

# Rare event simulation - part II

Josselin Garnier

ETICS 2020

# Interacting particle systems for the analysis of rare events

- Problem: estimation of the probability of occurrence of a rare event.
- Simulation by an Interacting Particle System.

Two versions:

- a rare event expressed in terms of the final state of a Markov chain,
- a rare event expressed in terms of a random variable, whose distribution is seen as the stationary distribution of a Markov chain.

## Rare events

- Description of the system:
  - $(\mathbf{X}_p)_{0 \leq p \leq M}$ : a  $E$ -valued Markov chain ( $E = \mathbb{R}, \mathbb{R}^d, \dots$ ):

$$\mathbb{P}(\mathbf{X}_p \in A \mid \mathbf{X}_{p-1} = \mathbf{x}_{p-1}, \dots, \mathbf{X}_0 = \mathbf{x}_0) = \mathbb{P}(\mathbf{X}_p \in A \mid \mathbf{X}_{p-1} = \mathbf{x}_{p-1})$$

- $V : E \rightarrow \mathbb{R}$ : the risk function.
- $a \in \mathbb{R}$ : the threshold level.
- Problem: estimation of the probability

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

when  $a$  is large  $\implies P \ll 1$ .

We know how to simulate the Markov chain  $(\mathbf{X}_p)_{0 \leq p \leq M}$ .

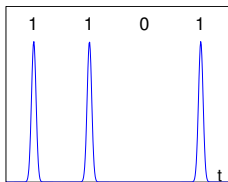
- *Example:*  $X_p = X_{p-1} + \theta_p$ ,  $X_0 = 0$ , where  $\theta_p$  is a sequence of i.i.d.

Gaussian random variables with mean zero and variance one. Here

- $E = \mathbb{R}$ ,
- $V(x) = x$ ,
- the solution is known:  $X_M = V(X_M) \sim \mathcal{N}(0, M)$ .

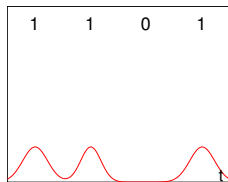
## Example: Communication in transoceanic optical fibers

- Optical fiber transmission principle:
  - a binary message is encoded as a train of short light pulses.
  - the pulse train propagates in a long optical fiber.
  - the message is decoded at the output of the fiber.



Input pulse train

Transmission  
→



Output pulse train

Transmission is perturbed by different random phenomena (amplifier noise, random dispersion, random birefringence, ...).

- Question: estimation of the bit-error-rate (probability of error), typically  $10^{-6}$  or  $10^{-8}$ .
- Answer: use of a big numerical code (but brute-force Monte Carlo too expensive).

## Example: Communication in transoceanic optical fibers

- Physical model:

$(u_0(t))_{t \in \mathbb{R}}$  = initial pulse profile.

$(u(z, t))_{t \in \mathbb{R}}$  = pulse profile after a propagation distance  $z$ .

$(u(Z, t))_{t \in \mathbb{R}}$  = output pulse profile (after a propagation distance  $Z$ ).

Propagation from  $z = 0$  to  $z = Z$  governed by two coupled nonlinear Schrödinger equations with randomly  $z$ -varying coefficients.

↔ black box.

→ Truncation of  $[0, Z]$  into  $M$  segments  $[z_{p-1}, z_p)$ ,  $z_p = pZ/M$ ,  $1 \leq p \leq M$ .

→  $\mathbf{X}_p = u(z_p, t)_{t \in \mathbb{R}}$  is the pulse profile at distance  $z_p$ .

Here  $(\mathbf{X}_p)_{0 \leq p \leq M}$  is Markov with state space  $E = H_0^2(\mathbb{R}) \cap L_2^2(\mathbb{R})$ .

## Example: Communication in transoceanic optical fibers

- Question: estimation of the probability of anomalous pulse spreading.  
Rms pulse width after propagation distance  $z$ :

$$\tau(z)^2 = \int |u(z, t)|^2 t^2 dt / \int |u(z, t)|^2 dt$$

The potential function is  $V : \begin{cases} E \rightarrow \mathbb{R} \\ V(\mathbf{X}) = \int t^2 |\mathbf{X}(t)|^2 dt / \int |\mathbf{X}(t)|^2 dt \end{cases}$

- Problem: estimation of the probability

$$P = \mathbb{P}(\tau(Z) \geq a) = \mathbb{P}(V(\mathbf{X}_M) \geq a)$$

## Monte Carlo method

- $n$  i.i.d copies  $((\mathbf{X}_0^{(k)}, \dots, \mathbf{X}_M^{(k)})_{k=1}^n$  distributed with the original  $\mathbb{P}$ .
- Proposed estimator:

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

- Unbiased estimator:

$$\mathbb{E} \left[ \hat{P}_n \right] = \mathbb{P}(V(\mathbf{X}_M) \geq a) = P$$

- Variance:

$$\mathbb{E} \left[ (\hat{P}_n - P)^2 \right] = \frac{1}{n} P (1 - P) \stackrel{P \ll 1}{\simeq} \frac{P}{n}$$

- Relative error:

$$\frac{\text{Std}(\hat{P}_n)}{P} \simeq \frac{1}{\sqrt{Pn}}$$

↪ We should have  $n > P^{-1}$  to get a relative error smaller than one.

