

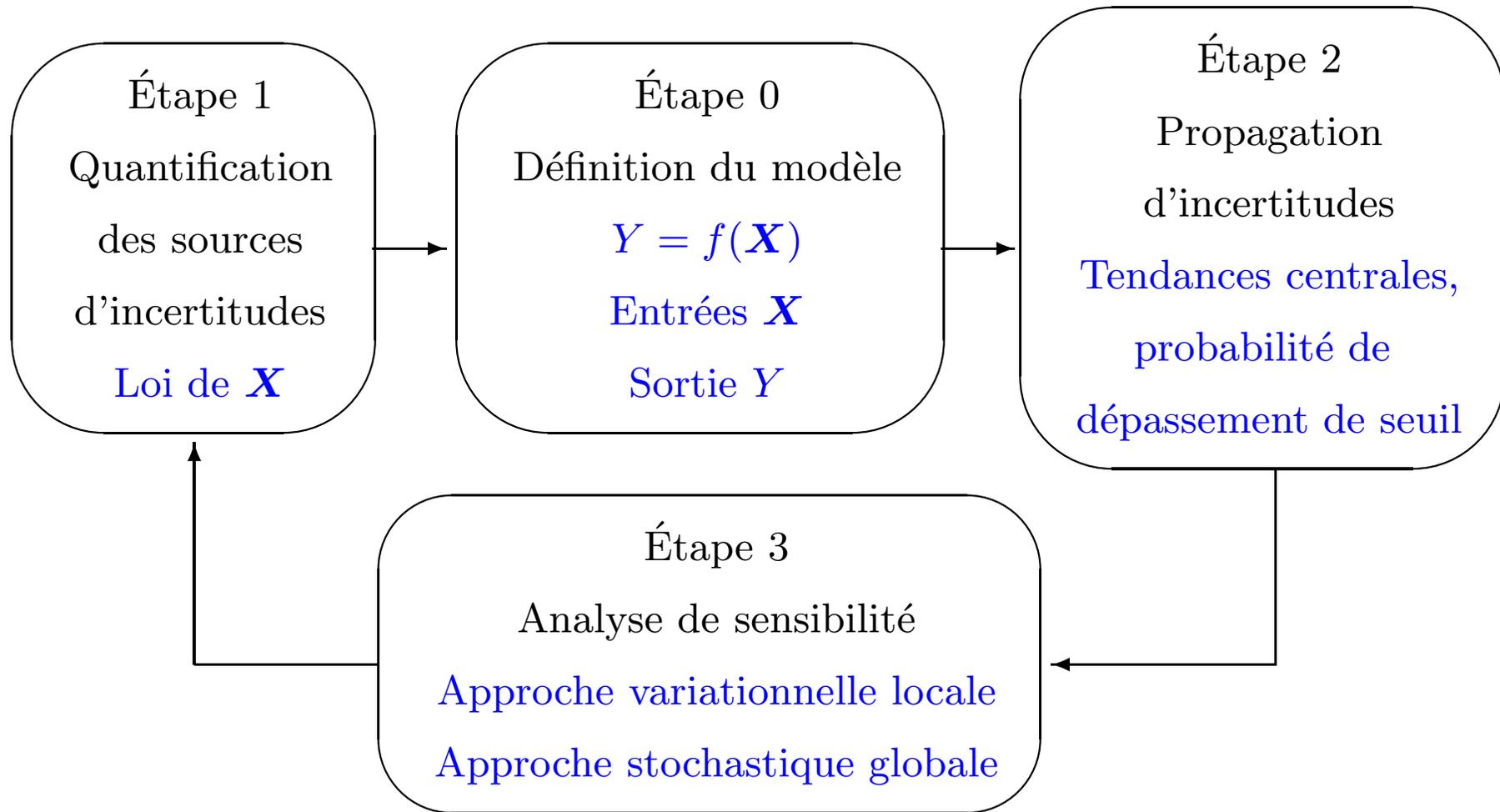
Modélisation et simulation d'événements rares

Josselin Garnier (Université Paris Diderot)

Traitement des incertitudes

En anglais : uncertainty quantification.

- Problème général :
- Comment modéliser les incertitudes et leur propagation dans les modèles physiques ou numériques ?
- Comment estimer (quantifier) la dispersion de la sortie d'un code ou d'une expérience en fonction de la dispersion des paramètres d'entrée ?
- Comment estimer (quantifier) la sensibilité de la sortie d'un code ou d'une expérience vis-à-vis d'un paramètre d'entrée particulier ?



Problème

- Contexte : code de calcul ou expérience modélisé par

$$Y = f(\mathbf{X})$$

avec Y =variable de sortie

$\mathbf{X} = (X_i)_{i=1,\dots,d}$ variables d'entrée, de loi donnée (à densité $p(\mathbf{x})$)

f = boîte plus ou moins noire

- But : estimation d'une quantité

$$\mathbb{E}[\psi(Y)]$$

avec une barre d'erreur et le minimum de simulations/expériences.

- Exemples (pour Y à valeurs réelles) :
 - $\psi(y) = y \rightarrow$ moyenne de Y , i.e. $\mathbb{E}[Y]$
 - $\psi(y) = y^2 \rightarrow$ variance de Y , i.e. $\text{Var}(Y) = \mathbb{E}[(Y - \mathbb{E}[Y])^2] = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2$
 - $\psi(y) = \mathbf{1}_{[a,\infty[} \rightarrow$ probabilité de dépasser le seuil a , i.e. $\mathbb{P}(Y \geq a)$

Méthode analytique

- La quantité à estimer est une intégrale d -dimensionnelle :

$$I = \mathbb{E}[\psi(Y)] = \mathbb{E}[F(\mathbf{X})] = \int_{\mathbb{R}^d} F(\mathbf{x})p(\mathbf{x})d\mathbf{x}$$

où $p(\mathbf{x})$ est la densité de probabilité de \mathbf{X} et $F(\mathbf{x}) = \psi(f(\mathbf{x}))$.

Dans des cas simples (quand la densité p et la fonction F sont connues de manière analytique), on peut parfois évaluer l'intégrale de manière exacte.

Situation "exceptionnelle".

Méthodes de quadrature

- La quantité à estimer est une intégrale d -dimensionnelle :

$$I = \mathbb{E}[\psi(Y)] = \mathbb{E}[F(\mathbf{X})] = \int_{\mathbb{R}^d} F(\mathbf{x})p(\mathbf{x})d\mathbf{x}$$

où $p(\mathbf{x})$ est la densité de probabilité de \mathbf{X} et $F(\mathbf{x}) = \psi(f(\mathbf{x}))$.

Si $p(\mathbf{x}) = \prod_{i=1}^d p_0(x_i)$, alors on peut appliquer une méthode de quadrature gaussienne avec grille pleine de n^d points :

$$\hat{I} = \sum_{j_1=1}^n \cdots \sum_{j_d=1}^n \rho_{j_1} \cdots \rho_{j_d} F(\xi_{j_1}, \dots, \xi_{j_d})$$

avec $(\rho_j)_{j=1, \dots, n}$ poids et $(\xi_j)_{j=1, \dots, n}$ nœuds de quadrature gaussienne associée à la densité p_0 .

Il existe aussi des méthodes de quadrature avec des grilles creuses (Smolyak).

- Les méthodes de quadrature sont efficaces avec :
 - des conditions de régularité sur $\mathbf{x} \rightarrow F(\mathbf{x})$ (et pas seulement f),
 - une dimension d petite (même avec des grilles creuses de type Smolyak).

Elles réclament sinon beaucoup d'appels au code.

Méthode de Monte Carlo

On tire un n -échantillon $(\mathbf{X}^{(k)})_{k=1,\dots,n}$ (des réalisations indépendantes de \mathbf{X}).

Estimateur de $I = \mathbb{E}[F(\mathbf{X})]$:

$$\hat{I}_n = \frac{1}{n} \sum_{k=1}^n F(\mathbf{X}^{(k)})$$

Estimation non-biaisée :

$$\mathbb{E}[\hat{I}_n] = I \quad \text{pour tout } n$$

Convergence :

$$\hat{I}_n \xrightarrow[n \rightarrow \infty]{} I \quad \text{avec probabilité 1}$$

Erreur (risque quadratique) :

$$\mathbb{E}[(\hat{I}_n - I)^2] = \text{Var}(\hat{I}_n) = \frac{1}{n} \text{Var}(F(\mathbf{X}))$$

Erreur relative :

$$\frac{\mathbb{E}[(\hat{I}_n - I)^2]^{1/2}}{I} = \frac{1}{\sqrt{n}} \frac{\text{Var}(F(\mathbf{X}))^{1/2}}{\mathbb{E}[F(\mathbf{X})]}$$

Caractéristiques :

- 1) possibilité d'obtenir des intervalles de confiance,
- 2) pas de régularité requise sur F (avec $F(\mathbf{x}) = \psi(f(\mathbf{x}))$),
- 3) vitesse de convergence indépendante de la dimension (mais lente).

Intervalles de confiance

Question : A partir de l'échantillon $(\mathbf{X}^{(k)})_{k=1,\dots,n}$, l'estimateur \hat{I}_n donne une valeur approchée de I , d'autant meilleure que n est grand. Comment quantifier précisément l'erreur ?

Réponse : On construit un intervalle de confiance au niveau 0.95, i.e. un intervalle $[\hat{a}_n, \hat{b}_n]$ tel que

$$\mathbb{P}\left(I \in [\hat{a}_n, \hat{b}_n]\right) \geq 0.95$$

Construction basée sur le **théorème central limite** :

$$\mathbb{P}\left(\left|\hat{I}_n - I\right| < c \frac{\text{Var}(F(\mathbf{X}))^{1/2}}{\sqrt{n}}\right) \xrightarrow{n \rightarrow \infty} \frac{2}{\sqrt{2\pi}} \int_0^c e^{-x^2/2} dx$$

Le membre de droite vaut 0.95 si $c = 1.96$. Donc

$$\mathbb{P}\left(I \in \left[\hat{I}_n - 1.96 \frac{\text{Var}(F(\mathbf{X}))^{1/2}}{\sqrt{n}}, \hat{I}_n + 1.96 \frac{\text{Var}(F(\mathbf{X}))^{1/2}}{\sqrt{n}}\right]\right) \simeq 0.95$$

$$\mathbb{P} \left(I \in \left[\hat{I}_n - 1.96 \frac{\text{Var}(F(\mathbf{X}))^{1/2}}{\sqrt{n}}, \hat{I}_n + 1.96 \frac{\text{Var}(F(\mathbf{X}))^{1/2}}{\sqrt{n}} \right] \right) \simeq 0.95$$

Les bornes sont encore inconnues car on ne connaît pas $\text{Var}(F(\mathbf{X}))$! Deux solutions :

- solution conservatrice, du type $\text{Var}(F(\mathbf{X})) \leq \|F\|_\infty^2$, et alors

$$\mathbb{P} \left(I \in \left[\hat{I}_n - 1.96 \frac{\|F\|_\infty}{\sqrt{n}}, \hat{I}_n + 1.96 \frac{\|F\|_\infty}{\sqrt{n}} \right] \right) \geq 0.95$$

- asymptotiquement, on remplace $\text{Var}(F(\mathbf{X}))$ dans les bornes par son estimateur empirique $\hat{\sigma}_n^2$:

$$\mathbb{P} \left(I \in \left[\hat{I}_n - 1.96 \frac{\hat{\sigma}_n}{\sqrt{n}}, \hat{I}_n + 1.96 \frac{\hat{\sigma}_n}{\sqrt{n}} \right] \right) \simeq 0.95$$

où

$$\hat{\sigma}_n = \left(\frac{1}{n} \sum_{k=1}^n F(\mathbf{X}^{(k)})^2 - \hat{I}_n^2 \right)^{1/2}$$

Conclusion : Il n'y a aucun intervalle borné de \mathbb{R} dont on puisse dire avec certitude qu'il contient I , mais il y a des intervalles, dits intervalle de confiance, dont on peut dire qu'ils contiennent I avec une probabilité proche de 1.

