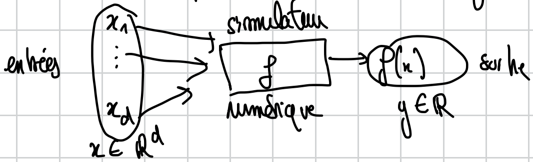
Cours Multi-fidélité.

Dans ce cours on cherche à comprendre comment sont propagées les incertitudes dans les simulateurs numériques.

I- Principe :

1.1 Simulation numérique :

On définit le simulateur par une fonction f qui à partir d'entrées $x \in \mathbb{R}^n$ permet de modéliser un phénomène et qui permet d'avoir un résultat $y \in \mathbb{R}$.



f peut par exemple modéliser un écoulement.

1.2 Simulateurs :

Le problème est que souvent f est très coûteuse.

On suppose qu'on a :

→ un simulateur haute fidélité $f : \mathbb{R}^d \to \mathbb{R}$.

→ un ensemble de simulateurs basse fidélité $\{g_\ell : \mathbb{R}^d \to \mathbb{R}\}_{\ell = 1, \dots, m}$
(ex : réseau de neurones, différents types de maillage).

On va alors chercher à comprendre comment sont propagées les incertitudes dans ces simulateurs.

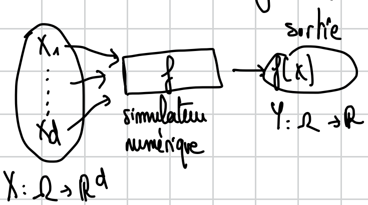
II - Incertitudes :

On se place maintenant dans un contexte probabiliste.

2.1 Incertitude dans le simulateur :

↳ On modélise les paramètres incertains par des variables aléatoires $X = (X_1, \dots, X_d)$

· On va s'intéresser à $Y = f(X)$ qui représente l'incertitude sur la sortie.



2.2 Estimation statistique :

On veut estimer un paramètre θ d'une variable aléatoire Y définie sur $(\Omega, \mathcal{A}, \mathbb{P})$

1) n-échantillons

Un n-échantillon de Y est un n-uplet $(Y_1, \dots, Y_n)$ de variable aléatoire indépendantes et i.i.d de même loi que Y.

2) Echantillon de Monte-Carlo:

Soit $(X_1, \dots, X_n)$ un n-échantillon de X alors $(Y_1, \dots, Y_n)$ avec $Y_i = f(X_i)$ est un-échantillon de Y.

3) Estimateur: Soit $(Y_1, \dots, Y_n)$ un n-échantillon de Y et $\hat{\varphi}_n : \mathbb{R}^n \to \mathbb{R}$ alors

$\hat{\theta}_n = \hat{\varphi}_n (Y_1, \dots, Y_n)$ est un estimateur de θ.

θ est un paramètre inconnu.

4) Estimation: Soit $(y_1, \dots, y_n)$ une réalisation de $(Y_1, \dots, Y_n)$ alors $\hat{\theta}_n (y_1, \dots, y_n)$ est une estimation de θ.

2.3 Qualité d'un estimateur:

1) Biais d'un estimateur:

$$\text{Biais}(\hat{\theta}_n, \theta) = \mathbb{E}[\hat{\theta}_n - \theta] = \mathbb{E}[\hat{\theta}_n] - \theta$$

→ si $\text{Biais}(\hat{\theta}_n, \theta) \neq 0 \Rightarrow \hat{\theta}_n$ estimateur biaisé de θ.

→ sinon il est sans biais.

2) Erreur quadratique moyenne :

- $EQM(\hat{\theta}_n, \theta) = \mathbb{E}[(\hat{\theta}_n - \theta)^2]$
- $REQM(\hat{\theta}_n, \theta) = \sqrt{EQM(\hat{\theta}_n, \theta)}$

3) Décomposition biais-variance:

$$EQM(\hat{\theta}_n, \theta) = \underbrace{\text{Biais}(\hat{\theta}_n, \theta)^2}_{\text{mesure la capacité du modèle à capturer la relation entre les entrées et la sortie}} + \underbrace{V[\hat{\theta}_n]}_{\text{mesure la sensibilité du modèle aux fluctuations des données}}.$$

3) Estimation de l'espérance:

Soit $(Y_1, \dots, Y_n)$ un n-échantillon de Y.

