Lecture 24: Hypothesis Testing and Bayes Factors

MATH 697

Omidali Jazi

November 22, 2018

McGill University

Goals for this Chapter

- Intoduction to Bayesian Hypothesis Testing
- Bayesian Hypothesis Testing
- Bayes Factor
- Bayes Factor for Hypothesis Assessment

Introduction

- Recall that the Neyman-Pearson in the frequentist hypothesis testing yields the (uniformly) most powerful test for the null hypothesis H_0 versus the alternative hypothesis H_a.
- A decision procedure is devised by which, on the basis of a set of collected data, the null hypothesis will either be rejected in favor of H_a, or not rejected.
- In Bayesian hypothesis testing, there can be more than two hypotheses under consideration, and they do not necessarily stand in an asymmetric relationship. Rather, Bayesian hypothesis testing works just like any other type of Bayesian inference.

Introduction

- Suppose now that we want to assess the evidence in the observed data concerning the hypothesis $H_0 : \psi(\theta) = \psi_0$
- It seems clear how we should assess this, namely, compute the posterior probability:

$$\Pi(\psi(\theta) = \psi_0 | s) \tag{1}$$

- If this is small, then conclude that we have evidence against H_0.

Example

- Suppose we want to assess the evidence concerning whether or not $\theta \in A$. If we let $\psi = I_A$, then we are assessing the hypothesis $H_0 : \psi(\theta) = 1$ and

$$\Pi(\psi(\theta) = 1|s) = \Pi(A|s)$$

- So, in this case, we simply compute the posterior probability that $\theta \in A$.

Introduction

- There can be a problem, however, with using this method. When the prior distribution of ψ is absolutely continuous, then $\Pi(\psi(\theta) = \psi_0|s) = 0$ for all data s. Therefore, we would always find evidence against H_0 no matter what is observed.
- In general, if the value ψ_0 is assigned small prior probability, then it can happen that this value also has a small posterior probability no matter what data are observed.
- To avoid this problem, there is an alternative approach to hypothesis assessment that is sometimes used.

Bayesian Hypothesis

- Recall that, if ψ_0 is a surprising value for the posterior distribution of ψ, then this is evidence that H_0 is false.
- The value ψ_0 is surprising whenever it occurs in a region of low probability for the posterior distribution of ψ. A region of low probability will correspond to a region where the posterior density $\omega(\cdot|s)$ is relatively low.
- Hence, one possible method for assessing this is by computing the (Bayesian) p-value.

$$\Pi(\{\theta : \omega(\psi(\theta)|s) \leq \omega(\psi_0|s)\}|s) \tag{2}$$

Remark

- When $\omega(\cdot|s)$ is unimodal, (2) corresponds to computing a tail probability. If the probability (2) is small, then ψ_0 is surprising, at least with respect to our posterior beliefs.
- When we decide to reject H_0 whenever the p-value is less than $1-\gamma$, then this approach is equivalent to computing a γ-HPD region for ψ and rejecting H_0 whenever ψ_0 is not in the region.
- Sometime, there is also some difficulties using (2) (Next Example).

Example

- Suppose that the posterior distribution of θ is $Beta(2, 1)$. That is, $\omega(\theta|s) = 2\theta$ when $0 \leq \theta \leq 1$ and we want to assess $H_0 : \theta = 3/4$. Then (2) is given by $9/16$.
- Suppose we make a 1–1 transformation to $\rho = \theta^2$ so that the hypothesis is now $H_0 : \rho = 9/16$. The posterior distribution of ρ is $Beta(1, 1)$. Since the posterior density of ρ is constant, this implies that the posterior density at every possible value is less than or equal to the posterior density evaluated at 9/16. Therefore, (2) equals 1, and we would never find evidence against H_0 using this parameterization.
- The difficulty in using (2) only occurs with continuous posterior distributions. So, to avoid this problem, it is often recommended that the hypothesis to be tested always be assigned a positive prior probability.

Bayesian Hypothesis

- In problems where it seems natural to use continuous priors, this is accomplished by taking the prior Π to be a mixture of probability distributions, namely, the prior distribution equals

$$\Pi = p\Pi_1 + (1-p)\Pi_2$$

  where $\Pi_1(\psi(\theta) = \psi_0) = 1$ and $\Pi_2(\psi(\theta) = \psi_0) = 0$. That is, Π_1 is degenerate at ψ_0 and Π_2 is continuous at ψ_0. Then,

$$\Pi(\psi(\theta) = \psi_0) = p > 0$$

  is the prior probability that H_0 is true.

- The prior predictive for the data S is then given by

$$m(s) = pm_1(s) + (1-p)m_2(s)$$

  where m_i is the prior predicitive obtained via prior Π_i. Then, it can be shown that

$$\Pi_1(\psi(\theta) = \psi_0 | s) = \frac{pm_1(s)}{pm_1(s) + (1-p)m_2(s)}$$

  and we use this probability to assess H_0.

Example : Normal Model (σ known)

- Suppose that $(x_1, \ldots, x_n)$ is a sample from an $N(\mu, \sigma_0^2)$ distribution, where $\mu \in \mathbb{R}$ is unknown and σ_0^2 is known.

