Reweighting samples under covariate shift using a Wasserstein distance criterion

Adrien Touboul
PhD Student, industrial thesis team AMC

December 5, 2020

IRT SystemX / CERMICS

Outline of the presentation

Introduction

Sample reweighting

Theoretical results

Following steps: propagating the weights through a graph of models

Introduction

Context: validation of complex industrial systems requirements by simulation

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.

Goal: decomposition-based approach of Uncertainty Propagation

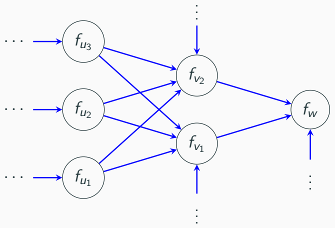
$$\mathbb{E}\left[\phi(f(X))\right] = \int f(x) p_X(x) dx$$
$$\simeq \frac{1}{n} \sum_{i=1}^{n} \phi(f(X_i))$$

Uncertainty quantification

- X random variable, f simulation code.
- Quantity of interest $\mathbb{E}[\phi(f(X))]$, computed by Monte-Carlo/other methods.

Goal: decomposition-based approach of Uncertainty Propagation

$$\mathbb{E}\left[\phi(f(X))\right] = \int f(x)p_X(x)dx$$
$$\simeq \frac{1}{n}\sum_{i=1}^{n}\phi(f(X_i))$$



Uncertainty quantification

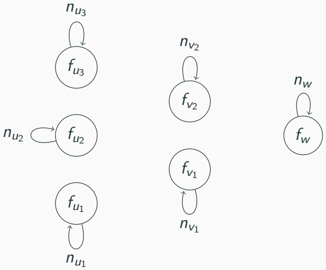
- X random variable, f simulation code.
- Quantity of interest $\mathbb{E}[\phi(f(X))]$, computed by Monte-Carlo/other methods.

Decomposition-based uncertainty quantification

$$f = f_w(f_{v_1}(f_{u_1}(\dots), f_{u_2}(\dots), \dots), \dots$$

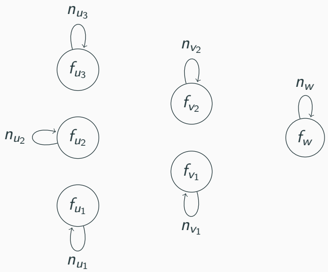
$f_{u_1}, f_{u_2} \dots, f_{v_1}, \dots$ cannot be computed **simultaneously**

Goal: decomposition-based approach of Uncertainty Propagation



Two phases:

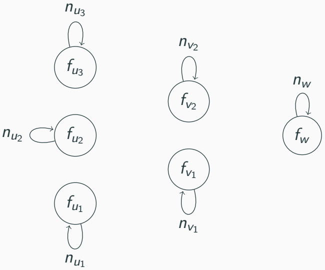
Goal: decomposition-based approach of Uncertainty Propagation



Two phases:

1. **Offline:** Each function computed separatley n_u times.

Goal: decomposition-based approach of Uncertainty Propagation



Two phases:

1. **Offline:** Each function computed separatley n_u times.
2. **Online:** No more computation of the functions. Samples gathered and used to estimate the QOI

$$\boldsymbol{X}'_u = \left(X'_{u,i}, f_u(X'_{u,i})\right)_{i \in [\![1, n_u]\!]}, u \in \text{Vertices}$$

$$\mathbb{E}\left[\phi(f(X))\right] \simeq \psi(\mathrm{X}'_{u_1}, \mathrm{X}'_{u_2}, \dots, \mathrm{X}'_w)$$

Goal: decomposition-based approach of Uncertainty Propagation

Online phase: No more computation of the functions. Samples are gathered and used to estimate the QOI

$$\mathbb{E}\left[\phi(f(X))\right] \simeq \psi(\mathrm{X}'_{u_1}, \mathrm{X}'_{u_2}, \ldots, \mathrm{X}'_{w})$$

Question: how to do it?

- Reconstruct locally each function $\rightarrow$ composition of metamodels.
- Reweight the samples so that they correspond to the true input distribution [1].

