

# Modèles & Méthodes Stochastiques pour l'Analyse de Risques.

P. Del Moral (Équipe INRIA ALEA)

INRIA & Bordeaux Mathematical Institute & X CMAP

Horizon Maths, Fondation Sciences Mathématiques de Paris  
AREVA, 12 Déc. , 2011

## Quelques refs et liens

- ▶ Feynman-Kac formulae, Genealogical & Interacting Particle Systems with appl., Springer (2004)
- ▶ Sequential Monte Carlo Samplers for Rare Events. (joint work with Johansen, Doucet) (2006)
- ▶ Branching and Interacting Particle Interpretations of Rare Event Probabilities Springer (2006). [+ Réfs]
- ▶ Plus de références <http://www.math.u-bordeaux1.fr/~delmoral/index.html> [+ Liens]

## Introduction

- Quelques exemples d'actualité
- La notion de risque
- Outils mathématiques
- Formalisation mathématique

## Méthodes de Monte Carlo

- Echantillonnage d'importance
- Méthodes particulières

## Illustrations

- Fiabilité de plateformes offshore en conditions extrêmes
- Sécurité et fiabilité de centrales nucléaires
- Fiabilité de codes de Tardos pour le fingerprinting
- Fiabilité réseaux de communication par fibres optiques
- Risques alimentaires et propagations d'épidémies

## Analyse des performances

- Modèles de simulation particulières
- Estimateurs particulières
- Taux d'erreurs exponentiellement faibles

## Perspectives et projets en cours

## Introduction

Quelques exemples d'actualité

La notion de risque

Outils mathématiques

Formalisation mathématique

## Méthodes de Monte Carlo

## Illustrations

## Analyse des performances

## Perspectives et projets en cours

# Exemples d'actualité

- ▶ **Risques sanitaires** : propagations d'épidémies, développement tumeurs, risques cardio-vasculaires, traitements et risques de rechutes, risques alimentaires, prolifération bactérienne.
- ▶ **Fiabilité de systèmes** : chaînes de production et d'exploitation, plateformes offshore, surcharge de réseaux de communication, sécurité informatique.
- ▶ **Sécurité du trafic aérien** : défaillances, risques de collision.
- ▶ **Sécurité nucléaire** : stockage, fissures de cuves de centrales.
- ▶ **Risques éco. et financiers** : krachs boursiers, défauts de paiement.
- ▶ **Risques naturels** : climatiques, inondations, tremblements de terre
- ▶ .../...

# La notion de risque

- ▶ **Phénomènes peu documentés :**
  - ▶ peu d'observation, manque de données statistiques.
  - ▶ modèles empiriques peu formalisés.
- ▶ **Systemes imprévisibles complexes :**
  - ▶ interaction multi-niveaux, multi-échelles.
  - ▶ systèmes chaotiques et instables.
  - ▶ combinaison d'aléa dans des espaces de grande dimension.

# Outils mathématiques

## ▶ Calibration de modèles :

1. Estimation statistique directe :
  - ▶ Étude empirique sur de faible jeux de données
  - ▶ Analyse et théorie des valeurs extrêmes.
2. Calibration de systèmes formalisés *en fonction d'observation partielles et des événements critiques étudiés* :
  - ▶ Quantification des incertitudes et des aléas.
  - ▶ Estimation des paramètres de modèles.
  - ▶ Formalisation des erreurs de modèles.

## ↪ Méthodes stochastiques :

Filtrage de processus, inférence bayésienne, méthodes de Monte Carlo séquentielles, algorithmes de simulation particulières, méthodes de gradients stochastique.

# Outils mathématiques (suite)

- ▶ **Analyse de propagation d'incertitudes dans ces modèles formalisés.**
  - ▶ Simulation d'événements rares :
    - ▶ estimation des probabilités de réalisation.
    - ▶ simulation des arbres de défauts.
  - ▶ Calcul de mesures de sensibilité :
    - ▶ stabilité et robustesse des estimations.
    - ▶ sensibilité par rapport à des conditions initiales,
    - ▶ sensibilité par rapport à des variations de paramètres.

## ↪ Méthodes stochastiques :

Simulation par des systèmes de particules en interaction (accept.-rejet+recyclage) généalogie des défaillances, branchement multi-niveaux, Calcul de gradient d'énergie libre, lois conditionnelles trajectorielles, échantillonnage d'importance non intrusif, modèles d'îlots particuliers de type génétiques (calcul parallèle).

# Formalisation mathématique

- ▶  $X$  = variable ou processus aléatoire  
(variable statique, trajectoire, excursion)
- ▶  $W$  = les aléas du phénomène  
(erreurs de modèles, incertitudes, variables externes au système)
- ▶  $\Theta$  = paramètre du modèle  
(condition initiale, paramètre cinétique, variable de contrôle)
- ▶  $A$  = un événement critique ( $\mathbb{P}(X \in A) \simeq 10^{-6} - 10^{-12}$ )



# Formalisation mathématique

- ▶  $X$  = variable ou processus aléatoire  
(variable statique, trajectoire, excursion)
- ▶  $W$  = les aléas du phénomène  
(erreurs de modèles, incertitudes, variables externes au système)
- ▶  $\Theta$  = paramètre du modèle  
(condition initiale, paramètre cinétique, variable de contrôle)
- ▶  $A$  = un événement critique ( $\mathbb{P}(X \in A) \simeq 10^{-6} - 10^{-12}$ )

Objectifs	$\mathbb{P}(X \in A)$	et	Loi ( $X \mid X \in A$ ) ?
			$\Updownarrow$
	$\mathbb{P}(W \text{ tels que } X \in A)$	et	Loi ( $W \mid X \in A$ ) ?

# Formalisation mathématique

- ▶  $X$  = variable ou processus aléatoire  
(variable statique, trajectoire, excursion)
- ▶  $W$  = les aléas du phénomène  
(erreurs de modèles, incertitudes, variables externes au système)
- ▶  $\Theta$  = paramètre du modèle  
(condition initiale, paramètre cinétique, variable de contrôle)
- ▶  $A$  = un événement critique ( $\mathbb{P}(X \in A) \simeq 10^{-6} - 10^{-12}$ )

Objectifs  $\mathbb{P}(X \in A)$  et Loi ( $X \mid X \in A$ ) ?



$\mathbb{P}(W \text{ tels que } X \in A)$  et Loi ( $W \mid X \in A$ ) ?

