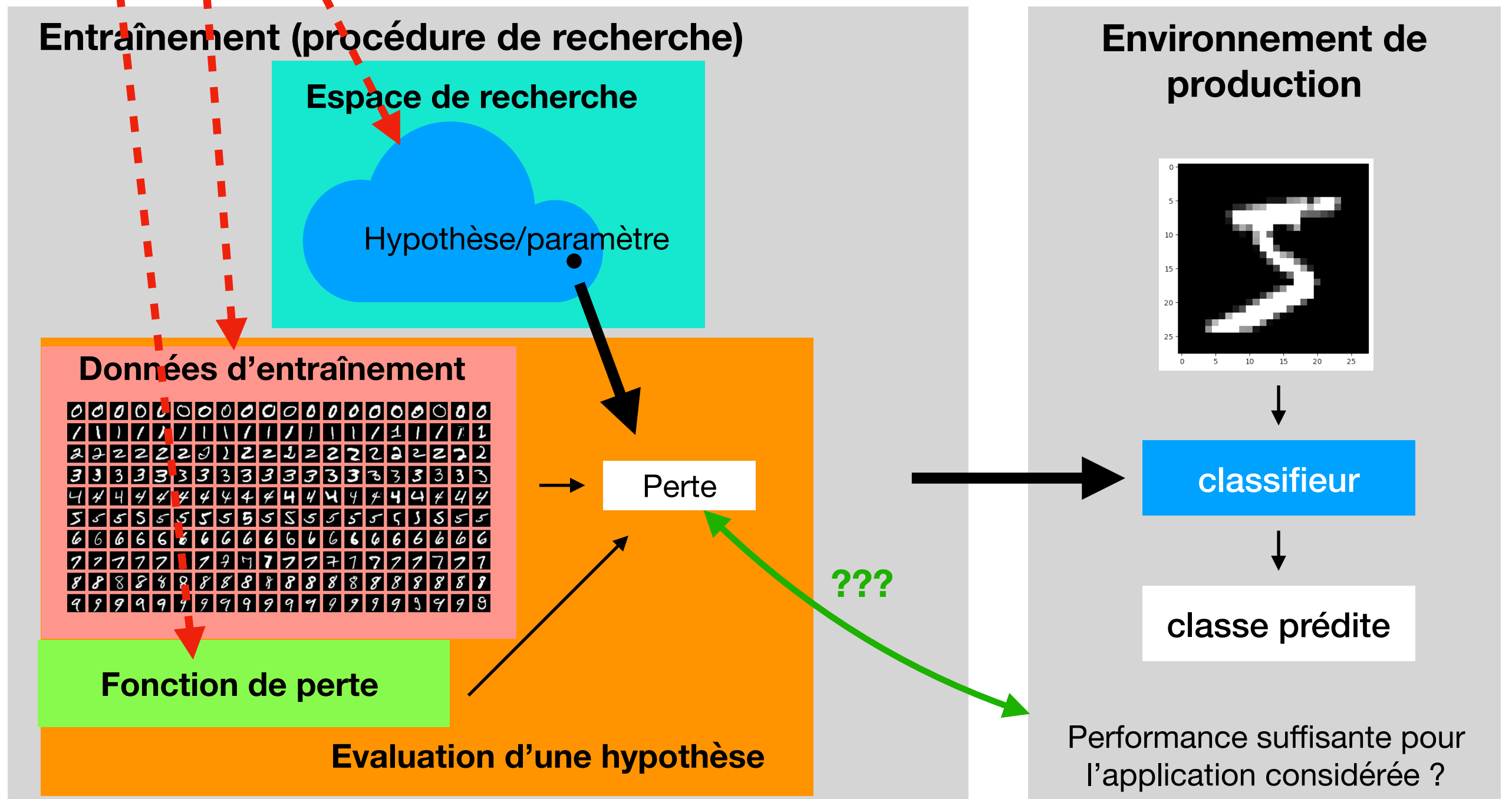


Probabilités, statistique et théorie de l'apprentissage

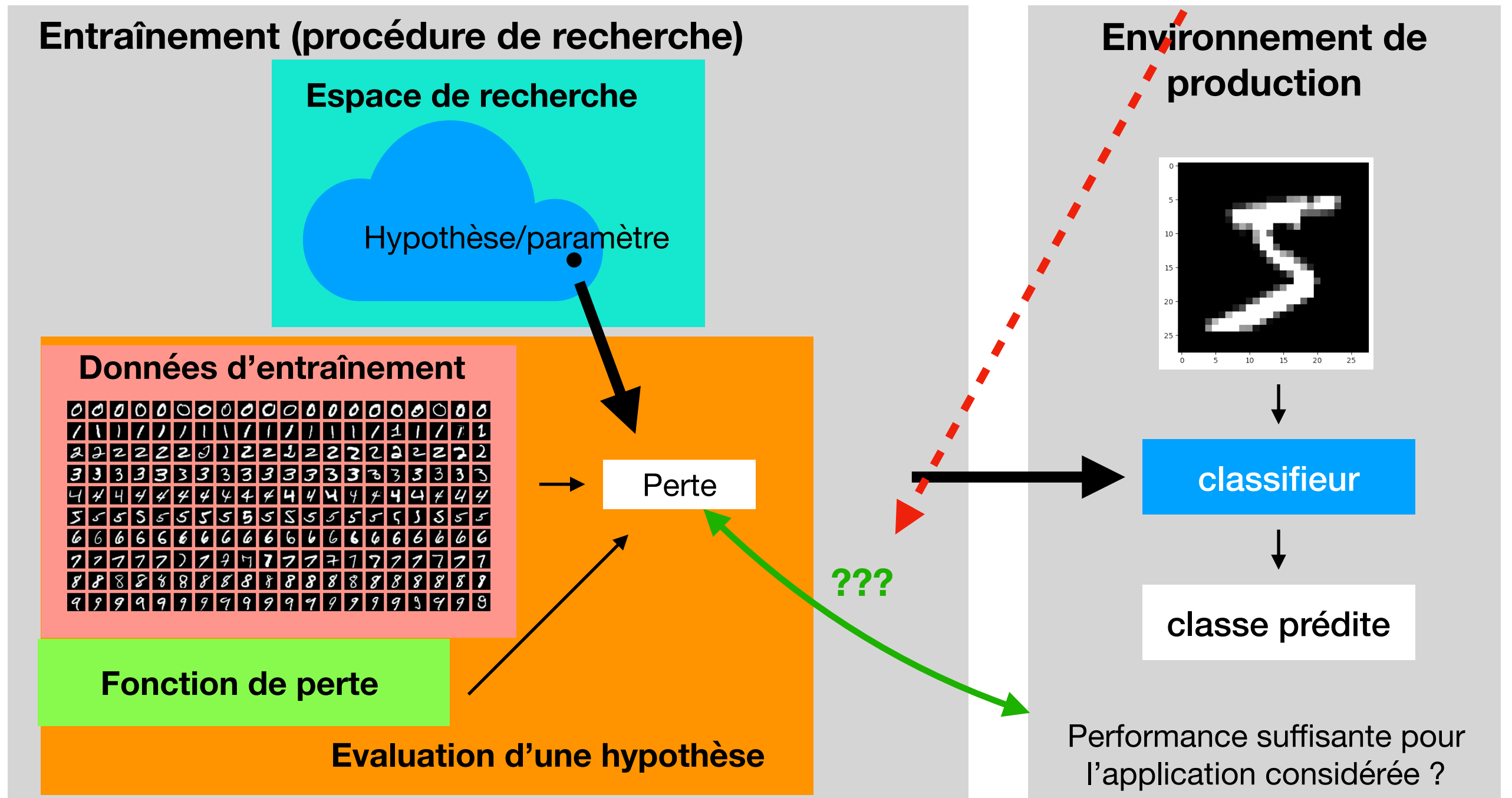
Introduction générale

Structure typique d'un problème d'apprentissage automatique



Introduction générale

Structure typique d'un problème d'apprentissage automatique



Espérance

$$E[X] = \int xp(x)dx \quad \text{Cas continu}$$

$$E[X] = \sum_x xp(x) \quad \text{Cas discret}$$

Espérance $f : \mathbf{R}^n \rightarrow \mathbf{R}$

Cas continu

$$E[f(\mathbf{X})] = \int_{x_1} \int_{x_2} \cdots \int_{x_n} f(x_1, x_2, \dots, x_n) p(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

Cas discret

$$E[f(\mathbf{X})] = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} f(x_1, x_2, \dots, x_n) p(x_1, x_2, \dots, x_n)$$

Moments

Variance $\text{Var}[X] = E[(X - E[X])^2]$

Moment $E[X^k]$

Moment centré $E[(X - E[X])^k]$

Covariance $\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$

Linéarité de l'espérance
$$E \left[\sum_{i=1}^n a_i X_i \right] = \sum_{i=1}^n a_i E(X_i)$$