[1] Amaral, S., Allaire, D., & Willcox, K. (2014). A decomposition-based approach to uncertainty analysis of feed-forward multicomponent systems. International Journal for Numerical Methods in Engineering, 100(13), 982-1005.

Goal: decomposition-based approach of Uncertainty Propagation

Online phase: No more computation of the functions. Samples are gathered and used to estimate the QOI

$$\mathbb{E}\left[\phi(f(X))\right] \simeq \psi(X'_{u_1}, X'_{u_2}, \ldots, X'_{w})$$

Question: how to do it?

- Reconstruct locally each function → composition of metamodels.
- **Reweight the samples so that they correspond to the true input distribution [1].**

[1] Amaral, S., Allaire, D., & Willcox, K. (2014). A decomposition-based approach to uncertainty analysis of feed-forward multicomponent systems. International Journal for Numerical Methods in Engineering, 100(13), 982-1005.

Sample reweighting

Sample reweighting - Principle

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

$$\boldsymbol{X}_{n_{\mathrm{on}}} = \begin{pmatrix} X_1 \\ \vdots \\ X_{n_{\mathrm{on}}} \end{pmatrix} \text{ and } \boldsymbol{X}'_{n_{\mathrm{off}}} = \begin{pmatrix} X'_1 \\ \vdots \\ X'_{n_{\mathrm{off}}} \end{pmatrix}$$

find weights $(w_1, \ldots, w_{n_{\mathrm{off}}})$ s.t, for any $\phi \circ f$ in a $\mathcal{F}$

$$\sum_{i=1}^{n_{\mathrm{off}}} w_i \phi(f(X'_i)) \xrightarrow[n \to +\infty]{} \mathbb{E}\left[\phi(f(X))\right] = \int \phi(f(x)) \mu_X(dx)$$

Sample reweighting - Principle

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

$$\boldsymbol{X}_{n_{\text{on}}} = \begin{pmatrix} X_1 \\ \vdots \\ X_{n_{\text{on}}} \end{pmatrix} \text{ and } \boldsymbol{X}'_{n_{\text{off}}} = \begin{pmatrix} X'_1 \\ \vdots \\ X'_{n_{\text{off}}} \end{pmatrix}$$

find weights $(w_1, \ldots, w_{n_{\text{off}}})$ s.t, for any $\phi \circ f$ in a $\mathcal{F}$

$$\sum_{i=1}^{n_{\text{off}}} w_i \phi(f(X'_i)) \xrightarrow[n \to +\infty]{} \mathbb{E}\left[\phi(f(X))\right] = \int \phi(f(x)) \mu_X(dx)$$

First idea: importance weighting

$$\int \phi(f(x)) \mu_X(dx) = \int \phi(f(x)) \frac{\mu_X}{\mu_{X'}}(x) \mu_{X'}(dx) \simeq \sum w_i \phi(f(X'_i))$$

with $w_i = \frac{\mu_X}{\mu_{X'}}(X'_i)$

First idea (importance weighting):

$$\int f(x)\mu_X(dx) = \int f(x)\frac{\mu_X}{\mu_{X'}}(x)\mu_{X'}(dx) \simeq \sum_{i=1}^{n_{\mathrm{off}}} w_i \phi(f(X'_i))$$

with $w_i = \frac{\mu_X}{\mu_{X'}}(X'_i)$.

Two limits:

First idea (importance weighting):

$$\int f(x)\mu_X(dx) = \int f(x)\frac{\mu_X}{\mu_{X'}}(x)\mu_{X'}(dx) \simeq \sum_{i=1}^{n_{\text{off}}} w_i \phi(f(X_i'))$$

with $w_i = \frac{\mu_X}{\mu_{X'}}(X_i')$.

Two limits:

- Computing μ_X and $\mu_{X'}$ required.
  We have only $(X_i)_{i\in[\![1,n_{\text{on}}]\!]}$ $(X')_{j\in[\![1,n_{\text{off}}]\!]}$.
  $\Rightarrow$ *Density ratio estimation* [2], various methods developed.