# Importance Sampling method

- $n$  i.i.d copies  $((\mathbf{X}_0^{(k)}, \dots, \mathbf{X}_M^{(k)}))_{k=1}^n$  with the biased distribution  $\mathbb{Q}$ .
- Proposed estimator:

$$\hat{P}_n = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{V(\mathbf{X}_M^{(k)}) \geq a} \frac{d\mathbb{P}}{d\mathbb{Q}}(\mathbf{X}_0^{(k)}, \dots, \mathbf{X}_M^{(k)})$$

Unbiased estimator:

$$\mathbb{E}_{\mathbb{Q}} \left[ \hat{P}_n \right] = \mathbb{E}_{\mathbb{Q}} \left[ \mathbf{1}_{V(\mathbf{X}_M) \geq a} \frac{d\mathbb{P}}{d\mathbb{Q}}(\mathbf{X}_0, \dots, \mathbf{X}_M) \right] = P$$

Variance:

$$\mathbb{E}_{\mathbb{Q}} \left[ (\hat{P}_n - P)^2 \right] = \frac{1}{n} \left\{ \mathbb{E}_{\mathbb{P}} \left[ \mathbf{1}_{V(\mathbf{X}_M) \geq a} \frac{d\mathbb{P}}{d\mathbb{Q}}(\mathbf{X}_0, \dots, \mathbf{X}_M) \right]^2 - P^2 \right\}$$

↪ With a proper choice of  $\mathbb{Q}$ , the error can be dramatically reduced.



# Importance Sampling method

- Optimal choice of  $\mathbb{Q}$ :  $d\mathbb{Q} = \frac{\mathbf{1}_{V(\mathbf{X}_M) \geq a}}{\mathbb{P}(V(\mathbf{X}_M) \geq a)} d\mathbb{P}$ .

Impossible to apply !

But this result gives ideas (adaptive strategies)

- Critical points in the choice of the biased distribution:
  - evaluation of the likelihood ratio,
  - simulation of the biased dynamics (intrusive method).

# Importance Sampling driven by Large Deviations Principle

- Consider the family of biased distributions,  $\lambda > 0$ :

$$d\mathbb{P}^{(\lambda)} = \frac{1}{\mathbb{E}_{\mathbb{P}}[e^{\lambda V(\mathbf{X}_M)}]} e^{\lambda V(\mathbf{X}_M)} d\mathbb{P}$$

$\mathbb{P}^{(\lambda)}$  favors random evolutions with high potential values  $V(\mathbf{X}_M)$ .

- $n$  i.i.d. copies  $(\mathbf{X}_0^{(k)}, \dots, \mathbf{X}_M^{(k)})_{1 \leq k \leq n}$  distributed with  $\mathbb{P}^{(\lambda)}$ .
- Estimator:

$$\hat{P}_{n,\lambda} = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{V(\mathbf{X}_M^{(k)}) \geq a} \frac{d\mathbb{P}}{d\mathbb{P}^{(\lambda)}}(\mathbf{X}_0^{(k)}, \dots, \mathbf{X}_M^{(k)})$$

- Variance:

$$\mathbb{E}_{\mathbb{P}^{(\lambda)}} \left[ (\hat{P}_{n,\lambda} - P)^2 \right] = \frac{1}{n} \left\{ \mathbb{E}_{\mathbb{P}} \left[ \mathbf{1}_{V(\mathbf{X}_M) \geq a} e^{-\lambda V(\mathbf{X}_M)} \right] \mathbb{E}_{\mathbb{P}}[e^{\lambda V(\mathbf{X}_M)}] - P^2 \right\}$$

# Importance Sampling driven by Large Deviations Principle

- Variance:

$$\begin{aligned} n\mathbb{E}_{\mathbb{P}^{(\lambda)}} \left[ (\hat{P}_{n,\lambda} - P)^2 \right] &= \mathbb{E}_{\mathbb{P}} \left[ \mathbf{1}_{V(\mathbf{X}_M) \geq a} e^{-\lambda V(\mathbf{X}_M)} \right] \mathbb{E}_{\mathbb{P}} [e^{\lambda V(\mathbf{X}_M)}] - P^2 \\ &\leq e^{-[\lambda a - \Lambda_M(\lambda)]} P - P^2 \end{aligned}$$

where  $\Lambda_M(\lambda) = \log \mathbb{E}_{\mathbb{P}} [e^{\lambda V(\mathbf{X}_M)}]$ .