Estimation particulière : probabilité d'un événement rare

- On cherche à estimer

$$P = \mathbb{P}(f(\mathbf{X}) \geq a)$$

avec a grand si bien que $P \ll 1$.

- Possible par Monte Carlo :

$$\hat{P}_n = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{f(\mathbf{X}^{(k)}) \geq a}$$

Erreur relative :

$$\frac{\mathbb{E}[(\hat{P}_n - P)^2]^{1/2}}{P} = \frac{1}{\sqrt{n}} \frac{\text{Var}(\mathbf{1}_{f(\mathbf{X}) \geq a})^{1/2}}{P} = \frac{1}{\sqrt{n}} \frac{\sqrt{1-P}}{\sqrt{P}} \stackrel{P \ll 1}{\approx} \frac{1}{\sqrt{nP}}$$

Il faut donc $nP > 1$ pour avoir une erreur relative plus petite que 1 (pas étonnant...).

↪ La "constante" est pénalisante.

Propagation d'incertitudes par métamodèles

On remplace f par un métamodèle (modèle réduit) f_r et on applique une des techniques précédentes (analytique, quadrature, Monte Carlo).

→ On peut faire beaucoup d'appels au métamodèle.

→ Le choix du métamodèle est critique.

→ Le contrôle de l'erreur n'est pas simple.

Développements de Taylor

- On approche la sortie $Y = f(\mathbf{X})$ par un développement de Taylor $Y_r = f_r(\mathbf{X})$.
- Exemple :
 - On souhaite estimer $\mathbb{E}[Y]$ et $\text{Var}(Y)$ pour $Y = f(\mathbf{X})$ avec X_i décorréliées, $\mathbb{E}[X_i] = \mu_i$ et $\text{Var}(X_i) = \sigma_i^2$ connus, σ_i^2 petits.
 - On approche $Y = f(\mathbf{X})$ par $Y_r = f_r(\mathbf{X}) = f(\boldsymbol{\mu}) + \nabla f(\boldsymbol{\mu}) \cdot (\mathbf{X} - \boldsymbol{\mu})$. On trouve :

$$\mathbb{E}[Y] \simeq \mathbb{E}[Y_r] = f(\boldsymbol{\mu}), \quad \text{Var}(Y) \simeq \text{Var}(Y_r) = \sum_{i=1}^d \partial_{x_i} f(\boldsymbol{\mu})^2 \sigma_i^2$$

On a juste besoin de calculer $f(\boldsymbol{\mu})$ et $\nabla f(\boldsymbol{\mu})$ (calcul du gradient par différences finies ou par différenciation automatique, donc en gros $d + 1$ appels à f).

- Rapide, analytique, permet de calculer approximativement des tendances centrales de la sortie (moyenne, variance).
- Convenable pour des petites variations des paramètres d'entrée et un modèle régulier (qu'on peut linéariser).
- Approche "locale". En général, pas de contrôle de l'erreur.

Probabilité à évaluer :

$$P = \mathbb{P}(f(\mathbf{X}) \geq a) = \mathbb{P}(\mathbf{X} \in \mathcal{F}) = \int_{\mathcal{F}} p(\mathbf{x})d\mathbf{x}, \quad \mathcal{F} = \{\mathbf{x} \in \mathbb{R}^d, f(\mathbf{x}) \geq a\}$$

- Méthode FORM-SORM, analytique mais approchée, sans contrôle d'erreur.
- on suppose que les X_i sont indépendants et de loi gaussienne de moyenne zéro et de variance un (ou on se ramène à ce cas par transformation isoprobabiliste).
- on trouve par optimisation (sous contrainte) le point \mathbf{x}_a de dépassement de seuil (i.e. $f(\mathbf{x}_a) = a$) le plus proche de l'origine.
- on approche la surface de défaillance $\{\mathbf{x} \in \mathbb{R}^d, f(\mathbf{x}) = a\}$ par une surface régulière $\hat{\mathcal{F}}$ qui permette de faire un calcul analytique $\hat{P} = \int_{\hat{\mathcal{F}}} p(\mathbf{x})d\mathbf{x}$:

- un hyperplan pour FORM

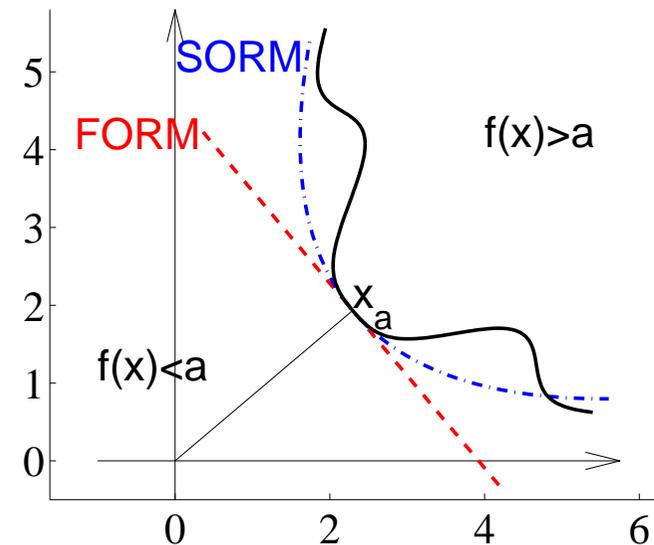
(et alors $\hat{P} = \frac{1}{2}\text{erfc}\left(\frac{|\mathbf{x}_a|}{\sqrt{2}}\right)$),

- une forme quadratique pour SORM

(et alors $\hat{P} =$ formule de Breitung).

Cf: O. Ditlevsen et H.O. Madsen,

Structural reliability methods, Wiley, 1996.



Techniques de réduction de variance

Objectif : réduire la variance de l'estimateur de Monte Carlo classique :

$$\mathbb{E}[(\hat{I}_n - I)^2] = \frac{1}{n} \text{Var}(F(\mathbf{X}))$$

Les méthodes

- Echantillonnage préférentiel
- Variables de contrôle
- Variables antithétiques
- Stratification

ont pour but de réduire la constante, en restant proches de l'esprit Monte Carlo (parallélisable).

Les méthodes

- Quasi-Monte Carlo

ont pour but de changer le $1/n$.

Les méthodes

- Systèmes de particules en interaction

s'éloignent de l'esprit Monte Carlo (séquentielle).

Echantillonnage préférentiel (importance sampling)

- Observation : la représentation de I comme espérance n'est pas unique. Si \mathbf{X} est à densité p :

$$I = \mathbb{E}_p[F(\mathbf{X})] = \int F(\mathbf{x})p(\mathbf{x})d\mathbf{x} = \int \frac{F(\mathbf{x})p(\mathbf{x})}{q(\mathbf{x})}q(\mathbf{x})d\mathbf{x} = \mathbb{E}_q\left[\frac{F(\mathbf{X})p(\mathbf{X})}{q(\mathbf{X})}\right]$$

L'utilisateur est libre dans le choix de la densité q .

- Idée : Dans la situation où on sait que $F(\mathbf{X})$ est surtout sensible à certaines valeurs de \mathbf{X} , au lieu de tirer les $\mathbf{X}^{(k)}$ selon la densité originale $p(\mathbf{x})$ de \mathbf{X} , on les tire selon une densité "biaisée" $q(\mathbf{x})$ qui favorise les valeurs de \mathbf{X} dans la zone d'importance.

- En considérant la représentation

$$I = \mathbb{E}_p[F(\mathbf{X})] = \mathbb{E}_q\left[F(\mathbf{X})\frac{p(\mathbf{X})}{q(\mathbf{X})}\right]$$

on propose l'estimateur :

$$\hat{I}_n = \frac{1}{n} \sum_{k=1}^n F(\mathbf{X}^{(k)}) \frac{p(\mathbf{X}^{(k)})}{q(\mathbf{X}^{(k)})}$$

où $(\mathbf{X}^{(k)})_{k=1,\dots,n}$ est un n -échantillon tiré selon q .

- Estimateur non-biaisé : $\mathbb{E}_q[\hat{I}_n] = I$.

- Estimateur convergent :

$$\hat{I}_n = \frac{1}{n} \sum_{k=1}^n F(\mathbf{X}^{(k)}) \frac{p(\mathbf{X}^{(k)})}{q(\mathbf{X}^{(k)})} \xrightarrow{n \rightarrow \infty} \mathbb{E}_q \left[F(\mathbf{X}) \frac{p(\mathbf{X})}{q(\mathbf{X})} \right] = \mathbb{E}_p [F(\mathbf{X})] = I$$

- Variance de l'estimateur :

$$\text{Var}(\hat{I}_n) = \frac{1}{n} \text{Var}_q \left(F(\mathbf{X}) \frac{p(\mathbf{X})}{q(\mathbf{X})} \right) = \frac{1}{n} \left(\mathbb{E}_p \left[F(\mathbf{X})^2 \frac{p(\mathbf{X})}{q(\mathbf{X})} \right] - \mathbb{E}_p [F(\mathbf{X})]^2 \right)$$

- En choisissant bien q , on peut fortement réduire la variance. En fait, en choisissant (en supposant $F \geq 0$)

$$q_{opt}(\mathbf{x}) = \frac{F(\mathbf{x})p(\mathbf{x})}{I}$$

on trouve

$$\text{Var}(\hat{I}_n) = 0$$

Mais $q_{opt}(\mathbf{x})$ dépend de I ! (\hookrightarrow méthodes séquentielles/adaptatives).

- Points pratiques importants pour pouvoir implémenter la méthode :
 - Il faut savoir simuler \mathbf{X} de loi de densité q .
 - Il faut savoir calculer le rapport de vraisemblance $\frac{p(\mathbf{x})}{q(\mathbf{x})}$.