[2] Sugiyama, Masashi, Taiji Suzuki, and Takafumi Kanamori. Density ratio estimation in machine learning. Cambridge University Press, 2012.

First idea (importance weighting):

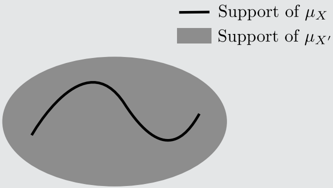
$$\int f(x)\mu_X(dx) = \int f(x)\frac{\mu_X}{\mu_{X'}}(x)\mu_{X'}(dx) \simeq \sum_{i=1}^{n_{\mathrm{off}}} w_i \phi(f(X_i'))$$

with $w_i = \frac{\mu_X}{\mu_{X'}}(X_i')$.

Two limits:

- Computing μ_X and $\mu_{X'}$ required.
  We have only $(X_i)_{i\in[\![1,n_{\mathrm{on}}]\!]}$ $(X')_{j\in[\![1,n_{\mathrm{off}}]\!]}$.
  $\Rightarrow$ *Density ratio estimation* [2], various methods developed.
- μ_X needs a density w.r.t $\mu_{X'}$ (absolute continuity)

Assumption not verified in practice



[2] Sugiyama, Masashi, Taiji Suzuki, and Takafumi Kanamori. Density ratio estimation in machine learning. Cambridge University Press, 2012.

Minimizing the Wasserstein distance

Reinterpretation with empirical measure

Empirical measure on $\boldsymbol{X}_{n_{\mathrm{on}}}$, weighted empirical measure on $\boldsymbol{X}'_{n_{\mathrm{off}}}$

$$\widehat{\mu}_{X,n_{\mathrm{on}}} = \frac{1}{n}\sum_{i=1}^{n_{\mathrm{on}}} \delta_{X_i}, \quad \widehat{\mu}^{\boldsymbol{w}}_{X',n_{\mathrm{off}}} = \sum_{j=1}^{n_{\mathrm{off}}} w_i \delta_{X'_j}$$

with $\sum_{i=1}^{n_{\mathrm{off}}} w_i = 1$.

Minimizing the distance between empirical measures

$$\mathrm{w}^* = \underset{\sum w_i = 1, w_i \geq 0}{\mathrm{argmin}} \quad d\left(\widehat{\mu}_{X,n_{\mathrm{on}}}, \widehat{\mu}^{\boldsymbol{w}}_{X',n_{\mathrm{off}}}\right)$$

Minimizing the Wasserstein distance

Reinterpretation with empirical measure

Empirical measure on $\boldsymbol{X}_{n_{\mathrm{on}}}$, weighted empirical measure on $\boldsymbol{X}'_{n_{\mathrm{off}}}$

$$\widehat{\mu}_{X,n_{\mathrm{on}}} = \frac{1}{n}\sum_{i=1}^{n_{\mathrm{on}}} \delta_{X_i}, \quad \widehat{\mu}^{\boldsymbol{w}}_{X',n_{\mathrm{off}}} = \sum_{j=1}^{n_{\mathrm{off}}} w_i \delta_{X'_j}$$

with $\sum_{i=1}^{n_{\mathrm{off}}} w_i = 1$.

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

Minimizing the distance between empirical measures

$$\mathrm{w}^* = \underset{\sum w_i = 1, w_i \geq 0}{\mathrm{argmin}} \quad d\left(\widehat{\mu}_{X,n_{\mathrm{on}}}, \widehat{\mu}^{\boldsymbol{w}}_{X',n_{\mathrm{off}}}\right)$$

Choice of distance

Wasserstein distances of order q [3]

$\mathcal{P}(\mathbb{R}^d)$ probability measures on $\mathbb{R}^d$

$$\mathcal{P}_q(\mathbb{R}^d) = \left\{ \nu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |x|^q \mathrm{d}\nu(x) < +\infty \right\}.$$

Wasserstein distance

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

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

No need for μ and ν to be absolutely continuous one w.r.t the other.