Mesures de sensibilité :

$$\frac{\partial}{\partial \theta} \mathbb{P}(X \in A) \quad \text{et} \quad \frac{\partial}{\partial \theta} \mathbb{E}(f(X) \mid X \in A) \quad ?$$

Introduction

Méthodes de Monte Carlo

Echantillonnage d'importance

Méthodes particulières

Illustrations

Analyse des performances

Perspectives et projets en cours

# Echantillonnage d'importance

$\mathbb{P}(X \in A) = \mathbb{P}_X(A) = 10^{-10} \rightsquigarrow$  Trouver  $\mathbb{P}_Y$  t.q.  $\mathbb{P}_Y(A) = \mathbb{P}(Y \in A) \simeq 1$

$\rightsquigarrow$  Simulation Monte Carlo classique  $Y^i$  i.i.d.  $\mathbb{P}_Y$

$$\int \frac{d\mathbb{P}_X}{d\mathbb{P}_Y}(y) \mathbf{1}_A(y) \mathbb{P}_Y(dy) = \mathbb{P}_X(A) \simeq \mathbb{P}_X^N(A) := \frac{1}{N} \sum_{1 \leq i \leq N} \frac{d\mathbb{P}_X}{d\mathbb{P}_Y}(Y^i) \mathbf{1}_A(Y^i)$$

$\Downarrow$

$$\text{Variance} = \int \frac{d\mathbb{P}_X}{d\mathbb{P}_Y}(x) \mathbf{1}_A(x) \mathbb{P}_X(dx) - \mathbb{P}_X(A)$$

# Echantillonnage d'importance

$\mathbb{P}(X \in A) = \mathbb{P}_X(A) = 10^{-10} \rightsquigarrow$  Trouver  $\mathbb{P}_Y$  t.q.  $\mathbb{P}_Y(A) = \mathbb{P}(Y \in A) \simeq 1$

$\rightsquigarrow$  Simulation Monte Carlo classique  $Y^i$  i.i.d.  $\mathbb{P}_Y$

$$\int \frac{d\mathbb{P}_X}{d\mathbb{P}_Y}(y) \mathbf{1}_A(y) \mathbb{P}_Y(dy) = \mathbb{P}_X(A) \simeq \mathbb{P}_X^N(A) := \frac{1}{N} \sum_{1 \leq i \leq N} \frac{d\mathbb{P}_X}{d\mathbb{P}_Y}(Y^i) \mathbf{1}_A(Y^i)$$

$\Downarrow$

$$\text{Variance} = \int \frac{d\mathbb{P}_X}{d\mathbb{P}_Y}(x) \mathbf{1}_A(x) \mathbb{P}_X(dx) - \mathbb{P}_X(A)$$

## Loi d'importance optimale

$$\text{Variance} = 0 \iff \mathbb{P}_Y(dx) = \frac{\mathbf{1}_A(x)}{\mathbb{P}_X(A)} \mathbb{P}_X(dx) = \mathbb{P}(X \in dx \mid X \in A)$$

# Echantillonnage d'importance (suite)

## Les défauts de la méthode

- ▶ Systèmes complexes  $\Rightarrow$  Choix difficile de la loi d'importance.
- ▶ Problèmes de robustesse : énorme variance si  $Y$  mal choisi.

Ex.: Processus aléatoire  $X = (X_0, \dots, X_n)$

$$\frac{d\mathbb{P}_n}{d\mathbb{Q}_n}(X) := \prod_{k=0}^n \frac{p_k(X_k|X_{k-1})}{q_k(X_k|X_{k-1})} \quad \text{martingale } \geq 0 \rightarrow \text{produit dégénéré}$$

- ▶ **Méthode intrusive : on simule  $Y$  et non  $X$ .**

# Méthodes particulières

**Un événement critique = Une cascade d'événements (moins rares)**

Exemples : *niveaux d'énergie croissants, passerelles physiques/chimiques, excursions de plus en plus critiques, configurations d'entrées de codes numériques conduisant à des sorties critiques, ...*

# Méthodes particulières

**Un événement critique = Une cascade d'événements (moins rares)**

Exemples : *niveaux d'énergie croissants, passerelles physiques/chimiques, excursions de plus en plus critiques, configurations d'entrées de codes numériques conduisant à des sorties critiques, ...*



**Mesures conditionnelles = Lois d'échantillonnage optimales**

$n \rightarrow \eta_n = \text{Loi}(\text{v.a., état ou trajectoire} \mid n \text{ événements intermédiaires})$



**Les proba d'événements rares = constantes de normalisation  $Z_n$**   
**=  $\mathbb{P}(n \text{ événements intermédiaires})$**



## Algorithme de type génétique avec $N$ individus $(\xi_n^i)_{1 \leq i \leq N}$

- ▶ **Explorations/propositions/prédiction/exploration locale** de l'espace des aléas du système (transitions, aléas internes, erreurs de modèles)
- ▶ **Branchements-Selection-Duplication** des prédictions & "individus" simulés conduisant le système dans les régimes critiques supérieurs.

## Algorithme de type génétique avec $N$ individus $(\xi_n^i)_{1 \leq i \leq N}$

- ▶ Explorations/propositions/prédiction/exploration locale de l'espace des aléas du système (transitions, aléas internes, erreurs de modèles)
- ▶ Branchements-Selection-Duplication des prédictions & "individus" simulés conduisant le système dans les régimes critiques supérieurs.

### Estimations particulières:

$$\begin{aligned}\eta_n &= \text{Loi(aléas, états ou trajectoire} \mid n \text{ événements)} \\ &\simeq_{N \uparrow \infty} \eta_n^N := \text{Mesure d'occupation d'un arbre généalogique}\end{aligned}$$

### Estimation non biaisée :

$$\mathcal{Z}_n = \mathbb{P}(n \text{ événements successifs})$$

$$= \mathbb{P}(n\text{-ième} \mid (n-1) \text{ précédents}) \times \dots \times \mathbb{P}(2\text{nd} \mid \text{le premier}) \times \mathbb{P}(\text{le premier})$$

$$\simeq_{N \uparrow \infty} \mathcal{Z}_n^N := [\% \text{ succès } (n-1) \rightsquigarrow n] \times \dots \times [\% \text{ succès } 1 \rightsquigarrow 2] \times [\% \text{ dans } 1]$$

Illustration graphique :  $\eta_n \simeq \eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i}$

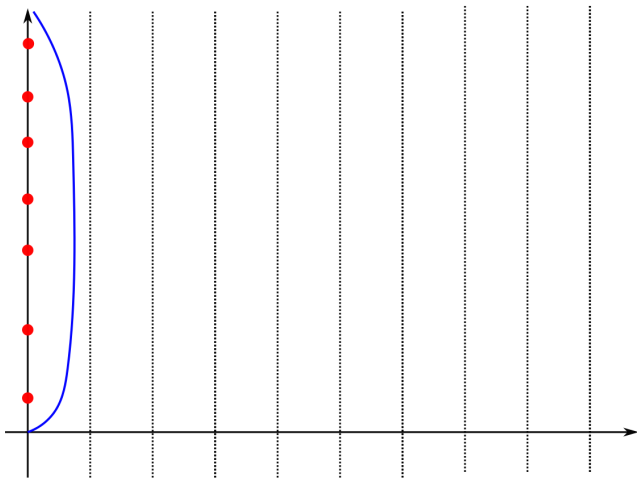


Illustration graphique :  $\eta_n \simeq \eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i}$

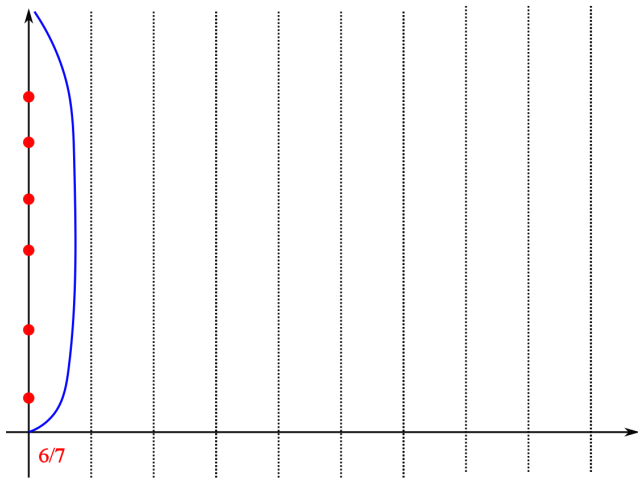


Illustration graphique :  $\eta_n \simeq \eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i}$

