Uncertainty Quantification in graphs of functions through sample reweighting

Julien Reygner[1], Adrien Touboul[1,2] (speaker)

June 3, 2021

CERMICS[1], IRT SystemX[2]

Outline of the presentation

Introduction & motivation

Decomposition-based Uncertainty Quantification

One node - A particular weighting scheme based on a Wasserstein distance criterion

Whole graph - Analysis of the general algorithm

Introduction & motivation

Motivation: simulation of complex industrial systems

Complex industrial systems: car, aircraft, rocket ship,...

- From hundreds to ten of thousands engineers working together
- Huge number of requirements to validate, from various sources (safety regulation, environmental regulation,...)
- Uncertainty during the whole design process.

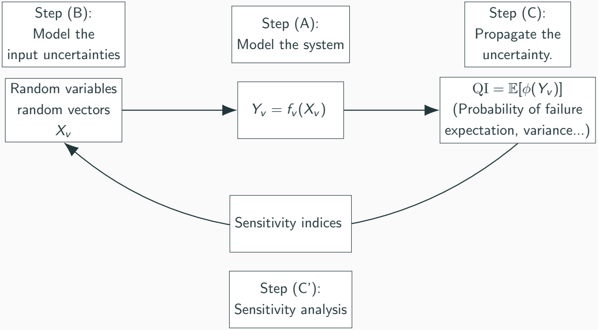
Motivation: simulation of complex industrial systems

Complex industrial systems: car, aircraft, rocket ship,...

- From hundreds to ten of thousands engineers working together
- Huge number of requirements to validate, from various sources (safety regulation, environmental regulation,...)
- Uncertainty during the whole design process.

Need to validate requirements by simulating large parts of systems

Uncertainty Quantification

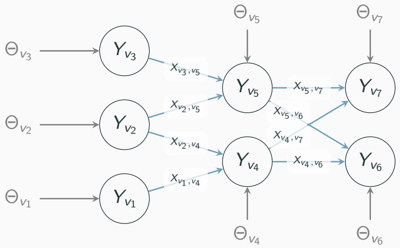


Model

- $Y_v = f_v(X_v)$ with X_v and Y_v random variables.
- ABCC' approach
- Probabilistic model of uncertainty

de Rocquigny, Etienne, Nicolas Devictor, and Stefano Tarantola, eds. Uncertainty in industrial practice: a guide to quantitative uncertainty management. John Wiley & Sons, 2008.

Uncertainty Quantification in complex sytems



Graph of computer codes.
$\mathcal{G} = (\mathcal{V}, \mathcal{E})$

Multidisciplinary system

Each discipline v has

- inputs X_v
- outputs Y_v
- a simulation code f_v s.t

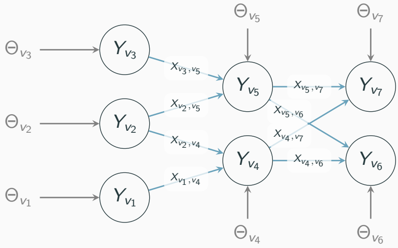
$$Y_v = f_v(X_v)$$

Inputs

X_v composed by

- external variables $X_{u,v}$, for u parents
- internal variables Θ_v

Disciplinary autonomy



Graph of computer codes.
$\mathcal{G} = (\mathcal{V}, \mathcal{E})$

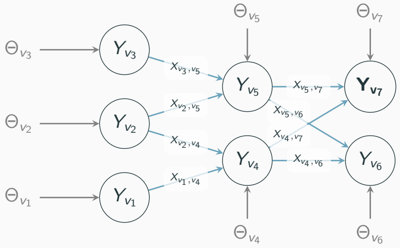
Goal

Given $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and $(\Theta_v, f_v)_{v \in \mathcal{V}}$, compute:

$\mathbb{E}\left[\phi(Y_{v_1}, \ldots, Y_{v_K})\right]$

Industrial constraints

Computations incompatible with graph order.



Graph of computer codes.
$\mathcal{G} = (\mathcal{V}, \mathcal{E})$

Goal

Given $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and $(\Theta_v, f_v)_{v \in \mathcal{V}}$, compute:

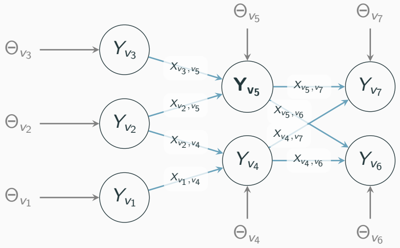
$$\mathbb{E}\left[\phi(Y_{v_1}, \ldots, Y_{v_K})\right]$$

Industrial constraints

Computations incompatible with graph order.

