STAT 431 — Applied Bayesian Analysis — Course Notes

Bayes Factors

- Substitute of p-values for Bayesian Stats (posterior prob. of 1-sided hypothesis)
- ① Simple-vs-Simple case

  $$BF = \frac{\text{Posterior Odds}}{\text{Prior Odds}}$$

- Proof
- example
- Interpretation at Scale
- ② Generic Case

  BF = Ratio of marginal prob.

- For Hypothesis Testing
- Proposition

Fall 2022

Recall: Bayesians don't use frequentist p-values.

What can we use instead?

Earlier, we used posterior probabilities of one-sided hypotheses (some of which happened to equal p-values in some noninformative cases).

But we would like a more general approach ...

Simple-vs-Simple Case

Consider data y which may follow one of two different models,

$$\mathcal{M}_0 \quad \text{and} \quad \mathcal{M}_1$$

Assume each of these models fully specifies a distribution for y, and the distributions have densities

$$f(y \mid \mathcal{M}_0) \quad \text{and} \quad f(y \mid \mathcal{M}_1)$$

Then

$$H_0 : \mathcal{M}_0 \text{ is true} \qquad\qquad H_1 : \mathcal{M}_1 \text{ is true}$$

are two **simple** hypotheses.

Let the models have prior probabilities

$$\mathrm{Prob}(\mathcal{M}_0) \;\; (= \mathrm{Prob}(H_0)) \qquad \mathrm{Prob}(\mathcal{M}_1) \;\; (= \mathrm{Prob}(H_1))$$

We assume

$$\mathrm{Prob}(\mathcal{M}_0) \; > \; 0 \qquad \text{and} \qquad \mathrm{Prob}(\mathcal{M}_1) \; > \; 0$$

but it is not necessary that they sum to 1.

The **prior odds in favor of** $\mathcal{M}_1$ are

$$\frac{\mathrm{Prob}(\mathcal{M}_1)}{\mathrm{Prob}(\mathcal{M}_0)}$$

By Bayes' rule,

$$\begin{aligned}\mathrm{Prob}(\mathcal{M}_0 \mid y) &\propto f(y \mid \mathcal{M}_0)\ \mathrm{Prob}(\mathcal{M}_0) \\ \mathrm{Prob}(\mathcal{M}_1 \mid y) &\propto f(y \mid \mathcal{M}_1)\ \mathrm{Prob}(\mathcal{M}_1)\end{aligned}$$

with the same normalizing constant $(1/m(y))$ in both cases.

The **posterior odds in favor of** $\mathcal{M}_1$ are

$$\begin{aligned}\frac{\mathrm{Prob}(\mathcal{M}_1 \mid y)}{\mathrm{Prob}(\mathcal{M}_0 \mid y)} &= \frac{f(y \mid \mathcal{M}_1)\ \mathrm{Prob}(\mathcal{M}_1)}{f(y \mid \mathcal{M}_0)\ \mathrm{Prob}(\mathcal{M}_0)} \\ &= \frac{f(y \mid \mathcal{M}_1)}{f(y \mid \mathcal{M}_0)} \times \text{ prior odds}\end{aligned}$$

The **Bayes factor in favor of $\mathcal{M}_1$ versus $\mathcal{M}_0$** is

$$BF_{1,0} \quad = \quad \frac{\text{posterior odds}}{\text{prior odds}} \quad = \quad \frac{f(y \mid \mathcal{M}_1)}{f(y \mid \mathcal{M}_0)}$$

Interpretation: $BF_{1,0}$ is the factor by which the "odds" of $\mathcal{M}_1$ (relative to $\mathcal{M}_0$) change due to the data.

The **Bayes factor in favor of $\mathcal{M}_1$ versus $\mathcal{M}_0$** is

$$BF_{1,0} \quad = \quad \frac{\text{posterior odds}}{\text{prior odds}} \quad = \quad \frac{f(y \mid \mathcal{M}_1)}{f(y \mid \mathcal{M}_0)}$$

Interpretation: $BF_{1,0}$ is the factor by which the "odds" of $\mathcal{M}_1$ (relative to $\mathcal{M}_0$) change due to the data.

