

# Méthodes de Monte Carlo avancées

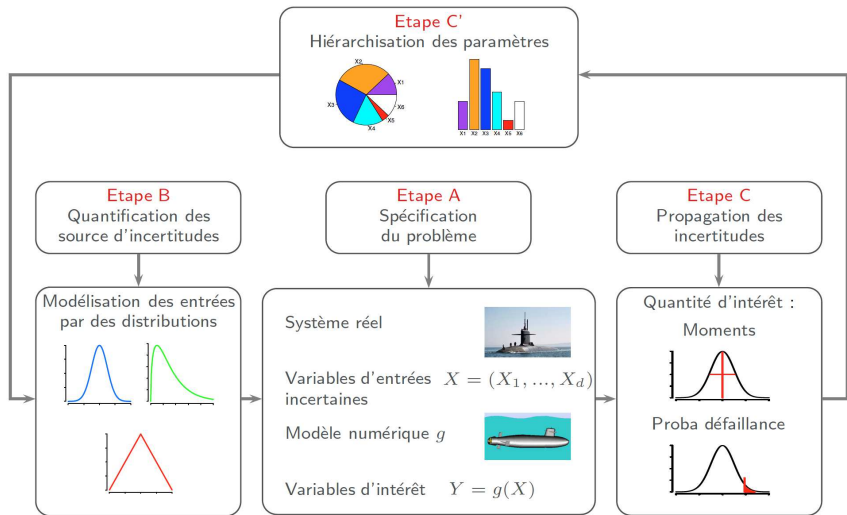
**G. Perrin**

guillaume.perrin@univ-eiffel.fr

Année 2022-2023



# Introduction - Schéma général



# Plan de la séance

- 1 Méthode de Monte Carlo - rappels
- 2 Méthodes de réduction de variance
- 3 Optimisation de plans d'expériences

# Monte Carlo - Objectif

La méthode de Monte Carlo permet le **calcul d'une intégrale d-dimensionnelle**  $I$  à partir de son **interprétation sous la forme d'une espérance** :

$$I = \mathbb{E}(\psi(Y)) = \mathbb{E}(\psi(g(X))) = \int_{\mathbb{R}^d} \psi(g(x)) f_X(x) dx$$

où  $Y = g(X)$  et  $f_X$  est la distribution jointe du vecteur  $X$ .

## Quelques exemples

- $\psi(Y) = Y$  ( $\leftrightarrow$  calcul de la moyenne de  $Y$ ),
- $\psi(Y) = (Y - \mathbb{E}(Y))^2$  ( $\leftrightarrow$  calcul de la variance de  $Y$ ),
- $\psi(Y) = \mathbf{1}_{[g_0, +\infty[}(Y)$  ( $\leftrightarrow$  calcul de la probabilité de dépasser le seuil  $g_0$ ).

# Monte Carlo - principe

- On cherche à estimer :

$$I = \mathbb{E}(\psi(Y)) = \mathbb{E}(\psi(g(X))) = \int_{\mathbb{R}^d} \psi(g(x)) f_X(x) dx$$

- On tire un  $n$ -échantillon de réalisations indépendantes de  $X$  de pdf jointe  $f_X$  :

$$(X^{(i)})_{i=1, \dots, n} = (X_1^{(i)}, \dots, X_d^{(i)})_{i=1, \dots, n}$$

- Estimateur Monte Carlo** de  $I$  :  $\hat{I}_n = \frac{1}{n} \sum_{i=1}^n \psi(g(X^{(i)}))$ .

## Outil universel !

Méthode générale de propagation des incertitudes, applicable à toutes quantités d'intérêt (moments, probabilité de dépasser un seuil) ... à un coût de calcul plus ou moins abordable.

# Monte Carlo - Propriétés de l'estimateur

Estimateur Monte Carlo de  $I$  :  $\hat{I}_n = \frac{1}{n} \sum_{i=1}^n g(X^{(i)})$ .

