Introduction to Bayesian linear regression with brms — Part II: Bayesian Inference

Stefano Coretta

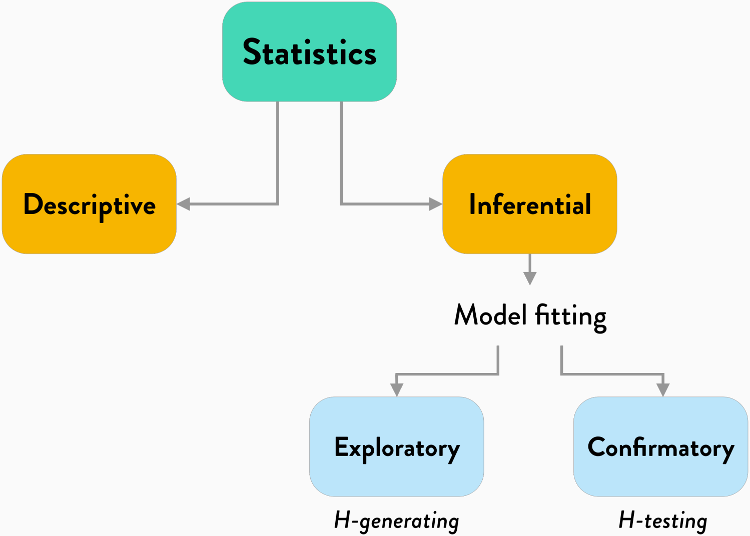
18/01/2020

Road map

1. Statistical inference.
2. Frequentist inference.
3. Bayesian inference:
    - 3.1 Posterior distributions.
    - 3.2 ROPE.
    - 3.3 Bayes factors.

STATISTICAL INFERENCE

Statistics


Statistics
Descriptive
Inferential
Model fitting
Exploratory
Confirmatory
H-generating
H-testing

Hypothesis testing

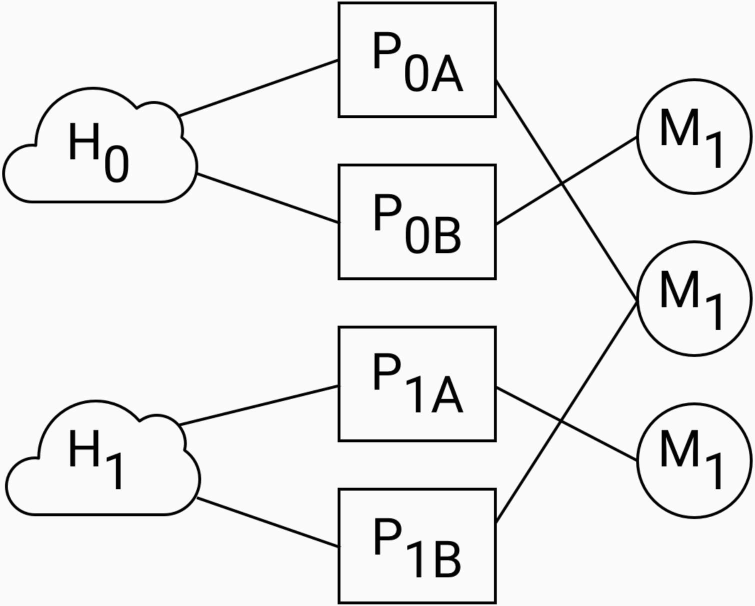
We want to know two (interrelated) things:

- Is there **evidence** for the hypothesis H?
- What is the **strength** of the evidence?

Hypothesis testing

- We test H against empirical data (**hypothesis testing**).
    - It is important to decide in advance the details of the analysis.
    - Even when you think you are not making decisions, the model is.
- Inference is ultimately a **long-term endeavour** (via accumulation of knowledge).

Hypothesis testing



FREQUENTIST INFERENCE

Frequentist statistics

- **Frequentist statistics** is based on the frequentist interpretation of probability.
- **Frequentist probability of an event** is the *relative frequency* of occurrence of an event within an infinite set of imagined repetitions of that event.
- Most of modern science is based on *frequentism*.
    - lme4 package.
    - **Null Hypothesis Significance Testing**.

Null Hypothesis Significance Testing

- NHST is based on falsification.
  - Rejection of the Null Hypothesis (H_0).
  - No direct “support/evidence” for hypotheses.
- *P*-value (between 0 and 1).
  - Probability of obtaining an estimate as extreme or more extreme, *assuming* H_0 *is true*.
  - Should be as low as possible.

Null Hypothesis Significance Testing

- Significance is dichotomous.
  - $p < \alpha$ = “significant”.
  - $p \geq \alpha$ = “non-significant”.
- No degrees of significance.
- “Significance” is a concept that makes sense only within frequentist statistics (NHST).

