

Repondération d'échantillons par minimisation de la distance de Wasserstein

Julien Reygner



École des Ponts

ParisTech

Collaboration avec A. Touboul (École des Ponts et IRT System'X),
P. Benjamin (Airbus), F. Mangeant (Renault), M. Yagoubi (IRT System'X).

Plan de l'exposé

Introduction

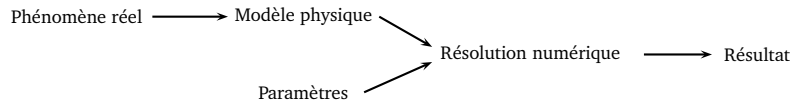
Repondération locale d'échantillons

Synthèse globale sur le graphe

Conclusion

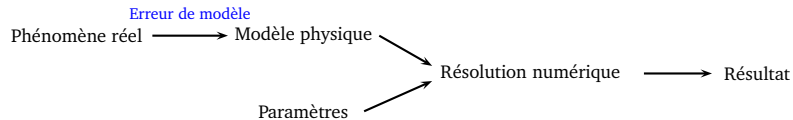
Quantification d'incertitude

Principe de la simulation numérique



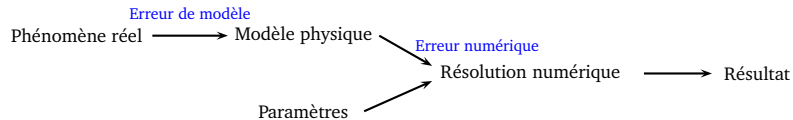
Quantification d'incertitude

Principe de la simulation numérique



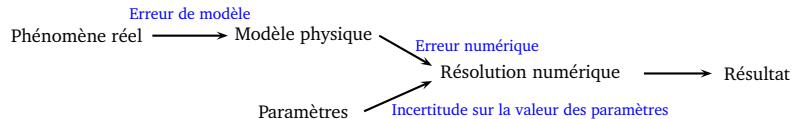
Quantification d'incertitude

Principe de la simulation numérique



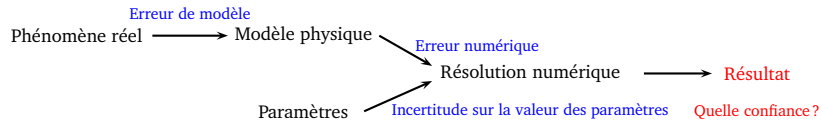
Quantification d'incertitude

Principe de la simulation numérique



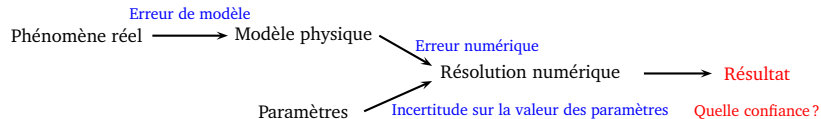
Quantification d'incertitude

Principe de la simulation numérique



Quantification d'incertitude

Principe de la simulation numérique



- Besoin d'outils mathématiques de **quantification d'incertitude**.

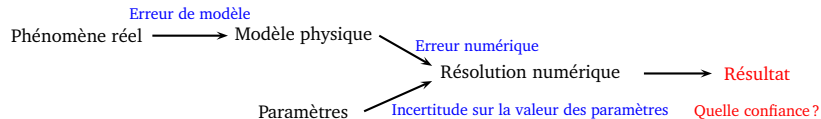
De Rocquigny, Devictor, Tarantola – *Uncertainty in industrial practice: a guide to quantitative uncertainty management* '08

Sullivan – *Introduction to uncertainty quantification* '15

Ghanem, Higdon, Owhadi (eds.) – *Handbook of uncertainty quantification* '17

Quantification d'incertitude

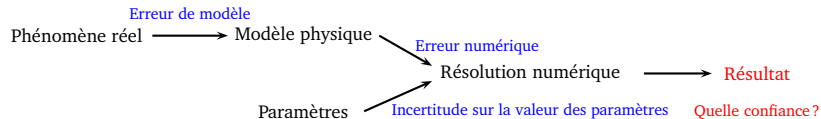
Principe de la simulation numérique



- ▶ Besoin d'outils mathématiques de **quantification d'incertitude**.
 De Rocquigny, Devictor, Tarantola – *Uncertainty in industrial practice: a guide to quantitative uncertainty management* '08
 Sullivan – *Introduction to uncertainty quantification* '15
 Ghanem, Higdon, Owhadi (eds.) – *Handbook of uncertainty quantification* '17
- ▶ Plusieurs problèmes standard : propagation directe, problème inverse, analyse de sensibilité...

Quantification d'incertitude

Principe de la simulation numérique



- ▶ Besoin d'outils mathématiques de **quantification d'incertitude**.
 De Rocquigny, Devictor, Tarantola – *Uncertainty in industrial practice: a guide to quantitative uncertainty management* '08
 Sullivan – *Introduction to uncertainty quantification* '15
 Ghanem, Higdon, Owhadi (eds.) – *Handbook of uncertainty quantification* '17
- ▶ Plusieurs problèmes standard : propagation directe, problème inverse, analyse de sensibilité...
- ▶ Dans cet exposé : méthode numérique pour la **propagation directe** d'incertitudes, dans le contexte de la conception de **systèmes complexes**.

Propagation directe d'incertitude

Modélisation d'un code de calcul :

$$f : \begin{cases} \mathbf{E} \times \mathbf{T} & \rightarrow \mathbf{F} \\ (X, \Theta) & \mapsto Y \end{cases}$$

- ▶ X : **variable d'entrée** (pour l'instant déterministe)
- ▶ Θ : **paramètre aléatoire**
- ▶ Y : **variable de sortie**

Propagation directe d'incertitude

Modélisation d'un code de calcul :

$$f : \begin{cases} \mathbf{E} \times \mathbf{T} & \rightarrow \mathbf{F} \\ (X, \Theta) & \mapsto Y \end{cases}$$

- ▶ X : **variable d'entrée** (pour l'instant déterministe)
- ▶ Θ : **paramètre aléatoire**
- ▶ Y : **variable de sortie**

Cadre permet de représenter différents paradigmes :

- ▶ **incertitude** (statistique ou systémique) sur la valeur de Θ ;
- ▶ **variabilité intrinsèque** à chaque exécution de f (« simulateur stochastique ») ;
- ▶ modèle de **régression** ;
- ▶ ...

Propagation directe d'incertitude

Modélisation d'un code de calcul :

$$f : \begin{cases} \mathbf{E} \times \mathbf{T} & \rightarrow \mathbf{F} \\ (X, \Theta) & \mapsto Y \end{cases}$$

- ▶ X : **variable d'entrée** (pour l'instant déterministe)
- ▶ Θ : **paramètre aléatoire**
- ▶ Y : **variable de sortie**

Cadre permet de représenter différents paradigmes :

- ▶ **incertitude** (statistique ou systémique) sur la valeur de Θ ;
- ▶ **variabilité intrinsèque** à chaque exécution de f (« simulateur stochastique ») ;
- ▶ modèle de **régression** ;
- ▶ ...