- Exemple : On veut estimer

$$I = \mathbb{E}[F(X)]$$

avec $X \sim \mathcal{N}(0, 1)$ et $F(x) = \mathbf{1}_{[4, \infty[}(x)$.

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathbf{1}_{[4, \infty[}(x) e^{-\frac{x^2}{2}} dx = \frac{1}{2} \operatorname{erfc}\left(\frac{4}{\sqrt{2}}\right) \simeq 3.17 \cdot 10^{-5}$$

Monte Carlo :

$$\hat{I}_n = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X_k \geq 4}, \quad X_k \sim \mathcal{N}(0, 1)$$

On a $\operatorname{Var}(\hat{I}_n) = \frac{1}{n} 3.17 \cdot 10^{-5}$.

Echantillonnage préférentiel : on tire X_k selon la loi $\mathcal{N}(4, 1)$.

$$\hat{I}_n = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X_k \geq 4} \frac{e^{-\frac{X_k^2}{2}}}{e^{-\frac{(X_k-4)^2}{2}}} = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X_k \geq 4} e^{-4X_k+8}, \quad X_k \sim \mathcal{N}(4, 1)$$

On a $\operatorname{Var}(\hat{I}_n) = \frac{1}{n} 5.53 \cdot 10^{-8}$.

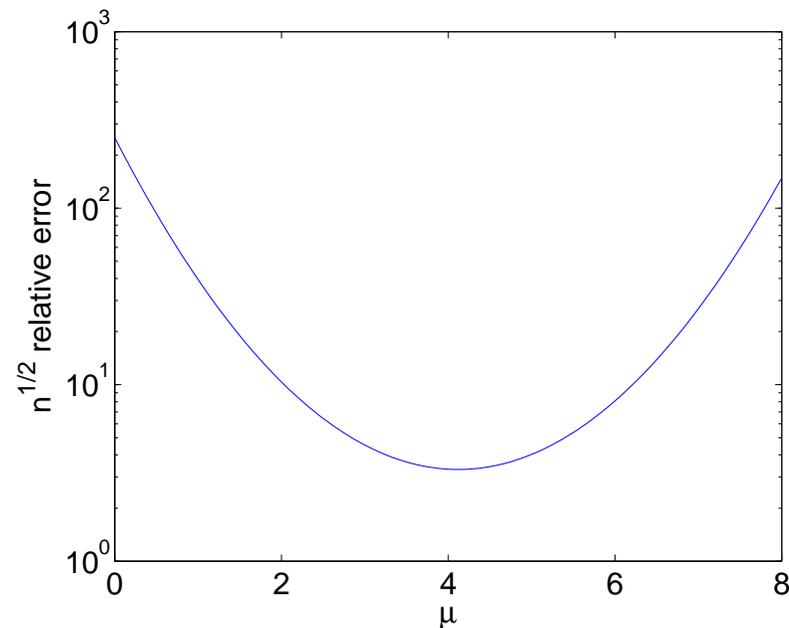
Il faut 1000 fois moins de simulations avec la méthode IS pour atteindre la même précision !

Attention cependant, il ne faut pas trop pousser.

Echantillonnage préférentiel : on tire X_k selon la loi $\mathcal{N}(\mu, 1)$.

$$\hat{I}_n = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X_k \geq 4} \frac{e^{-\frac{X_k^2}{2}}}{e^{-\frac{(X_k - \mu)^2}{2}}} = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X_k \geq 4} e^{-\mu X_k + \frac{\mu^2}{2}}, \quad X_k \sim \mathcal{N}(\mu, 1)$$

On a $\text{Var}(\hat{I}_n) = \frac{1}{n} \frac{e^{\mu^2}}{2} \text{erfc}\left(\frac{4+\mu}{\sqrt{2}}\right) - \frac{1}{n} I^2$, ce qui donne pour l'erreur relative normalisée $\sqrt{n} \mathbb{E}[(\hat{I}_n - I)^2]^{1/2} / I$:



Si on pousse trop, les fluctuations des rapports de vraisemblance deviennent trop grandes.

- Exemple : On veut estimer

$$I = \mathbb{E}[f(X)]$$

avec $X \sim \mathcal{N}(0, 1)$ et $f(x) = \exp(x)$.

$$I = \frac{1}{\sqrt{2\pi}} \int e^x e^{-\frac{x^2}{2}} dx = e^{\frac{1}{2}}$$

Ce sont les grandes valeurs de X qui sont importantes.

Echantillonnage préférentiel : on tire $X^{(k)}$ selon la loi $\mathcal{N}(\mu, 1)$, $\mu > 0$.

$$\hat{I}_n = \frac{1}{n} \sum_{k=1}^n f(X^{(k)}) \frac{e^{-\frac{[X^{(k)}]^2}{2}}}{e^{-\frac{[X^{(k)} - \mu]^2}{2}}} = \frac{1}{N} \sum_{k=1}^n f(X^{(k)}) e^{-\mu X^{(k)} + \frac{\mu^2}{2}}$$

$$\text{Var}(\hat{I}_n) = \frac{1}{n} \left(e^{\mu^2 - 2\mu + 2} - e^1 \right)$$

Monte Carlo $\mu = 0$: $\text{Var}(\hat{I}_n) = \frac{1}{n} (e^2 - e^1)$

Echantillonnage préférentiel optimal $\mu = 1$: $\text{Var}(\hat{I}_n) = 0$.

Variables de contrôle

- Rappel : Le but est d'estimer $I = \mathbb{E}[F(\mathbf{X})]$ pour \mathbf{X} un vecteur aléatoire et $F(\mathbf{x}) = \psi(f(\mathbf{x}))$ une fonction déterministe.
- Dans la situation où on dispose d'un modèle réduit $f_r(\mathbf{x})$.
- Méthode d'échantillonnage préférentielle : on calcule (on approche) la densité optimale $q_{opt,r}(\mathbf{x}) = \frac{\psi(f_r(\mathbf{x}))p(\mathbf{x})}{I_r}$, avec $I_r = \int \psi(f_r(\mathbf{x}))p(\mathbf{x})d\mathbf{x}$, puis on l'utilise comme densité d'importance.

- Méthode de variables de contrôle :

On note $F(\mathbf{x}) = \psi(f(\mathbf{x}))$, $F_r(\mathbf{x}) = \psi(f_r(\mathbf{x}))$.

On suppose qu'on connaît $I_r = \mathbb{E}[F_r(\mathbf{X})]$.

En considérant la représentation

$$I = \mathbb{E}[F(\mathbf{X})] = I_r + \mathbb{E}[F(\mathbf{X}) - F_r(\mathbf{X})]$$

on propose l'estimateur :

$$\hat{I}_n = I_r + \frac{1}{n} \sum_{k=1}^n F(\mathbf{X}^{(k)}) - F_r(\mathbf{X}^{(k)}),$$

où $(\mathbf{X}^{(k)})_{k=1,\dots,n}$ est un n -échantillon (tiré selon p).

Estimateur :

$$\hat{I}_n = I_r + \frac{1}{n} \sum_{k=1}^n F(\mathbf{X}^{(k)}) - F_r(\mathbf{X}^{(k)})$$

Cet estimateur est sans biais et convergent.

Sa variance est :

$$\text{Var}(\hat{I}_n) = \frac{1}{n} \text{Var}(F(\mathbf{X}) - F_r(\mathbf{X}))$$

↪ Cette méthode peut réduire la variance.

- Exemple : on souhaite estimer

$$I = \mathbb{E}[f(X)]$$

avec $X \sim \mathcal{U}(0, 1)$, $f(x) = \exp(x)$.

Résultat : $I = e - 1 \simeq 1.72$.

Monte Carlo.

$$\hat{I}_n = \frac{1}{n} \sum_{k=1}^n \exp[X^{(k)}]$$

Variance de l'estimateur MC = $\frac{1}{n}(2e - 1) \simeq \frac{1}{n}4.44$.

Variable de contrôle. Modèle réduit : $f_r(x) = 1 + x$ (ici $I_r = \frac{3}{2}$). Estimateur VC :

$$\hat{I}_n = I_r + \frac{1}{n} \sum_{k=1}^n \left\{ \exp[X^{(k)}] - 1 - X^{(k)} \right\}$$

Variance de l'estimateur VC = $\frac{1}{n}(3e - \frac{e^2}{2} - \frac{53}{12}) \simeq \frac{1}{n}0.044$.

Il faut donc 100 fois moins de simulations avec l'estimateur VC !

- Application : Méthodes de Romberg statistiques pour l'estimation de

$$I = \mathbb{E}[\psi(f(\mathbf{X}))]$$

On dispose d'un code léger f_r en plus du code lourd f . Le rapport du coût calcul entre un appel à f et un appel à f_r est $q > 1$.

Estimateur

$$\hat{I}_n = \frac{1}{n_r} \sum_{i=1}^{n_r} F_r(\tilde{\mathbf{X}}^{(k)}) + \frac{1}{n} \sum_{i=1}^n F(\mathbf{X}^{(k)}) - F_r(\mathbf{X}^{(k)})$$

avec $n_r \gg n$, $F(\mathbf{x}) = \psi(f(\mathbf{x}))$, $F_r(\mathbf{x}) = \psi(f_r(\mathbf{x}))$.

Allocation entre appels au code lourd et appels au code léger à optimiser sous la contrainte $n_r/q + n(1 + 1/q) = n_{\text{tot}}$.

Compromis classique entre erreur d'approximation et erreur d'estimation.

Utilisé dans le cas où $f(\mathbf{X})$ est la solution d'une équation différentielle discrétisée finement, avec $f_r(\mathbf{X})$ la solution avec un schéma de discrétisation grossier (MultiLevel Monte Carlo).

- Pas très utile pour l'estimation de probabilité d'événement rare.

Exemple : on souhaite estimer

$$I = \mathbb{P}(f(X) \geq 2.7)$$

avec $X \sim \mathcal{U}(0, 1)$, $f(x) = \exp(x)$.

Résultat : $I = 1 - \ln(2.7) \simeq 6.7 \cdot 10^{-3}$.

Le modèle réduit $f_r(x) = 1 + x$ ne sert à rien dans ce cas.

Il faudrait que le modèle réduit soit quantitativement bon dans la zone d'importance.

Suite à discrédance faible (quasi Monte Carlo)

on tire l'échantillon de manière moins aléatoire que MC, pour combler les trous qui se forment naturellement dans un échantillon aléatoire.