So, for example,

- $BF_{1,0} \approx 1$ means that the data do not distinguish between the models very well
- $BF_{1,0} \gg 1$ means that the data strongly support $\mathcal{M}_1$ over $\mathcal{M}_0$

Notice: In this simple-vs-simple case, the Bayes factor $BF_{1,0}$

- equals the likelihood ratio
- does not depend on the prior — it is the same for any valid values of $\mathrm{Prob}(\mathcal{M}_0)$ and $\mathrm{Prob}(\mathcal{M}_1)$

Example: Do I have COVID?

$$\mathcal{M}_1 \;=\; \text{have COVID} \qquad\qquad \mathcal{M}_0 \;=\; \text{don't have COVID}$$

$$y \;=\; \begin{cases} 1 & \text{if test positive} \\ 0 & \text{if not} \end{cases}$$

Based on published sensitivity and specificity,

$$f(y \mid \mathcal{M}_1) \;=\; \begin{cases} 0.95, & y = 1 \\ 0.05, & y = 0 \end{cases}$$

$$f(y \mid \mathcal{M}_0) \;=\; \begin{cases} 0.002, & y = 1 \\ 0.998, & y = 0 \end{cases}$$

If I test positive ($y = 1$),

$$BF_{1,0} \;=\; \frac{0.95}{0.002} \;=\; 475$$

and, if not ($y = 0$),

$$BF_{1,0} \;=\; \frac{0.05}{0.998} \;\approx\; 0.05$$

So testing positive increases the odds of having COVID, while testing negative decreases them (as expected).

An Interpretation Scale

$BF_{1,0}$	data evidence for $\mathcal{M}_1$ (H_1) vs. $\mathcal{M}_0$ (H_0)
1 to 3.2	Barely worth mentioning
3.2 to 10	Substantial
10 to 100	Strong
> 100	Decisive

More General Case

50% = 20hr/week

Consider modeling data $\boldsymbol{y}$.

Suppose models $\mathcal{M}_0$ and $\mathcal{M}_1$ have (unknown) parameters:

$$\boldsymbol{\theta}_0 \quad \text{for} \quad \mathcal{M}_0 \qquad\qquad \boldsymbol{\theta}_1 \quad \text{for} \quad \mathcal{M}_1$$

We will assume the models are "disjoint": They don't share any distributions for $\boldsymbol{y}$.

Suppose the models have (conditional) priors

$$\pi(\boldsymbol{\theta}_0 \mid \mathcal{M}_0) \qquad\qquad \pi(\boldsymbol{\theta}_1 \mid \mathcal{M}_1)$$

Then

$$\begin{aligned} m(\boldsymbol{y} \mid \mathcal{M}_0) &= \int f(\boldsymbol{y}, \boldsymbol{\theta}_0 \mid \mathcal{M}_0)\, d\boldsymbol{\theta}_0 \\ &= \int \underbrace{f(\boldsymbol{y} \mid \boldsymbol{\theta}_0, \mathcal{M}_0)}_{\mathcal{M}_0 \text{ data model}}\ \underbrace{\pi(\boldsymbol{\theta}_0 \mid \mathcal{M}_0)}_{\text{prior}}\, d\boldsymbol{\theta}_0 \end{aligned}$$

Suppose the models have (conditional) priors

$$\pi(\boldsymbol{\theta}_0 \mid \mathcal{M}_0) \qquad\qquad \pi(\boldsymbol{\theta}_1 \mid \mathcal{M}_1)$$

Then

$$\begin{aligned} m(\boldsymbol{y} \mid \mathcal{M}_0) &= \int f(\boldsymbol{y}, \boldsymbol{\theta}_0 \mid \mathcal{M}_0)\, d\boldsymbol{\theta}_0 \\ &= \int \underbrace{f(\boldsymbol{y} \mid \boldsymbol{\theta}_0, \mathcal{M}_0)}_{\mathcal{M}_0 \text{ data model}} \; \underbrace{\pi(\boldsymbol{\theta}_0 \mid \mathcal{M}_0)}_{\text{prior}}\, d\boldsymbol{\theta}_0 \end{aligned}$$

and similarly

$$m(\boldsymbol{y} \mid \mathcal{M}_1) = \int f(\boldsymbol{y} \mid \boldsymbol{\theta}_1, \mathcal{M}_1)\, \pi(\boldsymbol{\theta}_1 \mid \mathcal{M}_1)\, d\boldsymbol{\theta}_1$$

These are the **marginal likelihoods** of $\mathcal{M}_0$ and $\mathcal{M}_1$ (under their respective priors).