Objectif : calcul d'une **quantité d'intérêt statistique** qui dépend de la loi de Y .

Systèmes complexes

- ▶ X est la sortie d'un ou plusieurs **autres codes**.
- ▶ Exemple : conception d'un nouvel avion.

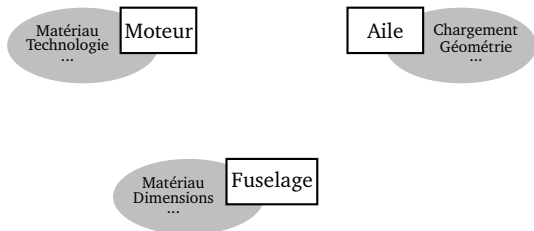
Moteur

Aile

Fuselage

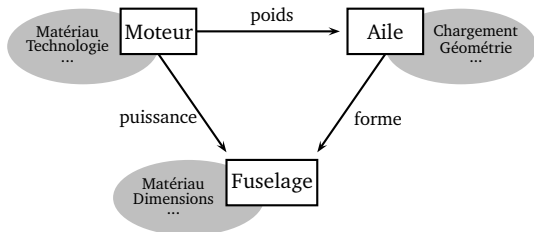
Systèmes complexes

- ▶ X est la sortie d'un ou plusieurs **autres codes**.
- ▶ Exemple : conception d'un nouvel avion.



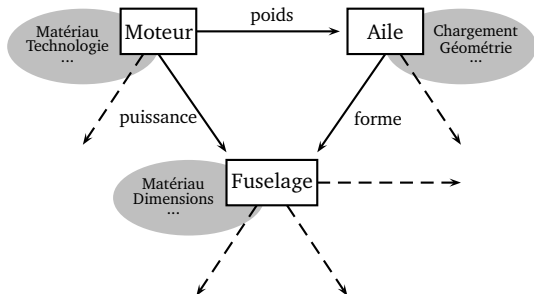
Systèmes complexes

- ▶ X est la sortie d'un ou plusieurs **autres codes**.
- ▶ Exemple : conception d'un nouvel avion.



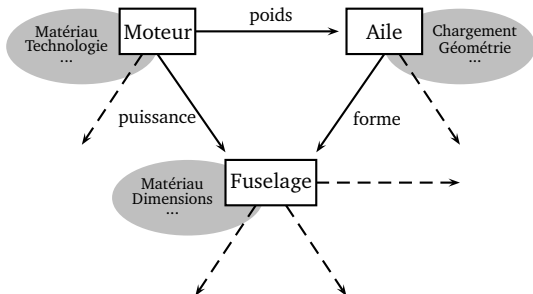
Systèmes complexes

- ▶ X est la sortie d'un ou plusieurs **autres codes**.
- ▶ Exemple : conception d'un nouvel avion.



Systèmes complexes

- ▶ X est la sortie d'un ou plusieurs **autres codes**.
- ▶ Exemple : conception d'un nouvel avion.



Objectif : calcul d'une quantité d'intérêt **globale**.

- ▶ L'évaluation de chaque code de calcul est **coûteuse** ;
- ▶ l'échange d'information entre opérateurs est **difficile**.

Modélisation graphique

- ▶ Le système est représenté par un **graphe orienté** $(\mathcal{V}, \mathcal{E})$.

Modélisation graphique

- ▶ Le système est représenté par un **graphe orienté** $(\mathcal{V}, \mathcal{E})$.
- ▶ Les sommets v et les arêtes (u, v) sont associés à des fonctions

$$\underbrace{f_v : \begin{cases} \mathbf{E}_v \times \mathbf{T}_v & \rightarrow \mathbf{F}_v \\ (\mathbf{X}_v, \Theta_v) & \mapsto \mathbf{Y}_v \end{cases}}_{\text{Code de calcul en } v} \quad \text{et} \quad \underbrace{g_{u,v} : \begin{cases} \mathbf{F}_u & \rightarrow \mathbf{E}_{u,v} \\ \mathbf{Y}_u & \mapsto \mathbf{X}_{u,v} \end{cases}}_{\text{Information transmise de } u \text{ vers } v}$$

$$\text{avec } \mathbf{E}_v = \prod_{(u,v) \in \mathcal{E}} \mathbf{E}_{u,v} \text{ et } \mathbf{X}_v = (\mathbf{X}_{u,v})_{(u,v) \in \mathcal{E}}.$$

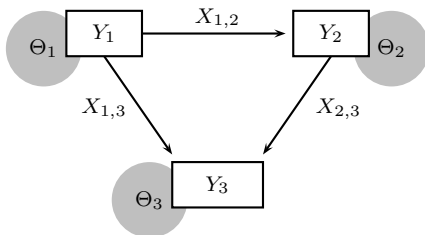
Modélisation graphique

- ▶ Le système est représenté par un **graphe orienté** $(\mathcal{V}, \mathcal{E})$.
- ▶ Les sommets v et les arêtes (u, v) sont associés à des fonctions

$$\underbrace{f_v : \begin{cases} \mathbf{E}_v \times \mathbf{T}_v & \rightarrow \mathbf{F}_v \\ (\mathbf{X}_v, \Theta_v) & \mapsto \mathbf{Y}_v \end{cases}}_{\text{Code de calcul en } v} \quad \text{et} \quad \underbrace{g_{u,v} : \begin{cases} \mathbf{F}_u & \rightarrow \mathbf{E}_{u,v} \\ \mathbf{Y}_u & \mapsto \mathbf{X}_{u,v} \end{cases}}_{\text{Information transmise de } u \text{ vers } v}$$

avec $\mathbf{E}_v = \prod_{(u,v) \in \mathcal{E}} \mathbf{E}_{u,v}$ et $\mathbf{X}_v = (\mathbf{X}_{u,v})_{(u,v) \in \mathcal{E}}$.

- ▶ Les paramètres internes $(\Theta_v)_{v \in \mathcal{V}}$ sont des variables aléatoires **indépendantes**.



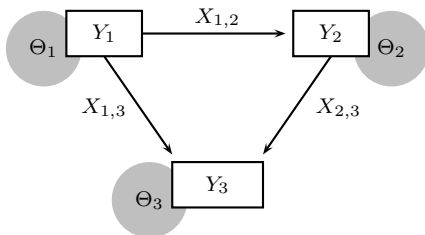
Modélisation graphique

- ▶ Le système est représenté par un **graphe orienté** $(\mathcal{V}, \mathcal{E})$.
- ▶ Les sommets v et les arêtes (u, v) sont associés à des fonctions

$$f_v : \underbrace{\left\{ \begin{array}{l} E_v \times T_v \rightarrow F_v \\ (X_v, \Theta_v) \mapsto Y_v \end{array} \right.}_{\text{Code de calcul en } v} \quad \text{et} \quad g_{u,v} : \underbrace{\left\{ \begin{array}{l} F_u \rightarrow E_{u,v} \\ Y_u \mapsto X_{u,v} \end{array} \right.}_{\text{Information transmise de } u \text{ vers } v}$$

avec $E_v = \prod_{(u,v) \in \mathcal{E}} E_{u,v}$ et $X_v = (X_{u,v})_{(u,v) \in \mathcal{E}}$.

