MT4531/5731: (Advanced) Bayesian Inference

Bayes Factors - Choice of hypothesis

Nicolò Margaritella

School of Mathematics and Statistics, University of St Andrews

–

University of St Andrews

Outline

Outline

Choice of hypothesis

- The underlying principle within classical hypothesis testing is that we wish to reject some hypothesis H_0, in favour of an alternative hypothesis H_1.
- Some Bayesian statisticians also sometimes refer to H_0 as the *null hypothesis* and H_1 as the *alternative hypothesis*, by convention, but within the Bayesian framework the two hypotheses are interchangeable.
- Suppose that we are interested in the parameter $\theta \in \Theta$. Then, our hypotheses are of the form,

$$H_0 : \theta \in \Theta_0; \qquad H_1 : \theta \in \Theta_1,$$

where Θ_0 and Θ_1 are disjoint and exhaustive subsets of the parameter space Θ.

Model choice as choice of hypothesis

- This can translate to a model comparison setting.
- For example, to choose between fitting a constant or a simple linear regression,

$$M_0 : \mathrm{E}(Y_i) = \beta_0; \qquad M_1 : \mathrm{E}(Y_i) = \beta_0 + \beta_1 x_i,$$

  one has to choose between the hypotheses,

$$H_0 : \beta_1 = 0; \qquad H_1 : \beta_1 \neq 0.$$

- More on this topic in the next lecture!

Outline

Prior odds

$$H_0 : \theta \in \Theta_0; \qquad H_1 : \theta \in \Theta_1,$$

- We place a prior on the hypotheses. Let,

$$p_i = \mathbb{P}(H_i) = \mathbb{P}(\theta \in \Theta_i),$$

  for $i = 0, 1$, such that $p_0 + p_1 = 1$.
- The prior odds for the null hypothesis, H_0, against the alternative hypothesis, H_1 are p_0/p_1.
- The prior odds specify prior beliefs on which of the two hypotheses is more likely, before observing any data.
- If $p_0 > p_1$ we would favour the null hypothesis;
- If $p_0 < p_1$, then we would favour the alternative hypothesis;
- If $p_0 \approx p_1$, we regard both hypotheses as roughly equally likely within the prior specification.

Posterior odds

- After we observe data $\boldsymbol{x}$, we calculate the corresponding *posterior odds*, given by,

$$\frac{\mathbb{P}(\theta \in \Theta_0|\boldsymbol{x})}{\mathbb{P}(\theta \in \Theta_1|\boldsymbol{x})} = \frac{f(\boldsymbol{x}|\theta \in \Theta_0)p_0/f(\boldsymbol{x})}{f(\boldsymbol{x}|\theta \in \Theta_1)p_1/f(\boldsymbol{x})} = \frac{p_0 f(\boldsymbol{x}|\theta \in \Theta_0)}{p_1 f(\boldsymbol{x}|\theta \in \Theta_1)}$$

- If the posterior odds are greater than one, we would favour the null hypothesis;
- If the posterior odds are less than one, we favour the alternative hypothesis;
- If they are equal each hypothesis is equally likely *a posteriori*.

The Bayes factor

- Another statistic that is often used is the *Bayes factor*.
- This is simply defined to be the ratio of posterior odds to prior odds.
- The Bayes factor for H_0 against H_1 is denoted by B_{01} and given by,

$$B_{01} = \frac{\mathbb{P}(\theta \in \Theta_0|\boldsymbol{x})/\mathbb{P}(\theta \in \Theta_1|\boldsymbol{x})}{p_0/p_1} = \frac{f(\boldsymbol{x}|\theta \in \Theta_0)}{f(\boldsymbol{x}|\theta \in \Theta_1)}.$$

- Some prefer to report odds ratios or Bayes factors rather than posterior hypothesis/model probabilities. This is because, for model choice, the true model may not be in the set of models under consideration.

Interpreting the Bayes factor (1)

- Kass and Raftery (1995) suggested the following ‘rule of thumb’ for Bayes factors.

Bayes Factor	Interpretation
< 3	No evidence of H_0 over H_1;
> 3	Positive evidence for H_0;
> 20	Strong evidence for H_0;
> 150	Very strong evidence for H_0.

- This general guideline is often used by practicing Bayesians in the interpretation of their results.

Interpreting the Bayes factor (2)

- Under the loss function,

$$L = \begin{cases} 0 & \text{choose } H_0 \text{ when } H_0 \text{ is true} \\ 0 & \text{choose } H_1 \text{ when } H_1 \text{ is true} \\ l_{10} & \text{choose } H_1 \text{ when } H_0 \text{ is true} \\ l_{01} & \text{choose } H_0 \text{ when } H_1 \text{ is true} \end{cases}$$

  the optimal decision is to choose H_1 over H_0 if-f $B_{01} < \frac{l_{01}}{l_{10}} \frac{P(H_1)}{P(H_0)}$. (Bernardo and Smith, 1994, p. 392).