Ex: v_7 does its computations before v_5

Disciplinary autonomy



Graph of computer codes.
$\mathcal{G} = (\mathcal{V}, \mathcal{E})$

Goal

Given $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and $(\Theta_v, f_v)_{v \in \mathcal{V}}$, compute:

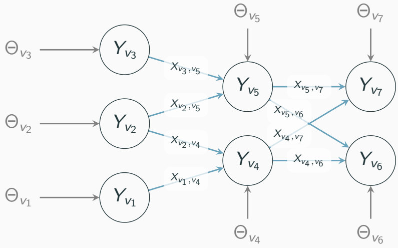
$\mathbb{E}\left[\phi(Y_{v_1}, \dots, Y_{v_K})\right]$

Industrial constraints

Computations incompatible with graph order.

Ex: v_7 does its computations before v_5

Disciplinary autonomy



Graph of computer codes.
$\mathcal{G} = (\mathcal{V}, \mathcal{E})$

Goal

Given $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and $(\Theta_v, f_v)_{v \in \mathcal{V}}$, compute:

$\mathbb{E}\left[\phi(Y_{v_1}, \ldots, Y_{v_K})\right]$

Industrial constraints

Computations incompatible with graph order.

Ex: v_7 does its computations before v_5

Classic Monte Carlo forbidden, need for

Disciplinary autonomy

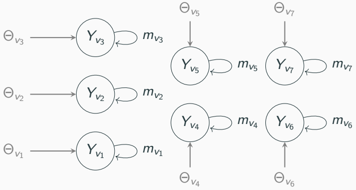
Decomposition-based Uncertainty Quantification

Decomposition-based Uncertainty Quantification

Disciplinary autonomy

Two phases

Decomposition-based Uncertainty Quantification

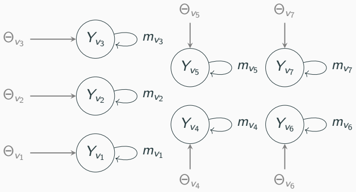


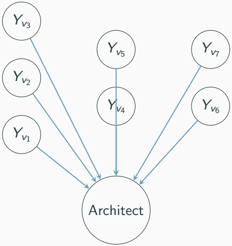
Disciplinary autonomy

Two phases

- *Offline* Computation of f_v, no samples exchange

Decomposition-based Uncertainty Quantification



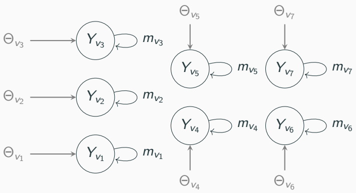


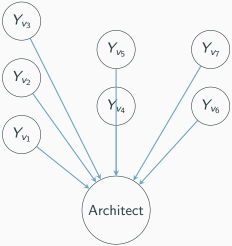
Disciplinary autonomy

Two phases

- *Offline* Computation of f_v, no samples exchange
- *Online* Gathering samples, no more f_v

Decomposition-based Uncertainty Quantification





Disciplinary autonomy

Two phases

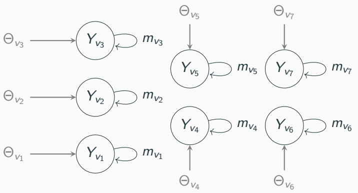
- *Offline* Computation of f_v, no samples exchange
- *Online* Gathering samples, no more f_v

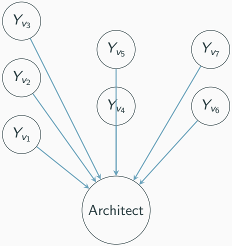
Question: how to “glue it back” together in the online phase?

Solutions

- Reconstruct each f_v with metamodels or

Decomposition-based Uncertainty Quantification





Amaral, Sergio, Douglas Allaire, and Karen Willcox. "A decomposition-based approach to uncertainty analysis of feed-forward multicomponent systems." International Journal for Numerical Methods in Engineering 100.13 (2014): 982-1005.

Disciplinary autonomy

Two phases

- *Offline* Computation of f_v, no samples exchange
- *Online* Gathering samples, no more f_v

Question: how to "glue it back" together in the online phase?

