# I07 - Posterior model probability

STAT 401 (Engineering) - Iowa State University

February 13, 2018

# Hypothesis test decisions

A pvalue can loosely be interpreted as "the probability of observing this data if the null hypothesis is true",i.e.

$$p(y|H_0),$$

But what we really want is "the probability the null hypothesis is true, given that we observed this data",i.e.

$$p(H_0|y) = \frac{p(y|H_0)p(H_0)}{p(y)}.$$

If there are only two hypotheses (say $H_0$ and $H_A$), then we have

$$p(H_0|y) = \frac{p(y|H_0)p(H_0)}{p(y|H_0)p(H_0) + p(y|H_A)p(H_A)} = \frac{1}{1 + \frac{p(y|H_A)p(H_A)}{p(y|H_0)p(H_0)}}.$$

# Point null hypotheses

If $H_0 : \theta = \theta_0$ and $H_A : \theta \neq \theta_0$, then

$$\begin{array}{rl} p(y|H_0) & = p(y|\theta_0) \\ p(y|H_A) & = \int p(y, \theta|H_A) d\theta \\ & = \int p(y|\theta, H_A) p(\theta|H_A) d\theta \\ & = \int p(y|\theta) p(\theta|H_A) d\theta \end{array}$$

where $p(\theta|H_A)$ is the distribution of the parameter $\theta$ when the alternative hypothesis is true.

## Example

If $Y_i \stackrel{ind}{\sim} N(\mu, 1)$ and we have the hypotheses $H_0 : \mu = 0$ vs $H_A : \mu \neq 0$ with $\theta|H_A \sim N(0, 1)$, then

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

# Relative frequency interpretation

Suppose you have a model $p(y|\theta)$, hypotheses $H_0 : \theta = \theta_0$ and $H_A : \theta \neq \theta_0$, and you observe a pvalue equal to 0.05. Now you want to understand what that means in terms of whether the null hypothesis is true or not. That is you want

$$p(H_0|pvalue = 0.05) = \left[1 + \frac{p(pvalue = 0.05|H_A)}{p(pvalue = 0.05|H_0)} \frac{p(H_A)}{p(H_0)}\right]^{-1}$$

If we are using a relative frequency interpretation of probability, then the answer depends on

- the relative frequency of the null hypothesis being true $p(H_0) = 1 - p(H_A)$and
- the ratio of the relative frequency of seeing $pvalue = 0.05$ under the null and the alternative which depends on the distribution for $\theta$ under the alternative because

$$p(pvalue = 0.05|H_A) = \int p(pvalue = 0.05|\theta)p(\theta|H_A)d\theta.$$

# Bayesian hypothesis tests

To conduct a Bayesian hypothesis test, you need to specify

- $p(H_j)$ and
- $p(\theta|H_j)$

for every hypothesis $j = 1, \ldots, J$. Then, you can calculate

$$p(H_j|y) = \frac{p(y|H_j)p(H_j)}{\sum_{k=1}^{J} p(y|H_k)p(H_k)} = \left[1 + \sum_{k \neq j} \frac{p(y|H_k)}{p(y|H_j)} \frac{p(H_k)}{p(H_j)}\right]^{-1}$$

where

$$BF(H_k : H_j) = \frac{p(y|H_k)}{p(y|H_j)}$$

are the Bayes factor for hypothesis $H_k$ compared to hypothesis $H_j$ and

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

for all $j$.

# Normal example

Let $Y \sim N(\mu, 1)$ and consider the hypotheses $H_0 : \mu = 0$ and $H_A : \mu \neq 0$ with $\mu|H_A \sim N(0, C)$ and, for simplicity, $p(H_0) = p(H_A) = 0.5$. Then the two hypotheses are really

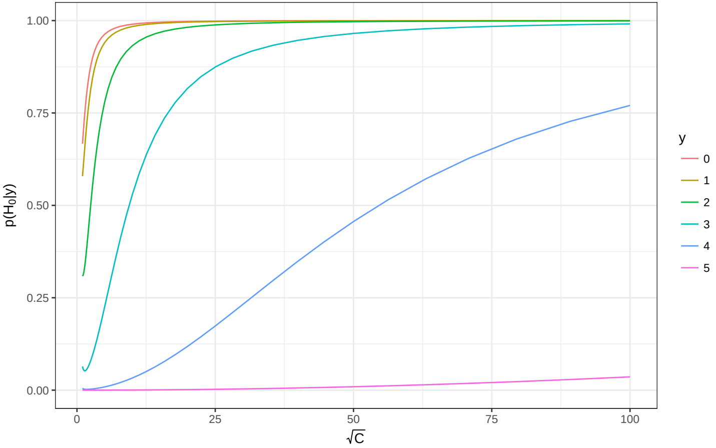
- $Y \sim N(0, 1)$ and
- $Y \sim N(0, 1 + C)$.

Thus

$$p(H_0|y) = \left[1 + \frac{p(y|H_A)}{p(y|H_0)}\right]^{-1} = \left[1 + \frac{N(y; 0, 1 + C)}{N(y; 0, 1)}\right]^{-1}$$

where $N(y; \mu, \sigma^2)$ is evaluating the probability density function for a normal distribution with mean $\mu$ and variance $\sigma^2$ at the value $y$.

# Normal example



# Do pvalues and posterior probabilities agree?

Suppose $\sim Bin(n, \theta)$ 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 pvalue is

$$pvalue \approx 2P(Y \leq 4900) = 0.0466$$

so we would reject $H_0$ at the 0.05 level.

The posterior probability of $H_0$ if we assume $\theta|H_A \sim Unif(0,1)$ and $p(H_0) = p(H_A) = 0.5$ 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 Bayesian and pvalue completely disagree!

# Jeffrey-Lindley Paradox

### Definition

The Jeffrey-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 $\approx 1$.

# No real paradox

Pvalues:

- Pvalues measure how incompatible your data are with the null hypothesis.
- The smaller the pvalue, the more incompatible.
- But they say nothing about how likely the alternative is.

Posterior model probabilities:

- Bayesian posterior probabilities measure how likely the data are under the predictive distribution for each hypothesis.
- The larger the posterior probability, the more predictive that hypothesis was compared to the other hypotheses.
- But this requires you to have at least two well-thought out models, i.e. no vague priors.

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