

Modélisation bayésienne

Rappel sur la modélisation fréquentiste

- une suite de variables aléatoires *iid* (indépendantes et identiquement distribuées) $\mathbf{Y} = (Y_1, \dots, Y_n)$
- on observe un échantillon $\mathbf{y} = (y_1, \dots, y_n)$

Rappel sur la modélisation fréquentiste

- une suite de variables aléatoires *iid* (indépendantes et identiquement distribuées) $\mathbf{Y} = (Y_1, \dots, Y_n)$
- on observe un échantillon $\mathbf{y} = (y_1, \dots, y_n)$
- on modélise leur distribution de probabilité comme $f(y|\theta)$, $\theta \in \Theta$

Rappel sur la modélisation fréquentiste

- une suite de variables aléatoires *iid* (indépendantes et identiquement distribuées) $\mathbf{Y} = (Y_1, \dots, Y_n)$
- on observe un échantillon $\mathbf{y} = (y_1, \dots, y_n)$
- on modélise leur distribution de probabilité comme $f(y|\theta)$, $\theta \in \Theta$

Ce modèle suppose qu'il existe une « vraie » distribution de Y caractérisée par la « vraie » valeur du paramètre θ^*

$\hat{\theta}?$

Présentation de l'application historique

Laplace

Quelle est la probabilité de naissance d'une fille (plutôt que d'un garçon).

⇒ **observations** : naissances observées à Paris entre 1745 et 1770
(241,945 filles & 251,527 garçons)

Quand un enfant naît, est-il plus probable que ce soit une fille
plutôt qu'un garçon ?

Trois composants

- 1 La question
- 2 Le modèle d'échantillonnage
- 3 La distribution *a priori*

Trois composants

1 La question

La première étape dans la construction d'un modèle est toujours d'identifier la question à laquelle on souhaite répondre

2 Le modèle d'échantillonnage

3 La distribution *a priori*

Trois composants

1 La question

La première étape dans la construction d'un modèle est toujours d'identifier la question à laquelle on souhaite répondre

2 Le modèle d'échantillonnage

Quelles **observations** sont disponibles pour répondre à cette question ?
Comment peuvent-elles être **décrites** ?

3 La distribution *a priori*

Trois composants

1 La question

La première étape dans la construction d'un modèle est toujours d'identifier la question à laquelle on souhaite répondre

2 Le modèle d'échantillonnage

Quelles **observations** sont disponibles pour répondre à cette question ?
Comment peuvent-elles être **décrites** ?

3 La distribution *a priori*

Une distribution de probabilité sur les paramètres θ du modèle d'échantillonnage

Le modèle d'échantillonnage

\mathbf{y} : les observations disponibles

⇒ **modèle probabiliste** (paramétrique) **génératif** :

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

La distribution *a priori*

Dans la modélisation bayésienne, par rapport à la modélisation fréquentiste, on ajoute une loi de probabilité sur les paramètres θ :

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

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

θ sera ainsi traité comme une variable aléatoire,
mais qui n'est jamais observée !

Retour à l'exemple historique de Laplace

① La question

...

② Modèle d'échantillonnage

...

③ Distribution *a priori*

...