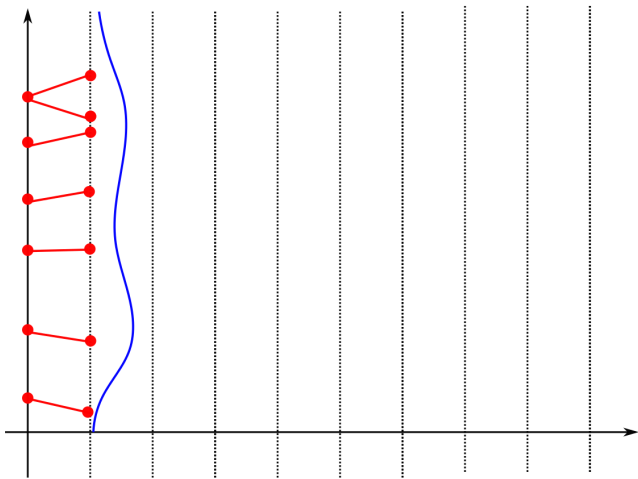


Illustration graphique :  $\eta_n \simeq \eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i}$

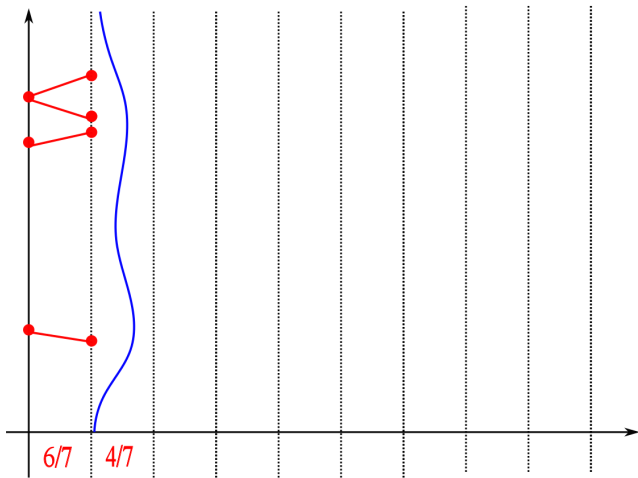


Illustration graphique :  $\eta_n \simeq \eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i}$

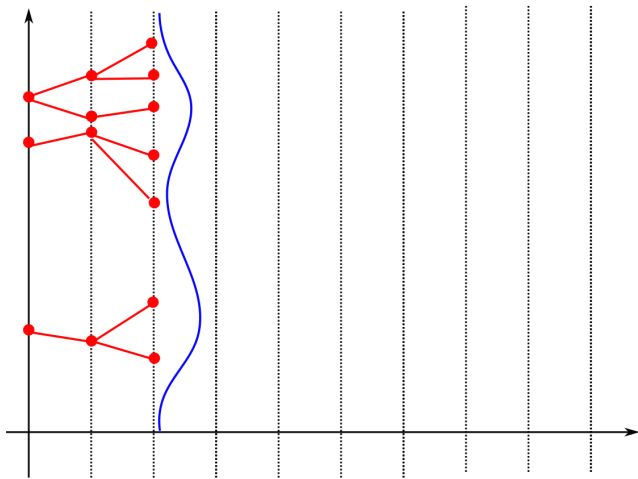


Illustration graphique :  $\eta_n \simeq \eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i}$

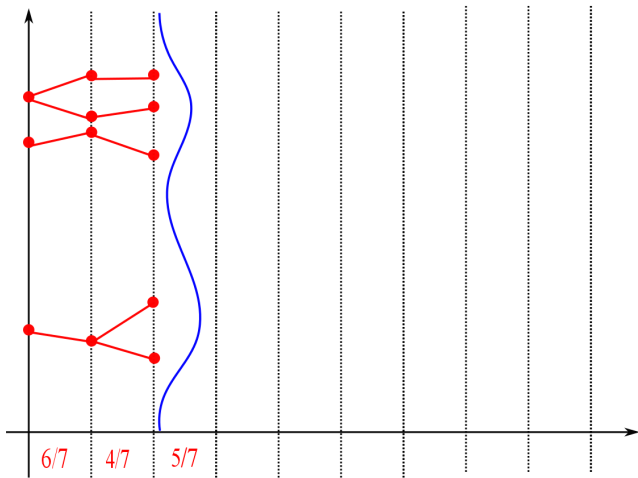




Illustration graphique :  $\eta_n \simeq \eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i}$

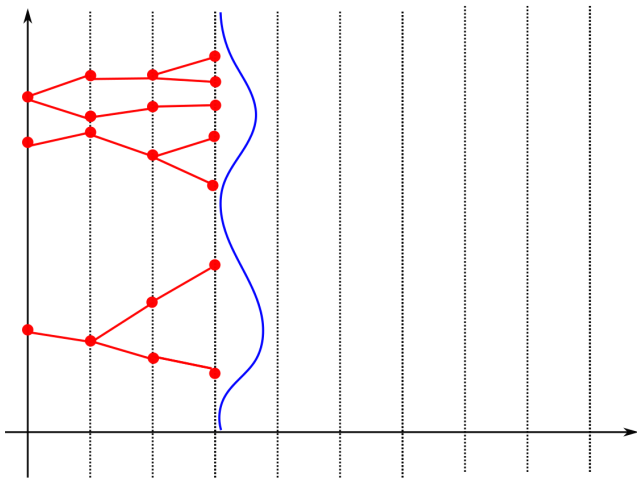




Illustration graphique :  $\eta_n \simeq \eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i}$

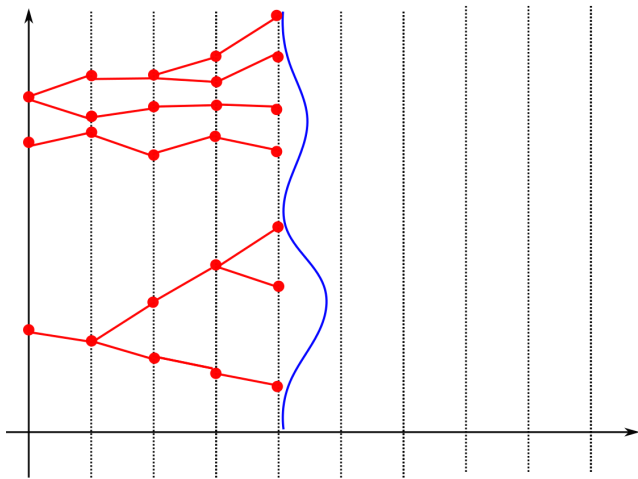


Illustration graphique :  $\eta_n \simeq \eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i}$