Solutions

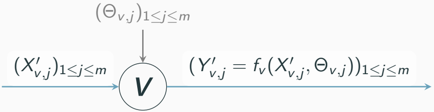
- Reconstruct each f_v with metamodels or
- **Sample reweighting**

Decomposition-based Uncertainty Quantification

Sample reweighting

Idea, at a given node v:

Decomposition-based Uncertainty Quantification

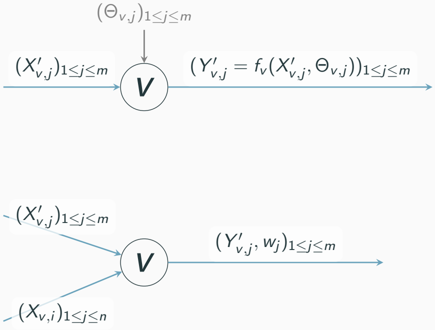


Sample reweighting

Idea, at a given node v:

- *Offline* External input sampled according to a proposal $p_{X'}$, compute $(X'_j, Y'_j)_{1\leq j\leq m}$.

Decomposition-based Uncertainty Quantification



Sample reweighting

Idea, at a given node v:

- *Offline* External input sampled according to a proposal $p_{X'}$, compute $(X'_j, Y'_j)_{1\leq j\leq m}$.
- *Online* When a true sample $(X_i)_{1\leq i\leq n}$ is available, reweight Y'_j to approximate the law $f(X)$

Sample reweighting

Goal

Given two i.i.d samples of law μ_X and $\mu_{X'}$

$$\boldsymbol{X}_n = (X_i)_{1\leq i\leq n}$$
$$\boldsymbol{X}'_m = \left(X'_j\right)_{1\leq j\leq m}$$

Sample reweighting

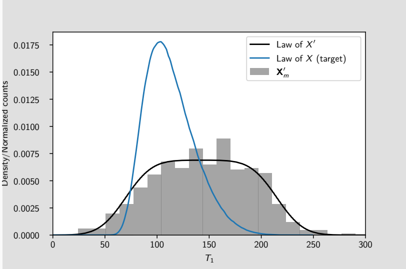
Goal

Given two i.i.d samples of law μ_X and $\mu_{X'}$

$$\boldsymbol{X}_n = (X_i)_{1 \leq i \leq n}$$
$$\boldsymbol{X}'_m = \left(X'_j\right)_{1 \leq j \leq m}$$

find weights $(w_1, \ldots, w_m)$

Sample reweighting



Goal

Given two i.i.d samples of law μ_X and $\mu_{X'}$

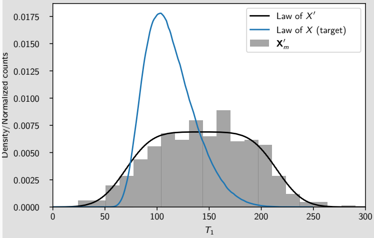
$$\boldsymbol{X}_n = (X_i)_{1\leq i\leq n}$$
$$\boldsymbol{X}'_m = \left(X'_j\right)_{1\leq j\leq m}$$

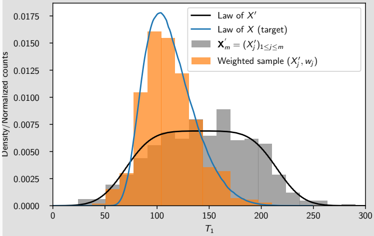
find weights $(w_1, \ldots, w_m)$ such that

$$\left(X'_j, w_j\right) \overset{\text{law}}{\simeq} \boldsymbol{X}_n$$

ie $\mathbb{E}[\psi(X)] \simeq 1/m \sum_{j=1}^m w_j \psi(X'_j)$

Sample reweighting





Goal

Given two i.i.d samples of law μ_X and $\mu_{X'}$

$$\boldsymbol{X}_n = (X_i)_{1\leq i\leq n}$$
$$\boldsymbol{X}'_m = \left(X'_j\right)_{1\leq j\leq m}$$

find weights $(w_1, \dots, w_m)$ such that

$$\left(X'_j, w_j\right) \overset{\text{law}}{\simeq} \boldsymbol{X}_n$$

ie $\mathbb{E}[\psi(X)] \simeq 1/m \sum_{j=1}^{m} w_j \psi(X'_j)$

One node - A particular weighting scheme based on a Wasserstein distance criterion

Sample reweighting, first idea: importance weighting

$$\int f(x)\mu_X(dx) = \int f(x)\frac{\mu_X}{\mu_{X'}}(x)\mu_{X'}(dx) \simeq \sum_{j=1}^{m} w_j f(X_j')$$

with $w_j = \frac{\mu_X}{\mu_{X'}}(X_j')$.