- Estimateur **non biaisé** :  $\mathbb{E}(\hat{I}_n) = I \quad \forall n$
- Convergence :  $\hat{I}_n \xrightarrow{n \rightarrow \infty} I$  avec probabilité 1 (**loi des grands nombres**)
- Variance de l'estimateur  $\hat{I}_n$  (i.e. **précision**) :

$$\mathbb{V}\text{ar}(\hat{I}_n) = \mathbb{E} \left( (\hat{I}_n - I)^2 \right) = \frac{1}{n} \mathbb{V}\text{ar}(g(X))$$

- Théorème central Limite (TCL) :

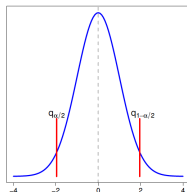
$$\sqrt{\frac{n}{\mathbb{V}\text{ar}(g(X))}} (\hat{I}_n - I) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

# Monte Carlo - Propriétés de l'estimateur

## Intervalle de confiance asymptotique :

$$\mathbb{P} \left( q_{\alpha/2} \leq \sqrt{\frac{n}{\widehat{\text{Var}}(g(X))}} \left( \hat{\text{I}}_n - \text{I} \right) \leq q_{1-\alpha/2} \right) = 1 - \alpha,$$

avec  $\widehat{\text{Var}}(g(X))$  un estimateur de  $\text{Var}(g(X))$ .



$$q_{95\%} \simeq 1.64 \mid q_{97.5\%} \simeq 1.96 \mid q_{99.5\%} \simeq 2.58$$

Attention ! Il n'existe pas d'intervalles contenant I avec une probabilité égale à 1, mais des IC contenant I avec une probabilité proche de 1.

# Monte Carlo accéléré

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^n g(X^{(i)}), \quad \text{Var}(\hat{I}_n) = \frac{1}{n} \text{Var}(g(X)).$$

Objectif : **réduire la variance** de l'estimateur de MC classique.

- Les méthodes dites de **réduction de variance** ont pour but de réduire la constante, en restant proches de l'esprit Monte Carlo.
- Les méthodes de **quasi-Monte Carlo** ont pour but de changer le  $1/n$ .



# Plan de la séance

- 1 Méthode de Monte Carlo - rappels
- 2 Méthodes de réduction de variance
- 3 Optimisation de plans d'expériences

# Méthodes de réduction de variance - Quelques exemples

## Principe de base

Ces méthodes sont toutes basées sur une réécriture judicieuse de l'intégrale  $d$ -dimensionnelle  $I$  (en apportant de l'information supplémentaire).

- **Monte Carlo conditionnel** : on décompose  $X$  en deux sous-vecteurs à partir desquels on peut introduire une espérance conditionnelle sachant l'autre et réduire la dimension de l'intégrale.
- **Tirage d'importance** : on échantillonne  $X$  suivant une densité *biaisée* favorisant les tirages dans la zone d'importance.
- **Variable de contrôle** : on utilise un modèle réduit que l'on connaît parfaitement, au maximum.
- **Statification** : on force l'échantillon à respecter exactement des proportions définies dans chaque strate d'un espace.
- ...

# Monte Carlo conditionnel (ou réduction de la dimension)

- Décomposition (si possible) de  $X$  en deux sous vecteurs  $X_1$  et  $X_2$  :

$$\begin{aligned} I &= \int \int g(x_1, x_2) f_X(x_1, x_2) dx_1 dx_2 \\ &= \int \mathbb{E}_{X_2} [g(X_1, X_2) | X_1 = x_1] f_{X_1}(x_1) dx_1, \end{aligned}$$

## Estimation

- $\hat{I}_n^{cond} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{X_2} [g(X_1^{(i)}, X_2) | X_1^{(i)}], X_1^{(i)} \text{ v.a.i.id } \sim f_{X_1}$
- $\mathbb{V}\text{ar}(\hat{I}_n^{cond}) = \frac{1}{n} \mathbb{V}\text{ar}(\mathbb{E}_{X_2} [g(X_1, X_2) | X_1])$

## Résultats

- Une réduction de variance garantie :  $\mathbb{V}\text{ar}(\hat{I}_n^{cond}) \leq \mathbb{V}\text{ar}(\hat{I}_n)$
- Facile à mettre en œuvre si une variable de fonction de répartition connue (ou un groupe de variables) est indépendante des autres variables.