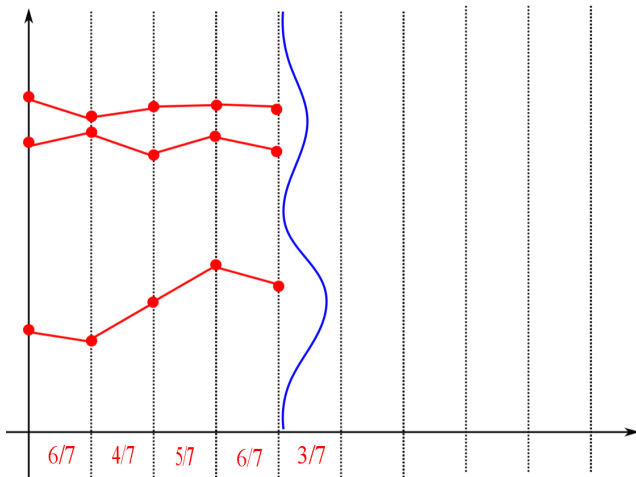


Illustration graphique :  $\eta_n \simeq \eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i}$

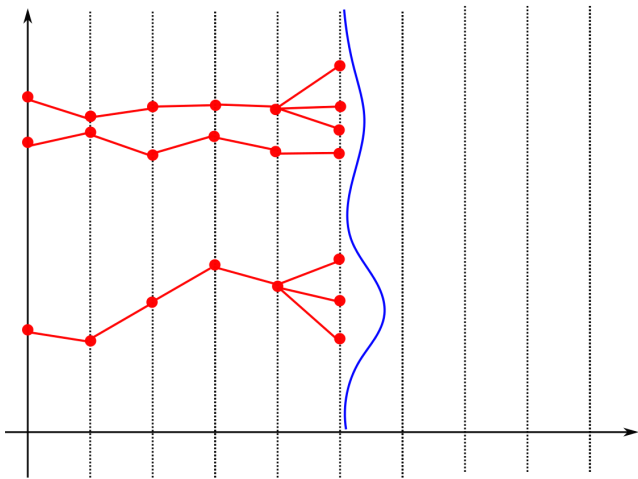




Illustration graphique :  $\eta_n \simeq \eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i}$

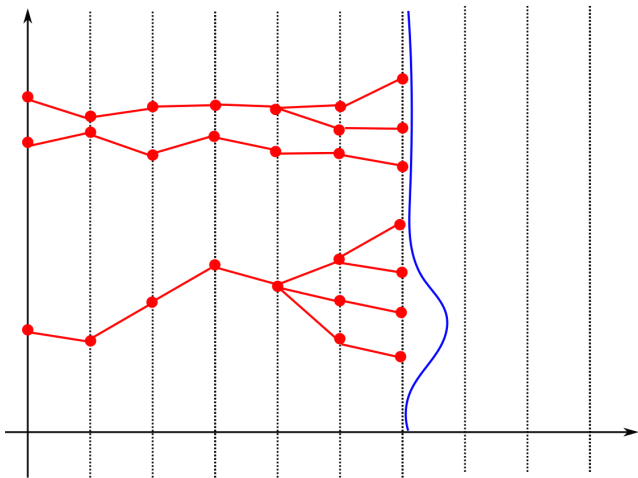


Illustration graphique :  $\eta_n \simeq \eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i}$

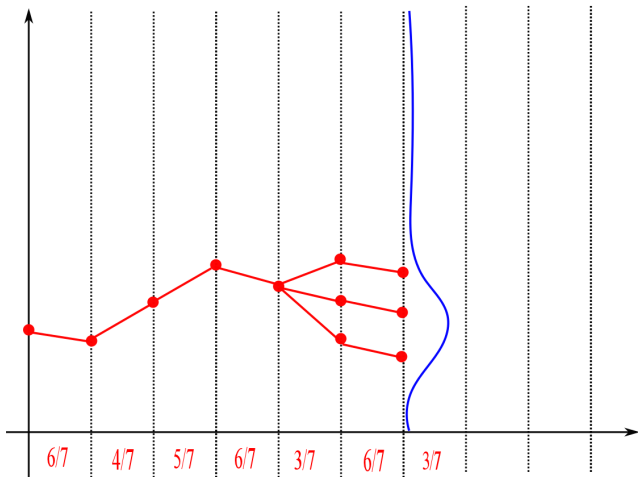




Illustration graphique :  $\eta_n \simeq \eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i}$

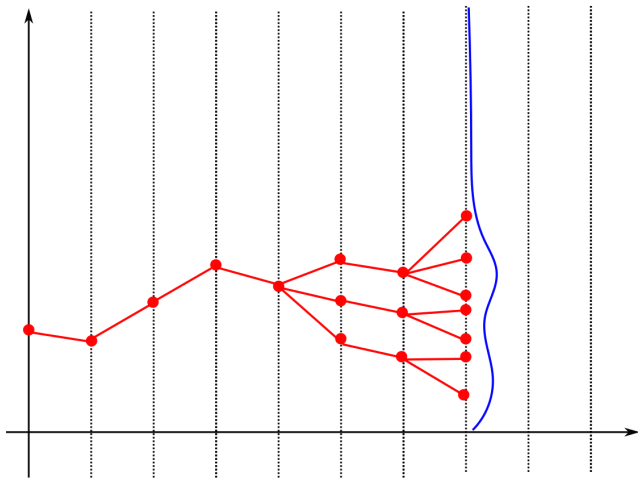


Illustration graphique :  $\eta_n \simeq \eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i}$

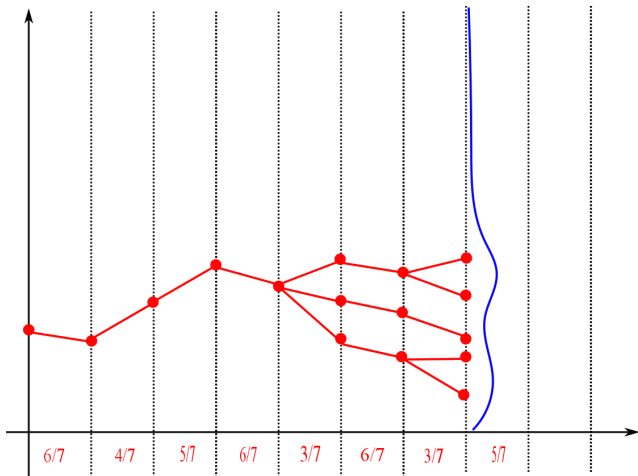


Illustration graphique :  $\eta_n \simeq \eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i}$

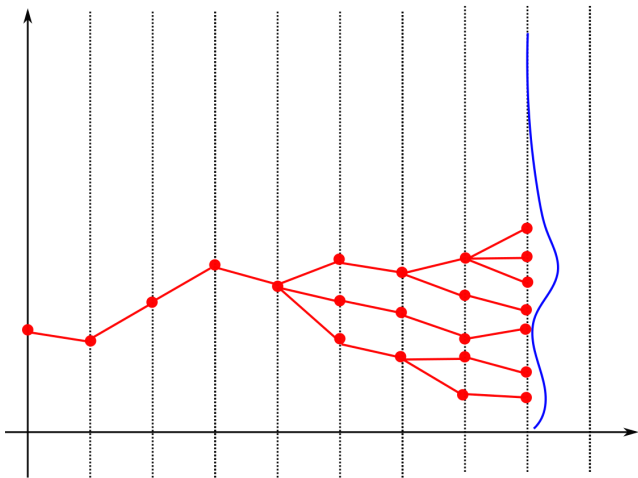


Illustration graphique :  $\eta_n \simeq \eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i}$