- For a judicious choice of  $\lambda$ ,

$$\lambda^* a - \Lambda_M(\lambda^*) = \sup_{\lambda > 0} [\lambda a - \Lambda_M(\lambda)] \simeq -\ln P$$

(Cramér's theorem, large deviations principle), so

$$\mathbb{E}_{\mathbb{P}^{(\lambda)}} [(\hat{P}_{n,\lambda} - P)^2] \lesssim \frac{P^2}{n}$$

- Almost optimal: the relative error is  $1/\sqrt{n}$  (compare with MC:  $1/\sqrt{Pn}$ ).

## Feynman-Kac path measures

*Question:* How to simulate the biased distribution  $\mathbb{P}^{(\lambda)}$  ?

*Answer:* We will show a way to simulate the distribution  $\mathbb{Q}$ :

$$d\mathbb{Q} = \frac{1}{\mathcal{Z}_M} \left\{ \prod_{p=0}^M G_p(\mathbf{X}_0, \dots, \mathbf{X}_p) \right\} d\mathbb{P}$$

where  $(G_p)_{1 \leq p \leq M}$  is a sequence of positive potential functions on the path spaces  $E^p$ , and  $\mathcal{Z}_M = \mathbb{E}_{\mathbb{P}}[\prod G_p(\mathbf{X}_0, \dots, \mathbf{X}_p)] > 0$  is a normalization constant.

Examples:

- $G_p(\mathbf{X}_0, \dots, \mathbf{X}_p) = 1$ ,  $p < M$ ,  $G_M(\mathbf{X}_0, \dots, \mathbf{X}_M) = e^{\lambda V(\mathbf{X}_M)}$ .
- $G_p(\mathbf{X}_0, \dots, \mathbf{X}_p) = e^{\lambda(V(\mathbf{X}_p) - V(\mathbf{X}_{p-1}))}$ .

- What is a “good” choice for  $G_p$  ?
- How to simulate  $\mathbb{Q}$  directly from  $\mathbb{P}$  ?

## Original measures

- $(\mathbf{X}_p)_{0 \leq p \leq M}$ : a  $E$ -valued Markov chain, starting from  $\mathbf{X}_0 = \mathbf{x}_0$ , with transition  $K_p(\mathbf{x}_{p-1}, d\mathbf{x}_p)$ :

$$\mathbb{P}(\mathbf{X}_p \in A \mid \mathbf{X}_{p-1} = \mathbf{x}_{p-1}, \dots, \mathbf{X}_0 = \mathbf{x}_0) = \int_A K_p(\mathbf{x}_{p-1}, d\mathbf{x}_p)$$

where  $K_p(\mathbf{x}_{p-1}, \cdot)$  is a probability measure for any  $\mathbf{x}_{p-1} \in E$ .

- Denote the (partial) path

$$\mathbf{Y}_p =_{\text{def.}} (\mathbf{X}_0, \dots, \mathbf{X}_p) \in E^{p+1}, \quad p = 0, \dots, M$$

The measure  $\mu_p$  on  $E^{p+1}$  is the distribution of  $\mathbf{Y}_p$ :

$$\mu_p(f_p) =_{\text{def.}} \int_{E^{p+1}} f_p(\mathbf{y}_p) \mu_p(d\mathbf{y}_p) = \mathbb{E}[f_p(\mathbf{Y}_p)], \quad f_p \in L^\infty(E^{p+1})$$

- Expression of  $P$  in terms of  $\mu_M$ :

$$P = \mu_M(f)$$

$$f(\mathbf{y}_M) = f(\mathbf{x}_0, \dots, \mathbf{x}_M) = \mathbf{1}_{V(\mathbf{x}_M) \geq a}$$

→ If one can compute/estimate  $\mu_M$ , then one can compute/estimate  $P$ .

# Unnormalized Feynman-Kac measures

$$Y_p \stackrel{\text{def.}}{=} (\mathbf{X}_0, \dots, \mathbf{X}_p) \in E^{p+1}, \quad p = 0, \dots, M$$

Feynman-Kac measure  $\gamma_p$  associated to the pair potential/transition  $(G_p, K_p)$ :

$$\gamma_p(f_p) = \mathbb{E} \left[ f_p(\mathbf{Y}_p) \prod_{0 \leq k < p} G_k(\mathbf{Y}_k) \right]$$

- Expression of  $P$  in terms of  $\gamma_M$ :

$$P = \gamma_M(g)$$

$$g(\mathbf{y}_M) = g(\mathbf{x}_0, \dots, \mathbf{x}_M) = \mathbf{1}_{V(\mathbf{x}_M) \geq a} \prod_{0 \leq p < M} G_p^{-1}(\mathbf{x}_0, \dots, \mathbf{x}_p)$$

→ If one can compute/estimate  $\gamma_M$ , then one can compute/estimate  $P$ .