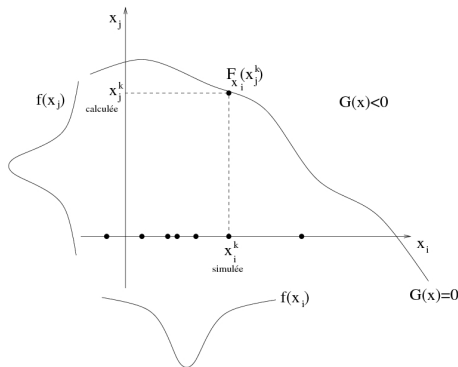
# Monte Carlo conditionnel - illustration en fiabilité

Isolement d'une variable de suppression de la fonction indicatrice :

$$\begin{aligned}
 P &= \mathbb{P}_{X,Z}(g(X) \geq Z) \\
 &= \int_{\mathcal{X}} \int_{\mathcal{Z}} \mathbf{1}_{Z \leq g(x)}(x, z) dz dx \\
 &= \int_{\mathcal{X}} \mathbb{P}_Z(Z \leq g(X)) dx \\
 &= \mathbb{E}_X(F_Z(g(X)))
 \end{aligned}$$

$$\text{D'où : } \hat{P} = \frac{1}{n} \sum_{i=1}^n F_Z(g(X^{(i)}))$$

où  $X^{(i)} \sim f_X$  et  $F_Z$  cdf de  $Z$ .



# Tirage d'importance (ou échantillonnage préférentiel)

- **Principe** : lorsque l'on sait que  $g(X)$  est surtout sensible à certaines valeurs de  $X$ , au lieu de tirer les  $X^{(i)}$  selon leur densité originale  $f_X(x)$ , on les tire selon une densité "biaisée"  $\tilde{f}_X(x)$  qui favorise les valeurs de  $X$  dans la zone d'importance :  $I = \mathbb{E}_{f_X} [g(X)] = \int g(x) f_X(x) dx = \int g(x) \frac{f_X(x)}{\tilde{f}_X(x)} \tilde{f}_X(x) dx = \mathbb{E}_{\tilde{f}_X} \left[ g(X) \frac{f_X(X)}{\tilde{f}_X(X)} \right]$
- **Estimateur** : avec  $(X^{(i)})_{i=1, \dots, n}$  tiré selon  $\tilde{f}_X$ ,

$$\hat{I}_n^{\text{TI}} = \frac{1}{n} \sum_{i=1}^n g(X^{(i)}) \frac{f_X(X^{(i)})}{\tilde{f}_X(X^{(i)})}$$

$$\mathbb{V}\text{ar} \left( \hat{I}_n^{\text{TI}} \right) = \frac{1}{n} \mathbb{V}\text{ar}_{\tilde{f}_X} \left( g(X) \frac{f_X(X)}{\tilde{f}_X(X)} \right)$$

- **Résultats** : la réduction de variance (non garantie) dépend du choix de la densité d'importance  $\tilde{f}_X$ .