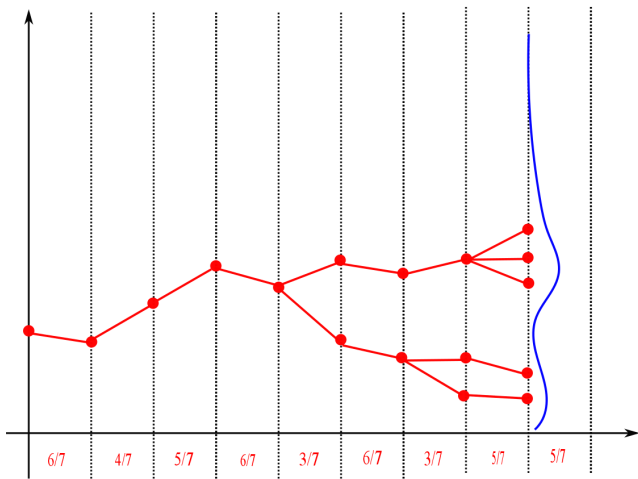
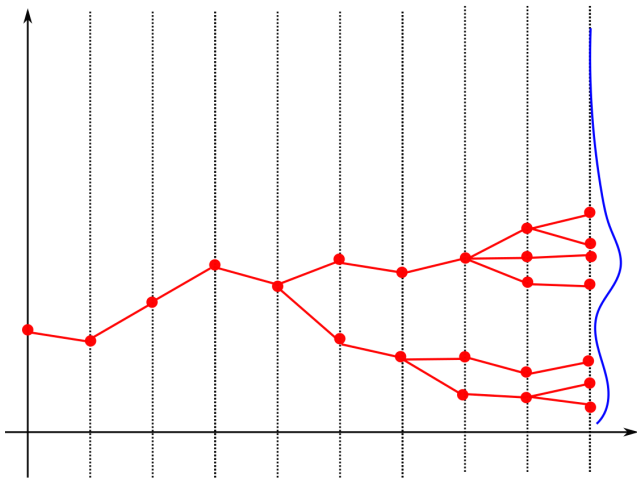


Illustration graphique :  $\eta_n \simeq \eta_n^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_n^i}$



Introduction

Méthodes de Monte Carlo

## Illustrations

Fiabilité de plateformes offshore en conditions extrêmes

Sécurité et fiabilité de centrales nucléaires

Fiabilité de codes de Tardos pour le fingerprinting

Fiabilité réseaux de communication par fibres optiques

Risques alimentaires et propagations d'épidémies

Analyse des performances

Perspectives et projets en cours

# Code IFREMER : hydro. + mécanique



(contrat ALEA-IFREMER 2010)

- ▶ **Entrées  $\mathcal{I}$**  : hauteur de vagues, données météo, courbes de température, incertitudes et paramètres du modèle.  
(ex. de formalisation :  $\mathcal{I} = 2000$  v.a. gaussiennes)
- ▶ **Sortie  $\mathcal{O}$**  : courbes de forces  $F_t(\mathcal{I})$  exercées sur une plateforme offshore (méthanier, plateforme pétrolière).
- ▶ **Événement rare** : dépassement d'un seuil/niveau de force critique  $a$  sur un intervalle de temps  $T$ .

$$A = B(a) = \left\{ \mathcal{I} \text{ telles que } \sup_{0 \leq t \leq T} F_t(\mathcal{I}) \geq a \right\} \rightsquigarrow \text{Loi (Entrées } \mathcal{I} \mid \mathcal{I} \in A \text{ )}$$

# Code IFREMER : hydro. + mécanique



(contrat ALEA-IFREMER 2010)

- ▶ **Entrées  $\mathcal{I}$**  : hauteur de vagues, données météo, courbes de température, incertitudes et paramètres du modèle.  
(ex. de formalisation :  $\mathcal{I} = 2000$  v.a. gaussiennes)
- ▶ **Sortie  $\mathcal{O}$**  : courbes de forces  $F_t(\mathcal{I})$  exercées sur une plateforme offshore (méthanier, plateforme pétrolière).
- ▶ **Événement rare** : dépassement d'un seuil/niveau de force critique  $a$  sur un intervalle de temps  $T$ .

$$A = B(a) = \left\{ \mathcal{I} \text{ telles que } \sup_{0 \leq t \leq T} F_t(\mathcal{I}) \geq a \right\} \rightsquigarrow \text{Loi (Entrées } \mathcal{I} \mid \mathcal{I} \in A \text{)}$$

**Algorithme de branchement multi-niveaux**  $A_n = B(a_n)$  avec  $a_n \uparrow$

- ▶ Exploration de type MCMC des niveaux  $A_n$ .
- ▶ Sélection des individus  $\in A_{n+1}$  ( $\oplus$  choix adaptatif de  $A_{n+1}$ )



# Sécurité et fiabilité de centrales nucléaires

(Contrat CIFRE, ALEA-EDF R&D Risques industriels)



- ▶ **Entrées  $\mathcal{I}$**  : grandeurs/défauts matériaux, température d'injection, incertitudes et paramètres du modèle.  
(ex. de formalisation :  $\mathcal{I} = 20$  v.a. gaussiennes)
- ▶ **Sortie codes neutroniques/mécaniques/hydrau.** :  $\mathcal{O} = \mathcal{F}(\mathcal{I})$
- ▶ **Événement rare :**

$\mathcal{G}(\mathcal{O}) \in \mathbb{R}$  fonct. de risque ( $\mathcal{G}^{-1}\{0\}$  = surface de défaillance)

$$A = \{\mathcal{I} \text{ tels que } \mathcal{G} \circ \mathcal{F}(\mathcal{I}) < 0\} \rightsquigarrow \text{Loi}(\mathcal{I} \mid \mathcal{I} \in A)$$

# Sécurité et fiabilité de centrales nucléaires

(Contrat CIFRE, ALEA-EDF R&D Risques industriels)



- ▶ **Entrées  $\mathcal{I}$**  : grandeurs/défauts matériaux, température d'injection, incertitudes et paramètres du modèle.  
(ex. de formalisation :  $\mathcal{I} = 20$  v.a. gaussiennes)
- ▶ **Sortie codes neutroniques/mécaniques/hydrau.** :  $\mathcal{O} = \mathcal{F}(\mathcal{I})$
- ▶ **Événement rare** :  
 $\mathcal{G}(\mathcal{O}) \in \mathbb{R}$  fonct. de risque ( $\mathcal{G}^{-1}\{0\}$ =surface de défaillance)

$$A = \{\mathcal{I} \text{ tels que } \mathcal{G} \circ \mathcal{F}(\mathcal{I}) < 0\} \rightsquigarrow \text{Loi}(\mathcal{I} \mid \mathcal{I} \in A)$$

**Même algorithme de branchement multi-niveaux  $A_n \downarrow$**

**⊕ Analyse de sensibilité (corrélation entre v.a.,  $\Delta$  paramètres)**

# Watermarking & Digital fingerprinting (détection de copies illégales)



(ANR Nebbiano 06-09, INRIA ALEA-ASPI-TEMICS)

- ▶ **Variables aléatoires** : message caché (empreinte)  
Formalisation : série de bits  $X = (X^1, \dots, X^d) \in \{0, 1\}$ , v.a. Gauss.
- ▶ **Événement rare** : Accuser un innocent (copie illégale).