- ▶ Les paramètres internes $(\Theta_v)_{v \in \mathcal{V}}$ sont des variables aléatoires **indépendantes**.



- ▶ **Quantité d'intérêt** : loi jointe de $(Y_v)_{v \in \mathcal{V}}$.

Autonomie disciplinaire

- ▶ Contraintes opérationnelles empêchent communication directe entre sommets.
- ▶ **Métamodélisation** : **Marque-Pucheu, Perrin, Garnier – ESAIM PS '19, Sanson, Le Maitre, Congedo – CMAME '19.**
- ▶ Approche par **décomposition** introduite par **Amaral, Allaire, Willcox – IJNME '14.**

Autonomie disciplinaire

- ▶ Contraintes opérationnelles empêchent communication directe entre sommets.
- ▶ **Métamodélisation** : **Marque-Pucheu, Perrin, Garnier – ESAIM PS '19, Sanson, Le Maitre, Congedo – CMAME '19.**
- ▶ Approche par **décomposition** introduite par **Amaral, Allaire, Willcox – IJNME '14.**

Deux phases :

- ▶ **Offline** : chaque sommet $v \in \mathcal{V}$ prépare un **plan d'expérience** $(X'_{j,v})_{1 \leq j \leq m_v}$ et utilise son code de calcul pour générer un **échantillon** $(X'_{j,v}, Y'_{j,v})_{1 \leq j \leq m_v}$, avec $Y'_{j,v} = f_v(X'_{j,v}, \Theta_{j,v})$.
- ▶ **Online** : les échantillons sont envoyés à un **architecte de simulation** qui doit les combiner pour reconstruire la **vraie loi** de $(Y_v)_{v \in \mathcal{V}}$, sans nouvel appel aux fonctions f_v .

Autonomie disciplinaire

- ▶ Contraintes opérationnelles empêchent communication directe entre sommets.
- ▶ **Métamodélisation** : **Marque-Pucheu, Perrin, Garnier – ESAIM PS '19, Sanson, Le Maitre, Congedo – CMAME '19.**
- ▶ Approche par **décomposition** introduite par **Amaral, Allaire, Willcox – IJNME '14.**

Deux phases :

- ▶ **Offline** : chaque sommet $v \in \mathcal{V}$ prépare un **plan d'expérience** $(X'_{j,v})_{1 \leq j \leq m_v}$ et utilise son code de calcul pour générer un **échantillon** $(X'_{j,v}, Y'_{j,v})_{1 \leq j \leq m_v}$, avec $Y'_{j,v} = f_v(X'_{j,v}, \Theta_{j,v})$.
- ▶ **Online** : les échantillons sont envoyés à un **architecte de simulation** qui doit les combiner pour reconstruire la **vraie loi** de $(Y_v)_{v \in \mathcal{V}}$, sans nouvel appel aux fonctions f_v .

Plan de l'exposé :

- ▶ repondération d'échantillon au niveau d'un sommet ;
- ▶ mise en œuvre globale sur le graphe.

Plan de l'exposé

Introduction

Repondération locale d'échantillons

Synthèse globale sur le graphe

Conclusion

Modélisation du problème

On veut estimer une **quantité d'intérêt**

$$\text{QI} := \mathbb{E}[\phi(Y)], \quad Y = f(X, \Theta), \quad (X, \Theta) \sim \mu_X \otimes \mu_\Theta.$$

On connaît/sait échantillonner μ_Θ , mais pas μ_X .

Modélisation du problème

On veut estimer une **quantité d'intérêt**

$$\text{QI} := \mathbb{E}[\phi(Y)], \quad Y = f(X, \Theta), \quad (X, \Theta) \sim \mu_X \otimes \mu_\Theta.$$

On connaît/sait échantillonner μ_Θ , mais pas μ_X .

On dispose de deux échantillons :

- ▶ échantillon **offline** $(X'_j)_{1 \leq j \leq m}$ **iid** selon une loi $\mu_{X'}$, à partir duquel on a calculé des réalisations $Y'_j = f(X'_j, \Theta_j)$;
- ▶ échantillon **online** $(X_i)_{1 \leq i \leq n}$ **iid** selon la **vraie loi** μ_X .

Modélisation du problème

On veut estimer une **quantité d'intérêt**

$$\text{QI} := \mathbb{E}[\phi(Y)], \quad Y = f(X, \Theta), \quad (X, \Theta) \sim \mu_X \otimes \mu_\Theta.$$

On connaît/sait échantillonner μ_Θ , mais pas μ_X .

On dispose de deux échantillons :

- ▶ échantillon **offline** $(X'_j)_{1 \leq j \leq m}$ **iid** selon une loi $\mu_{X'}$, à partir duquel on a calculé des réalisations $Y'_j = f(X'_j, \Theta_j)$;
- ▶ échantillon **online** $(X_i)_{1 \leq i \leq n}$ **iid** selon la **vraie loi** μ_X .

Difficulté : $\mu_X \neq \mu_{X'}$, « **covariate shift** ».

Quionero-Candela, Sugiyama, Schwaighofer, Lawrence – **Dataset shift in machine learning '09**

Échantillonnage préférentiel et repondération

Un premier calcul :

$$QI = \int \phi(f(x, \theta)) d\mu_{\Theta}(\theta) d\mu_X(x) = \int \phi(f(x, \theta)) d\mu_{\Theta}(\theta) \frac{d\mu_X}{d\mu_{X'}}(x) d\mu_{X'}(x).$$

↪ Estimateur naturel :

$$\widehat{QI}_{m,n}^{\text{IS}} = \frac{1}{m} \sum_{j=1}^m \widehat{w}_{m,n}(X'_j) \phi(Y'_j), \quad \widehat{w}_{m,n}(x) \simeq w(x) := \frac{d\mu_X}{d\mu_{X'}}(x).$$

↪ Comment estimer $w(x)$ à partir des échantillons $(X'_j)_{1 \leq j \leq m}$ et $(X_i)_{1 \leq i \leq n}$?