Outline

Simple hypotheses

- Suppose that we wish to test,

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta = \theta_1.$$

- Then, using Bayes Theorem,

$$\mathbb{P}(\theta = \theta_i | \boldsymbol{x}) = \frac{f(\boldsymbol{x}|\theta_i)p_i}{f(\boldsymbol{x})} \propto f(\boldsymbol{x}|\theta_i)p_i, \ \text{for } i = 0, 1.$$

- As $\mathbb{P}(\theta = \theta_0 | \boldsymbol{x}) + \mathbb{P}(\theta = \theta_1 | \boldsymbol{x}) = 1$, $f(\boldsymbol{x}) = \sum_{j=0}^{1} f(\boldsymbol{x}|\theta_j)p_j$.
- Then, the Bayes factor is equal to the likelihood ratio, i.e.,

$$B_{01} = \frac{f(\boldsymbol{x}|\theta_0)}{f(\boldsymbol{x}|\theta_1)}$$

and hence is based solely on the data.

Example (1)

- Assume $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ where, given μ, $X_i \overset{iid}{\sim} N(\mu, 1)$.
- For $n = 10$, we observe data $\boldsymbol{x}$:

$$3.4,\ 2.9,\ 3.0,\ 3.5,\ 3.3,\ 3.7,\ 2.7,\ 3.9,\ 2.7,\ 2.9,$$

  so that $\bar{x} = 3.2$.
- We test the simple hypothesis:

$$H_0 : \mu = 3, \quad \text{vs} \quad H_1 : \mu = 3.5.$$

- What is the corresponding Bayes factor of H_0 against H_1?

Example (2)

- Now,

$$f(\boldsymbol{x}|\mu = 3) = \prod_{i=1}^{10} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x_i - 3)^2}{2}\right) = \frac{\exp(-1)}{(2\pi)^5}.$$

- Similarly, we have that,

$$f(\boldsymbol{x}|\mu = 3.5) = \frac{\exp(-1.25)}{(2\pi)^5}$$

- Thus, we have that,

$$B_{01} = \frac{f(\boldsymbol{x}|\mu = 3)}{f(\boldsymbol{x}|\mu = 3.5)} = 1.28.$$

- There is no evidence (using Kass and Raftery (1995)) or only slight evidence to support model H_0 over H_1.

Outline

The Bayes factor - Composite hypotheses (1)

- For a continuous θ, we wish to test,

$$H_0 : \theta \in \Theta_0 \quad \text{vs} \quad H_1 : \theta \in \Theta_1.$$

- To calculate the prior probability $p_i = p(H_i)$, consider the prior for θ, denoted by $p(\theta)$.
- Then, $p_i = \int_{\theta \in \Theta_i} p(\theta) d\theta$.
- The posterior probabilites for the two hypotheses are,

$$\mathbb{P}(\theta \in \Theta_i | \boldsymbol{x}) = \int_{\theta \in \Theta_i} \pi(\theta | \boldsymbol{x}) d\theta.$$

- Thus, the corresponding Bayes' factor is given by,

$$B_{01} = \frac{\mathbb{P}(\theta \in \Theta_0 | \boldsymbol{x}) p_1}{\mathbb{P}(\theta \in \Theta_1 | \boldsymbol{x}) p_0}$$
$$\left(= \frac{f(\boldsymbol{x} | \theta \in \Theta_0)}{f(\boldsymbol{x} | \theta \in \Theta_1)} \right).$$

The prior's importance for composite hypotheses

- Consider the following derivation.
- Let $p_i(\theta) = p(\theta|H_i)$ denote the prior density restricted to Θ_i, renormalized to give a probability density over Θ_i.
- Then,

$$P(H_i|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|H_i)P(H_i)}{f(\boldsymbol{x})} = \frac{P(H_i)}{f(\boldsymbol{x})}\int_{\Theta_i} f(\boldsymbol{x}|\theta, H_i)p(\theta|H_i)d\theta$$
$$= \frac{P(H_i)}{f(\boldsymbol{x})}\int_{\Theta_i} f(\boldsymbol{x}|\theta, H_i)p_i(\theta)d\theta$$