# Watermarking & Digital fingerprinting (détection de copies illégales)



(ANR Nebbiano 06-09, INRIA ALEA-ASPI-TEMICS)

- ▶ **Variables aléatoires** : message caché (empreinte)  
Formalisation : série de bits  $X = (X^1, \dots, X^d) \in \{0, 1\}$ , v.a. Gauss.
- ▶ **Événement rare** : Accuser un innocent (copie illégale).

Le doc. trouvé contient un code  $y = (y^1, \dots, y^d)$ .

Degré d'accusation à tort  $\sim$  dépassement de niveau d'une fonction :

$$\mathcal{F}(X) = \sum_{1 \leq i \leq d} y^i f(X^i) \geq a \quad \text{Codes Tardos}$$

$$\mathcal{F}(X) = |\langle X, u \rangle| / \|X\| \geq a \quad \text{Zero-bit watermarking } (\|u\| = 1)$$

$$A = \{X \text{ tels que } \mathcal{F}(X) \geq a\} \rightsquigarrow \text{Loi}(X \mid \mathcal{F}(X) \geq a)$$

# Watermarking & Digital fingerprinting (détection de copies illégales)



(ANR Nebbiano 06-09, INRIA ALEA-ASPI-TEMICS)

- ▶ **Variables aléatoires** : message caché (empreinte)  
Formalisation : série de bits  $X = (X^1, \dots, X^d) \in \{0, 1\}^d$ , v.a. Gauss.
- ▶ **Événement rare** : Accuser un innocent (copie illégale).

Le doc. trouvé contient un code  $y = (y^1, \dots, y^d)$ .

Degré d'accusation à tort  $\sim$  dépassement de niveau d'une fonction :

$$\mathcal{F}(X) = \sum_{1 \leq i \leq d} y^i f(X^i) \geq a \quad \text{Codes Tardos}$$

$$\mathcal{F}(X) = |\langle X, u \rangle| / \|X\| \geq a \quad \text{Zero-bit watermarking } (\|u\| = 1)$$

$$A = \{X \text{ tels que } \mathcal{F}(X) \geq a\} \rightsquigarrow \text{Loi}(X \mid \mathcal{F}(X) \geq a)$$

$\Rightarrow$  **Même algorithme de branchement multi-niveaux**  $A_n \downarrow$

$\oplus$  **Analyse de sensibilité (corrélation entre v.a.,  $\Delta$  paramètres)**

# Communication par fibres optiques

- ▶ DM + J. Garnier. Simulations of rare events in fiber optics by interacting particle systems. Optics Communications (2006).
- ▶ DM + J. Garnier. Genealogical particle analysis of rare events. Ann. Appl. Probab. (2005).



## Modèle probabiliste :

- ▶ **Impulsions élémentaires  $X_k$  dans les  $k$ -ième segment de fibre (fictif)** : profils d'impulsions type solitons.
- ▶ **Aléas** :  $\omega_k$  = perturbations aléatoires dans chaque section. (fluctuations des dispersions de vitesses)

$$X_k = \underbrace{F_k(X_{k-1}, \omega_k)}_{\text{éq. de Schrödinger non linéaire}}$$

**Événement rare** : dépassement de niveau d'une grandeur caractéristique  $V(X_T)$  (largeur du profil d'impulsion, dispersion modale,  $\downarrow$  puissance).

$$\mathbb{P}(V(X_T) \geq a) \quad \& \quad \text{Loi}((X_t)_{0 \leq t \leq T} \mid V(X_T) \geq a)$$

# Communication par fibres optiques

**Algorithme particulière =  $N$  particules**  $\xi_t^i = (X_t^i, X_{t+1}^i)$

- ▶ Exploration libre selon les transitions  $X_{t-1} \rightsquigarrow X_t = F_t(X_{t-1}, \omega_t)$ .
- ▶ Sélection des transitions  $\uparrow$  niveaux (critère  $e^{\alpha(V(X_t) - V(X_{t-1}))}$ )

# Communication par fibres optiques

**Algorithme particulière =  $N$  particules**  $\xi_t^i = (X_t^i, X_{t+1}^i)$

- ▶ Exploration libre selon les transitions  $X_{t-1} \rightsquigarrow X_t = F_t(X_{t-1}, \omega_t)$ .
- ▶ Sélection des transitions  $\uparrow$  niveaux (critère  $e^{\alpha(V(X_t) - V(X_{t-1}))}$ )

$\Downarrow$

**Estimation non biaisée des constantes de normalisation**

$$\mathcal{Z}_T = \mathbb{P}(V_T(X_T) \geq a)$$

$$= \mathbb{E} \left( \mathbf{1}_{V_T(X_T) \geq a} e^{\alpha V(X_T)} \left\{ \prod_{0 \leq t \leq T} e^{\alpha(V(X_t) - V(X_{t-1}))} \right\} \right)$$

$$\simeq_{N \uparrow} \left( \frac{1}{N} \sum_{1 \leq i \leq N} \mathbf{1}_{V_T(X_T^i) \geq a} e^{\alpha V(X_T^i)} \right) \times \prod_{0 \leq t \leq T} \frac{1}{N} \sum_{1 \leq i \leq N} e^{\alpha(V(X_t^i) - V(X_{t-1}^i))}$$

$\oplus$

Arbres généalogiques  $\simeq_{N \uparrow}$  Arbre des défaillances



# Risques alimentaires et propagations d'épidémies

- ▶ ANR VIROSCOPY 08-11: Epidemic propagations analysis INRIA-ENST.
- ▶ ARC EPS INRA-INRIA : Eco-microbiologie previsionnelle (09-10).
- ▶ Projet CNRS : ENS Paris et Institut de Maths de Bordeaux (2011-2013).



## Modèles cinétiques d'évolution $\sim$ paramètres inconnus $\Theta$ :

$$X_k = F_k(X_{k-1}, \omega_k, \Theta)$$

- ▶ **Calibration** : observations partielles et bruitées (web, mesures stat.)  
 $\rightsquigarrow$  **filtres particulières, méthodes de Monte Carlo séquentielles.**
- ▶ **Analyse de risques** : dépassement de niveau de grandeurs caractéristiques  $V(X) > a$ , régimes et excursions critiques, ...  
 $\rightsquigarrow$  **Mêmes algorithmes particuliers (branchement, selection)**

Introduction

Méthodes de Monte Carlo

Illustrations

Analyse des performances

- Modèles de simulation particuliers

- Estimateurs particuliers

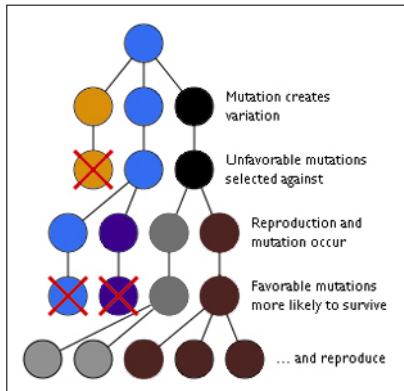
- Taux d'erreurs exponentiellement faibles

Perspectives et projets en cours

# Modèles particuliers de type génétique

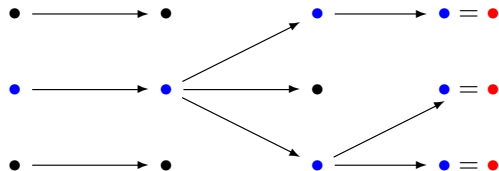
**Algorithme de simulation interactif + adaptatif + universel**

- ▶ **Mutation**-Propositions : transitions de Markov  $X_{n-1} \rightsquigarrow X_n \in E_n$ .
- ▶ **Selection**-Accept/Rejet-recyclage : critères locaux  $\rightsquigarrow$  fonction  $G_n$ .