Échantillonnage préférentiel et repondération

Un premier calcul :

$$QI = \int \phi(f(x, \theta)) d\mu_{\Theta}(\theta) d\mu_X(x) = \int \phi(f(x, \theta)) d\mu_{\Theta}(\theta) \frac{d\mu_X}{d\mu_{X'}}(x) d\mu_{X'}(x).$$

↪ Estimateur naturel :

$$\widehat{QI}_{m,n}^{IS} = \frac{1}{m} \sum_{j=1}^m \widehat{w}_{m,n}(X'_j) \phi(Y'_j), \quad \widehat{w}_{m,n}(x) \simeq w(x) := \frac{d\mu_X}{d\mu_{X'}}(x).$$

↪ Comment estimer $w(x)$ à partir des échantillons $(X'_j)_{1 \leq j \leq m}$ et $(X_i)_{1 \leq i \leq n}$?

- ▶ Estimateur à noyau proposé par **Amaral, Allaire, Willcox – IJNME '14**.

Échantillonnage préférentiel et repondération

Un premier calcul :

$$\text{QI} = \int \phi(f(x, \theta)) d\mu_{\Theta}(\theta) d\mu_X(x) = \int \phi(f(x, \theta)) d\mu_{\Theta}(\theta) \frac{d\mu_X}{d\mu_{X'}}(x) d\mu_{X'}(x).$$

↪ Estimateur naturel :

$$\widehat{\text{QI}}_{m,n}^{\text{IS}} = \frac{1}{m} \sum_{j=1}^m \widehat{w}_{m,n}(X'_j) \phi(Y'_j), \quad \widehat{w}_{m,n}(x) \simeq w(x) := \frac{d\mu_X}{d\mu_{X'}}(x).$$

↪ Comment estimer $w(x)$ à partir des échantillons $(X'_j)_{1 \leq j \leq m}$ et $(X_i)_{1 \leq i \leq n}$?

- ▶ Estimateur à noyau proposé par **Amaral, Allaire, Willcox – IJNME '14**.
- ▶ Toute une littérature : **Sugiyama, Suzuki, Kanamori – Density ratio estimation '12**.

Principe général :

$$\widehat{\text{QI}}_{m,n} = \sum_{j=1}^m w_j \phi(Y'_j),$$

avec des poids $(w_j)_{1 \leq j \leq m}$ choisis de sorte que

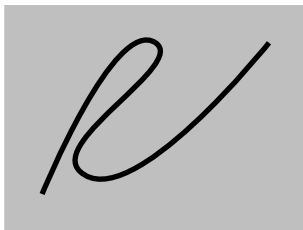
$$\sum_{j=1}^m w_j \delta_{X'_j} \simeq \mu_X \simeq \frac{1}{n} \sum_{i=1}^n \delta_{X_i}.$$

- ↪ **Critères** : distance L^2 entre densités, divergence de Kullback–Leibler, ...
- ↪ Souvent, $\mu_X \ll \mu_{X'}$ **nécessaire** pour analyse théorique.

Problème de singularité

Dans notre contexte :

- ▶ La loi de l'échantillon $\mu_{X'}$ est **choisie par l'utilisateur** en phase offline.
- ▶ Des **bornes** (support, moments) sur μ_X peuvent être connues à l'avance, mais l'absolue continuité est une hypothèse **trop forte**.



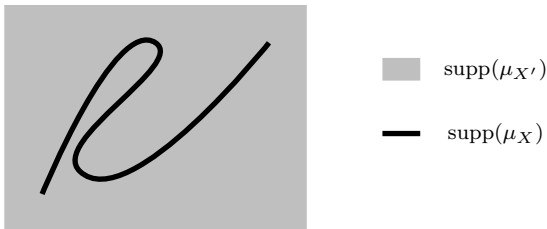
■ $\text{supp}(\mu_{X'})$

— $\text{supp}(\mu_X)$

Problème de singularité

Dans notre contexte :

- ▶ La loi de l'échantillon $\mu_{X'}$ est **choisie par l'utilisateur** en phase offline.
- ▶ Des **bornes** (support, moments) sur μ_X peuvent être connues à l'avance, mais l'absolue continuité est une hypothèse **trop forte**.



Critère indépendant de l'absolue continuité : **distance de Wasserstein**

$$\text{(pour } E = \mathbb{R}^d \text{)} \quad W_q(\mu, \mu') := \inf_{X \sim \mu, X' \sim \mu'} \mathbb{E} [|X - X'|^q]^{1/q}, \quad q \geq 1.$$

Minimisation de la distance de Wasserstein

Problème d'optimisation :

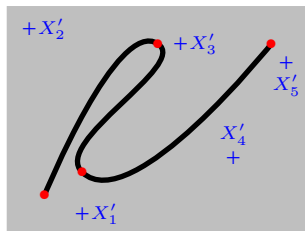
$$\min_{\mathbf{w}=(w_j)_{1 \leq j \leq m}} W_q \left(\underbrace{\sum_{j=1}^m w_j \delta_{X'_j}}_{\hat{\mu}_m^{\mathbf{w}}}, \underbrace{\frac{1}{n} \sum_{i=1}^n \delta_{X_i}}_{\hat{\mu}_n} \right), \quad \sum_{j=1}^m w_j = 1.$$

Minimisation de la distance de Wasserstein

Problème d'optimisation :

$$\min_{\mathbf{w}=(w_j)_{1 \leq j \leq m}} W_q \left(\underbrace{\sum_{j=1}^m w_j \delta_{X'_j}}_{\hat{\mu}_m^{\mathbf{w}}}, \underbrace{\frac{1}{n} \sum_{i=1}^n \delta_{X_i}}_{\hat{\mu}_n} \right), \quad \sum_{j=1}^m w_j = 1.$$

Solution explicite (et **indépendante de q**) :



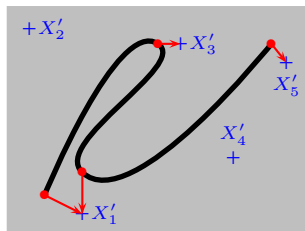
Minimisation de la distance de Wasserstein

Problème d'optimisation :

$$\min_{\mathbf{w}=(w_j)_{1 \leq j \leq m}} W_q \left(\underbrace{\sum_{j=1}^m w_j \delta_{X'_j}}_{\hat{\mu}_m^{\mathbf{w}}}, \underbrace{\frac{1}{n} \sum_{i=1}^n \delta_{X_i}}_{\hat{\mu}_n} \right), \quad \sum_{j=1}^m w_j = 1.$$

Solution explicite (et **indépendante de q**) :

► $\text{NN}^{(1)}(x)$ plus proche voisin de x parmi les X'_j .