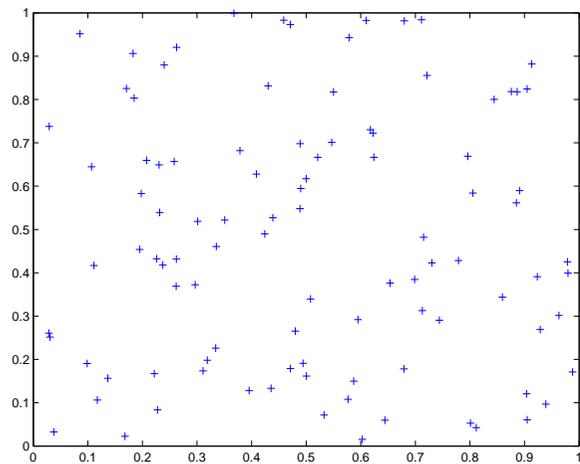
Cette technique

- réduit la variance si f a un peu de régularité et/ou de monotonie; on peut aller jusqu'à une variance en $C_d(\log n)^{s(d)}/n^2$,
- marche en dimension pas trop grande,
- représente un intermédiaire entre MC et quadrature usuelle,
- en compétition avec des méthodes de quadrature creuse (Smolyak).

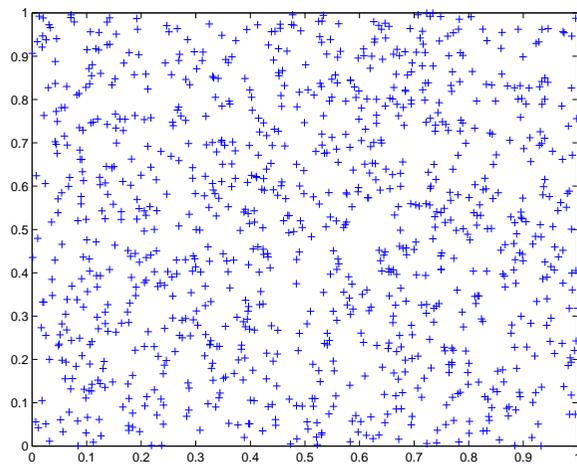
Attention :

- il n'est pas facile de rajouter des points,
- on n'a pas d'estimée d'erreur (sans hypothèse supplémentaire),
- la méthode n'est pas adaptée pour l'estimation de probabilité d'événement rare.

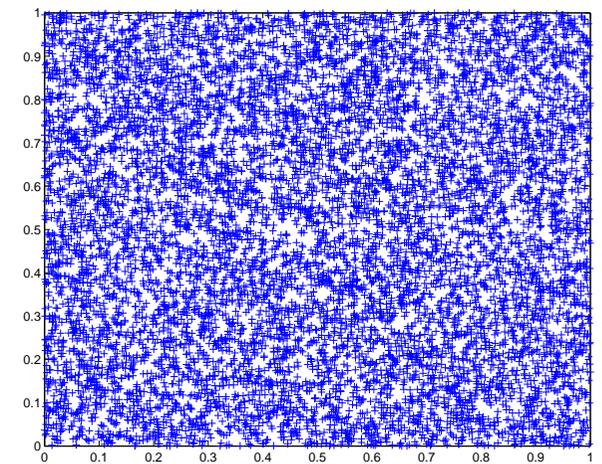
Exemple : échantillon Monte Carlo.



$n = 100$

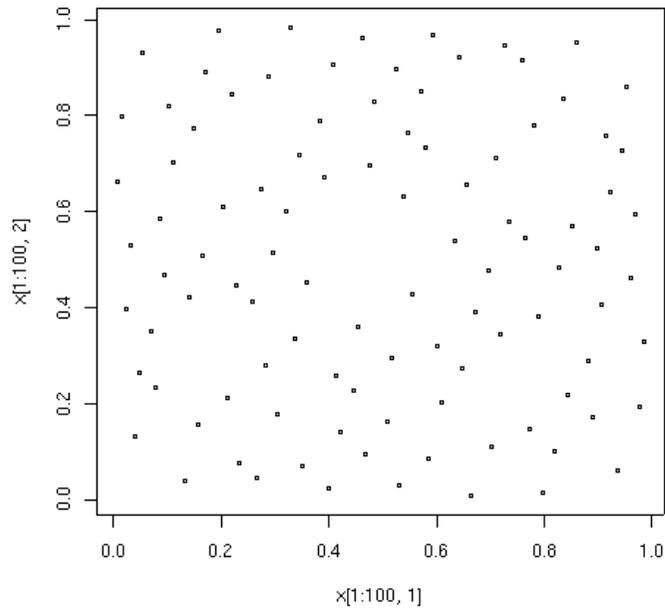


$n = 1000$

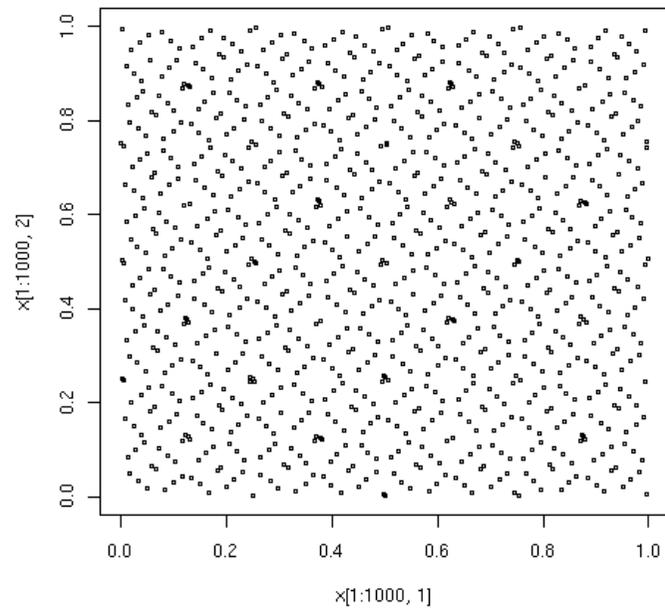


$n = 10000$

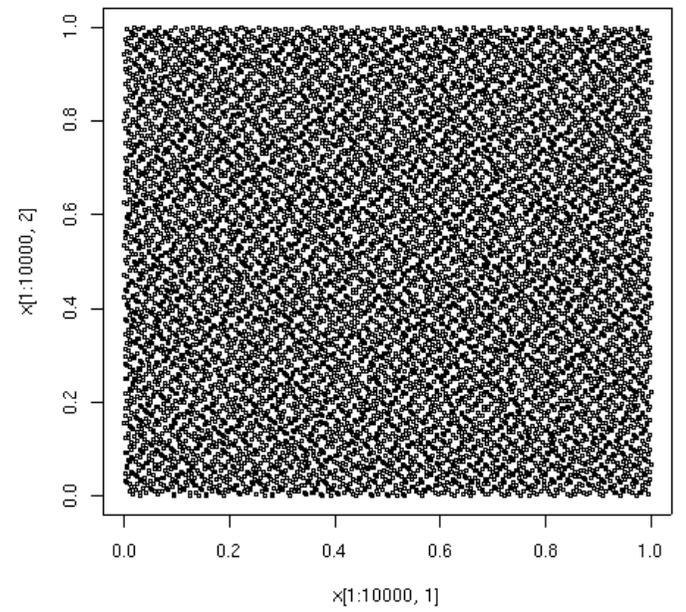
Exemple : suites de Sobol en dimension 2.



$n = 100$



$n = 1000$



$n = 10000$

Systèmes de particules en interaction

Problème : estimation de la probabilité d'un événement rare.

Deux versions :

- un événement rare pour un problème dynamique, s'exprimant en termes de l'état final d'une chaîne de Markov,
- un événement rare pour un problème statique, s'exprimant en termes d'une variable aléatoire, dont la loi est vue comme la loi stationnaire d'une chaîne de Markov.

Événements rares pour un problème dynamique

- Problème : estimation de la probabilité

$$P = \mathbb{P}(V(X_M) \geq a)$$

quand a est grand $\implies P \ll 1$.

- Description du système :

– $(X_p)_{0 \leq p \leq M}$: une chaîne de Markov à valeurs dans E :

$$X_p = \Phi_p(X_{p-1}, \Theta_p), \quad p = 1, \dots, M, \quad X_0 = x_0$$

où

- Θ_p : variables indépendantes,
 - Φ_p : fonction déterministe,
 - $V : E \rightarrow \mathbb{R}$: fonction de risque.
 - $a \in \mathbb{R}$: niveau de risque.
- *Exemple* : $X_p = X_{p-1} + \theta_p$, $X_0 = 0$, où θ_p est une suite de variables gaussiennes indépendantes de moyenne zéro et de variance un. Ici
 - $E = \mathbb{R}$, $\Phi_p(X, \theta) = X + \theta$,
 - $V(x) = x$,
 - solution connue : $X_M = V(X_M) \sim \mathcal{N}(0, M)$.

- *Second exemple* : Fibre optique avec des défauts aléatoires.

$u_0(t)$ = impulsion initiale.

$u(Z, t)$ = impulsion finale = impulsion après une distance de propagation Z .

Propagation régie par deux équations de Schrödinger Non-Linéaires couplées avec des coefficients aléatoires variant en z .

→ Troncation fictive en M segments $[z_{p-1}, z_p)$, $z_p = pZ/M$, $1 \leq p \leq M$.

→ $X_p = (u(z_p, t))_{t \in \mathbb{R}}$ satisfait la relation de récurrence aléatoire

$$X_p = \Phi_p(X_{p-1}, \Theta_p)$$

où

→ Θ_p modélise les défauts du p -ème segment,

→ $\Phi_p(X_{p-1}, \Theta_p)$ est la solution à $z = z_p$ de l'équation de Schrödinger avec conditions initiales X_{p-1} à $z = z_{p-1}$.