The **Bayes factor in favor of $\mathcal{M}_1$ versus $\mathcal{M}_0$** is

$$BF_{1,0} \;=\; \frac{m(\boldsymbol{y} \mid \mathcal{M}_1)}{m(\boldsymbol{y} \mid \mathcal{M}_0)}$$

Notes:

- Unlike in the simple-vs-simple case, this Bayes factor does depend on the priors — it is not purely a measure of the evidence in the data.
- Both priors must be proper — otherwise, the Bayes factor would depend on an arbitrary scaling.

Unfortunately, Bayes factors are generally difficult to compute, requiring specialized methods.

But they can be easily computed for certain types of hypothesis tests ...

For Hypothesis Testing

Consider a model with data densities $f(\boldsymbol{y} \mid \boldsymbol{\theta})$ and prior $\pi(\boldsymbol{\theta})$.

Consider testing

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \qquad\qquad H_1 : \boldsymbol{\theta} \in \Theta_1$$

where $\Theta_0 \cap \Theta_1 = \varnothing$, and both have positive prior probability.

For Hypothesis Testing

Consider a model with data densities $f(\boldsymbol{y} \mid \boldsymbol{\theta})$ and prior $\pi(\boldsymbol{\theta})$.

Consider testing

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \qquad\qquad H_1 : \boldsymbol{\theta} \in \Theta_1$$

where $\Theta_0 \cap \Theta_1 = \varnothing$, and both have positive prior probability.

Regard this as a test of two data model/prior combinations:

$$\mathcal{M}_0 : \qquad f(\boldsymbol{y} \mid \boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Theta_0$$

with prior $\pi(\boldsymbol{\theta})$ *restricted* to Θ_0

$$\mathcal{M}_1 : \qquad f(\boldsymbol{y} \mid \boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Theta_1$$

with prior $\pi(\boldsymbol{\theta})$ *restricted* to Θ_1

Proposition

In this case, the Bayes factor in favor of $\mathcal{M}_1$ versus $\mathcal{M}_0$ is

$$BF_{1,0} \;=\; \frac{\mathrm{Prob}(H_1 \mid \boldsymbol{y}) \,\big/\, \mathrm{Prob}(H_0 \mid \boldsymbol{y})}{\mathrm{Prob}(H_1) \,\big/\, \mathrm{Prob}(H_0)}$$

$$\left(\;=\; \frac{\textit{posterior odds}}{\textit{prior odds}} \right)$$

We call this the **Bayes factor in favor of H_1 (versus H_0)**.

Proof.

$$m(\boldsymbol{y} \mid \mathcal{M}_1) \;=\; \int_{\Theta_1} \underbrace{f(\boldsymbol{y} \mid \boldsymbol{\theta})}_{\mathcal{M}_1 \text{ on } \Theta_1} \; \underbrace{\frac{\pi(\boldsymbol{\theta})}{\mathrm{Prob}(H_1)}}_{\pi(\boldsymbol{\theta}) \text{ restricted to } \Theta_1} \; d\boldsymbol{\theta}$$

□

Proof.

$$
\begin{aligned}
m(\boldsymbol{y} \mid \mathcal{M}_1) &= \int_{\Theta_1} \underbrace{f(\boldsymbol{y} \mid \boldsymbol{\theta})}_{\mathcal{M}_1 \text{ on } \Theta_1} \underbrace{\frac{\pi(\boldsymbol{\theta})}{\mathrm{Prob}(H_1)}}_{\pi(\boldsymbol{\theta}) \text{ restricted to } \Theta_1} d\boldsymbol{\theta} \\
&= \frac{1}{\mathrm{Prob}(H_1)} \int_{\Theta_1} f(\boldsymbol{y} \mid \boldsymbol{\theta})\, \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}
\end{aligned}
$$

$\square$

Proof.

$$
\begin{aligned}
m(\boldsymbol{y} \mid \mathcal{M}_1) &= \int_{\Theta_1} \underbrace{f(\boldsymbol{y} \mid \boldsymbol{\theta})}_{\mathcal{M}_1 \text{ on } \Theta_1} \underbrace{\frac{\pi(\boldsymbol{\theta})}{\text{Prob}(H_1)}}_{\pi(\boldsymbol{\theta}) \text{ restricted to } \Theta_1} d\boldsymbol{\theta} \\
&= \frac{1}{\text{Prob}(H_1)} \int_{\Theta_1} f(\boldsymbol{y} \mid \boldsymbol{\theta})\, \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta} \\
&= \frac{m(\boldsymbol{y})}{\text{Prob}(H_1)} \int_{\Theta_1} p(\boldsymbol{\theta} \mid \boldsymbol{y})\, d\boldsymbol{\theta}
\end{aligned}
$$

