

Statistical hypotheses

Bayesian and non-Bayesian

STAT 5870 (Engineering)
Iowa State University

August 28, 2024

Statistical hypothesis

A **statistical hypothesis** is a model for data.

For example,

$$Y \sim \text{Ber}(\theta)$$

or

$$Y \sim \text{Bin}(10, 0.25)$$

or

$$Y_i \overset{\text{ind}}{\sim} N(0, \sigma^2)$$

or

$$Y_i \overset{\text{ind}}{\sim} N(\mu, \sigma^2).$$

Translating a scientific hypothesis into a statistical hypothesis

Scientific hypothesis: the coin is fair

Statistical hypothesis:

Let Y be an indicator that the coin is flipped heads.

$$Y \sim Ber(0.5)$$

Scientific hypothesis: the coin is biased, but we don't know the probability

Statistical hypothesis:

$$Y \sim Ber(\theta).$$

Null hypothesis

Wikipedia definition:

*the **null hypothesis**, H_0 , is the [model] that there is no relationship between two measured phenomena or no association among groups*

My definition:

*the **null hypothesis** is the straw man model that nobody believes is true*

For example, the coin is fair

$$H_0 : Y \sim \text{Ber}(0.5).$$

Alternative hypothesis

Wikipedia definition:

*the **alternative hypothesis**, H_A , is [the model] that states something is happening, a new theory is preferred instead of an old one (null hypothesis).*

My definition:

*the **alternative hypothesis** is the model that the researcher believes*

For example, the coin is biased, but we don't know the probability

$$H_A : Y \sim \text{Ber}(\theta)$$

Null vs alternative hypothesis

We typically simplify notation and write null and alternative hypotheses like this:

Model:

$$Y \sim Ber(\theta)$$

Hypotheses:

$$H_0 : \theta = 0.5 \quad \text{versus} \quad H_A : \theta \neq 0.5$$

I prefer

$$H_0 : Y \sim Ber(0.5) \quad \text{versus} \quad H_A : Y \sim Ber(\theta)$$

so that we remind ourselves that these hypotheses are models.

Bayesian hypotheses

Bayesian hypotheses are **full probability models** for the data.

For example,

$$Y \sim Ber(0.5)$$

or

$$Y|\theta \sim Ber(\theta) \quad \text{and} \quad \theta \sim Be(a, b)$$

for known values of a and b .

Prior predictive distribution

The **prior predictive distribution** is the distribution for the data with all the parameters integrated out, i.e.

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

For example, if

$$Y|\theta \sim \text{Ber}(\theta) \quad \text{and} \quad \theta \sim \text{Be}(a, b)$$

then

$$\begin{aligned} p(y) &= \int p(y|\theta)p(\theta)d\theta \\ &= \int_0^1 y^\theta (1-y)^{1-\theta} \frac{1}{\text{Beta}(a,b)} \theta^{a-1} (1-\theta)^{b-1} d\theta \\ &= \frac{1}{\text{Beta}(a,b)} \int_0^1 \theta^{a+y-1} (1-\theta)^{b+n-y-1} d\theta \\ &= \frac{\text{Beta}(a+y, b+n-y)}{\text{Beta}(a,b)} \end{aligned}$$

which is the probability mass function for the **beta-binomial** distribution

Comments

Three points about Bayesian hypotheses:

- Must use **proper** priors.
- No special hypotheses.
- Not restricted to 2 hypotheses.

Summary

- Model:

$$Y \sim \text{Ber}(\theta)$$

- Null hypothesis:

$$H_0 : \theta = 0.5$$

- Alternative hypothesis:

$$H_A : \theta \neq 0.5$$