Moyenne empirique: $\bar{Y}_n = \hat{\mu}_n (Y_1, \dots, Y_n) = \frac{1}{n} \sum_{i=1}^{n} Y_i$.

↳ $\mathbb{E}[\bar{Y}_n] = \mu$ (sans biais)

- $V[\bar{Y}_n] = \frac{\sigma^2}{n}$.

Question: Comment combiner les estimateurs ?

III. Méthodes avec variable de contrôle :

3.1 Une variable de contrôle :

Méthode utilisée pour améliorer l'efficacité et la précision des simulations de Monte-Carlo.

On a Y la variable aléatoire dont on veut estimer l'espérance et Z une variable aléatoire auxiliaire (variable de contrôle) dont on connaît l'espérance μ_2.

Il faut que Y et Z soient corrélés.

Estimateur : $\bar{Y}_n^{cv}(\alpha) = \bar{Y}_n - \alpha(\bar{Z}_n - \mu_2)$.

On cherche la valeur de α qui minimise $V[\bar{Y}_n^{cv}(\alpha)]$.

Démo : $V[\bar{Y}_n^{cv}(\alpha)] = V[\bar{Y}_n] + \alpha^2 V[\bar{Z}_n] - 2\alpha\, C[\bar{Y}_n, \bar{Z}_n] = J(\alpha)$.

$$J'(\alpha) = 2\alpha V[\bar{Z}_n] - 2\, C[\bar{Y}_n, \bar{Z}_n].$$

$\hookrightarrow J'(\alpha) = 0 \iff \boxed{\alpha^* = \dfrac{C[\bar{Y}_n, \bar{Z}_n]}{V[\bar{Z}_n]}}$.

$$J(\alpha^*) = V[\bar{Y}_n] + \frac{C^2[\bar{Y}_n, \bar{Z}_n]}{V[\bar{Z}_n]} - 2 \times \frac{C^2[\bar{Y}_n, \bar{Z}_n]}{V[\bar{Z}_n]} < J(\alpha)$$

$\hookrightarrow$ on a diminué le biais

$$J(\alpha^*) = V[\bar{Y}_n] - \frac{C^2[\bar{Y}_n, \bar{Z}_n]}{V[\bar{Z}_n]}$$

$$\Rightarrow V[\bar{Y}_n^{cv}(\alpha^*)] = V[\bar{Y}_n] \times \Big(1 - \underbrace{\frac{C^2[\bar{Y}_n, \bar{Z}_n]}{V[\bar{Y}_n] \cdot V[\bar{Z}_n]}}_{\rho^2}\Big)$$

$$= (1 - \rho^2)\, V[\bar{Y}_n].$$

- paramètre optimal : $\alpha^* = \dfrac{C[\bar{Y}_n, \bar{Z}_n]}{V[\bar{Z}_n]}$
- réduction de la variance : $V[\bar{Y}_n^{cv}(\alpha^*)] = (1 - \rho^2)\, V[\bar{Y}_n]$.

$\hookrightarrow$ on a bien réduit la variance par rapport à l'expression initiale.

3.2. Plusieurs variable de contrôle :

On a maintenant m variables de contrôle $Z^1, \dots, Z^m$.

On s'intéresse à l'estimateur : $\bar{Y}_n^{cv}(\alpha) = \bar{Y}_n - \sum_{k=1}^{m} \alpha_k (\bar{Z}_n^k - \mu^k)$

$$= \bar{Y}_n - (\bar{Z}_n - \mu_Z)^T \alpha.$$

- $\mathbb{C}[X,Y] = \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])^T]$, $X, Y : \Omega \to \mathbb{R}^n$
- $\mathbb{C}[X,Y] = \mathbb{E}[(Y-\mathbb{E}[Y])(X-\mathbb{E}[X])]$, $X : \Omega \to \mathbb{R}^m$, $Y : \Omega \to \mathbb{R}$

On cherche maintenant α_k qui minimise $V[\bar{Y}_n^{cv}(\alpha)]$.

$$V[\bar{Y}_n^{cv}(\alpha)] = V[\bar{Y}_n] + \alpha^T \mathbb{C}[\bar{Z}_n, \bar{Z}_n]\alpha - 2\mathbb{C}[\bar{Z}_n, \bar{Y}_n]^T \alpha.$$