## Normalized Feynman-Kac measures

Introduce the normalized measure  $\eta_p$ :

$$\eta_p(f_p) = \gamma_p(f_p) / \gamma_p(1), \quad p = 0, \dots, M$$

- Expression of  $P$  in terms of  $\eta_p$ :

$$P = \eta_M(g) \prod_{0 \leq p < M} \eta_p(G_p)$$

*Proof:*

$$P = \mathbb{E} \left[ g(\mathbf{Y}_M) \prod_{0 \leq k < M} G_k(\mathbf{Y}_k) \right] = \gamma_M(g) = \eta_M(g) \gamma_M(1)$$

Normalizing constant:

$$\gamma_M(1) = \gamma_{M-1}(G_{M-1}) = \eta_{M-1}(G_{M-1}) \gamma_{M-1}(1) = \prod_{0 \leq p < M} \eta_p(G_p)$$

□

→ If one can compute/estimate  $(\eta_p)_{p=0}^M$ , then one can compute/estimate  $P$ .

# Interacting path-particle system

- Goal: simulate the original measures  $\mu_p$

$$\mu_p(f_p) = \mathbb{E}[f_p(\mathbf{Y}_p)]$$

- Easy: Let  $(\mathbf{Y}_p^{(1)}, \dots, \mathbf{Y}_p^{(n)}) \in (E^{p+1})^n$  be i.i.d. Markov chains simulated with  $\mathbb{P}$ . Then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \delta_{\mathbf{Y}_p^{(i)}} = \mu_p$$

- Goal: simulate the normalized measures  $\eta_p$

$$\eta_p(f_p) = \frac{\mathbb{E}\left[f_p(\mathbf{Y}_p) \prod_{0 \leq k < p} G_k(\mathbf{Y}_k)\right]}{\mathbb{E}\left[\prod_{0 \leq k < p} G_k(\mathbf{Y}_k)\right]}$$

- Idea:  $\mathbb{Y}_p = (\mathbf{Y}_p^{(1)}, \dots, \mathbf{Y}_p^{(n)}) \in (E^{p+1})^n$  particle system s.t.

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \delta_{\mathbf{Y}_p^{(i)}} = \eta_p$$



# Interacting path-particle system

*Question:* How to simulate  $\eta_M$  directly from  $\mathbb{P}$  ?

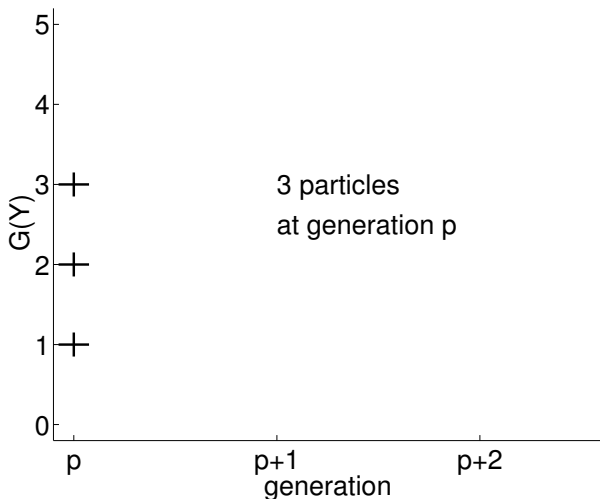
$$d\eta_M = \frac{1}{Z_M} \left\{ \prod_{p=0}^{M-1} G_p(\mathbf{X}_0, \dots, \mathbf{X}_p) \right\} d\mathbb{P}$$

*Answer:* System of path-particles, whose empirical measure will be approximately  $\mathbb{Q}$ .

- Path-particle:  $\mathbf{Y}_p = (\mathbf{X}_0, \dots, \mathbf{X}_p)$  taking values in  $E^{p+1}$ ,  $0 \leq p \leq M$ .
- System of  $n$  path-particles:  $\mathbb{Y}_p = (\mathbf{Y}_p^{(i)})_{1 \leq i \leq n}$  taking values in  $(E^{p+1})^n$ .
  - Initialization:  $p = 0$ :  $\mathbf{Y}_0^{(i)} = \mathbf{x}_0$  for all  $i = 1, \dots, n$ .
  - Dynamics: Evolution from generation  $p$  to  $p + 1$  as follows:

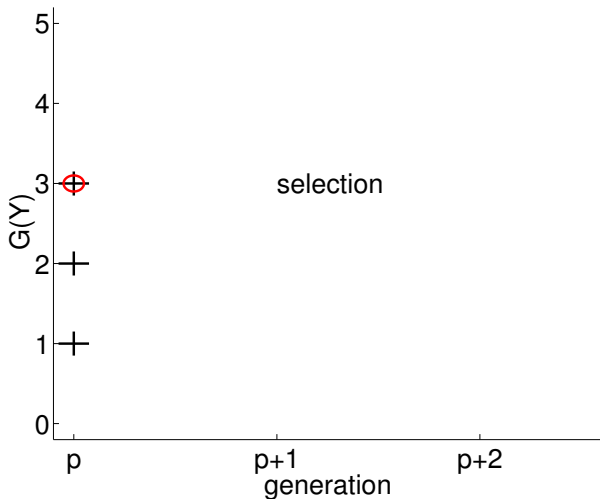
$$\mathbb{Y}_p \in (E^{p+1})^n \xrightarrow{\text{selection}} \widehat{\mathbb{Y}}_p \in (E^{p+1})^n \xrightarrow{\text{mutation}} \mathbb{Y}_{p+1} \in (E^{p+2})^n$$