Ici

→ $E = H_0^2(\mathbb{R}, \mathbb{C}^2) \cap L_2^2(\mathbb{R}, \mathbb{C}^2)$

→ $V : \left| \begin{array}{l} E \rightarrow \mathbb{R} \\ V(X) = \int t^2 |X(t)|^2 dt / \int |X(t)|^2 dt \end{array} \right.$

Problème : estimation de la probabilité

$$P = \mathbb{P}(V(X_M) \geq a)$$

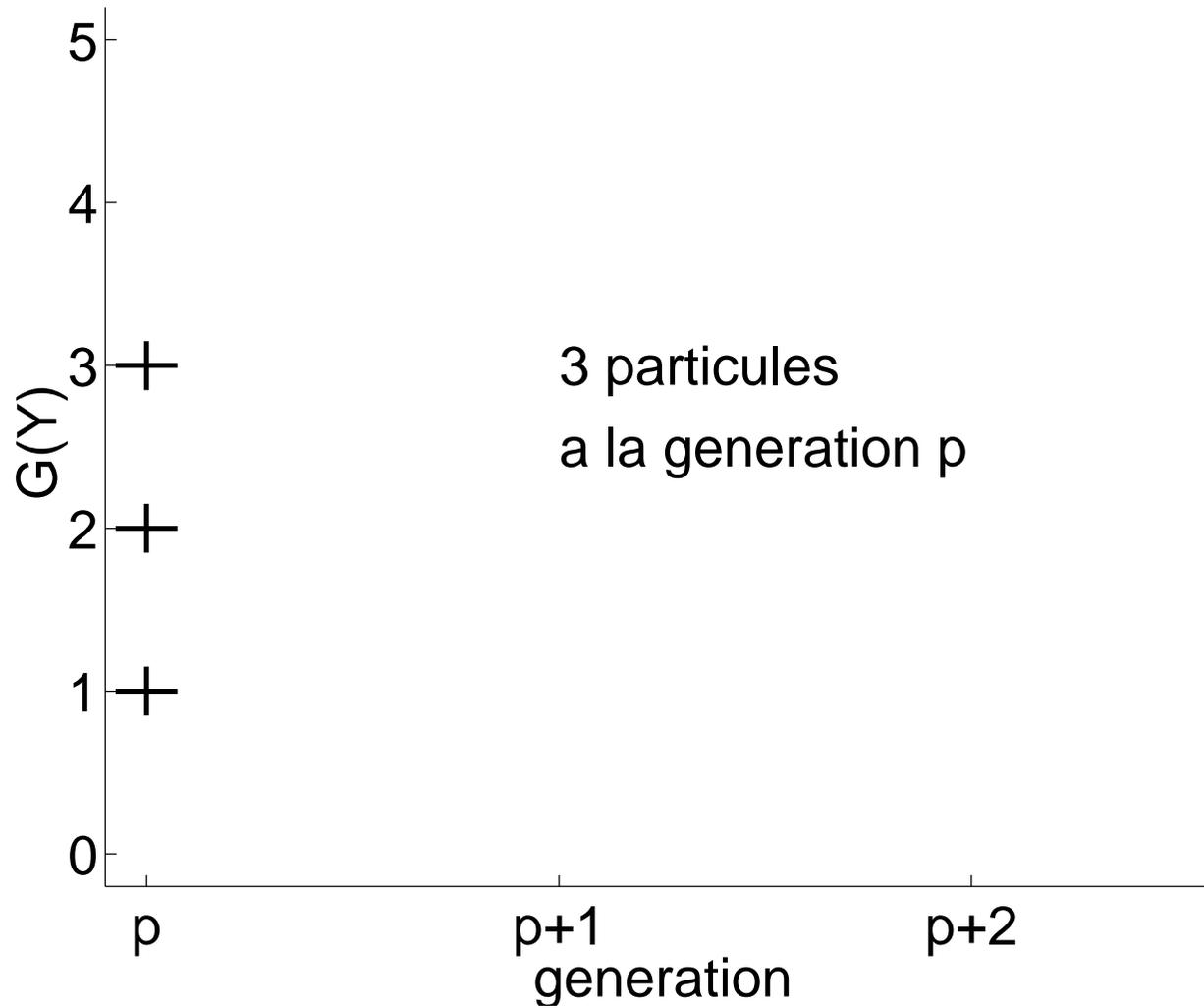
Système de particules-chemins

- Particule-chemin: $Y_p = (X_0, \dots, X_p)$ à valeurs dans E^{p+1} , $0 \leq p \leq M$.
- Système de n particules-chemins: $\mathbf{Y}_p = (Y_p^{(k)})_{1 \leq k \leq n}$ à valeurs dans $(E^{p+1})^n$.
- Initialisation : $p = 0$: $Y_0^{(k)} = x_0$ pour $k = 1, \dots, n$.
- Dynamique : Evolution de la génération p à la génération $p + 1$:

$$\mathbf{Y}_{p-1} \in (E^p)^n \xrightarrow{\textit{selection}} \widehat{\mathbf{Y}}_{p-1} \in (E^p)^n \xrightarrow{\textit{mutation}} \mathbf{Y}_p \in (E^{p+1})^n$$

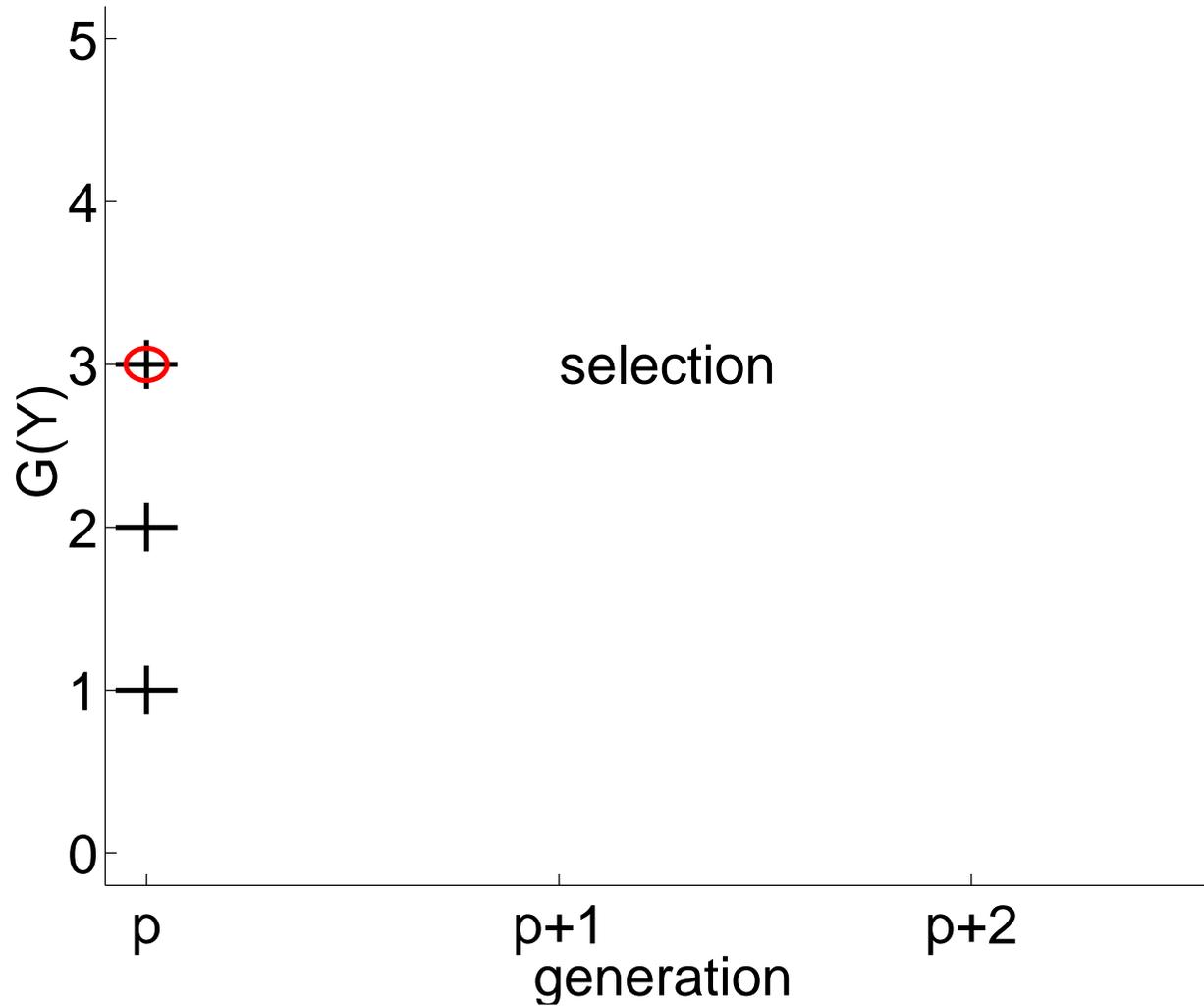
La sélection dépend de fonctions potentiels positives $G_p(Y_p)$, choisies par l'utilisateur, par exemple $G_p(Y_p) = \exp(V(X_p))$.

n=3 particules



3 particules $Y_p^{(1)}$, $Y_p^{(2)}$, $Y_p^{(3)}$ à la génération p ,
de potentiels $G(Y_p^{(1)}) = 1$, $G(Y_p^{(2)}) = 2$, $G(Y_p^{(3)}) = 3$.

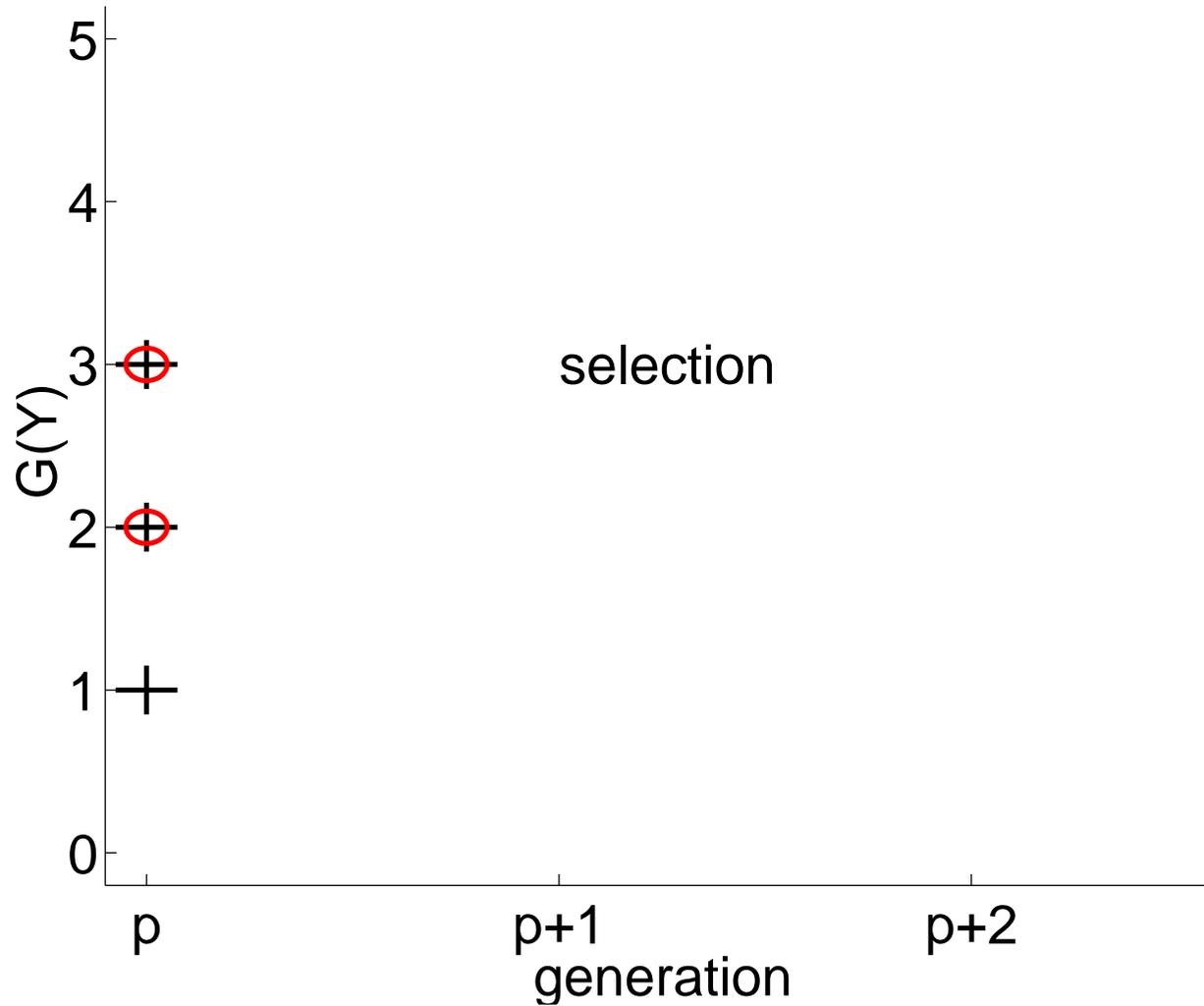
n=3 particules



Probabilité de sélectionner la particule j :

$$\frac{G(Y_p^{(j)})}{\sum_{k=1}^3 G(Y_p^{(k)})} = \begin{cases} \frac{1}{6} & \text{si } j = 1 \\ \frac{1}{3} & \text{si } j = 2 \\ \frac{1}{2} & \text{si } j = 3 \end{cases}$$

