Bayesian modeling

Refresher on frequentist modeling

- a series of *iid* (independent and identically distributed) random variables $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$

Refresher on frequentist modeling

- a series of *iid* (independent and identically distributed) random variables $\boldsymbol{Y} = (Y_1, \dots, Y_n)$
- we observe a sample $\boldsymbol{y} = (y_1, \dots, y_n)$

Refresher on frequentist modeling

- a series of *iid* (independent and identically distributed) random variables $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$
- we observe a sample $\boldsymbol{y} = (y_1, \ldots, y_n)$
- model their probability distribution as $f(y|\theta)$, $\theta \in \Theta$

Refresher on frequentist modeling

- a series of *iid* (independent and identically distributed) random variables $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$
- we observe a sample $\boldsymbol{y} = (y_1, \ldots, y_n)$
- model their probability distribution as $f(y|\theta)$, $\theta \in \Theta$

This model assumes there is a "true" distribution of Y characterized by the "true" value of the parameter θ^*

$$\widehat{\theta} \text{ ?}$$

Historical motivating example

Laplace

What is the probability of birth of girls rather than boys ?

⇒ **observations**: births observed in Paris between 1745 and 1770 ($241,945$ girls & $251,527$ boys)

When a child is born, is it equally likely to be a girl or a boy ?

Three building blocks

1. **the question**

2. **the sampling model**

3. **the prior**

Three building blocks

1. **the question**

   The first step in building a model is always to identify the question you want to answer

2. **the sampling model**

3. **the prior**

Three building blocks

1. **the question**

   The first step in building a model is always to identify the question you want to answer

2. **the sampling model**

   Which **observations** are available to inform our response to this ?
   How can they be **described**?

3. **the prior**

Three building blocks

1. **the question**

   The first step in building a model is always to identify the question you want to answer

2. **the sampling model**

   Which **observations** are available to inform our response to this ?
   How can they be **described**?

3. **the prior**

   A probability distribution on the parameters θ of the sampling model

The sampling model

$\boldsymbol{y}$: the observations available

⇒ (parametric) **probabilistic model** underlying their **generation**:

$$Y_i \overset{iid}{\sim} f(y|\theta)$$

The *prior* distribution

In Bayesian modeling, compared to frequentist modeling, we add a **probability distribution** on the **parameters** θ

$$\theta \sim \pi(\theta)$$

$$Y_i|\theta \overset{iid}{\sim} f(y|\theta)$$

θ will thus be treated like a random variable,
but which is never observed !

Back to Laplace's historical example

1. **The question**

2. **Sampling model**

3. ***prior***

Back to Laplace's historical example

1. **The question**

   ...

2. **Sampling model**

   ...

3. ***prior***

   ...