Bilinéarité de la covariance

$$\text{Cov} \left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j \right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}(X_i, Y_j)$$

Espérance conditionnelle

$$E[X|Y = y] = \int xp(x|y)dx \quad \text{Cas continu}$$

$$E[X|Y = y] = \sum_x xp(x|y) \quad \text{Cas discret}$$

$$p(x|y)p(y) = p(x, y)$$

Moments conditionnels

Covariance conditionnelle

$$\text{Cov}[X, Y \mid Z] = E[(X - E[X \mid Z])(Y - E[Y \mid Z]) \mid Z]$$

Variance conditionnelle

$$\text{Var}[X \mid Y] = \text{Cov}(X, X \mid Y)$$

Loi de l'espérance totale

$$E[X] = E[E[X \mid Y]]$$

Loi de la covariance totale

$$\text{Cov}[X, Y] = \text{Cov}[E[X \mid Z], E[Y \mid Z]] + E[\text{Cov}[X, Y \mid Z]]$$

Analyse statistique

	Modèle statistique	$\mathcal{P}$
Cadre général	Distribution	$P \in \mathcal{P}$
	Observations	$O \sim P$

- Estimation ponctuelle

Fonctionnelle $f : \mathcal{P} \rightarrow E$

Estimateur $\hat{f} : \mathcal{O} \rightarrow E$

But $\hat{f}(O) \approx f(P)$