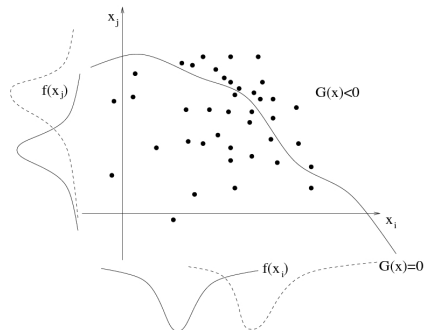
# Tirage d'importance - Propriétés

- Si  $\text{Supp}(f_X) \subset \text{Supp}(\tilde{f}_X)$ , l'estimateur  $\hat{I}_n^{TI}$  est sans biais.
- $\tilde{f}_X$  facile à manipuler :  $X \sim \tilde{f}_X$  simple à générer et  $\frac{f_X}{\tilde{f}_X}$  facile à calculer.
- Un mauvais choix de  $\tilde{f}_X$  peut augmenter la variance
- La densité optimale est :  $f^*(x) = \frac{|g(x)|f_X(x)}{\int |g(y)|f_X(y)dy}$ , où  $\int |g(y)|f_X(y)dy$  est malheureusement aussi difficile à évaluer que I...

⇒ Techniques adaptatives afin que  $\tilde{f}_X$  s'approche de  $f^*$ .

# Tirage d'importance - illustration en fiabilité

- Favoriser les tirages dans le domaine de défaillance.
- Pousser vers les défaillances, mais sans trop forcer car dans ce cas la variance de l'estimateur deviendrait trop importante...



⇒ Utiliser d'autres lois que les lois initiales afin de concentrer les tirages dans les régions de l'espace les plus intéressantes.

# Stratification (ou échantillonnage stratifié)

- Partition du support de  $X$  :  $\mathcal{X} = \bigcup_{i=1}^m \mathcal{X}_i$ ,  $\mathcal{X}_i \cap \mathcal{X}_j = \emptyset$ ,  $i \neq j$
- On force l'échantillon à respecter exactement les proportions théoriques dans certaines strates.
- Réécriture de  $I$  (formule des probabilités totale) :

$$I = \mathbb{E}(g(X)) = \sum_{i=1}^m \underbrace{\mathbb{E}(g(X)|X \in \mathcal{X}_i)}_{J_i} \underbrace{P(X \in \mathcal{X}_i)}_{p_i}$$

- $J_i = \int_{\mathbb{R}^d} \mathbf{1}_{x \in \mathcal{X}_i} \frac{g(x)}{p_i} f_X(x) dx$  estimée par Monte Carlo avec  $n_i$  simulations :  $\hat{J}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} g(X_{(i)}^{(j)})$   $X_{(i)} \sim \mathcal{L}(X|X \in \mathcal{X}_i)$
- $I : \hat{I}_n^S = \sum_{i=1}^m p_i \hat{J}_i$ ,  $\mathbb{V}\text{ar}(\hat{I}_n^S) = \sum_{i=1}^m \frac{p_i^2}{n_i} \sigma_i^2$ ,  $\sigma_i^2 = \mathbb{V}\text{ar}(g(X)|X \in \mathcal{X}_i)$ .



# Stratification

- La réduction (non garantie) de variance (par rapport à MC) dépend du choix des strates et du nombre de tirages dans chacune d'entre elles.

Choix des allocation  $n_i$  ( $n = \sum_{i=1}^m n_i$ )

- Stratification proportionnelle  $n_i = np_i : \mathbb{V}\text{ar} \left( \hat{I}_n^{SP} \right) \leq \mathbb{V}\text{ar} \left( \hat{I}_n^{MC} \right)$ .
- Stratification optimale

Trouver  $(n_i^*)_{i=1, \dots, m}$  qui minimise  $\sum_{i=1}^m \frac{p_i^2}{n_i} \sigma_i^2$  avec  $n = \sum_{i=1}^m n_i$

La solution est donnée par :  $n_i^* = n \frac{p_i \sigma_i}{\sum_{j=1}^m p_j \sigma_j}$

On a alors :  $\mathbb{V}\text{ar} \left( \hat{I}_n^{SO} \right) \leq \mathbb{V}\text{ar} \left( \hat{I}_n^{SP} \right) \leq \mathbb{V}\text{ar} \left( \hat{I}_n^{MC} \right)$ .