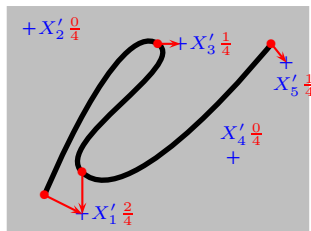


Minimisation de la distance de Wasserstein

Problème d'optimisation :

$$\min_{\mathbf{w}=(w_j)_{1 \leq j \leq m}} W_q \left(\underbrace{\sum_{j=1}^m w_j \delta_{X'_j}}_{\hat{\mu}_m^{\mathbf{w}}}, \underbrace{\frac{1}{n} \sum_{i=1}^n \delta_{X_i}}_{\hat{\mu}_n} \right), \quad \sum_{j=1}^m w_j = 1.$$

Solution explicite (et **indépendante de q**) :



- ▶ $\text{NN}^{(1)}(x)$ plus proche voisin de x parmi les X'_j .
- ▶ Optimum $\mathbf{w}^{(1)}$ donné par

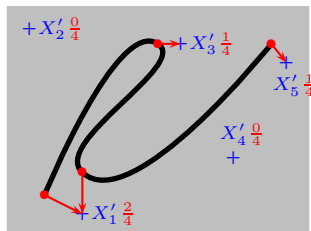
$$w_j^{(1)} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X'_j = \text{NN}^{(1)}(X_i)\}}.$$

Minimisation de la distance de Wasserstein

Problème d'optimisation :

$$\min_{\mathbf{w}=(w_j)_{1 \leq j \leq m}} W_q \left(\underbrace{\sum_{j=1}^m w_j \delta_{X'_j}}_{\hat{\mu}_m^{\mathbf{w}}}, \underbrace{\frac{1}{n} \sum_{i=1}^n \delta_{X_i}}_{\hat{\mu}_n} \right), \quad \sum_{j=1}^m w_j = 1.$$

Solution explicite (et **indépendante de q**) :



▶ $\text{NN}^{(1)}(x)$ plus proche voisin de x parmi les X'_j .

▶ Optimum $\mathbf{w}^{(1)}$ donné par

$$w_j^{(1)} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X'_j = \text{NN}^{(1)}(X_i)\}}.$$

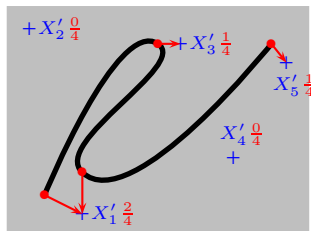
▶ $W_q \left(\hat{\mu}_m^{\mathbf{w}^{(1)}}, \hat{\mu}_n \right)^q = \frac{1}{n} \sum_{i=1}^n |X_i - \text{NN}^{(1)}(X_i)|^q.$

Minimisation de la distance de Wasserstein

Problème d'optimisation :

$$\min_{\mathbf{w}=(w_j)_{1 \leq j \leq m}} W_q \left(\underbrace{\sum_{j=1}^m w_j \delta_{X'_j}}_{\hat{\mu}_m^{\mathbf{w}}}, \underbrace{\frac{1}{n} \sum_{i=1}^n \delta_{X_i}}_{\hat{\mu}_n} \right), \quad \sum_{j=1}^m w_j = 1.$$

Solution explicite (et **indépendante de q**) :



▶ $\text{NN}^{(1)}(x)$ plus proche voisin de x parmi les X'_j .

▶ Optimum $\mathbf{w}^{(1)}$ donné par

$$w_j^{(1)} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X'_j = \text{NN}^{(1)}(X_i)\}}.$$

▶ $W_q \left(\hat{\mu}_m^{\mathbf{w}^{(1)}}, \hat{\mu}_n \right)^q = \frac{1}{n} \sum_{i=1}^n |X_i - \text{NN}^{(1)}(X_i)|^q.$

Lien avec la **quantisation optimale** Graf, Luschgy – Foundations of quantization for probability distributions '07.

Convergence de l'approximation de μ_X

$$\begin{aligned}
\mathbb{E} \left[W_q \left(\widehat{\mu}_m^{\mathbf{w}^{(1)}}, \widehat{\mu}_n \right)^q \right] &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[|X_i - \text{NN}^{(1)}(X_i)|^q \right] \\
&= \mathbb{E}_{X \sim \mu} \left[|X - \text{NN}^{(1)}(X)|^q \right] && \text{uniforme en } n \\
&= \mathbb{E}_{X \sim \mu} \left[\min_{1 \leq j \leq m} |X - X'_j|^q \right] \\
&\rightarrow \mathbb{E}_{X \sim \mu} \left[\text{dist}(X, \text{supp}(\mu_{X'}))^q \right] && m \rightarrow +\infty.
\end{aligned}$$

Convergence de l'approximation de μ_X

$$\begin{aligned}
 \mathbb{E} \left[W_q \left(\widehat{\mu}_m^{\mathbf{w}^{(1)}}, \widehat{\mu}_n \right)^q \right] &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[|X_i - \text{NN}^{(1)}(X_i)|^q \right] \\
 &= \mathbb{E}_{X \sim \mu} \left[|X - \text{NN}^{(1)}(X)|^q \right] && \text{uniforme en } n \\
 &= \mathbb{E}_{X \sim \mu} \left[\min_{1 \leq j \leq m} |X - X'_j|^q \right] \\
 &\rightarrow \mathbb{E}_{X \sim \mu} \left[\text{dist}(X, \text{supp}(\mu_{X'}))^q \right] && m \rightarrow +\infty.
 \end{aligned}$$

\rightsquigarrow Si $\text{supp}(\mu_X) \subset \text{supp}(\mu_{X'})$ alors $\widehat{\mu}_m^{\mathbf{w}^{(1)}}$ **converge vers** μ_X .

Convergence de l'approximation de μ_X

$$\begin{aligned}
 \mathbb{E} \left[W_q \left(\widehat{\mu}_m^{\mathbf{w}^{(1)}}, \widehat{\mu}_n \right)^q \right] &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[|X_i - \text{NN}^{(1)}(X_i)|^q \right] \\
 &= \mathbb{E}_{X \sim \mu} \left[|X - \text{NN}^{(1)}(X)|^q \right] && \text{uniforme en } n \\
 &= \mathbb{E}_{X \sim \mu} \left[\min_{1 \leq j \leq m} |X - X'_j|^q \right] \\
 &\rightarrow \mathbb{E}_{X \sim \mu} \left[\text{dist}(X, \text{supp}(\mu_{X'}))^q \right] && m \rightarrow +\infty.
 \end{aligned}$$

\rightsquigarrow Si $\text{supp}(\mu_X) \subset \text{supp}(\mu_{X'})$ alors $\widehat{\mu}_m^{\mathbf{w}^{(1)}}$ **converge vers** μ_X .

Taux de convergence :

- ▶ Si $\mu_{X'}$ a une densité $p_{X'} > 0$ sur un voisinage de $\text{supp}(\mu_X)$,

$$\forall x \in \text{supp}(\mu_X), \quad m^{1/d} \min_{1 \leq j \leq m} |x - X'_j| \rightarrow \text{Weibull.}$$

- ▶ Sous condition de moment, $\mathbb{E} \left[W_q \left(\widehat{\mu}_m^{\mathbf{w}^{(1)}}, \widehat{\mu}_n \right)^q \right]$ **d'ordre** $m^{-q/d}$.
- ▶ Taux usuel pour plus proches voisins **Biau, Devroye – Lectures on the nearest neighbor method '15, [malédiction de la dimension](#)**.