$n=3$  particles



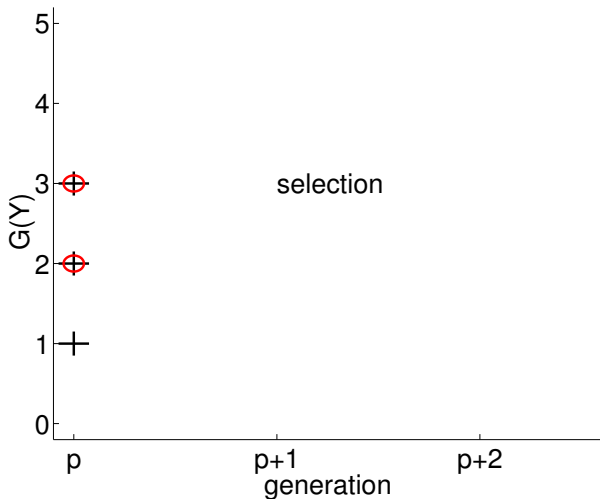
3 particles  $\mathbf{Y}_p^{(1)}$ ,  $\mathbf{Y}_p^{(2)}$ ,  $\mathbf{Y}_p^{(3)}$  at generation  $p$ ,  
with potential weights  $G(\mathbf{Y}_p^{(1)}) = 1$ ,  $G(\mathbf{Y}_p^{(2)}) = 2$ ,  $G(\mathbf{Y}_p^{(3)}) = 3$ .

n=3 particles



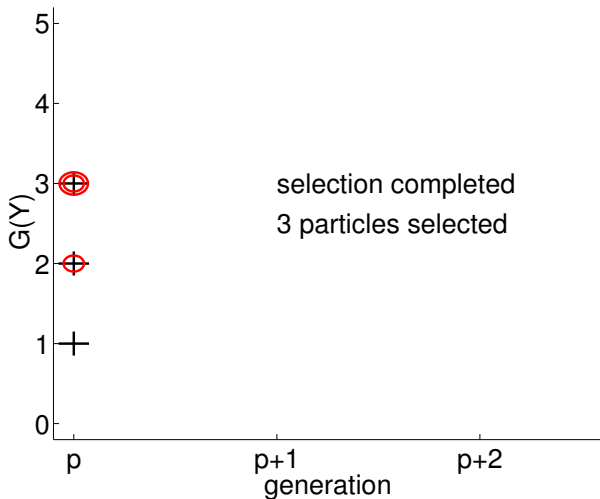
$$\text{Prob. to select part. } j: \frac{G(\mathbf{Y}_p^{(j)})}{G(\mathbf{Y}_p^{(1)}) + G(\mathbf{Y}_p^{(2)}) + G(\mathbf{Y}_p^{(3)})} = \begin{cases} \frac{1}{6} & \text{if } j = 1 \\ \frac{1}{3} & \text{if } j = 2 \\ \frac{1}{2} & \text{if } j = 3 \end{cases}$$

n=3 particles



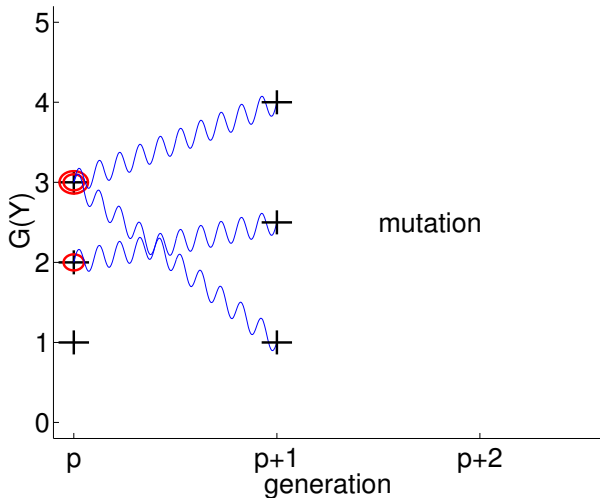
$$\text{Prob. to select part. } j: \frac{G(\mathbf{Y}_p^{(j)})}{G(\mathbf{Y}_p^{(1)}) + G(\mathbf{Y}_p^{(2)}) + G(\mathbf{Y}_p^{(3)})} = \begin{cases} \frac{1}{6} & \text{if } j = 1 \\ \frac{1}{3} & \text{if } j = 2 \\ \frac{1}{2} & \text{if } j = 3 \end{cases}$$