Démo :

$$V[\bar{Y}_n^{cv}(\alpha)] = Var[\bar{Y}_n] - Var\left[\sum_{k=1}^{m} \alpha_k (\bar{Z}_n^k - \mu^k)\right].$$

$$= \text{—} + Var\left(\sum_{k=1}^{m} \alpha_k (\bar{Z}_n^k - \mu_k)\right) - 2\,Cov\left(\bar{Y}_n, \sum_{k=1}^{m} \alpha_k (\bar{Z}_n^k - \mu_k)\right)$$

$$= \text{—} + Var\left(\sum_{k=1}^{m} \alpha_k \bar{Z}_n^k\right) - 2\,Cov\left(\bar{Y}_n, \sum_{k=1}^{m} \alpha_k \bar{Z}_n^k\right)$$

$$= \text{—} + \sum_{k=1}^{m}\sum_{j=1}^{m} \alpha_k \alpha_j\, Cov(\bar{Z}_n^k, \bar{Z}_n^j) - 2\sum_{k=1}^{m} \alpha_k\, Cov(\bar{Y}_n, \bar{Z}_n^k)$$

$$= \text{—} + \alpha^T Cov(\bar{Z}_n, \bar{Z}_n)\alpha - 2\alpha^T Cov(\bar{Y}_n, \bar{Z}_n).$$

- **Paramètre optimal :**

$$\alpha^* = \mathbb{C}[\bar{Z}_n, \bar{Z}_n]^{-1}\mathbb{C}[\bar{Z}_n, \bar{Y}_n].$$

Démo : le minimum est atteint en $\nabla_\alpha V[\bar{Y}_n^{cv}(\alpha)] = 0$

$$\nabla_\alpha V[\bar{Y}_n^{cv}(\alpha)] = \mathbb{C}[\bar{Z}_n, \bar{Z}_n]\alpha + \mathbb{C}[\bar{Z}_n, \bar{Z}_n]^T\alpha - 2\mathbb{C}[\bar{Z}_n, \bar{Y}_n]$$

$\mathbb{C}[\bar{Z}_n, \bar{Z}_n]$ est symétrique donc :

$$\begin{cases} \nabla_\alpha V[\bar{Y}_n^{cv}(\alpha)] = 0 \\ \Sigma \text{ inversible} \end{cases} \Rightarrow \mathbb{C}[\bar{Z}_n, \bar{Z}_n]\,\alpha^* = \mathbb{C}[\bar{Z}_n, \bar{Y}_n] \Rightarrow \alpha^* = \mathbb{C}[\bar{Z}_n, \bar{Z}_n]^{-1}\mathbb{C}[\bar{Z}_n, \bar{Y}_n]$$

- **Réduction de la variance :**

$$V[\bar{Y}_n^{cv}(\alpha)] = V[\bar{Y}_n] + \alpha^T \mathbb{C}[\bar{Z}_n, \bar{Z}_n]\alpha - 2\mathbb{C}[\bar{Z}_n, \bar{Y}_n]^T \alpha.$$

$$V[\bar{Y}_n^{cv}(\alpha^*)] = (1 - R^2)V[\bar{Y}_n] \quad \text{avec } R^2 = \frac{\mathbb{C}[\bar{Z}_n, \bar{Y}_n]^T \mathbb{C}[\bar{Z}_n, \bar{Z}_n]^{-1}\mathbb{C}[\bar{Z}_n, \bar{Y}_n]}{V[\bar{Y}_n]}.$$

Démo :

$$V[\bar{Y}_n^{cv}(\alpha^*)] = V[\bar{Y}_n] + \mathbb{C}[\bar{Z}_n, \bar{Z}_n]^{-1} \times \mathbb{C}[\bar{Z}_n, \bar{Y}_n] \times \cancel{\mathbb{C}[\bar{Z}_n, \bar{Z}_n]} \times \cancel{\mathbb{C}[\bar{Z}_n, \bar{Z}_n]^{-1}} \times \mathbb{C}[\bar{Z}_n, \bar{Y}_n]$$
$$- 2\,\mathbb{C}[\bar{Z}_n, \bar{Y}_n]^T \times \mathbb{C}[\bar{Z}_n, \bar{Z}_n]^{-1}\mathbb{C}[\bar{Z}_n, \bar{Y}_n].$$

IV- Couplage:

Lien avec les modèles de haute/basse fidélité:

On pose $Y = f(x)$ (le modèle de haute fidélité).
$Z^k = g_k(x)$ (le modèle de basse fidélité).

