I07 - Posterior model probability

STAT 5870 (Engineering)
Iowa State University

November 22, 2024

One-sided alternative hypotheses

For "one-sided alternative hypotheses" just calculate posterior probabilities.

For example, with hypotheses

$$H_0 : \theta \le \theta_0 \qquad \text{versus} \qquad H_A : \theta > \theta_0$$

Calculate

$$p(H_0|y) = P(\theta \le \theta_0|y)$$

and

$$p(H_A|y) = P(\theta > \theta_0|y).$$

Posterior probabilities

Let $Y \sim Bin(n, \theta)$ with hypotheses

$$H_0 : \theta \le 0.5 \qquad \text{and} \qquad H_A : \theta > 0.5.$$

Assume $\theta \sim Unif(0, 1)$ and obtain the posterior i.e.

$$\theta | y \sim Be(1 + y, 1 + n - y).$$

Then calculate

$$p(H_0|y) = P(\theta \le 0.5|y) = 1 - p(H_A|y).$$

```
n = 10
y = 3
probH0 = pbeta(0.5, 1+y, 1+n-y)
probH0   # p(H_0|y)

[1] 0.8867188

1-probH0 # p(H_A|y)
```

Posterior model probabilities

Calculate the posterior model probabilities over some set of J models i.e,

$$p(M_j|y) = \frac{p(y|M_j)p(M_j)}{p(y)} = \frac{p(y|M_j)p(M_j)}{\sum_{k=1}^{J} p(y|M_k)p(M_k)}.$$

In order to accomplish this, we need to determine

- prior model probabilities:

$$p(M_j) \quad \text{for all} \quad j = 1, \ldots, J$$

  and

- priors over parameters in each model:

$$p(y|M_j) = \int p(y|\theta)p(\theta|M_j)d\theta.$$

Prior predictive distribution

The prior predictive distribution for model M_j is

$$p(y|M_j) = \int p(y|\theta)p(\theta|M_j)d\theta.$$

For example, let

$$y|\mu, M_j \sim N(\mu, 1)$$

and

$$\mu|M_j \sim N(0, C),$$

then

$$y|M_j \sim N(0, 1 + C).$$

Bayes Factor

In the context of a null hypothesis (H_0) and an alternative hypothesis (H_A) we have

$$\begin{aligned} p(H_0|y) &= \frac{p(y|H_0)p(H_0)}{p(y|H_0)p(H_0)+p(y|H_A)p(H_A)} \\ &= \left[1 + \frac{p(y|H_A)}{p(y|H_0)}\frac{p(H_A)}{p(H_0)}\right]^{-1} \\ &= \left[1 + BF(H_A:H_0)\frac{p(H_A)}{p(H_0)}\right]^{-1} \end{aligned}$$

where

$$BF(H_A:H_0) = \frac{p(y|H_A)}{p(y|H_0)}$$

is the Bayes Factor for H_A over H_0.

Normal model

Let $Y \sim N(\mu, 1)$ and $H_0 : \mu = 0$ vs $H_A : \mu \neq 0$.

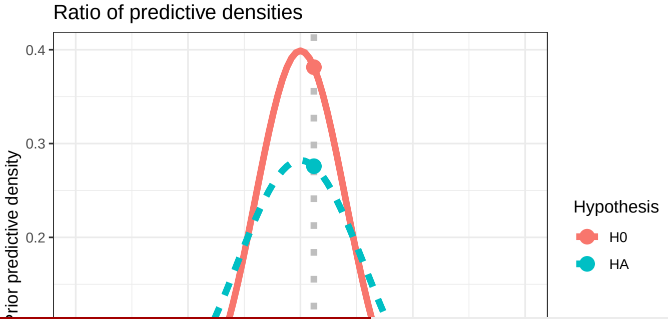
Assume $p(H_0) = p(H_A)$ and $\mu | H_A \sim N(0, 1)$, then

