05 Model comparison and hypothesis testing uding Bayes factors

Shravan Vasishth

September 10, 2021

Introduction

Bayes' rule can be written with reference to a specific statistical model M_1. D refers to the data. θ is the parameter, or vector of parameters.

$$P(\theta \mid D, M_1) = \frac{P(D \mid \theta, M_1)P(\theta \mid M_1)}{P(D \mid M_1)} \tag{1}$$

Introduction

$P(D \mid M_1)$ is the likelihood, and is a single number that tells you the likelihood of the observed data D given the model M_1.

Introduction

Obviously, you would prefer a model that gives a higher likelihood. For example, and speaking informally, if you have data that were generated from a Normal(0,1) distribution, then the likelihood of the data given that $\mu = 0$ will be higher than the likelihood given some other value like $\mu = 10$.

Introduction

The higher likelihood is telling us that the underlying model is more likely to have produced the data. So we would prefer the model with the higher likelihood: we would prefer Normal(0,1) over Normal(10,1) as the presumed distribution that generated the data.

Introduction

Assume for simplicity that $\sigma = 1$.

```
## sample 100 iid data points:
x<-rnorm(100)
## compute log likelihood under mu=0
(loglikmu0<-sum(dnorm(x,mean=0,sd=1,log=TRUE)))
```

```
## [1] -157.0761
```

```
## compute log likelihood under mu=10
(loglikmu10<-sum(dnorm(x,mean=10,sd=1,log=TRUE)))
```

```
## [1] -5153.903
```

```
## the likelihood ratio is a difference of logliks
## on the log scale:
loglikmu0-loglikmu10
```

```
## [1] 4996.827
```

Introduction

One way to compare two models M_1 and M_2 is to use the Bayes factor:

$$BF_{12} = \frac{P(D \mid M_1)}{P(D \mid M_2)} \quad (2)$$

The Bayes factor is similar to the frequentist likelihood ratio test (or ANOVA), with the difference that in the Bayes factor, the likelihood is integrated over the parameter space, not maximized (shown below).

Introduction

How to compute the likelihood? Consider the simple binomial case where we have a subject answer 10 questions, and they get 9 right. That's our data.

Introduction

Discrete example

Assuming a binomial likelihood function, $Binomial(n, \theta)$, the two models we will compare are

- M_1, the parameter has a point value $\theta = 0.5$ with probability 1 (a very sharp prior), and
- M_2, the parameter has a vague prior $\theta \sim Beta(1, 1)$. Recall that this $Beta(1, 1)$ distribution is $Uniform(0, 1)$.

Introduction

Discrete example

The likelihood under M_1 is:

$$\binom{n}{k}\theta^9(1-\theta)^1 = \binom{10}{9}0.5^{10} \tag{3}$$

We already know how to compute this:

```
(probDataM1<-dbinom(9,p=0.5,size=10))
```

```
## [1] 0.009765625
```

Introduction

Discrete example

The marginal likelihood under M_2 involves solving the following integral:

$$P(D \mid M_2) = \int P(D \mid \theta, M_2) P(\theta \mid M_2)\, d\theta \tag{4}$$

The integral is simply integrating out ("summing over'') all possible values of the parameter θ.

Introduction

Discrete example

To see what summing over all possible values means, first consider a discrete version of this:
suppose we say that our θ can take on only these three values: $\theta_1 = 0, \theta_2 = 0.5, \theta_3 = 1$, and each has probability $1/3$. Then, the marginal likelihood of the data given this prior specification of θ would be:

$$
\begin{aligned}
P(D \mid M) =& P(\theta_1)P(D \mid \theta_1) + P(\theta_2)P(D \mid \theta_2) + P(\theta_3)P(D \mid \theta_3) \\
=& \sum P(D \mid \theta_i, M)P(\theta_i \mid M)
\end{aligned}
\tag{5}
$$

Introduction
Discrete example

In our discrete example, this evaluates to:

```
res<-(1/3)* (choose(10,9)* (0)^9 * (1-0)^1) + (1/3)*
  (choose(10,9)* (0.5)^9 * (1-0.5)^1) +
  (1/3)* (choose(10,9)* (1)^9 * (1-1)^1)
res
```

```
## [1] 0.003255208
```

This may be easier to read in mathematical form:

$$\begin{aligned} P(D \mid M) =& P(\theta_1)P(D \mid \theta_1) + P(\theta_2)P(D \mid \theta_2) + P(\theta_3)P(D \mid \theta_3) \\ =& \frac{1}{3}\left(\binom{10}{9} 0^9 (1-0)^1\right) + \frac{1}{3}\left(\binom{10}{9} 0.5^9 (1-0.5)^1\right) \\ +& \frac{1}{3}\left(\binom{10}{9} 1^9 (1-1)^1\right) \\ =& 0.003 \end{aligned} \tag{6}$$

Introduction

Discrete example

Essentially, we are computing the marginal likelihood $P(D \mid M)$ by averaging the likelihood across possible parameter values (here, only three possible values), with the prior probabilities for each parameter value serving as a weight.