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

Theoretical results

Explicit expression of the optimal weights

$$\boldsymbol{w}^* = \underset{\sum w_i=1, w_i \geq 0}{\operatorname{argmin}} W_q \left(\widehat{\mu}_{X, n_{\mathrm{on}}}, \widehat{\mu}^{\boldsymbol{w}}_{X', n_{\mathrm{off}}} \right)$$

Optimal weights

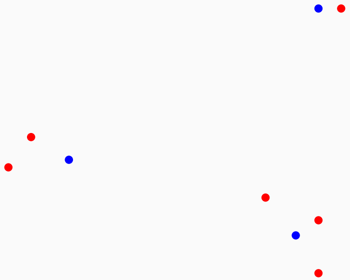
$$w_j^* = \frac{1}{n_{\mathrm{on}}} \sum_{i=1}^{n_{\mathrm{on}}} \mathbb{1}_{\{X'_j = \mathrm{NN}(X_i)\}}$$

Nb of X_i for which X'_j is the NN.

$$\mathrm{NN}(x) = \underset{x' \in \{X'_1, \dots, X'_{n_{\mathrm{off}}}\}}{\operatorname{argmin}} |x - x'|$$

Explicit expression of the optimal weights

Proof (idea)



- X'_j weight w_j, tbd
- X_i weight $1/n_{\mathrm{off}}$

$$\boldsymbol{w}^* = \underset{\sum w_i = 1, w_i \geq 0}{\operatorname{argmin}} W_q \left(\widehat{\mu}_{X, n_{\mathrm{on}}}, \widehat{\mu}^{\boldsymbol{w}}_{X', n_{\mathrm{off}}} \right)$$

Optimal weights

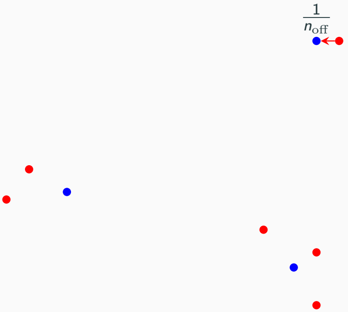
$$w_j^* = \frac{1}{n_{\mathrm{on}}} \sum_{i=1}^{n_{\mathrm{on}}} \mathbb{1}_{\{X'_j = \mathrm{NN}(X_i)\}}$$

Nb of X_i for which X'_j is the NN.

$$\mathrm{NN}(x) = \underset{x' \in \{X'_1, \dots, X'_{n_{\mathrm{off}}}\}}{\operatorname{argmin}} |x - x'|$$

Explicit expression of the optimal weights

Proof (idea)



- X'_j weight w_j, tbd
- X_i weight $1/n_{\mathrm{off}}$

$$\boldsymbol{w}^* = \underset{\sum w_i = 1, w_i \geq 0}{\operatorname{argmin}} W_q \left(\widehat{\mu}_{X, n_{\mathrm{on}}}, \widehat{\mu}^{\boldsymbol{w}}_{X', n_{\mathrm{off}}} \right)$$

Optimal weights

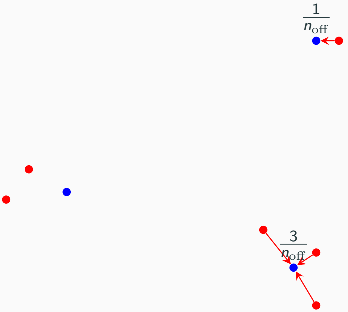
$$w_j^* = \frac{1}{n_{\mathrm{on}}} \sum_{i=1}^{n_{\mathrm{on}}} \mathbb{1}_{\{X'_j = \mathrm{NN}(X_i)\}}$$

Nb of X_i for which X'_j is the NN.

$$\mathrm{NN}(x) = \underset{x' \in \{X'_1, \ldots, X'_{n_{\mathrm{off}}}\}}{\operatorname{argmin}} |x - x'|$$

Explicit expression of the optimal weights

Proof (idea)



- X'_j weight w_j, tbd
- X_i weight $1/n_{\mathrm{off}}$

$$\boldsymbol{w}^* = \underset{\sum w_i = 1, w_i \geq 0}{\mathrm{argmin}} \; W_q\left(\widehat{\mu}_{X, n_{\mathrm{on}}}, \widehat{\mu}^{\boldsymbol{w}}_{X', n_{\mathrm{off}}}\right)$$