On a alors: $\mathbb{C}[\bar{Y}_n, \bar{Z}_n^k] = \frac{1}{n}\mathbb{C}[Y, Z^k]$ et $\mathbb{C}[\bar{Z}_n^k, \bar{Z}_n^l] = \frac{1}{n}\mathbb{C}[Z^k, Z^l]$.

- Paramètre optimal + réduction de variance:

$\mathrm{Corr}[X,X] = \mathrm{Var}[X]^{-1/2}\mathbb{C}[X,X]\mathrm{Var}[X]^{-1/2}$

- $\alpha^* = \mathbb{C}[Z,Z]^{-1} \times \mathbb{C}[Z,Y]$.

- $R^2 = \dfrac{\mathbb{C}[Z,Y]^T\mathbb{C}[Z,Z]^{-1}\mathbb{C}[Z,Y]}{V[Y]} = \mathrm{Corr}[Z,Y]^T \times \mathrm{Corr}[Z,Z]^{-1} \times \mathrm{Corr}[Z,Y]$.

(Si $m=1$ (une var de contrôle)) : $\alpha^* = \dfrac{\mathbb{C}[Y,Z]}{V[Z]}$ et $R^2 = \dfrac{\mathbb{C}[Y,Z]^2}{V[Y]V[Z]} = \mathrm{Corr}[Y,Z]^2 = \rho^2$.

V- Ajout de variable de contrôle:

inversion par bloc:

Soit M la matrice par bloc définie par: $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$.
Si A et M inversibles alors $S = D - CA^{-1}B$ est inversible et:

$$M^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BS^{-1}CA^{-1} & -A^{-1}BS^{-1} \\ -S^{-1}CA^{-1} & S^{-1} \end{bmatrix}.$$

- **Théorème:** Soit $R_+ = \begin{bmatrix} R & u \\ u^T & 1 \end{bmatrix}$ et $r_+ = \begin{bmatrix} r \\ b \end{bmatrix}$ qui correspondent à l'ajout d'une variable Z^{m+1}.

On suppose que R et R_+ sont inversibles :

$$R_+ = R^2 + \frac{(b - u^T R^{-1} r)^2}{1 - u^T R^{-1} u} \geq R^2.$$

et $R_+^2 = R^2 \Leftrightarrow b = u^T R^{-1} r.$

Démonstration:

$R = Corr[Z, Z]$ $\quad u = Corr[Z, Z^{m+1}]$. $\quad r = Corr[Y, Z^{m+1}]$

$R_+ = Corr[Z_+, Z_+]$ $\quad r_+ = Corr[Y, Z_+]$.

$$R_+^2 = r_+^T R_+^{-1} r_+.$$

$$= \underbrace{(r \quad b)}_{(1 \times n)} \times \underbrace{\begin{pmatrix} R^{-1} + R^{-1}US^{-1}U^TR^{-1} & -R^{-1}U^TS^{-1} \\ -S^{-1}U^TR^{-1} & S^{-1} \end{pmatrix}}_{(2\times 2)} \times \begin{pmatrix} r \\ b \end{pmatrix}$$

$$= \left(r(R^{-1} + R^{-1}US^{-1}U^TR^{-1}) - bS^{-1}U^TR^{-1} \quad -rR^{-1}U^TS^{-1} + bS^{-1} \right) \times \begin{pmatrix} r \\ b \end{pmatrix}.$$

$$= \left(r(R^{-1} + R^{-1}US^{-1}U^TR^{-1}) - bS^{-1}U^TR^{-1} \right) r + b\left(-rR^{-1}U^TS^{-1} + bS^{-1}\right).$$

$$= r^T(R^{-1} + R^{-1}US^{-1}U^TR^{-1})r - bS^{-1}U^TR^{-1}r - brR^{-1}U^TS^{-1} + b^2S^{-1}$$

$$= R^2 + r^TR^{-1}US^{-1}U^TR^{-1}r - r^TR^{-1}U^TS^{-1}b - bS^{-1}U^TR^{-1}r + bS^{-1}b.$$

$$= R^2 + \beta^2S^{-1} - 2b\beta S^{-1} + b^2S^{-1}$$

$$= R^2 + (\beta^2 - 2b\beta + b^2)S^{-1}$$

$$= R^2 + (b - \beta)^2 S^{-1}$$

$$= R^2 + (b - u^T R^{-1} r)^2 S^{-1}.$$

R^+ est inversible, R est symétrique définie positive : $\forall v,\ v^T R_+ v > 0$.