Introduction

Discrete example

The Bayes factor for Model 1 vs Model 2 would then be

```
0.0097/0.003
```

```
## [1] 3.233333
```

Model 1, which assumes that θ has a point value 0.5, is approximately three times more likely than the Model 2 with the discrete prior over θ ($\theta_1 = 0, \theta_2 = 0.5, \theta_3 = 1$, each with probability $1/3$).

Introduction

Continuous example

The integral shown above does essentially the calculation we show above, but summing over the entire continuous space that is the range of possible values of θ:

$$P(D \mid M_2) = \int P(D \mid \theta, M_2) P(\theta \mid M_2) \, d\theta \tag{7}$$

Introduction

Continuous example

Let's solve this integral analytically. We need to know only one small detail from integral calculus:

$$\int_a^b x^9 \, dx = \left[\frac{x^{10}}{10}\right]_a^b \tag{8}$$

Similarly:

$$\int_a^b x^{10} \, dx = \left[\frac{x^{11}}{11}\right]_a^b \tag{9}$$

Having reminded ourselves of how to solve this simple integral, we proceed as follows.

Introduction

Continuous example

Our prior for θ is *Beta*$(\alpha = 1, \beta = 1)$:

$$
\begin{aligned}
P(\theta \mid M_2) =& \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}\theta^{\beta-1} \\
=& \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} \theta^{1-1}\theta^{1-1} \\
=& 1
\end{aligned}
\tag{10}
$$

Introduction

Continuous example

So, our integral simplifies to:

$$
\begin{aligned}
P(D \mid M_2) &= \int_0^1 P(D \mid \theta, M_2)\, d\theta \\
&= \int_0^1 \binom{10}{9} \theta^9 (1-\theta)^1\, d\theta \\
&= \int_0^1 \binom{10}{9} (\theta^9 - \theta^{10})\, d\theta \\
&=10 \left[\frac{\theta^{10}}{10} - \frac{\theta^{11}}{11} \right]_0^1 \\
&=10 \times \frac{1}{110} = \frac{1}{11}
\end{aligned}
\tag{11}
$$

Introduction

Continuous example

So, when Model 1 assumes that the θ parameter is 0.5, and Model 2 has a vague prior $Beta(1,1)$ on the θ parameter, our Bayes factor will be:

$$BF_{12} = \frac{P(D \mid M_1)}{P(D \mid M_2)} = \frac{0.00977}{1/11} = 0.107 \qquad (12)$$

Introduction

Continuous example

Thus, the model with the vague prior (M2) is about 9 times more likely than the model with $\theta = 0.5$:

$$\frac{1}{0.10742} = 9.309 \tag{13}$$

Introduction

Continuous example

We could conclude that we have some evidence against the guessing model M1 in this case. Jeffreys (1939/1998) has suggested the following decision criterion using Bayes factors. Here, we are comparing two models, labeled 1 and 2.

- $BF_{12} > 100$: Decisive evidence
- $BF_{12} = 32 - 100$: Very strong
- $BF_{12} = 10 - 32$: Strong
- $BF_{12} = 3 - 10$: Substantial
- $BF_{12} = 2 - 3$: Not worth more than a bare mention

Do not interpret these as absolute divisions.

Introduction

Prior sensitivity

The Bayes factor is sensitive to the choice of prior. It is therefore important to do a sensitivity analysis with different priors.
Read the article Schad et al. (2021).

Introduction

Prior sensitivity

For the model M_2 above, consider the case where we have a prior on θ such that there are 10 possible values for θ, 0.1, 0.2, 0.3,...,1, and the probabilities of each value of θ are $1/10$.

```
theta<-seq(0.1,1,by=0.1)
w<-rep(1/10,10)

prob<-rep(NA,length(w))
for(i in 1:length(theta)){
prob[i]<-(w[i])*choose(10,9)*theta[i]^9*(1-theta[i]^1)
}
## Likelihood for model M2 with
## new prior on theta:
sum(prob)
```

```
## [1] 0.08287079
```

Introduction

Prior sensitivity

Now the Bayes factor for M1 compared to M2 is:

```
0.0097/sum(prob)
```

```
## [1] 0.1170497
```

Now, model M2 is about 8.5 times more likely compared to model M1:

```
1/(0.0097/sum(prob))
```

```
## [1] 8.543381
```

This toy example illustrates the effect of prior specification on the Bayes factor. It is therefore very important to display the Bayes factor under both uninformative and informative priors for the parameter that we are interested in.

