propto \quad \sqrt{|\boldsymbol{I}(\boldsymbol{\theta})|}$$

where $\boldsymbol{I}(\boldsymbol{\theta})$ is the Fisher information *matrix*, and $|\boldsymbol{I}(\boldsymbol{\theta})|$ is its *determinant*.

However, the JP is not always considered satisfactory in the multi-parameter case. *Reference priors* (BSM, Section 2.3.2) are often recommended instead.

Example: Mean-Only Normal Sample

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, with

$$Y_1, \ldots, Y_n \mid \mu \quad \sim \quad iid \ \mathrm{Normal}(\mu, \sigma^2)$$

where σ^2 is known.

Recall:

$$f(\boldsymbol{y} \mid \mu) \quad \propto \quad e^{-\frac{1}{2\sigma^2}\sum_i (y_i - \mu)^2}$$

(where the proportionality is in μ).

Then

$$\begin{aligned}\ln f(\boldsymbol{y} \mid \mu) &= -\frac{1}{2\sigma^2}\sum_i (y_i - \mu)^2 + \text{constant (no } \mu) \\ &= \text{a strictly concave quadratic in } \mu\end{aligned}$$

and the coefficient of μ^2 doesn't depend on $\boldsymbol{y}$.

Thus

$$\begin{aligned}\frac{\partial^2}{\partial \mu^2} \ln f(\boldsymbol{y} \mid \mu) &= \text{twice the coefficient of } \mu^2 \\ &= \text{a negative constant (no } \mu \text{ or } \boldsymbol{y})\end{aligned}$$

So the Fisher information is

$$
\begin{aligned}
I(\mu) &= -\mathrm{E}\left(\frac{\partial^2}{\partial \mu^2} \ln f(\boldsymbol{Y} \mid \mu) \;\middle|\; \mu \right) \\
&= \text{some positive constant (no } \mu \text{)}
\end{aligned}
$$

and the JP is

$$
\pi(\mu) \;\propto\; \sqrt{I(\mu)} \;\propto\; 1 \qquad (-\infty < \mu < \infty)
$$

That is, the JP for μ is the flat prior seen previously.

Example: Variance-Only Normal Sample

Again, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$, with

$$Y_1, \ldots, Y_n \mid \sigma^2 \quad \sim \quad iid \ \mathrm{Normal}(\mu, \sigma^2)$$

but now μ is known.

Recall:

$$f(\boldsymbol{y} \mid \sigma^2) \quad \propto \quad \frac{1}{(\sigma^2)^{n/2}} \, e^{-\frac{SSE}{2\sigma^2}}$$

(where the proportionality is in σ^2), with

$$SSE \quad = \quad \sum_i (y_i - \mu)^2$$

To derive the Fisher information ...

$$\ln f(\boldsymbol{y} \mid \sigma^2) \;\; = \;\; -\frac{n}{2} \ln \sigma^2 \; - \; \frac{SSE}{2\sigma^2} \; + \; \text{constant}$$

so

$$\frac{\partial^2}{\partial (\sigma^2)^2} \, \ln f(\boldsymbol{y} \mid \sigma^2) \;\; = \;\; \frac{n}{2(\sigma^2)^2} \; - \; \frac{SSE}{(\sigma^2)^3}$$

and

$$\begin{aligned} I(\sigma^2) \;\; &= \;\; -\,\mathrm{E}\left(\frac{\partial^2}{\partial (\sigma^2)^2} \, \ln f(\boldsymbol{Y} \mid \sigma^2) \;\middle|\; \sigma^2 \right) \\ &= \;\; -\,\frac{n}{2(\sigma^2)^2} \; + \; \frac{\mathrm{E}(SSE \mid \sigma^2)}{(\sigma^2)^3} \end{aligned}$$

Since

$$\begin{aligned}
\mathrm{E}(SSE \mid \sigma^2) &= \mathrm{E}\left(\sum_i (Y_i - \mu)^2 \;\middle|\; \sigma^2\right) \\
&= \sum_i \mathrm{E}\big((Y_i - \mu)^2 \mid \sigma^2\big) \\
&= \sum_i \sigma^2 \;=\; n\sigma^2
\end{aligned}$$

we obtain

$$I(\sigma^2) \;=\; -\frac{n}{2(\sigma^2)^2} + \frac{n\sigma^2}{(\sigma^2)^3} \;=\; \frac{n}{2(\sigma^2)^2}$$

so the JP is

$$\pi(\sigma^2) \;\propto\; \sqrt{I(\sigma^2)} \;\propto\; \frac{1}{\sigma^2} \qquad \sigma^2 > 0$$

This JP is improper:

$$\int_0^\infty \frac{1}{\sigma^2}\, d\sigma^2 \;\;=\;\; \infty$$

[Draw prior curve area ...]

Therefore, one must verify that the posterior will be proper.

Recall that the inverse gamma prior density for σ^2 has kernel

$$\frac{1}{(\sigma^2)^{\alpha+1}}\, e^{-\beta/\sigma^2}$$

We obtain the JP by formally setting "$\alpha = 0$" and "$\beta = 0$."

Thus the inverse gamma prior becomes less informative (more "vague") as we let α and β approach zero.

Since the JP is parameterization-invariant, we can use the JP for σ^2 to derive the JP for $\tau^2 = 1/\sigma^2$.

We use the change-of-variables formula (which still works for improper densities):

$$\frac{d\sigma^2}{d\tau^2} \;=\; \frac{d}{d\tau^2}\left(\frac{1}{\tau^2}\right) \;=\; -\frac{1}{(\tau^2)^2}$$

so

$$\pi(\tau^2) \;=\; \pi(\sigma^2)\left|\frac{d\sigma^2}{d\tau^2}\right| \;=\; \frac{1}{\sigma^2}\cdot\frac{1}{(\tau^2)^2} \;=\; \frac{1}{\tau^2}$$

(Note: This is like a conjugate gamma prior density for τ^2 after formally setting "$\alpha = 0$" and "$\beta = 0$.")

Notation:

$$\sigma^2 \sim \frac{1}{\sigma^2}\, d\sigma^2$$

$$\tau^2 \sim \frac{1}{\tau^2}\, d\tau^2$$

Writing the differential $d\sigma^2$ (or $d\tau^2$) is important!
It emphasizes that the parameter is σ^2 (or τ^2), rather than σ (or τ).