Two limits:

Sample reweighting, first idea: importance weighting

$$\int f(x)\mu_X(dx) = \int f(x)\frac{\mu_X}{\mu_{X'}}(x)\mu_{X'}(dx) \simeq \sum_{j=1}^{m} w_j f(X'_j)$$

with $w_j = \frac{\mu_X}{\mu_{X'}}(X'_j)$.

Two limits:

- Computing μ_X and $\mu_{X'}$ required.
  Only $(X_i)_{i\in[\![1,n]\!]}$ $(X'_j)_{j\in[\![1,m]\!]}$, not μ_X and $\mu_{X'}$.
  $\Rightarrow$ *Density ratio estimation* [1], various methods developed.

[1] Sugiyama, M., Taiji S., and Takafumi K. Density ratio estimation in machine learning. Cambridge University Press, 2012.

Sample reweighting, first idea: importance weighting

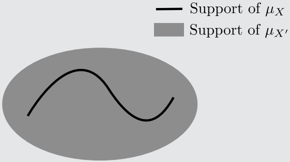
$$\int f(x)\mu_X(dx) = \int f(x)\frac{\mu_X}{\mu_{X'}}(x)\mu_{X'}(dx) \simeq \sum_{j=1}^{m} w_j f(X'_j)$$

with $w_j = \frac{\mu_X}{\mu_{X'}}(X'_j)$.

Two limits:

- Computing μ_X and $\mu_{X'}$ required.
  Only $(X_i)_{i\in[\![1,n]\!]}$ $(X'_j)_{j\in[\![1,m]\!]}$, not μ_X and $\mu_{X'}$.
  $\Rightarrow$ *Density ratio estimation* [1], various methods developed.
- μ_X needs a density w.r.t $\mu_{X'}$ (absolute continuity)

Assumption not verified in practice



[1] Sugiyama, M., Taiji S., and Takafumi K. Density ratio estimation in machine learning. Cambridge University Press, 2012.

Another approach: reinterpretation with empirical measures

Empirical measure on $\boldsymbol{X}_n$, weighted empirical measure on $\boldsymbol{X}'_m$

$$\widehat{\mu}_{X,n} = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}, \quad \widehat{\mu}^{\boldsymbol{w}}_{X',m} = \sum_{j=1}^{m} w_j \delta_{X'_j}$$

with $\sum_{j=1}^{m} w_j = m$.

$$\langle \widehat{\mu}_{X,n}, \phi \circ f \rangle \xrightarrow{n \to +\infty} \mathbb{E}[\phi(f(X))](L.L.N)$$

Minimization of the distance between empirical measures

$$\mathbf{w}^* = \underset{\sum w_i = 1, w_i \geq 0}{\operatorname{argmin}} \quad d\left(\widehat{\mu}_{X,n}, \widehat{\mu}^{\boldsymbol{w}}_{X',m}\right)$$

Choice of distance

Wasserstein distances of order q [2]

μ and ν two probability measures on $\mathbb{R}^d$, with first q moments.

Wasserstein distance

$$W_q(\mu,\nu) = \inf\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-x'|^q \mathrm{d}\gamma(x,x') : \gamma \in \Pi(\mu,\nu)\right\}^{1/q},$$

$\Pi(\mu,\nu)$: proba measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν.

Optimal weights. (Reygner J., T.A 2020)

Let the optimal weights be $\boldsymbol{w}^* = \underset{\sum w_j = 1, w_j \geq 0}{\operatorname{argmin}} W_q\left(\widehat{\mu}_{X,n}, \widehat{\mu}^{\boldsymbol{w}}_{X',m}\right)$.

Then

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

[2] Villani, Cédric. Optimal transport: old and new. Vol. 338. Springer Science & Business Media, 2008.

Consistency of the weighting scheme

A1. Support condition

We have $\mathrm{Supp}(\mu_X) \subset \mathrm{Supp}(\mu_{X'})$.