Convergence de l'estimateur $\widehat{QI}_{m,n}$

$$\widehat{QI}_{m,n} = \sum_{j=1}^m w_j^{(1)} \phi(Y'_j) = \frac{1}{n} \sum_{i=1}^n \phi(Y'_{j_i}), \quad X'_{j_i} = \text{NN}^{(1)}(X_i).$$

(Combinaison d'estimation de Monte-Carlo et de régression au plus proche voisin)

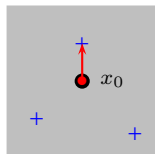
Convergence de l'estimateur $\widehat{QI}_{m,n}$

$$\widehat{QI}_{m,n} = \sum_{j=1}^m w_j^{(1)} \phi(Y'_j) = \frac{1}{n} \sum_{i=1}^n \phi(Y'_{j_i}), \quad X'_{j_i} = \text{NN}^{(1)}(X_i).$$

(Combinaison d'estimation de Monte-Carlo et de régression au plus proche voisin)

Un cas « extrême » : $\mu_X = \delta_{x_0}$.

- ▶ $\widehat{QI}_{m,n} = \phi(f(X'_j, \Theta_j))$ pour $X'_j = \text{NN}^{(1)}(x)$.
- ▶ Converge (en loi) vers $\phi(f(x_0, \Theta)) \neq \text{QI} = \mathbb{E}[\phi(f(x_0, \Theta))]$.
- ▶ **Pas de moyennage en Θ .**



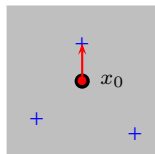
Convergence de l'estimateur $\widehat{QI}_{m,n}$

$$\widehat{QI}_{m,n} = \sum_{j=1}^m w_j^{(1)} \phi(Y'_j) = \frac{1}{n} \sum_{i=1}^n \phi(Y'_{j_i}), \quad X'_{j_i} = \text{NN}^{(1)}(X_i).$$

(Combinaison d'estimation de Monte-Carlo et de régression au plus proche voisin)

Un cas « extrême » : $\mu_X = \delta_{x_0}$.

- ▶ $\widehat{QI}_{m,n} = \phi(f(X'_j, \Theta_j))$ pour $X'_j = \text{NN}^{(1)}(x)$.
- ▶ Converge (en loi) vers $\phi(f(x_0, \Theta)) \neq \text{QI} = \mathbb{E}[\phi(f(x_0, \Theta))]$.
- ▶ **Pas de moyennage en Θ .**



Pour retrouver la convergence :

- ▶ Il faut prendre **k plus proches voisins**, avec $k \rightarrow +\infty$ et $k/m \rightarrow 0$.
- ▶ Poids associés :

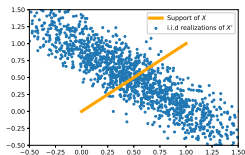
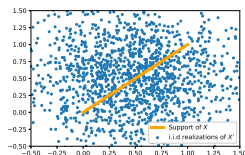
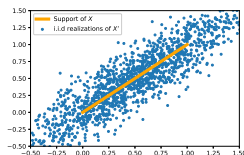
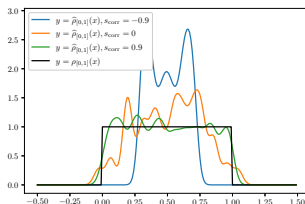
$$w_j^{(k)} = \frac{1}{k} \sum_{\ell=1}^k \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X'_j = \text{NN}^{(\ell)}(X_i)\}}.$$

- ▶ Choix optimal : $k \sim m^{2/(d+2)}$.
- ▶ Convergence de $\widehat{QI}_{m,n}^{(k)}$ vers QI en $n^{-1/d} + m^{-1/(d+2)}$.

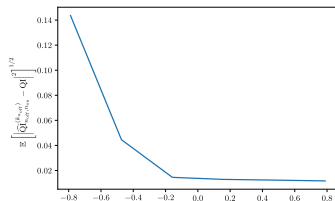
Détails : R., Touboul – arXiv 2010.09267.

Illustration numérique

$$X = \begin{pmatrix} U \\ U \end{pmatrix}, \quad U \sim \text{Unif}([0, 1]); \quad X' \sim \text{Norm} \left(\begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & s \\ s & 1 \end{pmatrix} \right).$$


 $s = -0.09$

 $s = 0$

 $s = 0.09$


Approximation de μ_X par $\hat{\mu}_m^w(1)$
(+lissage à noyau)



Erreur L^2 entre $\widehat{QI}_{m,n}^{s_{\text{score}}}$ et QI
($n = m = 900$)

Une dernière remarque

- Les poids $w_j^{(k)}$ s'écrivent de manière abstraite sous la forme

$$w_j^{(k)} = \frac{1}{n} \sum_{i=1}^n \frac{1}{k} \sum_{\ell=1}^k \mathbb{1}_{\{X'_j = \text{NN}^{(\ell)}(X_i)\}} = \int W_j^m(\mathbf{X}'_m, x) d\widehat{\mu}_n(x),$$

avec $\mathbf{X}'_m := (X'_j)_{1 \leq j \leq m}$.

Une dernière remarque

- ▶ Les poids $w_j^{(k)}$ s'écrivent de manière abstraite sous la forme

$$w_j^{(k)} = \frac{1}{n} \sum_{i=1}^n \frac{1}{k} \sum_{\ell=1}^k \mathbb{1}_{\{X'_j = \text{NN}^{(\ell)}(X_i)\}} = \int W_j^m(\mathbf{X}'_m, x) d\widehat{\mu}_n(x),$$

avec $\mathbf{X}'_m := (X'_j)_{1 \leq j \leq m}$.

- ▶ La quantité $W_j^m(\mathbf{X}'_m, x)$ est le poids associé à Y'_j dans l'approximation de la **loi conditionnelle** de Y sachant $X = x$:

$$\ell(x, dy) := \text{Loi}(Y|X = x) = \text{Loi}(f(x, \Theta)) \simeq \sum_{j=1}^m W_j^m(\mathbf{X}'_m, x) \delta_{Y'_j}.$$

Une dernière remarque

- ▶ Les poids $w_j^{(k)}$ s'écrivent de manière abstraite sous la forme

$$w_j^{(k)} = \frac{1}{n} \sum_{i=1}^n \frac{1}{k} \sum_{\ell=1}^k \mathbb{1}_{\{X'_j = \text{NN}^{(\ell)}(X_i)\}} = \int W_j^m(\mathbf{X}'_m, x) d\widehat{\mu}_n(x),$$

avec $\mathbf{X}'_m := (X'_j)_{1 \leq j \leq m}$.