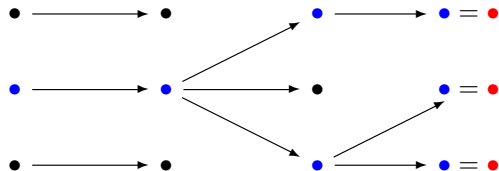
# Modèles d'arbres généalogiques

(Taille population, horizon de temps) =  $(N, n) = (3, 3)$



# Modèles d'arbres généalogiques

(Taille population, horizon de temps) =  $(N, n) = (3, 3)$



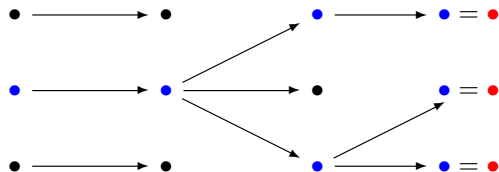
## Théorème (96-98'):

Les lignes ancestrales  $\simeq$  v.a. i.i.d. de loi = mesure de Feynman-Kac

$$Q_n := \frac{1}{Z_n} \left\{ \prod_{0 \leq p < n} G_p(X_p) \right\} \mathbb{P}_n \quad \text{avec} \quad \mathbb{P}_n := \text{Loi}(X_0, \dots, X_n)$$

# Modèles d'arbres généalogiques

(Taille population, horizon de temps) =  $(N, n) = (3, 3)$



## Théorème (96-98'):

Les lignes ancestrales  $\simeq$  v.a. i.i.d. de loi = mesure de Feynman-Kac

$$\mathbb{Q}_n := \frac{1}{Z_n} \left\{ \prod_{0 \leq p < n} G_p(X_p) \right\} \mathbb{P}_n \quad \text{avec} \quad \mathbb{P}_n := \text{Loi}(X_0, \dots, X_n)$$

## Exemple

$$G_n = 1_{A_n} \rightarrow \mathbb{Q}_n = \text{Loi}((X_0, \dots, X_n) \mid X_p \in A_p, p < n)$$

Mesures de sensibilité  $\theta \in \mathbb{R}^{d=1} \rightsquigarrow (X_n^\theta, G_n^\theta, Q_n^\theta)$

**Probabilités de risques = Constantes de normalisation**

$$\mathcal{Z}_n(\theta) = \mathbb{P}(\text{événement critique} \sim \theta) = \mathbb{E} \left( \prod_{0 \leq p < n} G_p^\theta(X_p^\theta) \right)$$

Mesures de sensibilité  $\theta \in \mathbb{R}^{d=1} \rightsquigarrow (X_n^\theta, G_n^\theta, Q_n^\theta)$

**Probabilités de risques = Constantes de normalisation**

$$\mathcal{Z}_n(\theta) = \mathbb{P}(\text{événement critique} \sim \theta) = \mathbb{E} \left( \prod_{0 \leq p < n} G_p^\theta(X_p^\theta) \right)$$

$$\frac{\partial}{\partial \theta} \log \mathcal{Z}_n(\theta) = Q_n(\mathbf{f}_n) = \int_{\mathbf{x}=\text{trajectoires}} Q_n^\theta(\mathbf{d}\mathbf{x}) \mathbf{f}_n(\mathbf{x})$$

avec la fonctionnelle additive

$$\mathbf{f}_n(x_0, \dots, x_n) = \frac{1}{n} \sum_{0 \leq p < n} \frac{\partial}{\partial \theta} \log (p_\theta(x_{p+1}|x_p) G_p^\theta(x_p))$$



Mesures de sensibilité  $\theta \in \mathbb{R}^{d=1} \rightsquigarrow (X_n^\theta, G_n^\theta, Q_n^\theta)$

Probabilités de risques = Constantes de normalisation

$$\mathcal{Z}_n(\theta) = \mathbb{P}(\text{événement critique} \sim \theta) = \mathbb{E} \left( \prod_{0 \leq p < n} G_p^\theta(X_p^\theta) \right)$$

$$\frac{\partial}{\partial \theta} \log \mathcal{Z}_n(\theta) = Q_n(\mathbf{f}_n) = \int_{\mathbf{x}=\text{trajectoires}} Q_n^\theta(\mathbf{d}\mathbf{x}) \mathbf{f}_n(\mathbf{x})$$

avec la fonctionnelle additive

$$\mathbf{f}_n(x_0, \dots, x_n) = \frac{1}{n} \sum_{0 \leq p < n} \frac{\partial}{\partial \theta} \log (p_\theta(x_{p+1}|x_p) G_p^\theta(x_p))$$

↓

**Approximations particulières**

$$\frac{\partial}{\partial \theta} \log \mathcal{Z}_n(\theta) = Q_n^\theta(\mathbf{f}_n) \simeq_{N \uparrow} \frac{1}{N} \sum_{1 \leq i \leq N} \mathbf{f}_n(\text{ligne ancestrale}_n(i))$$

# En résumé

## ⊕ *Arbre généalogique*

$$\begin{aligned} \mathbb{A}_n^N(f_n) &= \frac{1}{N} \sum_{1 \leq i \leq N} f_n(\text{ligne ancestrale}_n(i)) \\ &\simeq_{N \rightarrow \infty} \mathbb{Q}_n(f_n) = \mathbb{E}(f_n(X_0, \dots, X_n) \mid \text{événement critique}) \end{aligned}$$

# En résumé

## ⊕ *Arbre généalogique*

$$\begin{aligned} \mathbb{A}_n^N(f_n) &= \frac{1}{N} \sum_{1 \leq i \leq N} f_n(\text{ligne ancestrale}_n(i)) \\ &\simeq_{N \rightarrow \infty} \mathbb{Q}_n(f_n) = \mathbb{E}(f_n(X_0, \dots, X_n) \mid \text{événement critique}) \end{aligned}$$

## ⊕ *Population courante*

$$\eta_n^N(f_n) := \frac{1}{N} \sum_{1 \leq i \leq N} f_n(\xi_n^i) \xrightarrow{N \rightarrow \infty} \eta_n(f_n) = \mathbb{E}(f_n(X_n) \mid \text{événement critique})$$