One should never use a single 'default' prior or report a single Bayes factor. Example: Nicenboim, Vasishth, and Rösler (2020).

Introduction

The Bayes factor is the ratio of posterior to prior odds

The Bayes factor is really the ratio of posterior odds vs prior odds for any given pair of models:
$BF = \frac{\text{posterior odds}}{\text{prior odds}}$
In the context of our problem:

$$\underbrace{\frac{P(M_1 \mid D)}{P(M_2 \mid D)}}_{\uparrow \atop \textit{posterior odds}} = \underbrace{\frac{P(D \mid M_1)}{P(D \mid M_2)}}_{\uparrow \atop BF_{12}} \underbrace{\frac{P(M_1)}{P(M_2)}}_{\uparrow \atop \textit{prior odds}} \tag{14}$$

Introduction

The Bayes factor is the ratio of posterior to prior odds

So, when the prior odds for M_1 vs M_2 are 1 (i.e., when both models are a priori equi-probable), then we are just interested in computing the posterior odds for the two models.

Bayes factors with brms

`brms` has a function for computing Bayes factors:

- bayes_factor(m0,m1)

Bayes factors with brms

Set up data

```
library(bcogsci)
data("df_gg05_rc")
df_gg05_rc$so<-ifelse(df_gg05_rc$condition=="objgap",1,-1)
```

Bayes factors with brms

Define priors for full model

```
priors <- c(set_prior("normal(6, 0.6)", class = "Intercept"),
            set_prior("normal(0.12, 0.04)", class = "b", coef
            set_prior("normal(0, 0.1)", class = "sd"),
            set_prior("normal(0, 0.5)", class = "sigma"),
            set_prior("lkj(2)", class = "cor"))
```

Bayes factors with brms

```
brm1 <- brm(RT ~ so +
              (1+so|subj) + (1+so|item), df_gg05_rc,
                   family=lognormal(), prior=priors,
                   warmup=2000,
                   iter=10000,
                   cores=4,
                   save_pars = save_pars(all = TRUE),
                   control=list(adapt_delta=0.99, max_treedep
```

Bayes factors with brms

```
priorsNULL <- c(set_prior("normal(6, 0.6)", class = "Intercept
                #set_prior("normal(0, 0.05)", class = "b", coef =
                set_prior("normal(0, 0.1)", class = "sd"),
                set_prior("normal(0, 0.5)", class = "sigma"),
                set_prior("lkj(2)", class = "cor"))
```

Bayes factors with brms

```
brm0 <- brm(RT ~ 1 +
              (1+so|subj) + (1+so|item), df_gg05_rc,
                   family=lognormal(), prior=priorsNULL,
                   warmup=2000,
                   iter=10000,
                   cores=4,
                   save_pars = save_pars(all = TRUE),
                   control=list(adapt_delta=0.99, max_treedep
```

Bayes factors with brms

```
## Iteration: 1
## Iteration: 2
## Iteration: 3
## Iteration: 4
## Iteration: 5
## Iteration: 1
## Iteration: 2
## Iteration: 3
## Iteration: 4
## Iteration: 5
## Iteration: 6

## Estimated Bayes factor in favor of brm0 over brm1: 0.16363
```

Bayes factors with brms

```
## Iteration: 1
## Iteration: 2
## Iteration: 3
## Iteration: 4
## Iteration: 5
## Iteration: 6
## Iteration: 1
## Iteration: 2
## Iteration: 3
## Iteration: 4
## Iteration: 5
## Iteration: 6

## Estimated Bayes factor in favor of brm1 over brm0: 5.63373
```

Bayes factors with brms

Run the command several times to check stability.

Class Exercise 1

Refit the above models with a different prior for σ than the one used. Does the Bayes Factor change when the priors are changed?

Class Exercise 2

In the above example, how does the Bayes factor change when the prior for the slope for *so* is changed to a Normal(0,0.05) to Normal(0,1)?

References

Jeffreys, Harold. 1939/1998. *The Theory of Probability*. Third. Oxford University Press.

Nicenboim, Bruno, Shravan Vasishth, and Frank Rösler. 2020. "Are Words Pre-Activated Probabilistically During Sentence Comprehension? Evidence from New Data and a Bayesian Random-Effects Meta-Analysis Using Publicly Available Data." *Neuropsychologia* 142. https://doi.org/10.1016/j.neuropsychologia.2020.107427.

Schad, Daniel J., Bruno Nicenboim, Paul-Christian Bürkner, Michael J. Betancourt, and Shravan Vasishth. 2021. "Workflow Techniques for the Robust Use of Bayes Factors."