- ▶ La quantité $W_j^m(\mathbf{X}'_m, x)$ est le poids associé à Y'_j dans l'approximation de la **loi conditionnelle** de Y sachant $X = x$:

$$\ell(x, dy) := \text{Loi}(Y|X = x) = \text{Loi}(f(x, \Theta)) \simeq \sum_{j=1}^m W_j^m(\mathbf{X}'_m, x) \delta_{Y'_j}.$$

- ▶ La formule de désintégration

$$\mathbb{P}(Y \in dy) = \int \ell(x, dy) d\mu_X(x)$$

implique que si $\mu_X \simeq \tilde{\mu}$ (empirique avec poids, ou quelconque), alors

$$\text{Loi}(Y) \simeq \sum_{j=1}^m \left(\int W_j^m(\mathbf{X}'_m, x) d\tilde{\mu}(x) \right) \delta_{Y'_j}.$$

Plan de l'exposé

Introduction

Repondération locale d'échantillons

Synthèse globale sur le graphe

Conclusion

Cas du graphe linéaire

$$\begin{array}{ccccccc} \Theta_1 & & & \Theta_2 & & & \Theta_N \\ \downarrow & & & \downarrow & & & \downarrow \\ Y_1 = f_1(\Theta_1) & \xrightarrow{X_2 = g_{1,2}(Y_1)} & & Y_2 = f_2(X_2, \Theta_2) & \cdots \cdots \cdots \rightarrow & & Y_N = f_N(X_N, \Theta_N) \end{array}$$

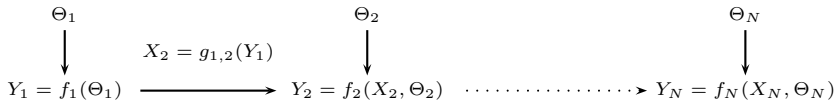
Cas du graphe linéaire

$$\begin{array}{ccccc}
 \Theta_1 & & \Theta_2 & & \Theta_N \\
 \downarrow & & \downarrow & & \downarrow \\
 Y_1 = f_1(\Theta_1) & \xrightarrow{X_2 = g_{1,2}(Y_1)} & Y_2 = f_2(X_2, \Theta_2) & \cdots \cdots \cdots \rightarrow & Y_N = f_N(X_N, \Theta_N)
 \end{array}$$

Phase offline :

- ▶ Le sommet 1 génère un échantillon $(Y'_{j,1})_{1 \leq j \leq m_1}$ **(non biaisé)**.
- ▶ Les autres sommets v génèrent des échantillons $(X'_{j,v}, Y'_{j,v})_{1 \leq j \leq m_v}$ **(biaisés)**.

Cas du graphe linéaire



Phase offline :

- ▶ Le sommet 1 génère un échantillon $(Y'_{j,1})_{1 \leq j \leq m_1}$ (**non biaisé**).
- ▶ Les autres sommets v génèrent des échantillons $(X'_{j,v}, Y'_{j,v})_{1 \leq j \leq m_v}$ (**biaisés**).

Phase online « séquentielle » :

- ▶ Au sommet 1 on pose $w_{j,1} = 1/m_1$ pour tout $j = 1, \dots, m_1$.
- ▶ Supposons calculés des poids $(w_{i,v-1})_{1 \leq i \leq m_{v-1}}$ pour $(Y'_{i,v-1})_{1 \leq i \leq m_{v-1}}$.

$$\rightsquigarrow \mu_{X_v} \simeq \sum_{i=1}^{m_{v-1}} w_{i,v-1} \delta_{X_{i,v}}, \text{ avec } X_{i,v} = g_{v-1,v}(Y'_{i,v-1}).$$

\rightsquigarrow On a donc, pour tout $j = 1, \dots, m_v$,

$$w_{j,v} = \sum_{i=1}^{m_{v-1}} W_{j,v}^{m_v} (X'_{m_v,v}, X_{i,v}) w_{i,v-1}.$$

\rightsquigarrow C'est un **produit matriciel**.

Cadre général

- ▶ QI peut dépendre de **plusieurs** Y_v : besoin d'estimer des **lois jointes**.
- ▶ Hypothèse : graphe **sans cycle orienté**, possibilité de dépendances en amont.

Cadre général

- ▶ QI peut dépendre de **plusieurs** Y_v : besoin d'estimer des **lois jointes**.
- ▶ Hypothèse : graphe **sans cycle orienté**, possibilité de dépendances en amont.

On note \mathcal{R} l'ensemble des racines du graphe.

Formule de factorisation

$$\text{Loi}((Y_v)_{v \in \mathcal{V}}) = \prod_{v \in \mathcal{R}} \ell_v(dy_v) \prod_{v \in \mathcal{V} \setminus \mathcal{R}} \ell_v(x_v, dy_v),$$

où :

- ▶ $\ell_v(dy_v) = \text{Loi}(Y_v) = \text{Loi}(f_v(\Theta_v))$ si $v \in \mathcal{R}$,
- ▶ $\ell_v(x_v, dy_v) = \text{Loi}(Y_v | X_v = x_v) = \text{Loi}(f_v(x_v, \Theta_v))$ si $v \in \mathcal{V} \setminus \mathcal{R}$,
- ▶ $x_v = (g_{u,v}(y_u))_{(u,v) \in \mathcal{E}}$.

Estimateur de Loi $((Y_v)_{v \in \mathcal{V}})$

$$\text{Loi}((Y_v)_{v \in \mathcal{V}}) = \prod_{v \in \mathcal{R}} \ell_v(dy_v) \prod_{v \in \mathcal{V} \setminus \mathcal{R}} \ell_v(x_v, dy_v).$$

Estimateur de Loi $((Y_v)_{v \in \mathcal{V}})$

$$\text{Loi}((Y_v)_{v \in \mathcal{V}}) = \prod_{v \in \mathcal{R}} \ell_v(dy_v) \prod_{v \in \mathcal{V} \setminus \mathcal{R}} \ell_v(x_v, dy_v).$$

On sait estimer chaque terme par

$$\widehat{\ell}_{m_v, v}(dy_v) := \sum_{j=1}^{m_v} \frac{1}{m_v} \delta_{Y_{j,v}'} , \quad v \in \mathcal{R},$$

$$\widehat{\ell}_{m_v, v}(x_v, dy_v) := \sum_{j=1}^{m_v} W_{j,v}^{m_v}(\mathbf{X}'_{m_v, v}, x_v) \delta_{Y_{j,v}'} , \quad v \in \mathcal{V} \setminus \mathcal{R}.$$

On en déduit l'estimateur de la loi jointe

$$\widehat{\mu}_{(m_v)_{v \in \mathcal{V}}} := \prod_{v \in \mathcal{R}} \widehat{\ell}_{m_v, v}(dy_v) \prod_{v \in \mathcal{V} \setminus \mathcal{R}} \widehat{\ell}_{m_v, v}(x_v, dy_v)$$

Estimateur de Loi $((Y_v)_{v \in \mathcal{V}})$

$$\text{Loi}((Y_v)_{v \in \mathcal{V}}) = \prod_{v \in \mathcal{R}} \ell_v(dy_v) \prod_{v \in \mathcal{V} \setminus \mathcal{R}} \ell_v(x_v, dy_v).$$