$\square$

Proof.

$$m(\boldsymbol{y} \mid \mathcal{M}_1) = \int_{\Theta_1} \underbrace{f(\boldsymbol{y} \mid \boldsymbol{\theta})}_{\mathcal{M}_1 \text{ on } \Theta_1} \underbrace{\frac{\pi(\boldsymbol{\theta})}{\mathrm{Prob}(H_1)}}_{\pi(\boldsymbol{\theta}) \text{ restricted to } \Theta_1} d\boldsymbol{\theta}$$

$$= \frac{1}{\mathrm{Prob}(H_1)} \int_{\Theta_1} f(\boldsymbol{y} \mid \boldsymbol{\theta})\, \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}$$

$$= \frac{m(\boldsymbol{y})}{\mathrm{Prob}(H_1)} \int_{\Theta_1} p(\boldsymbol{\theta} \mid \boldsymbol{y})\, d\boldsymbol{\theta} = m(\boldsymbol{y}) \frac{\mathrm{Prob}(H_1 \mid \boldsymbol{y})}{\mathrm{Prob}(H_1)}$$

$\square$

Proof.

$$m(\boldsymbol{y} \mid \mathcal{M}_1) \;=\; \int_{\Theta_1} \underbrace{f(\boldsymbol{y} \mid \boldsymbol{\theta})}_{\mathcal{M}_1 \text{ on } \Theta_1} \; \underbrace{\frac{\pi(\boldsymbol{\theta})}{\mathrm{Prob}(H_1)}}_{\pi(\boldsymbol{\theta}) \text{ restricted to } \Theta_1} \; d\boldsymbol{\theta}$$

$$= \frac{1}{\mathrm{Prob}(H_1)} \int_{\Theta_1} f(\boldsymbol{y} \mid \boldsymbol{\theta})\, \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}$$

$$= \frac{m(\boldsymbol{y})}{\mathrm{Prob}(H_1)} \int_{\Theta_1} p(\boldsymbol{\theta} \mid \boldsymbol{y})\, d\boldsymbol{\theta} \;=\; m(\boldsymbol{y}) \frac{\mathrm{Prob}(H_1 \mid \boldsymbol{y})}{\mathrm{Prob}(H_1)}$$

and similarly

$$m(\boldsymbol{y} \mid \mathcal{M}_0) \;=\; m(\boldsymbol{y}) \frac{\mathrm{Prob}(H_0 \mid \boldsymbol{y})}{\mathrm{Prob}(H_0)}$$

so the result follows by taking the ratio. $\square$

Example: Are shark attacks becoming more frequent?

Recall model for yearly number of attacks:

$$Y \mid \lambda \;\sim\; \text{Poisson}(\lambda)$$

$$\ln(\lambda) \;=\; \beta_1 + (\mathsf{year} - \overline{\mathsf{year}})\,\beta_2$$

so we consider

$$H_0 : \beta_2 \leq 0 \quad \textsf{(no)} \qquad\qquad H_1 : \beta_2 > 0 \quad \textsf{(yes)}$$

Recall prior on β_2:

$$\beta_2 \;\sim\; \text{Normal}(0, 100^2)$$

So

$$\text{Prob}(H_0) \;=\; 0.5 \qquad\qquad \text{Prob}(H_1) \;=\; 0.5$$

From JAGS (Example 4.1):

$$\begin{aligned} \text{Prob}(H_1 \mid \boldsymbol{y}) &\approx 0.99999 \\ \text{Prob}(H_0 \mid \boldsymbol{y}) &\approx 1 - 0.99999 \;=\; 0.00001 \end{aligned}$$

So

$$BF_{1,0} \;\approx\; \frac{0.99999/0.00001}{0.5/0.5} \;=\; 99999$$

representing decisive evidence that shark attacks are becoming more frequent.

What about *point null hypotheses*, like

$$H_0 : \beta_2 = 0$$

(perhaps corresponding to a two-sided test)?

Historically, Bayesians generally have considered this to be problematic: If β_2 is a continuous parameter (with a continuous prior), it should have posterior probability 0 of being any specified value.

Also, Bayes factors for point null hypotheses can lead to paradoxes — see discussion in BSM at end of Sec. 5.2.