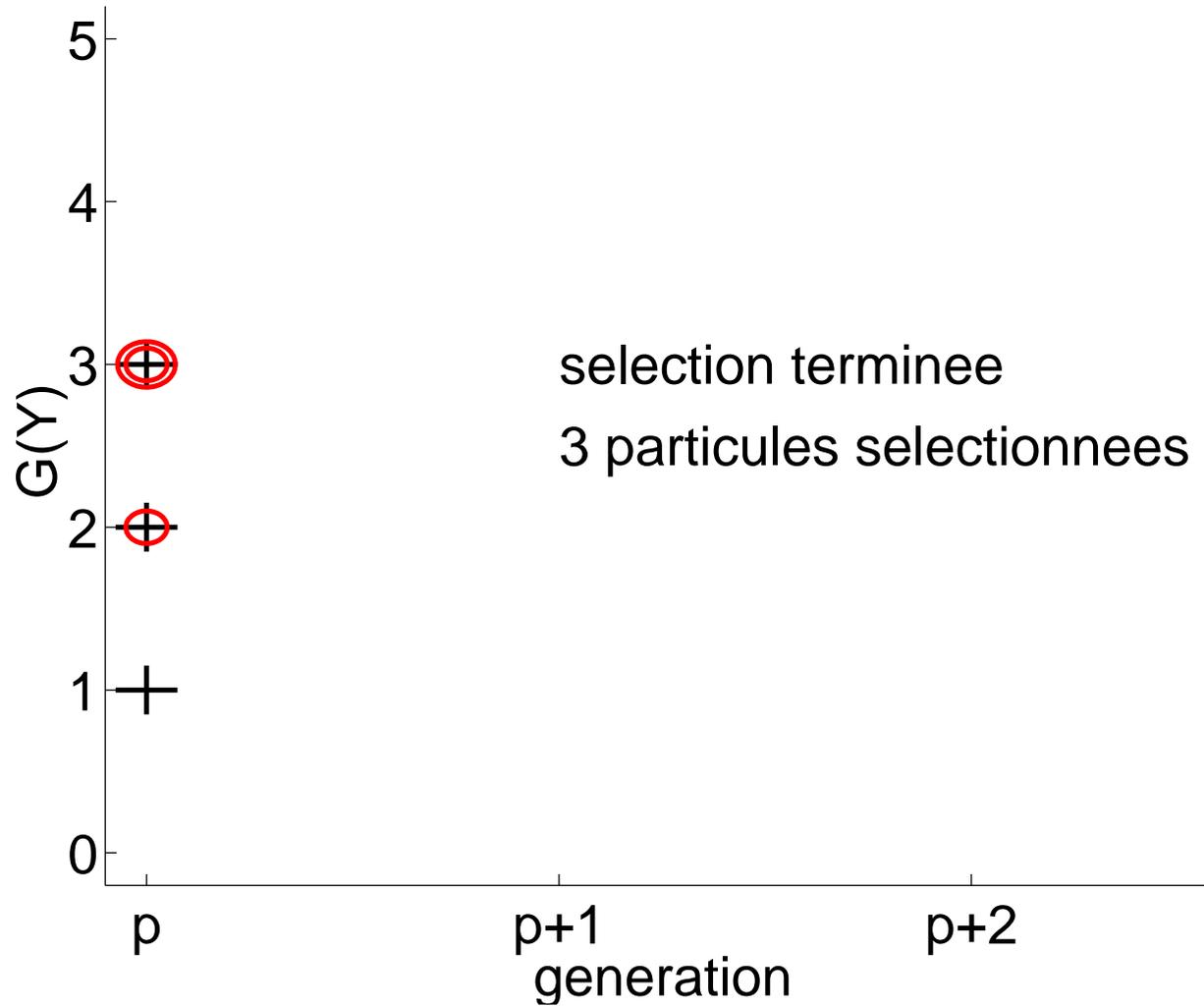
n=3 particules



Probabilité de sélectionner la particule j :

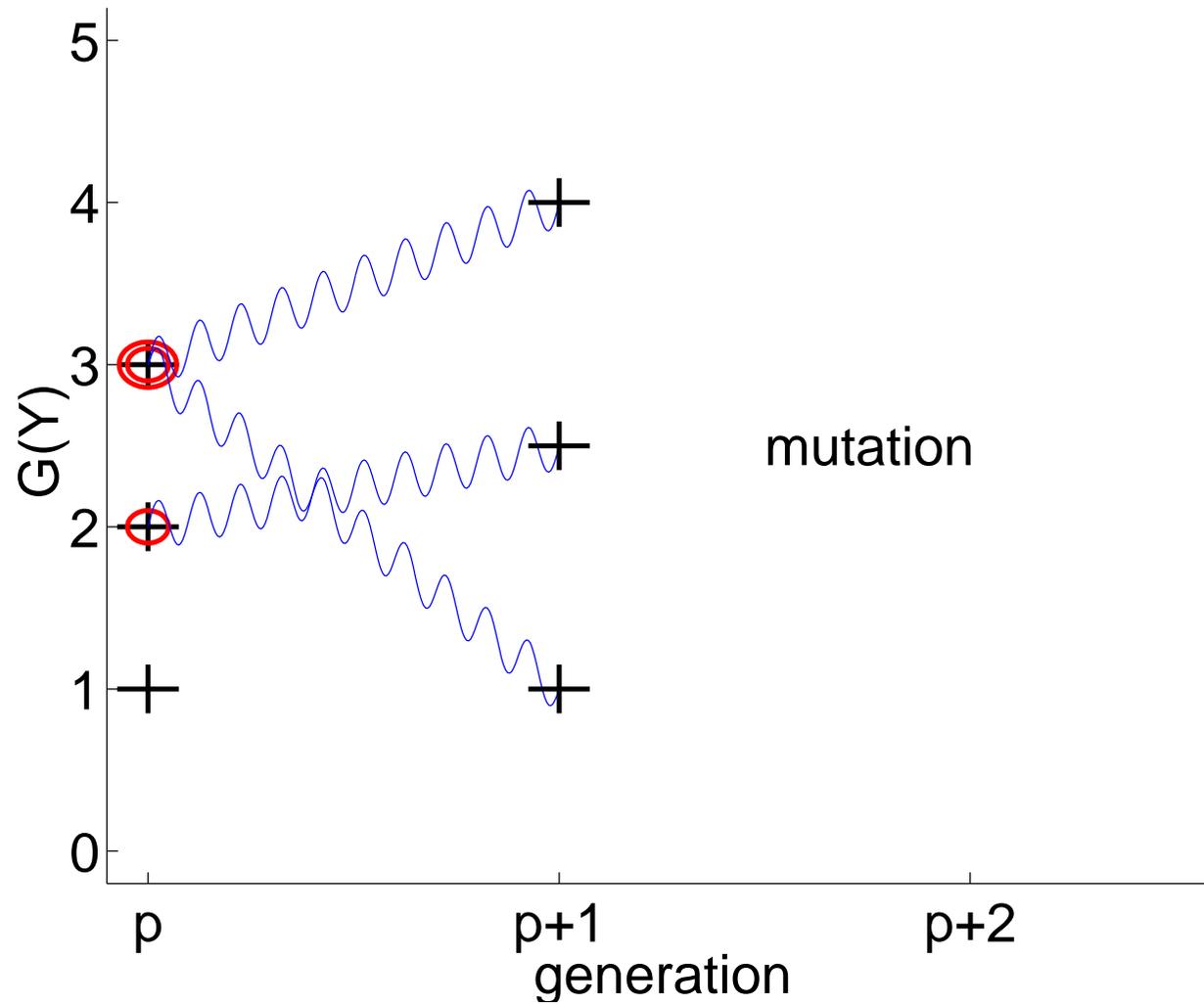
$$\frac{G(Y_p^{(j)})}{\sum_{k=1}^3 G(Y_p^{(k)})} = \begin{cases} \frac{1}{6} & \text{si } j = 1 \\ \frac{1}{3} & \text{si } j = 2 \\ \frac{1}{2} & \text{si } j = 3 \end{cases}$$