n=3 particles



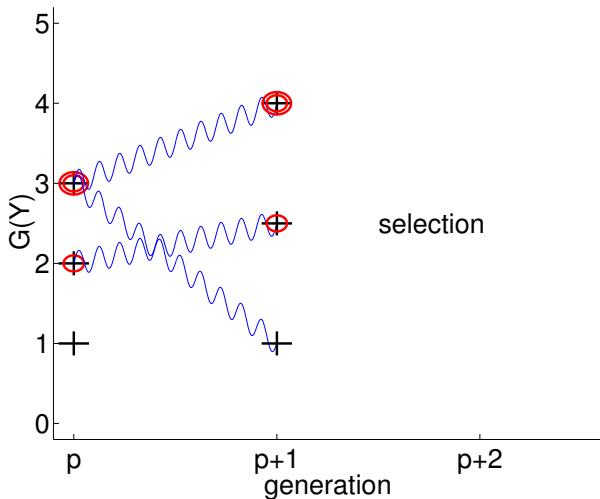
$$\text{Prob. to select part. } j: \frac{G(\mathbf{Y}_p^{(j)})}{G(\mathbf{Y}_p^{(1)}) + G(\mathbf{Y}_p^{(2)}) + G(\mathbf{Y}_p^{(3)})} = \begin{cases} \frac{1}{6} & \text{if } j = 1 \\ \frac{1}{3} & \text{if } j = 2 \\ \frac{1}{2} & \text{if } j = 3 \end{cases}$$

$n=3$  particles

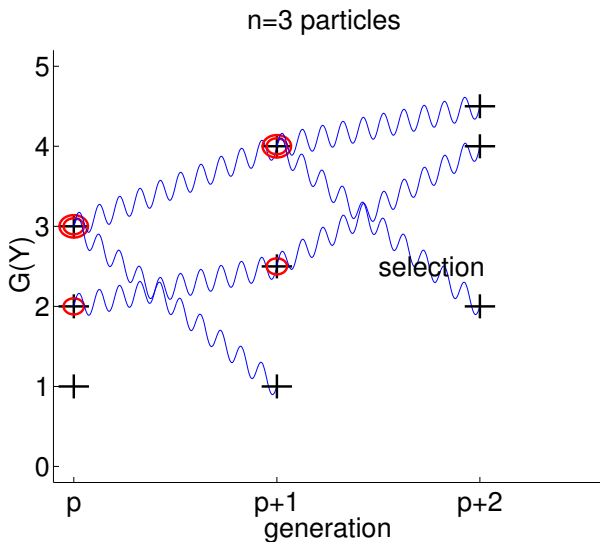


Each particle evolves independently from  $p$  to  $p + 1$ .

$n=3$  particles



3 particles are selected at generation  $p + 1$ .



Each particle evolve independently from  $p + 1$  to  $p + 2$ .



At each generation  $p = 0, \dots, M - 1$ :

**Selection:** from the system  $\mathbb{Y}_p = (\mathbf{Y}_p^{(i)})_{1 \leq i \leq n}$ , choose randomly and independently  $n$  path-particles

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

according to the Boltzmann-Gibbs particle measure

$$\sum_{i=1}^n \frac{G_p(\mathbf{Y}_p^{(i)})}{\sum_{j=1}^n G_p(\mathbf{Y}_p^{(j)})} \delta_{\mathbf{Y}_p^{(i)}}$$

**Mutation:** each selected path-particle  $\widehat{\mathbf{Y}}_p^{(i)}$  is extended by an elementary unbiased  $K_p$ -transition:

$$\begin{aligned} \mathbf{Y}_{p+1}^{(i)} &= ( (\mathbf{Y}_{0,p+1}^{(i)}, \dots, \mathbf{Y}_{p,p+1}^{(i)}) , \mathbf{Y}_{p+1,p+1}^{(i)} ) \\ &= ((\widehat{\mathbf{Y}}_{0,p}^{(i)}, \dots, \widehat{\mathbf{Y}}_{p,p}^{(i)}), \mathbf{Y}_{p+1,p+1}^{(i)}) \in E^{p+1} \end{aligned}$$

where  $\mathbf{Y}_{p+1,p+1}^{(i)}$  is a random variable with distribution  $K_p(\widehat{\mathbf{Y}}_{p,p}^{(i)}, \cdot)$ . The mutations are performed independently.

- The occupation measures of the ancestral lines converge to the desired measures:

$$\eta_p^n =_{\text{def.}} \frac{1}{n} \sum_{i=1}^n \delta_{(\mathbf{Y}_{0,p}^{(i)}, \dots, \mathbf{Y}_{p,p}^{(i)})} \xrightarrow{n \rightarrow \infty} \eta_p$$

In addition, several propagation-of-chaos estimates ensure that the ancestral lines  $\mathbf{Y}_p^{(i)} = (\mathbf{Y}_{0,p}^{(i)}, \dots, \mathbf{Y}_{p,p}^{(i)})$  are asymptotically i.i.d. with common distribution  $\eta_p$ .