- The posterior odds are given by,

$$\frac{\mathbb{P}(\theta \in \Theta_0|\boldsymbol{x})}{\mathbb{P}(\theta \in \Theta_1|\boldsymbol{x})} = \frac{p_0 \int_{\theta\in\Theta_0} f(\boldsymbol{x}|\theta)p_0(\theta)d\theta}{p_1 \int_{\theta\in\Theta_1} f(\boldsymbol{x}|\theta)p_1(\theta)d\theta}.$$

- Therefore, the corresponding Bayes' factor is given by,

$$B_{01} = \frac{\int_{\theta\in\Theta_0} f(\boldsymbol{x}|\theta)p_0(\theta)d\theta}{\int_{\theta\in\Theta_1} f(\boldsymbol{x}|\theta)p_1(\theta)d\theta},$$

Important Notes (1)

- So, the Bayes factor also depends on the parameter prior.
- When calculating the Bayes factor, only proper prior distributions should be used for the model parameters, otherwise the Bayes factor becomes arbitrary.

Important Notes (2)

- Remember that,

$$B_{01} = \frac{\int_{\theta \in \Theta_0} f(\boldsymbol{x}|\theta) p_0(\theta) d\theta}{\int_{\theta \in \Theta_1} f(\boldsymbol{x}|\theta) p_1(\theta) d\theta}.$$

- If improper priors were allowed, one could set $p_0(\theta) = c_0 h_0(\theta)$, where the integral of $h_0(\theta)$ diverges (say, $h_0(\theta) = 1$) and c_0 is any arbitrary constant.
- Similarly, set $p_1(\theta) = c_1 h_1(\theta)$, where the integral of $h_1(\theta)$ diverges (say, $h_1(\theta) = 1$) and c_1 is any arbitrary constant.
- By assigning equal prior probabilites for the two hypotheses, one could then set c_0/c_1 to be arbitrarily large or small, and thus control the derived Bayes factor.

Important Notes (3)

- This is quite obvious in the case where $H_0 : \beta = 0$, $H_1 : \beta \neq 0$.
- Then,

$$B_{01} = \frac{f(\mathbf{x}|\beta = 0)}{\int_{\beta \neq 0} f(\mathbf{x}|\beta) p_1(\beta) d\beta},$$

and one could set that, $p_1(\beta) = p(\beta|H_1) \propto c_1$, where c_1 is arbitrarily large or small.

Example (1)

- Revisit the Example in Section 1.2.2, where, given λ, $x_i \sim Exp(\lambda)$, and $1/\lambda$ denotes the average lifetime of a laptop in years.
- Assume, as in Section 1.2.2, that prior beliefs are described by $\lambda \sim \Gamma(0.2, 0.6)$, so that $\mathrm{E}(\lambda) = 1/3 = 0.33$ (the median is 0.03).
- Consider,

$$H_0 : \lambda > 1/4 \quad \text{vs} \quad H_1 : \lambda \leq 1/4.$$

- $p_0 = \int_{1/4}^{\infty} p(\lambda) d\lambda = 0.27$, and $p_1 = \int_0^{1/4} p(\lambda) d\lambda = 0.73$.
- ... using [1-pgamma(0.25,0.2,0.6)] and [pgamma(0.25,0.2,0.6)], respectively.

Example (2)

- 20 laptops are tested, and their average lifetime turns out to be $\bar{x} = 5$.
- As $\lambda|\boldsymbol{x} \sim \Gamma(n + \alpha, n\bar{x} + \beta)$,

$$\lambda|\boldsymbol{x} \sim \Gamma(20 + 0.2, 20 \times 5 + 0.6) = \Gamma(20.2, 100.6),$$

so that $\mathrm{E}(\lambda|\boldsymbol{x}) = 20.2/100.6 = 0.2007$.Then,

- $P(H_0|\boldsymbol{x}) = \int_{1/4}^{\infty} \pi(\lambda|\boldsymbol{x}) d\lambda = 0.14$
- Also, $P(H_1|\boldsymbol{x}) = \int_{0}^{1/4} \pi(\lambda|\boldsymbol{x}) d\lambda = 0.86$

Example (3)

- Then,

$$B_{01} = \frac{P(H_0|\boldsymbol{x})/P(H_1|\boldsymbol{x})}{p_0/p_1} = 0.4, \ (B_{10} = 2.5).$$

- The posterior odds,

$$\frac{P(H_0|\boldsymbol{x})}{P(H_1|\boldsymbol{x})} = 0.15, \qquad \frac{P(H_1|\boldsymbol{x})}{P(H_0|\boldsymbol{x})} = 6.6.$$

- ... updated the prior odds,

$$\frac{p_0}{p_1} = 0.36, \qquad \frac{p_1}{p_0} = 2.7.$$