- Estimation par intervalle, tests d'hypothèses

Analyse statistique

- Problème de l'induction, inséparable de la démarche scientifique dans son ensemble
- Bayésien vs fréquentiste et autres (“best systems” etc.)
- Question d'**interprétation** de la modélisation probabiliste dans un cadre applicatif donné (que représentent les probabilités considérées ?)
 - Une méthode formelle n'est donc pas intrinsèquement “Bayésienne” ou “fréquentiste”, tout dépend de son utilisation
 - Référence sur le sujet : <https://plato.stanford.edu/entries/probability-interpret/>

Qualité statistique d'un estimateur ponctuel

- biais $b_P(\hat{f}) = E_{O \sim P}[\hat{f}(O)] - f(P)$ $\|b_P(\hat{f})\|_2$
- variance $\text{Var}_P(\hat{f}) = \text{Var}_{O \sim P}[\hat{f}(O)]$
- risque $R_P(\hat{f}) = E_{O \sim P}[\ell(f(P), \hat{f}(O))]$

ℓ Fonction de coût

Décomposition biais-variance du risque dans le cas $\ell : x, y \mapsto \|x - y\|_2^2$

$$R_P(\hat{f}) = \text{Var}_P(\hat{f}) + \|b_P(\hat{f})\|_2^2$$