- Estimator of  $P = \eta_M(g) \prod_{0 \leq p < M} \eta_p(G_p)$ :

$$\hat{P}_n = \eta_M^n(g) \prod_{0 \leq p < M} \eta_p^n(G_p)$$

$$g(\mathbf{x}_0, \dots, \mathbf{x}_M) = \mathbf{1}_{V(\mathbf{x}_M) \geq a} \prod_{0 \leq p < M} G_p^{-1}(\mathbf{x}_0, \dots, \mathbf{x}_p)$$

[cf. P. Del Moral and J. Garnier, Ann. Appl. Probab. **15** (2005), 2496-2534.]

## Efficient implementation of the selection step

- Let  $(x_j)_{j=1}^n$  be points and  $(w_j)_{j=1}^n$  be weights such that  $\sum_{i=1}^n w_i = 1$ . We want to sample  $(Y_j)_{j=1}^n$  i.i.d. with the distribution  $\sum_{i=1}^n w_i \delta_{x_i}$ .
- Let  $(Z_j)_{j=1}^{n+1}$  be i.i.d. random variables with distribution  $\mathcal{E}(1)$ .

Set  $U_j = \sum_{i=1}^j Z_i / \sum_{i=1}^{n+1} Z_i$ .

Result 1: The vector  $(U_1, \dots, U_n)$  has the same distribution as the order statistics of a vector of i.i.d. r.v. with distribution  $\mathcal{U}([0, 1])$ .

- Set  $W_0 = 0$ ,  $W_k = \sum_{i=1}^k w_i$ .

Let  $Y_j = x_k$  if  $W_{k-1} < U_j \leq W_k$  for  $j = 1, \dots, n$ .

Result 2 : The  $(Y_j)_{j=1}^n$  are i.i.d. with the distribution  $\sum_{i=1}^n w_i \delta_{x_i}$ .

k=1;

for j=1:n

    Y(j)=x(k);

    while W(k)<U(j)

        k=k+1;

        Y(j)=x(k);

    end

end

# Estimator of the probability of the rare event

- Let

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

- $\hat{P}_n$  is an unbiased estimator of  $P$ :

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

such that

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

# Central limit theorem

- The estimator  $\hat{P}_n$  satisfies the central limit theorem

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

with the asymptotic variance

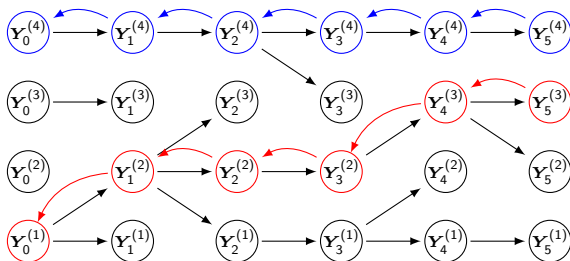
$$\sigma^2 = \sum_{p=0}^M \mathbb{E} \left[ \prod_{j=0}^{p-1} G_j \right] \mathbb{E} \left[ \prod_{j=0}^{p-1} G_j^{-1} (P_{p,M}^a)^2 \right] - P^2$$

Here the functions  $P_{p,M}^a$  are defined by

$$\mathbf{x}_p \in E \mapsto P_{p,M}^a(\mathbf{x}_p) = \mathbb{P}(V(\mathbf{X}_M) \geq a \mid \mathbf{X}_p = \mathbf{x}_p)$$

- Useful for
  - 1) the choice of “good” functions  $G_p$  (variance reduction)
  - 2) the design of an estimator of the asymptotic variance.

## Empirical estimator of the asymptotic variance

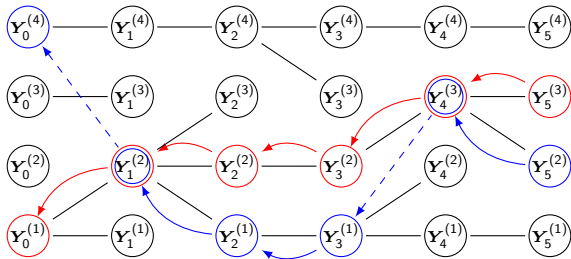


An IPS with two lineages with distinct parents  $a(4) = 4$  and  $a(3) = 1$ .

- Unbiased and weakly consistent estimator of  $\sigma^2$ :

$$\hat{\sigma}_n^2 = \left\{ \prod_{0 \leq p < M} \left[ \frac{1}{n} \sum_{i=1}^n G_p(\mathbf{Y}_p^{(i)}) \right]^2 \right\} \left\{ \frac{1}{n} \left( \sum_{i=1}^n g(\mathbf{Y}_M^{(i)}) \right)^2 - \frac{n^M}{(n-1)^{M+1}} \sum_{i,j, a(i) \neq a(j)} g(\mathbf{Y}_M^{(i)}) g(\mathbf{Y}_M^{(j)}) \right\}$$

with  $g(\mathbf{y}_M) = \mathbf{1}_{V(\mathbf{y}_{M,M}) \geq a} \prod_{0 \leq p < M} G_p(\mathbf{y}_p)^{-1}$ .