# En résumé

## ⊕ *Arbre généalogique*

$$\begin{aligned} \mathbb{A}_n^N(f_n) &= \frac{1}{N} \sum_{1 \leq i \leq N} f_n(\text{ligne ancestrale}_n(i)) \\ &\simeq_{N \rightarrow \infty} \mathbb{Q}_n(f_n) = \mathbb{E}(f_n(X_0, \dots, X_n) \mid \text{événement critique}) \end{aligned}$$

## ⊕ *Population courante*

$$\eta_n^N(f_n) := \frac{1}{N} \sum_{1 \leq i \leq N} f_n(\xi_n^i) \longrightarrow_{N \rightarrow \infty} \eta_n(f_n) = \mathbb{E}(f_n(X_n) \mid \text{événement critique})$$

## ⊕ *Approximation particulière non biaisée*

$$\mathbb{Z}_n^N = \prod_{0 \leq p < n} \frac{1}{N} \sum_{1 \leq i \leq N} G_p(\xi_p^i) \longrightarrow_{N \rightarrow \infty} \mathbb{Z}_n = \mathbb{P}(\text{événement critique})$$

# En résumé

## ⊕ *Arbre généalogique*

$$\begin{aligned} \mathbb{A}_n^N(f_n) &= \frac{1}{N} \sum_{1 \leq i \leq N} f_n(\text{ligne ancestrale}_n(i)) \\ &\underset{N \rightarrow \infty}{\simeq} \mathbb{Q}_n(f_n) = \mathbb{E}(f_n(X_0, \dots, X_n) \mid \text{événement critique}) \end{aligned}$$

## ⊕ *Population courante*

$$\eta_n^N(f_n) := \frac{1}{N} \sum_{1 \leq i \leq N} f_n(\xi_n^i) \xrightarrow{N \rightarrow \infty} \eta_n(f_n) = \mathbb{E}(f_n(X_n) \mid \text{événement critique})$$

## ⊕ *Approximation particulière non biaisée*

$$\mathbb{Z}_n^N = \prod_{0 \leq p < n} \frac{1}{N} \sum_{1 \leq i \leq N} G_p(\xi_p^i) \xrightarrow{N \rightarrow \infty} \mathbb{Z}_n = \mathbb{P}(\text{événement critique})$$

Ex.:  $G_n = 1_{A_n} \rightsquigarrow \mathbb{Z}_n^N = \prod \% \text{ de succès} \xrightarrow{} \mathbb{P}(X_p \in A_p, p < n)$

⊕ Arbre généalogique complet =  $[(\xi_p^i)_{1 \leq i \leq N}, 0 \leq p \leq n]$   
Formulation markovienne à rebours

$$Q_n^N(\xi_0^{i_0}, \xi_1^{i_1}, \dots, \xi_n^{i_n}) := \frac{1}{N} \times \mathbb{M}_n^N(i_n, i_{n-1}) \times \dots \times \mathbb{M}_2^N(i_2, i_1) \times \mathbb{M}_1^N(i_1, i_0)$$

avec les transitions aléatoires : ( $p_n$ =densité des transitions  $x_{n-1} \rightsquigarrow x_n$ )

$$\mathbb{M}_{n+1}^N(i_{n+1}, i_n) = \frac{p_{n+1}(\xi_{n+1}^{i_{n+1}} | \xi_n^{i_n}) G_n(\xi_n^{i_n})}{\sum_{1 \leq k \leq N} p_{n+1}(\xi_{n+1}^{i_{n+1}} | \xi_n^k) G_n(\xi_n^k)}$$

⊕ Arbre généalogique complet =  $[(\xi_p^i)_{1 \leq i \leq N}, 0 \leq p \leq n]$   
 Formulation markovienne à rebours

$$\mathbb{Q}_n^N(\xi_0^{i_0}, \xi_1^{i_1}, \dots, \xi_n^{i_n}) := \frac{1}{N} \times \mathbb{M}_n^N(i_n, i_{n-1}) \times \dots \times \mathbb{M}_2^N(i_2, i_1) \times \mathbb{M}_1^N(i_1, i_0)$$

avec les transitions aléatoires : ( $p_n$ =densité des transitions  $x_{n-1} \rightsquigarrow x_n$ )

$$\mathbb{M}_{n+1}^N(i_{n+1}, i_n) = \frac{p_{n+1}(\xi_{n+1}^{i_{n+1}} | \xi_n^{i_n}) G_n(\xi_n^{i_n})}{\sum_{1 \leq k \leq N} p_{n+1}(\xi_{n+1}^{i_{n+1}} | \xi_n^k) G_n(\xi_n^k)}$$

*Exemple: Intégration de fonctionnelles additives normalisées*

$$\mathbf{f}_n(x_0, \dots, x_n) = \frac{1}{n+1} \sum_{0 \leq p \leq n} f_p(x_p)$$

⇓

$$\mathbb{Q}_n(\mathbf{f}_n) \simeq_{N \uparrow} \mathbb{Q}_n^N(\mathbf{f}_n) := \frac{1}{n+1} \sum_{0 \leq p \leq n} \left[ \frac{1}{N}, \dots, \frac{1}{N} \right] \mathbb{M}_n^N \mathbb{M}_{n-1}^N \dots \mathbb{M}_{p+1}^N \begin{bmatrix} f_p(\xi_p^1) \\ \vdots \\ f_p(\xi_p^N) \end{bmatrix}$$

# Inégalités de concentration

Constantes  $(c_1, c_2)$  liées au (biais, variance)  $\perp n$ ,  $c$  cte universelle,  $\|f_n\| \leq 1$ , et  $\mathbf{F}_n$  fonctionnelle additive normalisée

$\forall (x \geq 0, n \geq 0, N \geq 1, \epsilon \in \{+1, -1\})$ , la probabilité des inégalités suivante est supérieure à  $1 - e^{-x}$  :

$$[\eta_n^N - \eta_n](f_n) \leq \frac{c_1}{N} (1 + x + \sqrt{x}) + \frac{c_2}{\sqrt{N}} \sqrt{x}$$

$$[\mathbb{A}_n^N - \mathbb{Q}_n](f_n) \leq c_1 \frac{n+1}{N} (1 + x + \sqrt{x}) + c_2 \sqrt{\frac{(n+1)}{N}} \sqrt{x}$$

$$[\mathbb{Q}_n^N - \mathbb{Q}_n](\mathbf{F}_n) \leq c_1 \frac{1}{N} (1 + (x + \sqrt{x})) + c_2 \sqrt{\frac{x}{N(n+1)}}$$

$$\frac{\epsilon}{n} \log \frac{\mathcal{Z}_n^N}{\mathcal{Z}_n} \leq \frac{c_1}{N} (1 + x + \sqrt{x}) + \frac{c_2}{\sqrt{N}} \sqrt{x}$$



Introduction

Méthodes de Monte Carlo

Illustrations

Analyse des performances

Perspectives et projets en cours

# Perspectives et projets en cours

- ▶ Arbres généalogiques = Arbres de défauts  
= Trajectoires typiques liées aux événements rares
- ▶ Estimation de paramètres/aléas favorisant l'événement critique



- ▶  $\rightsquigarrow$  Contrôle de processus évoluant dans des régimes critiques
- ▶  $\rightsquigarrow$  Augmentation du degré de fiabilité de systèmes
- ▶  $\rightsquigarrow$  Prédiction des phénomènes en fonction d'observations réelles