En pratique,  $\sigma_i$  n'est pas connu et donc  $(n_i^*)_{i=1, \dots, m}$  incalculables

→ **techniques adaptatives.**

## Méthodes de réduction de variance

# Variables de contrôle - Principe

- Utilisation dans le cas où on dispose d'un modèle réduit  $g_r$  de  $g$ .
- On suppose que l'on connaît  $I_r = \mathbb{E}[g_r(X)]$
- $I$  peut se réécrire ( $b > 0$ ) :  $I = \mathbb{E}(g(X) - bg_r(X)) + b\mathbb{E}(g_r(X))$ .

## Estimation

- Estimateur de  $I$  :  $\hat{I}_n^{\text{VC}} = \frac{1}{n} \sum_{i=1}^n \left( g(X^{(i)}) - bg_r(X^{(i)}) \right) + bI_r$
- Et sa variance :  $\mathbb{V}\text{ar} \left( \hat{I}_n^{\text{VC}} \right) = \frac{1}{n} \mathbb{V}\text{ar} (g(X) - bg_r(X))$

## Resultats

- Peut beaucoup réduire la variance.
- Le  $b$  optimal est donné par :  $b^* = \frac{\text{cov}(g(X), g_r(X))}{\mathbb{V}\text{ar}(g_r(X))}$

Dans ce cas, on a :  $\mathbb{V}\text{ar} \left( \hat{I}_n^{\text{VC}} \right) = (1 - \rho_{gg_r}^2) \mathbb{V}\text{ar} \left( \hat{I}_n \right)$ .

où  $\rho_{gg_r}$  est le coefficient de corrélation entre  $g$  et  $g_r$ .

# Synthèse des méthodes de réduction de variance

- Les méthodes, complémentaires plus que concurrentielles, peuvent être **combinées** entre-elles (en particulier pour le MC conditionnel).
- Les méthodes de Monte Carlo peuvent être très gourmandes en nombre de simulations.
- Le gain en nombre de simulations **n'est pas toujours garanti** (stratification, tirage d'importance).
- Ces techniques nécessitent un choix réfléchi des **paramètres** les régissant (le choix des strates, la densité d'importance).
- Le calcul des **variances des estimateurs** de  $I$  est indispensable.

# Plan de la séance

- 1 Méthode de Monte Carlo - rappels
- 2 Méthodes de réduction de variance
- 3 Optimisation de plans d'expériences

# Monte Carlo accéléré

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^n g(X^{(i)}), \quad \mathbb{V}\text{ar}(\hat{I}_n) = \frac{1}{n} \mathbb{V}\text{ar}(g(X)).$$

Objectif : **réduire la variance de l'estimateur** de MC classique.

- Les méthodes dites de réduction de variance ont pour but de réduire la constante, en restant proches de l'esprit Monte Carlo.
- Les méthodes de **quasi-Monte Carlo** ont pour but de changer le  $1/n$ .

# Méthodes de quasi-Monte Carlo - principe

- On cherche à calculer :  $I = \mathbb{E}(g(X)) = \int_{[0,1]^d} g(x) dx.$
- Estimateur **quasi-Monte Carlo** de  $I$  :  $\tilde{I}_n = \frac{1}{n} \sum_{i=1}^n g(\tilde{X}^{(i)})$

avec  $D_n = (\tilde{X}^{(i)})_{i=1,\dots,n}$  un échantillon **non**-aléatoire de points tirés selon une **suite à discrépance faible**.

- Propriété générale (**inégalité de Koksma-Hlawka**) :

$$\varepsilon := \left| I - \tilde{I}_n \right| \leq V(g) \times \text{discrepance}(D_n),$$

avec  $V(g)$  une constante contrôlant les variations de  $g$  (borne de Hardy-Krause) et  $\text{discrepance}(D_n)$  la **discrépance** de  $D_n$ .