$v_+ = \begin{bmatrix} v \\ \beta \end{bmatrix}$ $\qquad [v^T \quad \beta] \times \begin{bmatrix} R & u \\ u^T & 1 \end{bmatrix} \times \begin{bmatrix} v \\ \beta \end{bmatrix} = 1 - u^T R^{-1} u.$

$[v^TR + \beta u^T]\, v + (v^Tu + \beta)\beta = 1 - u^TR^{-1}u$ $\qquad u^TR^{-1}RR^{-1}u - u^TR^{-1}u - u^TRu = -u^TR^{-1}u$

$v^TRv + \beta u^Tv + v^Tu\beta + \beta^2 = 1 - u^TR^{-1}u$ $\qquad u^TR^{-1}u - u^TR^{-1}u - u^TR^{-1}u = -u^TR^{-1}u.$

$\begin{cases} \beta = -1 \\ v^TRv - u^Tv - v^Tu - 1 = 1 - u^TR^{-1}u \end{cases}$ $\qquad v_+ \neq 0 \mid q \quad v_+ R_+^{-1} v_+ > 0$

VI - Questions pour l'application à la simulation numérique :

1) Quand la méthode est-elle rentable ?

On a l'estimateur $\varphi_n^{cv}(\alpha^*) = \overline{\varphi_n} - \alpha^*(\overline{Z}_n - \mu_Z)$.

Coût de l'estimateur :

- Soit $C_f(X)$ le coût (temps CPU) d'évaluer $f(X)$ et $c_f = \mathbb{E}[C_f(X)]$.
- Soit $C_g(X)$ le coût (temps CPU) d'évaluer $g(X)$ et $c_g = \mathbb{E}[C_g(X)] < c_f$.

On note $\omega = \frac{c_g}{c_f} > 1$.

Le coût total moyen est $c = n(c_f + c_g)$ et on note $\tilde{n} = \frac{c}{c_f} = n(1+\omega)$.
(= nombre d'échantillons haute fidélité).

- Variance de l'estimateur :

$$V[\overline{\varphi}_n^{cv}(\alpha^*)] = V[\overline{\varphi}_{\tilde{n}}](1+\omega)(1-\rho^2)$$

Réduction de variance $\Leftrightarrow$ $\rho^2 > \frac{\omega}{1+\omega}$.

2) Comment approximer α^* ?

1) Estimateur indépendant à l'aide d'échantillons pilotes :

On a $(\tilde{X}_1, \dots, \tilde{X}_p)$ un p-échantillon indépendant de $(X_1, \dots, X_n)$, $\tilde{Y}_i = f(\tilde{X}_i)$ et $\tilde{Z}_i = g(\tilde{X}_i)$.

Estimateur : $$\hat{\alpha}^* = \frac{\widehat{Cov}_p(\tilde{Y}_1, \dots, \tilde{Y}_p ; \tilde{Z}_1, \dots, \tilde{Z}_p)}{\widehat{Var}_p(\tilde{Z}_1, \dots, \tilde{Z}_p)}.$$

↳ $\bar{Y}_n^{cv}(\hat{\alpha}^*)$ non biaisé.

2) Estimateur avec les mêmes échantillons $\bar{Y}_n$ et $\bar{Z}_n$:

$$\hat{\alpha}^* = \frac{\widehat{Cov}_n(Y_1, \dots, Y_n ; Z_1, \dots, Z_n)}{Var(Z_1, \dots, Z_n)}.$$

↳ $\bar{Y}_n^{cv}(\hat{\alpha}^*)$ est biaisé.

↳ Dans les 2 cas on a une détérioration de la réduction de la variance.