(Frequentist) confidence intervals

https://rpsychologist.com/d3/ci/

BAYESIAN INFERENCE

Bayesian inference

Bayesian statistics is based on the Bayesian interpretation of the **Bayes theorem**.

$$P(\theta \mid d) = \frac{P(d \mid \theta) \times P(\theta)}{P(d)}$$

Bayesian inference

- Evidence for any H (even the null).
- Strength of evidence.
- Capitalise on previous knowledge.

Bayesian inference

Three ways of doing hypothesis testing with Bayesian statistics:

- Inference from the **posterior**.
- Inference using a **Region Of Practical Equivalence** (ROPE).
- Inference using the **Bayes factor**.

Inference from the posterior

1. Formulate a **hypothesis**:
    - H: Condition B decreases reaction times relative to Condition A.
2. Choose **model specification** (including priors).
3. Collect data.
4. Calculate the **posterior** (fit the model):
    - Condition B 95% CI = [-80, -15] ms.
5. **Inference**:
    - The posterior suggests that Condition B decreases reaction times by 15 to 80 ms at 95% confidence.
    - The posterior is **compatible with H**.

Inference from the posterior

1. Formulate a **hypothesis**:
    - H: Condition B decreases reaction times relative to Condition A *by 100 ms*.
2. Choose **model specification** (including priors).
3. Collect data.
4. Calculate the **posterior** (fit the model):
    - Condition B 95% CI = [-80, -15] ms.
5. **Inference**:
    - The posterior suggests that Condition B decreases reaction times by 15 to 80 ms at 95% confidence.
    - The posterior is **not compatible with H**.

Inference with a ROPE

H0 vs H1

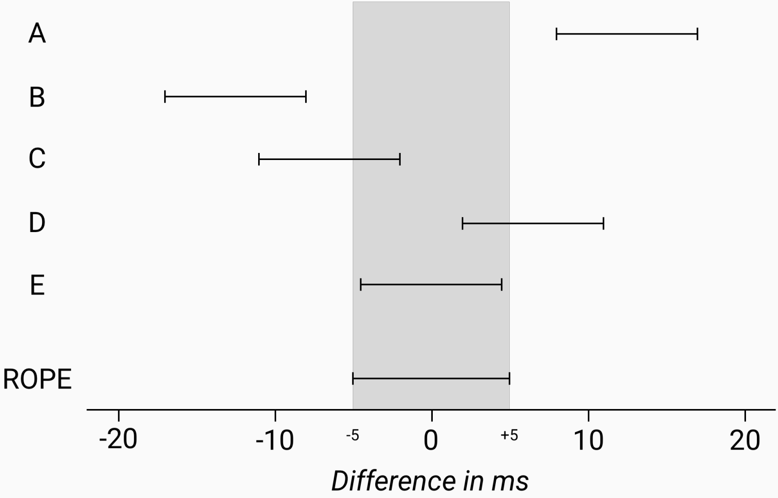
- H1 states that Condition B increases reaction times (alternative hypothesis), while H0 states that Condition B does not increase reaction times (null hypothesis, null effect).
    - $H_1 : \beta > 0$
    - $H_0 : \beta = 0$

Inference with a ROPE

Region Of Practical Equivalence (ROPE):

- Define a region around $\beta = 0$ that practically corresponds to a null effect.
    - For example: [-5, +5] ms ($-5 \geq \beta \leq +5$ = null effect).
    - This ROPE has a width of 10 ms.
- Choose a minimal sample size (ideally based on a prospective power analysis).
- Collect data until the 95% CI of β has a width equal to or smaller than the width of the ROPE.

Inference with a ROPE



Inference with a ROPE

- We focus on the estimate **precision** of β.
- Independent from the value of β.
- Higher precision means greater confidence in the estimated value of β.

Bayes Factor

The Bayes factor is the ratio of the likelihood of H1 to the likelihood of H2.

$$BF_{12} = \mathcal{L}(H_1)/\mathcal{L}(H_2)$$

Bayes Factor

BF	p(M1|D)	evidence
1–3	0.5–0.75	weak
3–20	0.75–0.95	positive
20–150	0.95–0.99	strong
> 150	> 0.99	very strong

Bayes Factor

```r
priors <- c(
  prior(normal(40, 10), class = Intercept),
  prior(cauchy(0, 10), class = sigma),
  prior(normal(10, 10), class = b, coef = "c1_placecoronal"),
  prior(normal(20, 10), class = b, coef = "c1_placevelar"),
  prior(normal(0, 25), class = sd),
  prior(lkj(2), class = cor)
)
```