A very small intro to Bayesian Statistics

Rosana Zenil-Ferguson, Will Freyman, and Jordan Koch

University of Minnesota

Botany 2018

Goals of Bayesian statistics

1. Set up a (probabilistic) model based on hypothesis of interest

Goals of Bayesian statistics

1. Set up a (probabilistic) model based on hypothesis of interest

2. Condition that model on observed data

Goals of Bayesian statistics

1. Set up a (probabilistic) model based on hypothesis of interest
2. Condition that model on observed data
3. Draw inferences, evaluate its fit and implications
   Gelman et al. 2014 Bayesian Data Analysis. Third Edition

In a Bayesian framework

We are always interested in knowing the posterior distribution

$$P(\theta|D) \propto P(D|\theta)P(\theta)$$

What is a prior distribution $P(\theta)$?

In simple terms it is our hypothesis

What is a prior distribution $P(\theta)$?

In simple terms it is our hypothesis

- It is **subjective** because it is an **informed** assumption

What is a prior distribution $P(\theta)$?

In simple terms it is our hypothesis

- It is **subjective** because it is an **informed** assumption
- We need clarify how it is set up (**elicit priors**)

What is a prior distribution $P(\theta)$?

In simple terms it is our hypothesis

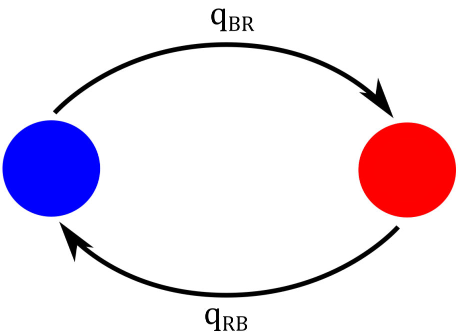
- It is **subjective** because it is an **informed** assumption
- We need clarify how it is set up (**elicit priors**)
- We usually set our hypothesis via **parameters** that are unknown and random

Hypothesis: *Red flowers evolve into purple and viceversa*

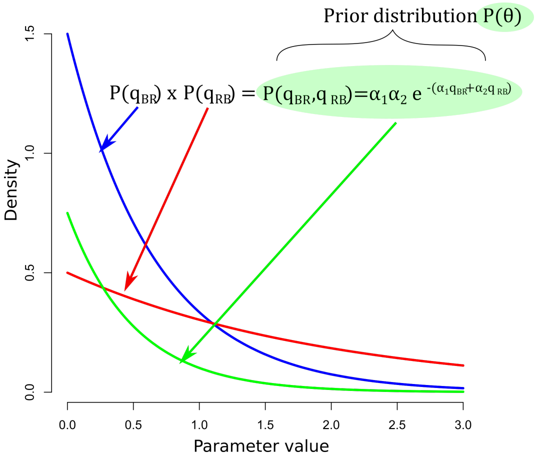
$$\theta = (q_{BR}, q_{RB})$$

Hypothesis: *Red flowers evolve into purple and viceversa*

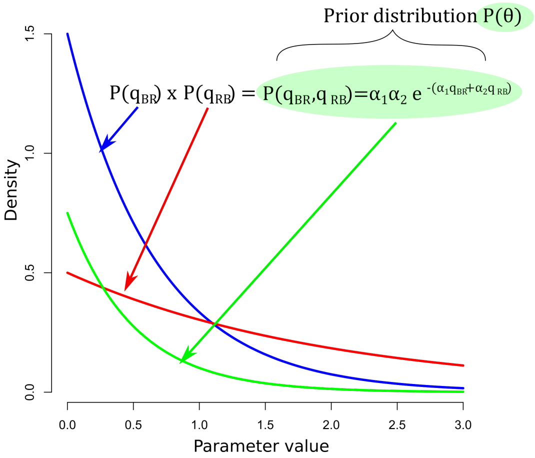
$$\theta = (q_{BR}, q_{RB})$$



The prior distribution: $P(\theta)$



The prior distribution: $P(\theta)$



By selecting a blue exponential faster than the red we are implicitly saying that evolution from blue to red has happened more frequently than red to blue

D is our data

We go into our favorite herbarium, field site, or green house and we collect color of multiple species

How do we integrate our model θ and our data D ?

Calculating the likelihood $P(D|\theta)$

- **Likelihood function**: Probability of the sample given the hypothesis

Calculating the likelihood $P(D|\theta)$

- **Likelihood function**: Probability of the sample given the hypothesis
- Is it a probability?

Calculating the likelihood $P(D|\theta)$

- **Likelihood function**: Probability of the sample given the hypothesis
- Is it a probability?

Calculating the likelihood $P(D|\theta)$

- **Likelihood function**: Probability of the sample given the hypothesis
- Is it a probability? Yes for the sample. **BUT NO! for the parameters**
- In likelihood framework then the parameters are **unknown but fixed**

Calculating the likelihood $P(D|\theta)$

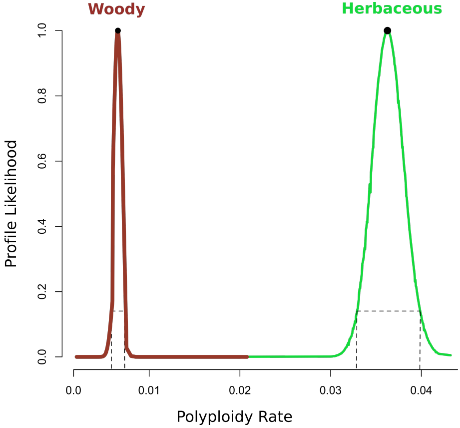
- **Likelihood function**: Probability of the sample given the hypothesis
- Is it a probability? Yes for the sample. **BUT NO! for the parameters**
- In likelihood framework then the parameters are **unknown but fixed**
- **Implications**: parameters do not have a probability distribution, and it is more complicated to assess their uncertainty

Calculating the likelihood is computationally challenging

Thousands of optimizations to find maximum likelihood estimates and confident intervals

Calculating the likelihood is computationally challenging

Thousands of optimizations to find maximum likelihood estimates and confident intervals



Where does the posterior come from?

Bayes theorem (conditional probabilities)

$$P(\theta|D) = \frac{P(\theta, D)}{P(D)} = \frac{P(D|\theta)P(\theta)}{P(D)}$$

Where does the posterior come from?

Bayes theorem (conditional probabilities)

$$P(\theta|D) = \frac{P(\theta, D)}{P(D)} = \frac{P(D|\theta)P(\theta)}{P(D)}$$

Why do we ignore $P(D)$ and put a symbol $\propto$?

$$P(\theta|D) \propto P(D|\theta) \; P(\theta)$$

Where does the posterior come from?

Bayes theorem (conditional probabilities)

$$P(\theta|D) = \frac{P(\theta, D)}{P(D)} = \frac{P(D|\theta)P(\theta)}{P(D)}$$

Why do we ignore $P(D)$ and put a symbol $\propto$?

$$P(\theta|D) \propto P(D|\theta)\, P(\theta)$$

Because $P(D)$ is the probability of the sample and does not contain information about θ

Making inferences with the posterior distribution

- The posterior distribution is **probability.**

Making inferences with the posterior distribution

- The posterior distribution is **probability**.
- .Measures the uncertainty of the hypothesis after being confronted to data (update of my hypothesis)

Making inferences with the posterior distribution

- The posterior distribution is **probability**.
- .Measures the uncertainty of the hypothesis after being confronted to data (update of my hypothesis)
- We need explore it thoroughly (MCMC quality).