n=3 particules



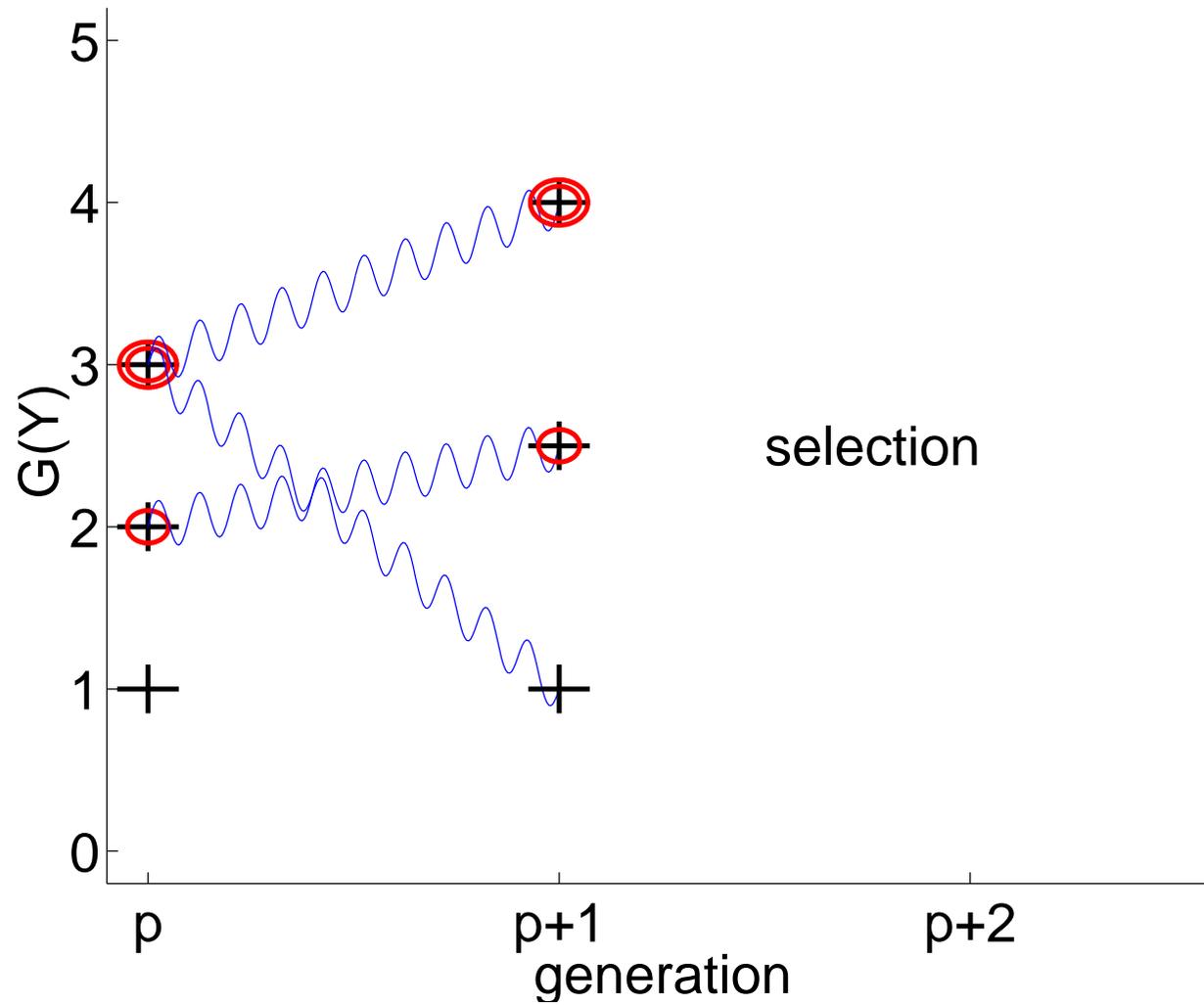
Probabilité de sélectionner la particule j :
$$\frac{G(Y_p^{(j)})}{\sum_{k=1}^3 G(Y_p^{(k)})} = \begin{cases} \frac{1}{6} & \text{si } j = 1 \\ \frac{1}{3} & \text{si } j = 2 \\ \frac{1}{2} & \text{si } j = 3 \end{cases}$$

n=3 particules



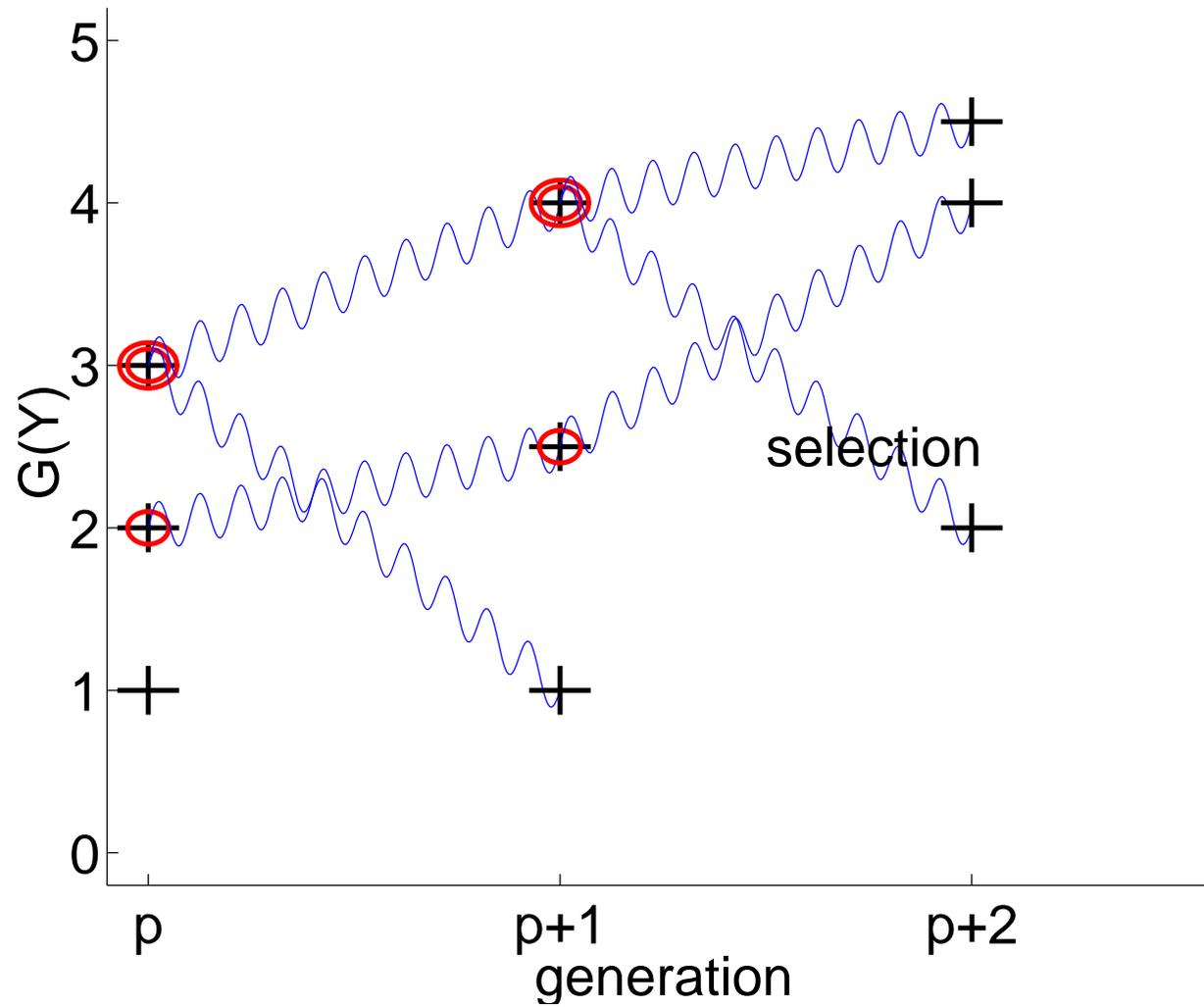
Chaque particule évolue de manière indépendante de p à $p + 1$.

n=3 particules



3 particules sont sélectionnées à la génération $p + 1$.

n=3 particules



Chaque particule évolue de manière indépendante de $p + 1$ à $p + 2$.

A chaque génération $p = 0, \dots, M - 1$:

Sélection : à partir du système $\mathbf{Y}_p = (Y_p^{(k)})_{1 \leq k \leq n}$, on choisit aléatoirement et indépendamment n particules-chemins

$$\widehat{Y}_p^{(k)} = (\widehat{X}_{0,p}^{(k)}, \widehat{X}_{1,p}^{(k)}, \dots, \widehat{X}_{p,p}^{(k)}) \in E^{p+1}$$

selon la loi de Boltzmann-Gibbs

$$\sum_{k=1}^n \frac{G_p(Y_p^{(k)})}{\sum_{j=1}^n G_p(Y_p^{(j)})} \delta_{Y_p^{(k)}}$$

Mutation : chaque particule-chemin sélectionnée $\widehat{Y}_p^{(k)}$ est prolongée par une transition **non-biaisée** :

$$Y_{p+1}^{(k)} = ((\widehat{X}_{0,p}^{(k)}, \dots, \widehat{X}_{p,p}^{(k)}), X_{p+1,p+1}^{(k)}) \in E^{p+1}$$

où $X_{p+1,p+1}^{(k)} = \Phi_{p+1}(\widehat{X}_{p,p}^{(k)}, \Theta_{p+1}^{(k)})$. Les mutations sont indépendantes.

Estimateur de la probabilité de l'événement rare

Soit

$$\hat{P}_n = \prod_{1 \leq p < M} \left[\frac{1}{n} \sum_{k=1}^n G_p(X_{0,p}^{(k)}, \dots, X_{p,p}^{(k)}) \right] \times \left[\frac{1}{n} \sum_{k=1}^n \mathbf{1}_{V(X_{M,M}^{(k)}) \geq a} \prod_{1 \leq p < M} G_p^{-1}(X_{0,p}^{(k)}, \dots, X_{p,p}^{(k)}) \right]$$

\hat{P}_n est un estimateur sans biais de P :

$$\mathbb{E}[\hat{P}_n] = P$$

et convergent :

$$\hat{P}_n \xrightarrow{n \rightarrow \infty} P \quad \text{p.s.}$$

Théorème de la limite centrale

- L'estimateur \hat{P}_n satisfait le théorème de la limite centrale

$$\sqrt{n} \left[\hat{P}_n - P \right] \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, \sigma^2)$$

avec la variance normalisée

$$\sigma^2 = \sum_{p=1}^M \mathbb{E} \left[\prod_{j=1}^p G_j(X_0, \dots, X_j) \right] \mathbb{E} \left[\prod_{j=1}^p G_j^{-1}(X_0, \dots, X_j) (P_p^a)^2(X_p) \right] - P^2$$

Ici les fonctions P_p^a sont définies par

$$x_p \in E \mapsto P_p^a(x_p) = \mathbb{P}(V(X_M) \geq a \mid X_p = x_p)$$

- Utile pour
 - 1) le choix de "bons" potentiels G_p (réduction de variance),
 - 2) la mise au point d'un estimateur de la variance.

Comparaisons de variances pour le modèle gaussien $X_p = X_{p-1} + \theta_p$

où $(\theta_p)_{1 \leq p \leq M}$ indépendants, gaussiens, de moyenne 0, de variance 1,

$$V(x) = x.$$

Ici X_M est gaussien, de moyenne 0 et de variance M :

$$P = \mathbb{P}(X_M \geq a) = \frac{1}{\sqrt{2\pi M}} \int_a^\infty \exp\left(-\frac{s^2}{2M}\right) ds \sim \exp\left(-\frac{a^2}{2M}\right)$$

On considère $a \gg \sqrt{M}$ de telle sorte que $P \ll 1$.