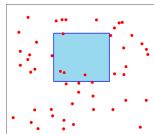
- Avec une suite à discrépance faible :  $\varepsilon = O\left(\frac{(\ln n)^d}{n}\right).$

# Méthodes de quasi-Monte Carlo - discrédance

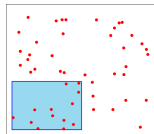
- La **discrédance** est un critère statistique qui mesure la déviation maximale entre la répartition des points de l'échantillon et une répartition uniforme.
- Interprétation géométrique : comparaison entre le volume de sous-domaines et le nombre de points contenus dans ces domaines.

$$Q(t) \subset \mathcal{X} = [0, 1]^d, \quad Q(t) = [0, t_1[ \times [0, t_2[ \times \cdots \times [0, t_d[$$

$$\text{discrepancy}(\text{plan}) = \sup_{Q(t) \in [0, 1]^d} \left| \frac{n_{Q(t)}}{n} - \prod_{i=1}^d t_i \right|$$



- Disprédance faible : bonne répartition uniforme des points dans l'espace
- En pratique : on choisit une discrédance avec une norme  $L^2$  pour avoir une formule **analytique**.

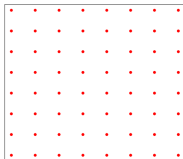




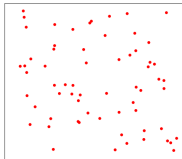
# Suites à discrépances faibles

- Il existe de nombreuses suites à faible discrépance (en général déterministes) : suite de Halton, suite de Van der Corput, suite de Sobol ...
- On parle de méthodes de **quasi-Monte-Carlo**  $\Rightarrow$  "bon" remplissage de l'espace, rapide à construire, flexible (séquentialité)...
- Mêmes cadres d'utilisation que les méthodes de Monte Carlo (calcul d'intégrale) pour accélérer la convergence.

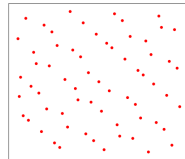
Grille régulière



Monte Carlo



Suite de Sobol



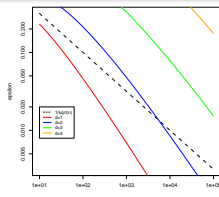
# Suites à discrécances faibles - réduction de la variance

## Monte Carlo

- $\hat{I}_n = \frac{1}{n} \sum_{i=1}^n g(X^{(i)})$ ,  $(X^{(i)})_{i=1,\dots,n}$  échantillon **aléatoire (i.i.d)**.
- Erreur d'estimation :  $\varepsilon = O\left(\frac{1}{\sqrt{n}}\right)$  ( $\text{Var}(\hat{I}_n) = \frac{1}{n} \text{Var}(g(X))$ ).

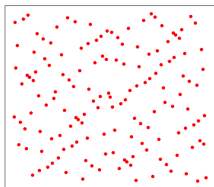
## quasi-Monte Carlo

- $\tilde{I}_n = \frac{1}{n} \sum_{i=1}^n g(\tilde{X}^{(i)})$ ,  $(\tilde{X}^{(i)})_{i=1,\dots,n}$  issu d'une suite à **faible discrécance**.
- Erreur d'estimation :  $\varepsilon = O\left(\frac{(\ln n)^d}{n}\right)$ .

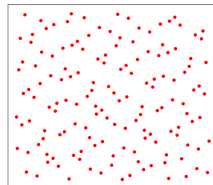


# Suites à discrépances faibles - Exemple et pathologies

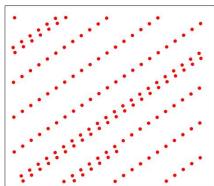
Suite de Sobol ( $n = 150$ ,  $d = 2$ )



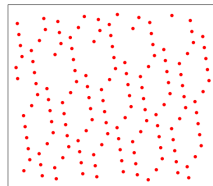
Suite de Halton ( $n = 150$ ,  $d = 2$ )



Projection 2D Sobol ( $n = 150$ ,  $d = 10$ )



Projection 2D Halton ( $n = 150$ ,  $d = 10$ )



# Plan de la séance

- 1 Méthode de Monte Carlo - rappels
- 2 Méthodes de réduction de variance
- 3 Optimisation de plans d'expériences