- We want to assess the hypothesis $H_0 : \mu = \mu_0$. We will take the prior for μ to be an $N(\mu_0, \tau_0^2)$ distribution. Given that we are assessing whether or not $\mu = \mu_0$, it seems reasonable to place the mode of the prior at the hypothesized value. The choice of the hyperparameter τ_0^2 then reflects the degree of our prior belief that H_0 is true.

- We let Π_2 denote this prior probability measure, i.e., Π_2 is the $N(\mu_0, \sigma_0^2)$ probability measure. If we use Π_2 as our prior, then the posterior distribution of μ is absolutely continuous.

- This implies that (1) is 0. So, following the preceding discussion, we consider instead the prior $\Pi = p\Pi_1 + (1-p)\Pi_2$ obtained by mixing Π_2 with a probability measure Π_1 degenerate at μ_0.

- Therefore, $\Pi_1(\{\mu_0\}) = 1$ and under Π_2, the posterior distribution of μ is $N(\mu_*, \sigma_*^2)$.

Odds Ratio

Definition 1

In a probability model with sample space S and probability measure P, the odds in favor of event $A \subset S$ is defined to be $P(A)/P(A^c)$, namely, the ratio of the probability of A to the probability of A^c.

Obviously, large values of the odds in favor of A indicate a strong belief that A is true. Odds represent another way of presenting probabilities that are convenient in certain contexts, e.g., horse racing. Bayes factors compare posterior odds with prior odds.

Bayes Factor

Definition 2

The Bayes factor BF_{H_0} in favor of the hypothesis $H_0 : \psi(\theta) = \psi_0$ is defined, whenever the prior probability of H_0 is not 0 or 1, to be the ratio of the posterior odds in favor of H_0 to the prior odds in favor of H_0, or

$$BF_{H_0} = \left\{ \frac{\Pi(\psi(\theta) = \psi_0)|s)}{1 - \Pi(\psi(\theta) = \psi_0)|s)} \right\} / \left\{ \frac{\Pi(\psi(\theta) = \psi_0))}{1 - \Pi(\psi(\theta) = \psi_0))} \right\} \tag{3}$$

Remarks

- The Bayes factor in favor of H_0 is measuring the degree to which the data have changed the odds in favor of the hypothesis. If BF_{H_0} is small, then the data are providing evidence against H_0. When BF_{H_0} is large, the data are providing evidence in favor of H_0 .
- There is a relationship between the posterior probability of H_0 being true and BF_{H_0}. From (3), we obtain

$$\Pi(\psi(\theta) = \psi_0|s) = \frac{rBF_{H_0}}{1 + rBF_{H_0}}$$

  where

$$r = \frac{\Pi(\psi(\theta) = \psi_0)}{1 - \Pi(\psi(\theta) = \psi_0)}$$

  is the prior odds in favor of H_0. So, when BF_{H_0} is small, then $\Pi(\psi(\theta) = \psi_0|s)$ is small and conversely.

Example

- The incidence of a disease in the population is 1%. A medical test for the disease is 90% accurate in the sense that it produces a false reading 10% of the time, both: (a) when the test is applied to a person with the disease; and (b) when the test is applied to a person without the disease.
- A person is randomly selected from population and given the test. The test result is positive (i.e. it indicates that the person has the disease). Calculate the Bayes factor for testing that the person has the disease versus that they do not have the disease.

Example

- $A=$ 'person has disease' and $B=$ 'person tests positive', the relevant probabilities are $P(A) = 0.01$, $P(B|A) = 0.9$ and $P(B|A) = 0.9$ from which can be deduced that $P(A|B) = 1/12$.
- We now wish to test $H_0 : A$ vs $H_a : A^c$. So, we calculate:
  - the prior odds in favour of the null hypothesis: $PRO = \pi_0/\pi_1 = 1/99$
  - the posterior odds in favour of the null hypothesis: $POO = p_0/p_1 = 1/11$
- Hence the required Bayes factor is $BF = POO/PRO = (1/11)/(1/99) = 9$.
- This means the positive test result has multiplied the odds of the person having the disease relative to not having it by a factor of 9 or 900%. Another way to say this is that those odds have increased by 800%.
- We could also work out the Bayes factor here as the ratio of the probability that the person tests positive given they have the disease to the probability that they test positive given they do not have the disease.

Bayes Factors to Assess Hypotheses

One reason for using Bayes factors to assess hypotheses is the following result. This establishes a connection with likelihood ratios.

Theorem 3

If the prior Π *is a mixture* $\Pi = p\Pi_1 + (1-p)\Pi_2$, *where* $\Pi_1(A) = 1$, $\Pi_2(A^c) = 1$, *and we want to assess the hypothesis* $H_0 : \theta \in A$, *then*

$$BF_{H_0} = m_1(s)/m_2(s)$$

where m_i *is the prior predictive of the data under* Π_i.

What We Have Learned

- Intoduction to Bayesian Hypothesis Testing
- Bayesian Hypothesis Testing
- Bayes Factor
- Bayes Factor for Hypothesis Assessment