A2. Min-integrability

There exists an integer $m_0 \geq 1$ such that

$$\mathbb{E}\left[\min_{j \in [\![1, m_0]\!]} |X'_j|\right] < +\infty.$$

Theorem Consistency (Reygner J., T. A. , 2020)

Let (A1) and (A2) hold. For all $q \in [1, +\infty)$ s.t $\mathbb{E}[|X|^q] < +\infty$, then

$$\lim_{m \to +\infty} \mathbb{E}\left[W_q^q\left(\widehat{\mu}_{X,n}, \widehat{\mu}^{w^*}_{X',m}\right)\right] = 0,$$

uniformly in n.

A3: Strong support condition [3]

There exists an open set $U \subset \mathbb{R}^d$ which contains $\mathrm{Supp}(\mu_X)$ and such that:

1. the measure $\mu_{X'}(\cdot \cap U)$ has a density $p_{X'}$ with respect to the Lebesgue measure; $p_{X'}$ is continuous and positive on U;
2. there exist $\kappa \in (0, 1]$ and $r_\kappa > 0$ such that, for any $x \in U$, for any $r \in [0, r_\kappa]$,

$$\mathbb{P}\left(X' \in B(x, r)\right) \geq \kappa p_{X'}(x) v_d r^d.$$

A4: Moments

$$\mathbb{E}\left[\frac{1 + |X|^q}{p_{X'}(X)^{q/d}}\right] < +\infty.$$

[3] Sébastien Gadat, Thierry Klein, Clément Marteau, et al. Classification in general finite dimensional spaceswith the k-nearest neighbor rule.The Annals of Statistics, 44(3):982–1009, 2016.

Theorem Convergence rates (Reygner J., T. A. , 2020)

Let Assumptions A2 and A3 hold, and let $q \in [1, +\infty)$ be such that Assumption A4 holds. Then we have

$$\lim_{m \to +\infty} m^{q/d} \mathbb{E}\left[W_q^q\left(\widehat{\mu}_{X,n}, \widehat{\mu}^{\boldsymbol{w}^*}_{X',m}\right)\right] = c_{q,d} \mathbb{E}\left[\frac{1}{p_{X'}(X)^{q/d}}\right].$$

- Curse of the dimensionality $m^{-q/d}$ (similar NNR)
- Can be reinterpreted in terms of NNR under covariate shift

$$\mathbb{E}[W_q^q(\widehat{\mu}_{X,n}, \widehat{\mu}^{\boldsymbol{w}^*}_{X',m})] = \mathbb{E}[|X - \mathrm{NN}_{\boldsymbol{X}'_m}(X)|^q]$$

Reygner, Julien, and T.A . "Reweighting samples under covariate shift using a Wasserstein distance criterion."

arXiv preprint arXiv:2010.09267 (2020).

Conclusion of the part

Results

For a given node, a weighting method has been analyzed

- Weights given by minimization of Wasserstein distance
- Weights expressed in terms of 1-Nearest Neighbor
- Expected rate of convergence (under appropriate assumptions)

$$\mathbb{E}[W_q^q(\mu_X, \widehat{\mu}^{\boldsymbol{w}^*}_{X',m})] \leq \mathbb{E}[W_q^q(\mu_X, \widehat{\mu}_{X,n})] + \mathbb{E}[W_q^q(\widehat{\mu}_{X,n}, \widehat{\mu}^{\boldsymbol{w}^*}_{X',m})]$$
$$= O(n^{-q/d}) + O(m^{-q/d})$$

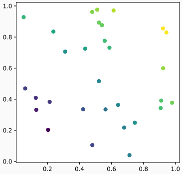
[4] + Reygner J., T. A.(2020)

- Application to estimation of a quantity of interest

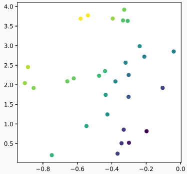
[4] Fournier, Nicolas, and Arnaud Guillin. "On the rate of convergence in Wasserstein distance of the empirical measure." Probability Theory and Related Fields 162.3 (2015): 707-738.