On sait estimer chaque terme par

$$\widehat{\ell}_{m_v, v}(dy_v) := \sum_{j=1}^{m_v} \frac{1}{m_v} \delta_{Y_{j,v}'} , \quad v \in \mathcal{R},$$

$$\widehat{\ell}_{m_v, v}(x_v, dy_v) := \sum_{j=1}^{m_v} W_{j,v}^{m_v}(\mathbf{X}'_{m_v, v}, x_v) \delta_{Y_{j,v}'} , \quad v \in \mathcal{V} \setminus \mathcal{R}.$$

On en déduit l'estimateur de la loi jointe

$$\widehat{\mu}_{(m_v)_{v \in \mathcal{V}}} := \prod_{v \in \mathcal{R}} \widehat{\ell}_{m_v, v}(dy_v) \prod_{v \in \mathcal{V} \setminus \mathcal{R}} \widehat{\ell}_{m_v, v}(x_v, dy_v)$$

Remarque :

- ▶ $\widehat{\ell}_{m_v, v}(x_v, dy_v)$ forme canonique en **régression non-paramétrique**.
- ▶ D'autres choix possibles (ex : Nadaraya–Watson).

Tsybakov – Introduction to nonparametric estimation '09

Interprétation en termes de réseau bayésien

Notons à nouveau $\mathcal{V} = \{1, \dots, N\}$.

$$\hat{\mu}_{(m_v)_{v \in \mathcal{V}}} = \sum_{j_1=1}^{m_1} \cdots \sum_{j_N=1}^{m_N} Q_{j_1, \dots, j_N} \delta_{(Y'_{j_1,1}, \dots, Y'_{j_N,N})}.$$

Chaque poids Q_{j_1, \dots, j_N} est donné par la formule

$$Q_{j_1, \dots, j_N} = \prod_{v \in \mathcal{R}} q_v(j_v) \prod_{v \in \mathcal{V} \setminus \mathcal{R}} q_v(j_v | (j_u)_{(u,v) \in \mathcal{E}}),$$

avec :

- ▶ $q_v(j_v)$ le poids associé à l'observation $Y'_{j_v, v}$ dans l'estimation de $\text{Loi}(Y_v)$,
- ▶ $q_v(j_v | (j_u)_{(u,v) \in \mathcal{E}})$ le poids associé à l'observation $Y'_{j_v, v}$ dans l'estimation de $\text{Loi}(Y_v | X_v = (g_{(u,v)}(Y'_{j_u}))_{(u,v) \in \mathcal{E}})$.

Interprétation en termes de réseau bayésien

Notons à nouveau $\mathcal{V} = \{1, \dots, N\}$.

$$\hat{\mu}_{(m_v)_{v \in \mathcal{V}}} = \sum_{j_1=1}^{m_1} \cdots \sum_{j_N=1}^{m_N} Q_{j_1, \dots, j_N} \delta_{(Y'_{j_1,1}, \dots, Y'_{j_N,N})}.$$

Chaque poids Q_{j_1, \dots, j_N} est donné par la formule

$$Q_{j_1, \dots, j_N} = \prod_{v \in \mathcal{R}} q_v(j_v) \prod_{v \in \mathcal{V} \setminus \mathcal{R}} q_v(j_v | (j_u)_{(u,v) \in \mathcal{E}}),$$

avec :

- ▶ $q_v(j_v)$ le poids associé à l'observation $Y'_{j_v, v}$ dans l'estimation de $\text{Loi}(Y_v)$,
- ▶ $q_v(j_v | (j_u)_{(u,v) \in \mathcal{E}})$ le poids associé à l'observation $Y'_{j_v, v}$ dans l'estimation de $\text{Loi}(Y_v | X_v = (g_{(u,v)}(Y'_{j_u}))_{(u,v) \in \mathcal{E}})$.

↪ C'est un **réseau bayésien** !

↪ Selon l'observable Φ que l'on cherche à estimer et la structure du graphe, des algorithmes d'optimisation combinatoire permettent d'évaluer

$$\widehat{Q\mathbb{I}} = \sum_{j_1=1}^{m_1} \cdots \sum_{j_N=1}^{m_N} Q_{j_1, \dots, j_N} \Phi(Y'_{j_1,1}, \dots, Y'_{j_N,N})$$

sans calculer les $m_1 \times \cdots \times m_N$ termes. **Koller, Friedman – Probabilistic graphical models '09**

Plan de l'exposé

Introduction

Repondération locale d'échantillons

Synthèse globale sur le graphe

Conclusion

Synthèse de la méthode

Objectif : estimer $QI = \mathbb{E}[\Phi((Y_v)_{v \in \mathcal{V}})]$.

- ▶ Phase **offline** : chaque sommet v génère un échantillon $(X'_{j,v}, Y'_{j,v})_{1 \leq j \leq m_v}$.
 ↪ Sélection d'un plan d'expérience.
- ▶ Phase **online I** : pour chaque sommet v , l'architecte de simulation construit le tableau des $m_v \times \prod_{(u,v) \in \mathcal{E}} m_u$ quantités

$$q_v(j_v | (j_u)_{(u,v) \in \mathcal{E}}) = \text{poids à attribuer à } Y'_{j_v, v} \text{ sachant } (Y'_{j_u, u})_{(u,v) \in \mathcal{E}}.$$

↪ Choix d'une méthode de régression non-paramétrique.

- ▶ Phase **online II** : pour chaque observable Φ , l'architecte de simulation estime QI en effectuant une requête au réseau bayésien.
 ↪ Emploi d'algorithmes combinatoires (ou heuristiques).

Synthèse de la méthode

Objectif : estimer $QI = \mathbb{E}[\Phi((Y_v)_{v \in \mathcal{V}})]$.

- ▶ Phase **offline** : chaque sommet v génère un échantillon $(X'_{j,v}, Y'_{j,v})_{1 \leq j \leq m_v}$.
 ↪ Sélection d'un plan d'expérience.
- ▶ Phase **online I** : pour chaque sommet v , l'architecte de simulation construit le tableau des $m_v \times \prod_{(u,v) \in \mathcal{E}} m_u$ quantités

$$q_v(j_v | (j_u)_{(u,v) \in \mathcal{E}}) = \text{poids à attribuer à } Y'_{j_v, v} \text{ sachant } (Y'_{j_u, u})_{(u,v) \in \mathcal{E}}.$$

↪ Choix d'une méthode de régression non-paramétrique.

- ▶ Phase **online II** : pour chaque observable Φ , l'architecte de simulation estime QI en effectuant une requête au réseau bayésien.
 ↪ Emploi d'algorithmes combinatoires (ou heuristiques).

Remarques finales :

- ▶ pas encore d'étude quantitative très poussée ;
- ▶ beaucoup de degrés de liberté pour optimiser la performance.

Référence : **Touboul – Thèse de doctorat '21.**