Optimal weights

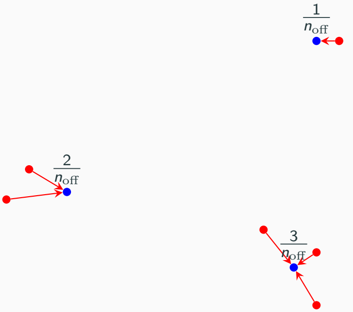
$$w_j^* = \frac{1}{n_{\mathrm{on}}} \sum_{i=1}^{n_{\mathrm{on}}} \mathbb{1}_{\{X'_j = \mathrm{NN}(X_i)\}}$$

Nb of X_i for which X'_j is the NN.

$$\mathrm{NN}(x) = \underset{x' \in \{X'_1, \ldots, X'_{n_{\mathrm{off}}}\}}{\mathrm{argmin}} \; |x - x'|$$

Explicit expression of the optimal weights

Proof (idea)



- X_j' weight w_j, tbd
- X_i weight $1/n_{\text{off}}$

$$\boldsymbol{w}^* = \underset{\sum w_i = 1, w_i \geq 0}{\operatorname{argmin}} W_q\left(\widehat{\mu}_{X,n_{\text{on}}}, \widehat{\mu}^{\boldsymbol{w}}_{X',n_{\text{off}}}\right)$$

Optimal weights

$$w_j^* = \frac{1}{n_{\text{on}}} \sum_{i=1}^{n_{\text{on}}} \mathbb{1}_{\{X_j' = \mathrm{NN}(X_i)\}}$$

Nb of X_i for which X_j' is the NN.

$$\mathrm{NN}(x) = \underset{x' \in \{X_1', \dots, X_{n_{\text{off}}}'\}}{\operatorname{argmin}} |x - x'|$$

Consistency

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

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

uniformly in n_{on}.

A3: Strong support condition [4]

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

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

[4] 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

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

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

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

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

arXiv preprint arXiv:2010.09267 (2020).

Application

Back to our initial problem

$$\mathrm{QI} = \mathbb{E}[f(X)], \quad \widehat{\mathrm{QI}}_{n_{\mathrm{on}},n_{\mathrm{off}}} = \sum_{j=1}^{n_{\mathrm{off}}} w_j^* f(X_j')$$

Rate of convergence

If $\mathbb{E}[|X|^{2+\epsilon}] < +\infty$ for $\epsilon > 0$, $d \geq 2$, $(A2), (A3), (A4)$, f Lipschitz-continuous

$$\mathbb{E}\left[\left|\mathrm{QI} - \widehat{\mathrm{QI}}_{n_{\mathrm{on}},n_{\mathrm{off}}}\right|\right] = O(n_{\mathrm{on}}^{1/d} + n_{\mathrm{off}}^{1/d}).$$

L_q, $q \geq 1$ rates of convergences are given and the case of noisy observations is also handled.

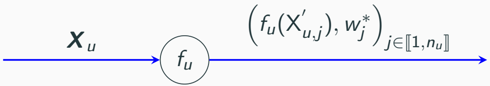
We do not need the absolute continuity of μ_X w.r.t $\mu_{X'}$.

Following steps: propagating the weights through a graph of models

So far, weighting scheme is computed for one node

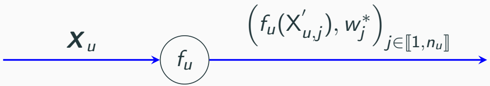
Work in progress.

So far, weighting scheme is computed for one node



Work in progress.

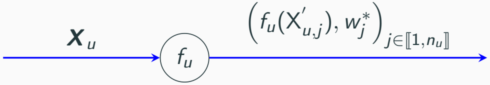
So far, weighting scheme is computed for one node



Propagation through the graph? Propagate weighted samples.

Work in progress.