Whole graph - Analysis of the general algorithm

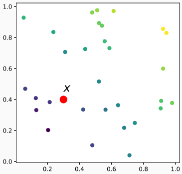
Weighted Linear Approximation Method



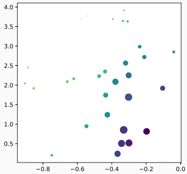
Sample X'_j



$Y'_j = f(X'_j, \Theta_j)$



New input
$x = (0.3, 0.4)$



Gives weights
to Y'_j

Question

What is a weighting method in general?

Weighted Linear Approximation Method (WLAM) (Reygner J., T. A. 2021)

- $\mathbf{S}_m = (X'_j, Y'_j)_{1 \le j \le m}$
- $\mathbf{W}_m = (W_j)_{1 \le j \le m}$: $(\mathrm{E} \times \mathrm{F})^m \times \mathrm{E} \to [0, +\infty)^m$ s.t $\forall \mathbf{S}_m \in (\mathrm{E} \times \mathrm{F})^m, x \in \mathrm{E}$,

$$\sum_{j=1}^{m} W_j(\mathbf{S}_m, x) = 1.$$

Approximate image measure associated to $\mathbf{W}_m$

$$\widehat{\ell}_m(x, \mathrm{d}y) = \sum_{j=1}^{m} W_j(\mathbf{S_m}, x)\delta_{Y'_j}(\mathrm{d}y)$$

approximates the law of Y' conditioned to $X' = x$ (Markov kernel)

$\mathfrak{B}$-consistency (definition)

A WLAM is consistent i.i.f for all $\phi \in \mathfrak{B}$

$$\begin{aligned}\lim_{m\to+\infty} \sum_{j=1}^{m} W_j(\mathbf{S}_m, x)\phi(Y'_j) &= \mathbb{E}[\phi(Y')|X' = x] \\ &= \mathbb{E}[\phi(f(x, \Theta_v))],\end{aligned}$$

in probability.

Examples of WLAMs

- k_m-Nearest-neighbor:

$$W_j(\mathbf{S}_m, x) = 1/k_m \sum_{=1}^{k_m} \mathbb{1}_{\{X'_j = \mathrm{NN}^{(k_m)}(x)\}}$$

  $\mathfrak{B}$: ϕ bounded for which $x \mapsto \mathbb{E}\left[\phi(f(x,\Theta))\right]$ is Lipschitz-continuous (previous section).

- Nadayara-Watson:

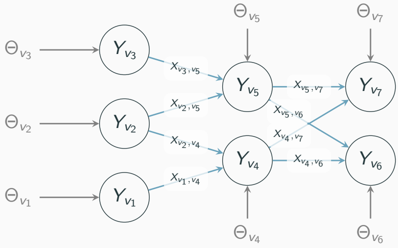
$$W_j(\mathbf{S}_m, x) = K_{j,m}(x)/(\sum_{k=1}^{m} K_{k,m}(x))$$

  $\mathfrak{B}$: ϕ bounded for which $x \mapsto \mathbb{E}\left[\phi(f(x,\Theta))\right]$ is continuous.

- Regression tree: $W_j(\mathbf{S}_m, x) = 1/m(x)\mathbb{1}_{\{X'_j \in L(x)\}}$

$W_j(\mathbf{S}_m, x)$ needs not to be linear in x.

Back to the whole graph



The graph is a Bayesian Network

$(\mathcal{G}, (Y_v)_{v \in \mathcal{V}})$ verifies a Markov Property.

Each node is conditionally independent from its nondescendants, given its direct parents.

Proof:

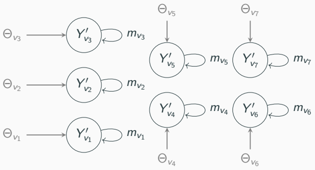
$Y_v = f_v((X_{u,v})_{u \in \mathrm{Par}(v)}, \Theta_v)$

Factorization property of a B.N

$$\mu_{\mathcal{V}}((\mathrm{d}y_v)_{v \in \mathcal{V}}) = \prod_{v \in \mathcal{V}} \ell_v(y_v, (\mathrm{d}y_u)_{u \in \mathrm{Par}(v)})$$

Final algorithm (Reygner J. T.A, 2021)

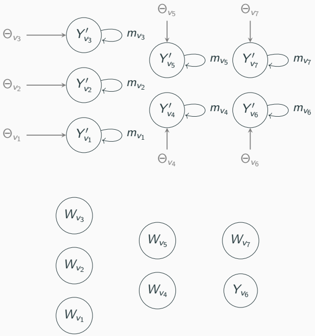
Propagation



Final algorithm (Reygner J. T.A, 2021)