Premier choix pour le potentiel :

$$G_p(x_0, \dots, x_p) = \exp(\alpha x_p), \quad \text{pour un certain } \alpha > 0$$

Les calculs montrent

$$\sigma^2 \simeq \sum_{p=1}^M \left[e^{-\frac{a^2}{n}} e^{\frac{p}{M(M+p)} [a - \alpha M(p-1)/2]^2 + \frac{1}{12} \alpha^2 (p-1)p(p+1)} - P^2 \right]$$

En optimisant, on prend $\alpha = 2a/[M(M-1)]$ et on trouve

$$\sigma^2 \simeq e^{-\frac{a^2}{M}} \frac{2}{3} \left(1 - \frac{1}{M-1}\right)$$

↪ la variance est d'ordre $P^{4/3}/n$ (rappel pour Monte Carlo: P/n)

→ l'erreur relative est $\sim 1/\sqrt{nP^{2/3}}$ (rappel pour Monte Carlo: $\sim 1/\sqrt{nP}$).

On considère le même modèle.

Second choix pour le potentiel :

$$G_p(x_0, \dots, x_p) = \exp[\alpha(x_p - x_{p-1})], \quad \text{pour un certain } \alpha > 0$$

On obtient :

$$\sigma^2 \simeq \sum_{0 \leq p < M} \left[e^{-\frac{a^2}{M}} e^{\frac{p+1}{M(M+p+1)}} \left[a - \alpha \frac{Mp}{p+1} \right]^2 + \alpha^2 \frac{p}{p+1} - P^2 \right]$$

En optimisant, on prend $\alpha = a/M$ et on trouve

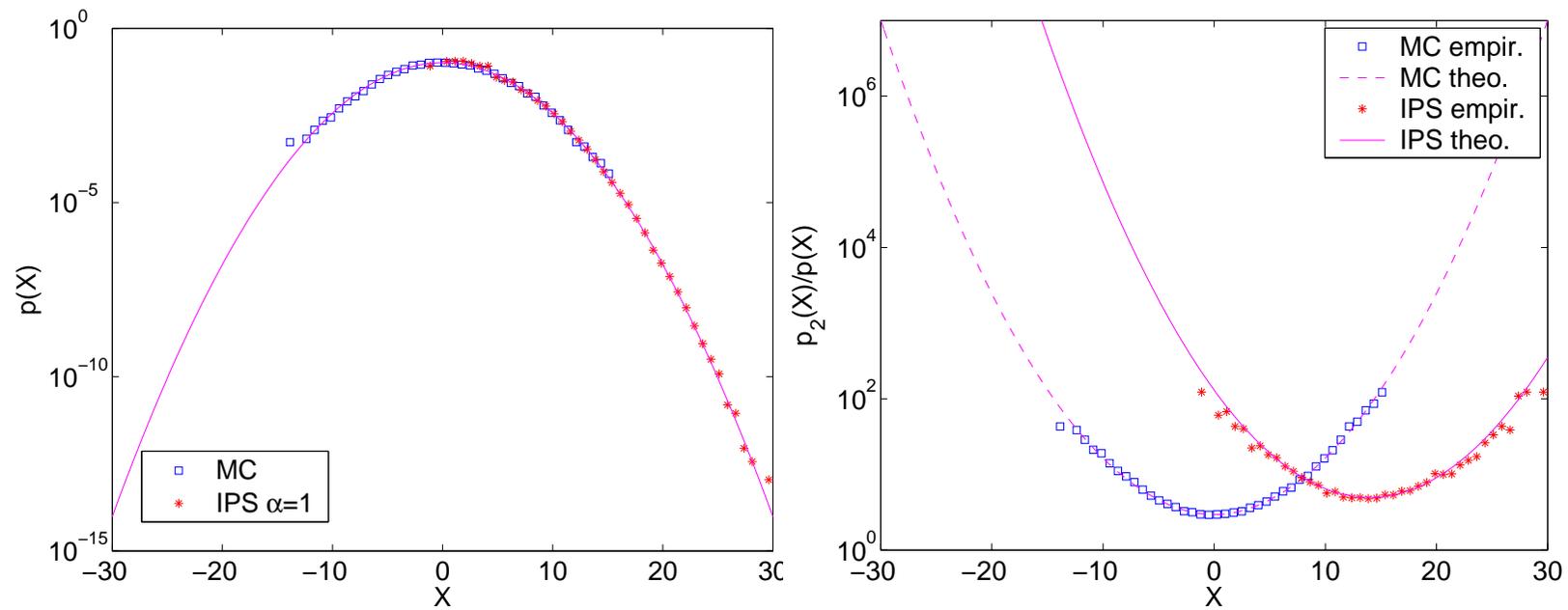
$$\sigma^2 \sim e^{-\frac{a^2}{M}} \left(1 - \frac{1}{M} \right)$$

↪ la variance normalisée est d'ordre P^2/n .

→ l'erreur relative est $\sim 1/\sqrt{n}$.

En comparant avec le cas précédent : **une sélection dépendant de l'état n'est pas efficace !**

Simulations avec le modèle gaussien



$M = 15, n = 2 \cdot 10^4$ particules, $\alpha = 1$.

Exemple de la fibre optique :

$(u(z, t)_{t \in \mathbb{R}})$ = champ électrique après la distance de propagation z .

$\tau(z)^2 = \int |u(z, t)|^2 t^2 dt / \int |u(z, t)|^2 dt$ largeur rms de l'impulsion.

Propagation régie par deux équations de Schrödinger couplées avec des coefficients variant aléatoirement en z (défauts de la fibre).

1) modèle asymptotique (technique de séparation d'échelles)

→ La largeur de l'impulsion $\tau(z)$ est un processus de diffusion

$$\mathcal{L} = 8\sigma^2 \tau \frac{\partial^2}{\partial \tau^2} + 12\sigma^2 \frac{\partial}{\partial \tau}$$

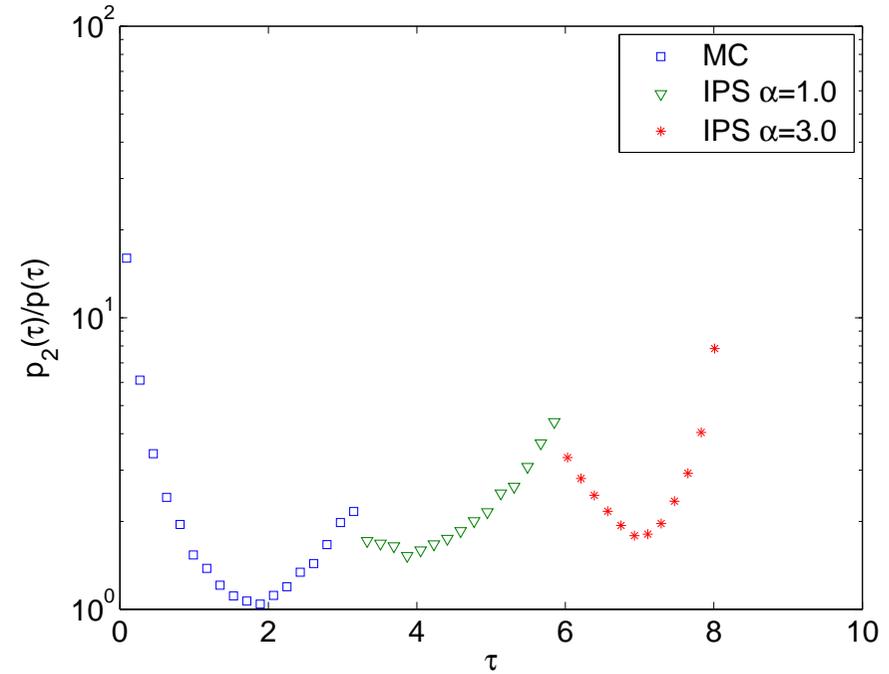
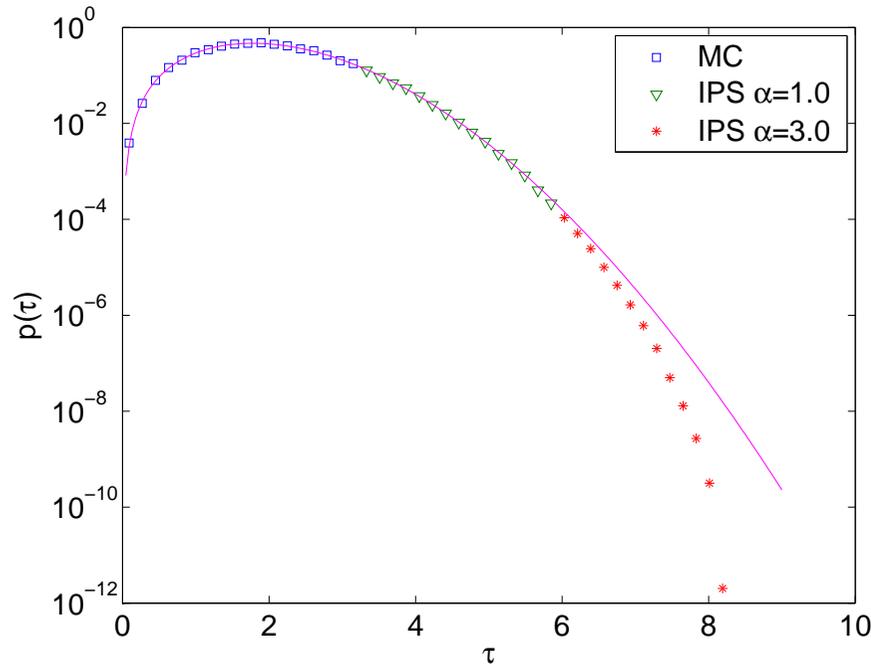
la pdf de $\tau(z)$ est :

$$p_z(\tau) = \frac{\tau^{1/2}}{\sqrt{2\pi}(4\sigma^2 z)^{3/2}} \exp\left(-\frac{\tau}{8\sigma^2 z}\right) \mathbf{1}_{[0, \infty)}(\tau)$$

2) modèle réaliste : pdf impossible à calculer théoriquement.

Observations expérimentales : la queue de la pdf de $\tau(Z)$ à la sortie de la fibre $z = Z$ ne suit pas la distribution Maxwellienne dans des configurations réelles.

Simulations numériques



$n = 15$, $M = 2 \cdot 10^4$ particules, $\alpha = 1$ et $\alpha = 3$.

Trait plein : pdf Maxwellienne prédite par le modèle asymptotique.

Conclusions

- Echantillonnage préférentiel: on biaise l'entrée.

Système de particules en interaction: on sélectionne les particules selon la sortie.

- L'échantillonnage préférentiel est à privilégier quand on a une bonne connaissance du système, mais il faut changer les lois des entrées (méthode intrusive). La méthode est facilement parallélisable.
- Le système est simulé sous sa loi originale (méthode non-intrusive).
 - Aucune intuition (ou étude) n'est nécessaire pour deviner une bonne loi biaisée.
 - Pas besoin de changer le code de simulation.
 - Nombre de particules fixé, coût calcul fixé.
 - La méthode permet de simuler des lois conditionnelles.
 - La méthode est partiellement parallélisable.