$$\begin{aligned} y|H_0 &\sim N(0,1) \\ y|H_A &\sim N(0,2). \end{aligned}$$

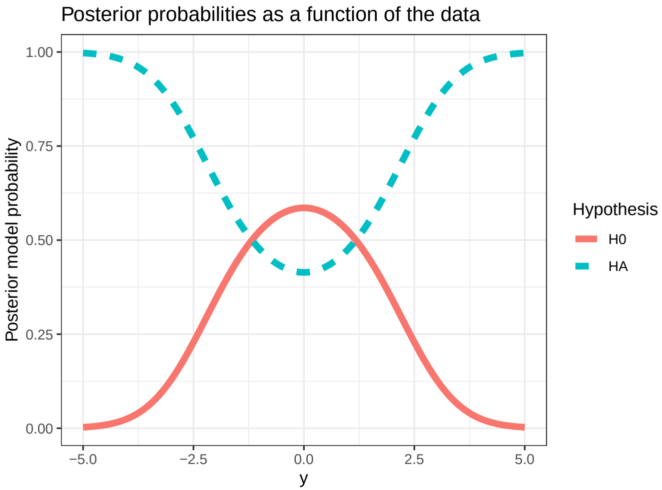
```
y = 0.3
probH0 = 1/(1+dnorm(y, 0, sqrt(2))/dnorm(y, 0, 1))
probH0   # p(H_0|y)

[1] 0.5803167

1-probH0 # p(H_A|y)

[1] 0.4196833
```

Ratio of predictive densities

```
Warning:  Using `size` aesthetic for lines was deprecated in
ggplot2 3.4.0.
i Please use `linewidth` instead.
This warning is displayed once every 8 hours.
Call `lifecycle::last_lifecycle_warnings()` to see where
this warning was generated.
```



Normal model



Prior impact

Let $Y \sim N(\mu, 1)$ and $H_0 : \mu = 0$ vs $H_A : \mu \neq 0$.

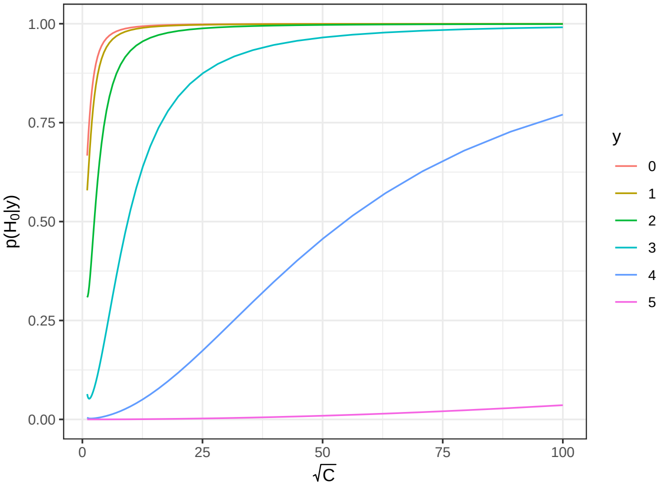
Assume $p(H_0) = p(H_A)$ and $\mu | H_A \sim N(0, C)$, then

$$\begin{aligned} y|H_0 &\sim N(0,1) \\ y|H_A &\sim N(0,1+C) \end{aligned}$$

and

$$p(H_0|y) = \left[1 + \frac{p(y|H_A)}{p(y|H_0)}\right]^{-1}.$$

Prior impact



Interpretation

Since posterior model probabilities depend on the prior predictive distribution

$$p(y|M_j) = \int p(y|\theta)p(\theta|M_j)d\theta$$

posterior model probabilities tell you which model does a better job of prediction and priors, $p(\theta|M_j)$, must be informative.

Do pvalues and posterior probabilities agree?

Suppose $Y \sim Bin(n, \theta)$ and we have the hypotheses $H_0 : \theta = 0.5$ and $H_A : \theta \neq 0.5$ We observe $n = 10,000$ and $y = 4,900$ and find the p-valueis

$$p\text{-value} \approx 2P(Y \leq 4900) = 0.0466$$

so we would reject H_0 at the 0.05 level.

If we assume $p(H_0) = p(H_A) = 0.5$ and $\theta | H_A \sim Unif(0, 1)$, then the posterior probability of H_0, is

$$p(H_0|y) \approx \frac{1}{1 + 1/10.8} = 0.96,$$

so the probability of H_0 being true is 96%.

It appears the posterior probability of H_0 and p-value completely disagree!

Jeffreys-Lindley Paradox

The Jeffreys-Lindley Paradox concerns a situation when comparing two hypotheses H_0 and H_1 given data y and find

- a frequentist test result is significant leading to rejection of H_0, but
- the posterior probability of H_0 is high.

This can happen when

- the effect size is small,
- n is large,
- H_0 is relatively precise,
- H_1 is relative diffuse, and
- the prior model odds is ≈ 1.

No real paradox

p-values:

- a p-value measure how incompatible your data are with the null hypothesis, but
- it says nothing about how incompatible your data are with the alternative hypothesis.

Posterior model probabilities are

- a measure of the (prior) predictive ability of a model relative to the other models, but
- this requires you to have at least two (or more) well-thought out models with informative priors.

Thus, these two statistics provide completely different measures of model adequecy.

Summary

- Use posterior probabilities for one-sided alternative hypotheses.
- Posterior model probabilities evaluate relative predictive ability.