1. *Offline* Computation of $(X'_v, f_v(X'_v, \Theta_v))$

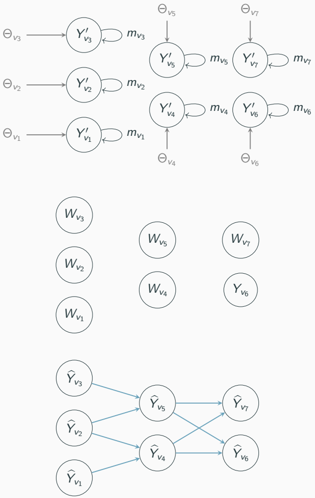
Propagation



Final algorithm (Reygner J. T.A, 2021)

1. *Offline* Computation of $(X'_v, f_v(X'_v, \Theta_v))$
2. *Online 1* Weighting and choice of a WLAM at each node.

Propagation



[5] Koller, D., & Friedman, N. (2009). Probabilistic graphical models: principles and techniques. MIT press.

Final algorithm (Reygner J. T.A, 2021)

1. *Offline* Computation of $(X'_v, f_v(X'_v, \Theta_v))$
2. *Online 1* Weighting and choice of a WLAM at each node.
3. *Online 2* Propagation - Replace each ℓ_v by $\widehat{\ell}_{v,m_v}$.

$$\mu_{\mathcal{V}}((\mathrm{d}y_v)_{v\in\mathcal{V}}) \simeq \widehat{\mu}((\mathrm{d}y_v)_{v\in\mathcal{V}}) = \prod_{v\in\mathcal{V}} \widehat{\ell}_v(y_v, (\mathrm{d}y_u)_{u\in\mathrm{Par}(v)})$$

Discrete Bayesian Network.

Classic propagation methods[5].

Consistency (Reygner J., T.A 2021)

Assume that $\phi((y_v)_{v\in\mathcal{V}})$ is bounded continuous and at each v, the WLAM is $\mathfrak{B}$-consistent, for a compatible family, then

$$\lim_{m_{v_1}\to+\infty}\cdots\lim_{m_{v_N}\to+\infty}\langle\phi,\widehat{\mu}((\mathrm{d}y_v)_{v\in\mathcal{V}})\rangle=\mathbb{E}\left[\phi((Y_v)_{v\in\mathcal{V}})\right],$$

with the $m_{v_1},\ldots,m_{v_n}$ chosen in an order compatible with the graph.

Conclusion

Three key points

1. A graph of composition of functions is a Bayesian Network $(\mathcal{G}, \mathcal{L}(Y_v|(X_{u,v})_{u\in\mathrm{Par}(v)}))$.
   Conditional Probability Tables (C.P.T) $\mathcal{L}(Y_v|(X_{u,v})_{u\in\mathrm{Par}(v)})$ given by
   $$Y_v = f((X_{u,v})_{u\in\mathrm{Par}(v)}, \Theta_v)$$
2. Some weighting methods (NNR, Nadayara-Watson, regression trees...) approximate naturally a C.P.T $\widehat{\mathcal{L}}(Y_v|X_{u,v})$
3. The Bayesian network $(\mathcal{G}, \widehat{\mathcal{L}}(Y_v|(X_{u,v})_{u\in\mathrm{Par}(v)}))$ is discrete (computations available). Its law approximates $(\mathcal{G}, \mathcal{L}(Y_v|(X_{u,v})_{u\in\mathrm{Par}(v)}))$. The weights can be computed numerically.

Perspectives

- Rates of convergences *in the graph* + a posteriori error estimates.
- Efficient computations in Bayesian inference (weights propagation). Sparse propagation?
- Application to other problems than disciplinary autonomy?
- Numerical benchmark of various WLAMs methods, in terms of law approximation.

Resources

- Reygner, Julien, and Touboul, Adrien. "Reweighting samples under covariate shift using a Wasserstein distance criterion." arXiv preprint arXiv:2010.09267 (2020).
- Reygner, Julien "Theoretical analysis and numerical methods for conservation laws, metastability and uncertainty propagation", HDR, (2021), to be defended.
- Touboul, Adrien "Model of margin, margin sensitivity analysis and uncertainty quantification in graphs of functions in complex industrial systems", PhD Thesis (2021), to be defended.