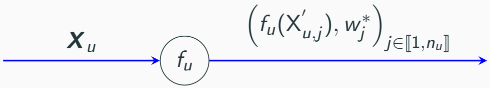
So far, weighting scheme is computed for one node



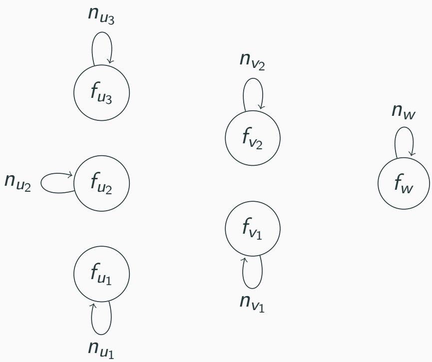
Propagation through the graph? Propagate weighted samples.

Work in progress.

So far, weighting scheme is computed for one node

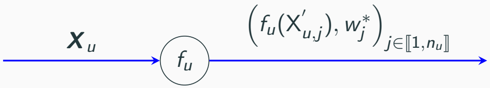


Propagation through the graph? Propagate weighted samples.

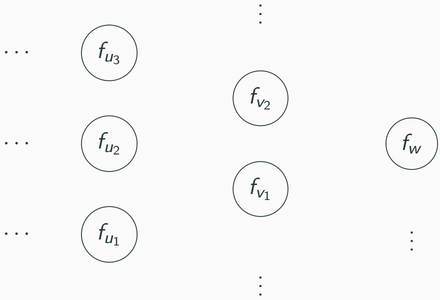


Work in progress.

So far, weighting scheme is computed for one node

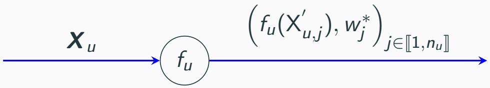


Propagation through the graph? Propagate weighted samples.

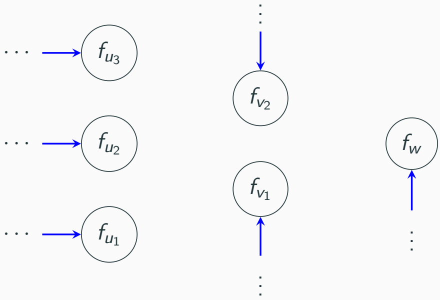


Work in progress.

So far, weighting scheme is computed for one node

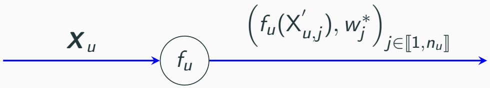


Propagation through the graph? Propagate weighted samples.

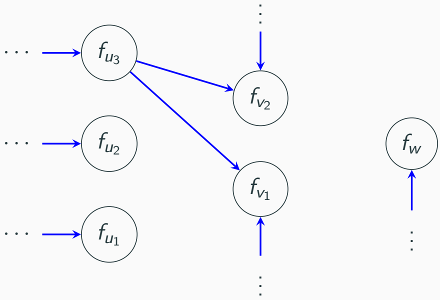


Work in progress.

So far, weighting scheme is computed for one node

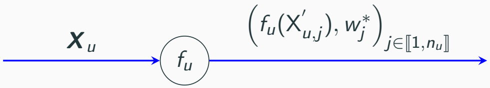


Propagation through the graph? Propagate weighted samples.

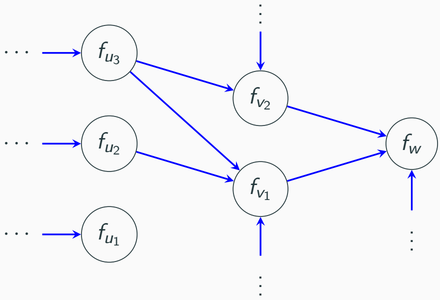


Work in progress.

So far, weighting scheme is computed for one node

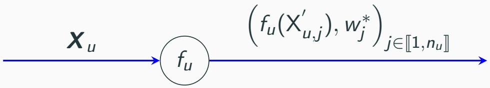


Propagation through the graph? Propagate weighted samples.

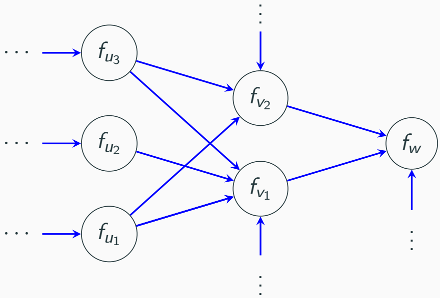


Work in progress.

So far, weighting scheme is computed for one node

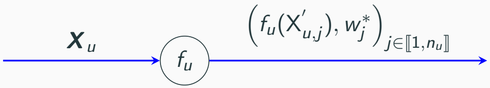


Propagation through the graph? Propagate weighted samples.

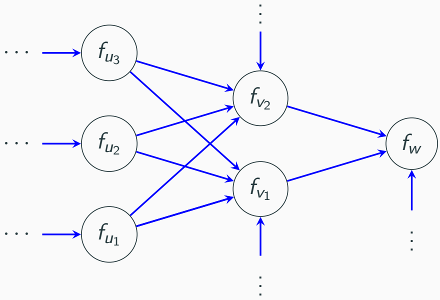


Work in progress.

So far, weighting scheme is computed for one node

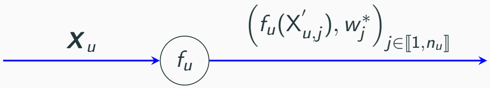


Propagation through the graph? Propagate weighted samples.

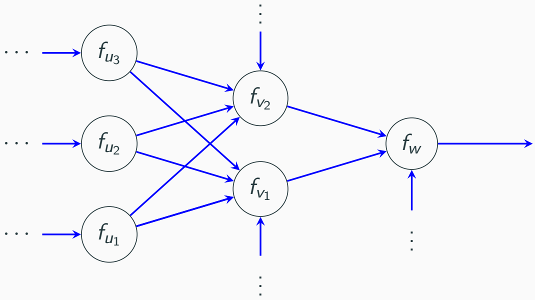


Work in progress.

So far, weighting scheme is computed for one node

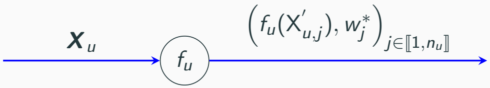


Propagation through the graph? Propagate weighted samples.

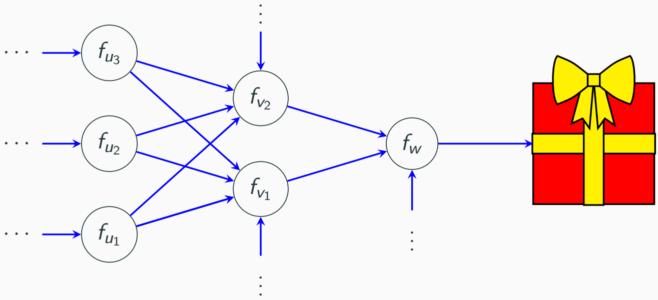


Work in progress.

So far, weighting scheme is computed for one node

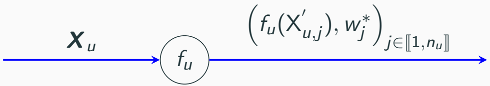


Propagation through the graph? Propagate weighted samples.

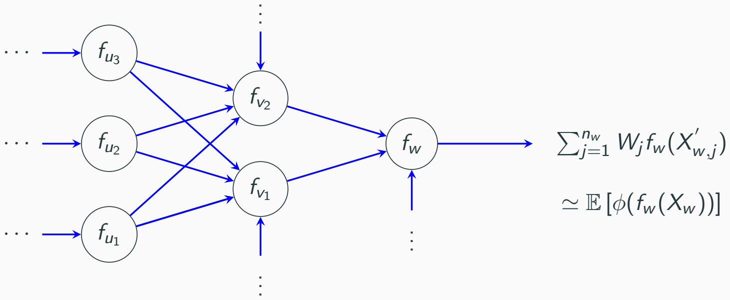


Work in progress.

So far, weighting scheme is computed for one node



Propagation through the graph? Propagate weighted samples.



Work in progress.

Thank you for your attention!

Do you have any question?