An IPS with a lineage  $\mathbf{Y}^{\mathbf{L}^1}$  in red and a type-II lineage  $\mathbf{Y}^{\mathbf{L}^2}$  in blue.

- Unbiased and weakly consistent estimator of  $\sigma^2$ :

$$\hat{\sigma}_n^2 = \sum_{p=0}^{M-1} (\mu_{n,p} - \mu_{n,\emptyset})$$

$$\mu_{n,p} = \frac{n^M}{(n-1)^M} \sum_{\mathbf{L}^1} \frac{1}{\#(\mathbf{L}^2, \mathbf{L}^1 \cap \mathbf{L}^2 = \{\mathbf{L}_p^1\})} \sum_{\mathbf{L}^2, \mathbf{L}^1 \cap \mathbf{L}^2 = \{\mathbf{L}_p^1\}} g(\mathbf{Y}^{\mathbf{L}^1}) g(\mathbf{Y}^{\mathbf{L}^2}),$$

$$\mu_{n,\emptyset} = \frac{n^M}{(n-1)^{M+1}} \sum_{\mathbf{L}^1} \frac{1}{\#(\mathbf{L}^2, \mathbf{L}^1 \cap \mathbf{L}^2 = \emptyset)} \sum_{\mathbf{L}^2, \mathbf{L}^1 \cap \mathbf{L}^2 = \emptyset} g(\mathbf{Y}^{\mathbf{L}^1}) g(\mathbf{Y}^{\mathbf{L}^2}).$$

[A. Lee and N. Whiteley, *Biometrika* **105** (2018), 609-625.]

## Optimal potentials

The optimal potential functions are defined up to a multiplicative constant. For  $p \geq 1$ , let  $G_p^*$  be defined by:

$$G_p^*(\mathbf{y}_p)^2 = \begin{cases} \frac{\mathbb{E} \left[ \mathbb{E} [h(\mathbf{Y}_M) | \mathbf{Y}_{p+1}]^2 | \mathbf{Y}_p = \mathbf{y}_p \right]}{\mathbb{E} \left[ \mathbb{E} [h(\mathbf{Y}_M) | \mathbf{Y}_p]^2 | \mathbf{Y}_{p-1} = \mathbf{y}_{0:p-1} \right]} & \text{if } \mathbb{E} \left[ \mathbb{E} [h(\mathbf{Y}_M) | \mathbf{Y}_p]^2 | \mathbf{Y}_{p-1} = \mathbf{y}_{0:p-1} \right] \neq 0 \\ 0 & \text{if } \mathbb{E} \left[ \mathbb{E} [h(\mathbf{Y}_M) | \mathbf{Y}_p]^2 | \mathbf{Y}_{p-1} = \mathbf{y}_{0:p-1} \right] = 0 \end{cases}$$

- The potential functions minimizing  $\sigma^2$  are proportional to  $G_p^*$ .

The optimal variance of the IPS method is then

$$\begin{aligned} \sigma_{G^*}^2 &= \mathbb{E} \left[ \mathbb{E} [h(\mathbf{Y}_M) | \mathbf{Y}_0]^2 \right] - P^2 \\ &+ \sum_{p=1}^M \left\{ \mathbb{E} \left[ \sqrt{\mathbb{E} \left[ \mathbb{E} [h(\mathbf{Y}_M) | \mathbf{Y}_p]^2 | \mathbf{Y}_{p-1} \right]} \right]^2 - P^2 \right\}. \end{aligned}$$

- Two observations:
  - The optimal asymptotic variance is positive.
  - The optimal potential  $G_p^*$  depends only on  $x_p = y_{p,p}$  and  $x_{p-1} = y_{p,p-1}$ .



## Example: the Gaussian walk

$$X_p = X_{p-1} + \theta_p$$

where  $(\theta_p)_{1 \leq p \leq M}$  i.i.d. with the distribution  $\mathcal{N}(0, 1)$ ,

$$V(x) = x.$$

Here  $X_M$  is Gaussian, has zero-mean and variance  $M$ :

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

Consider  $a \gg \sqrt{M}$  so that  $P \ll 1$ .

- MC for the Gaussian walk

Consider the unbiased dynamics with no selection.

The MC estimator

$$\hat{P}_n = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X_M^{(k)} \geq a}, \quad X_M^{(k)} \text{ i.i.d. distributed as } X_M$$

has the asymptotic variance  $P(1 - P) \simeq P$ .

→ relative error  $\sim 1/\sqrt{nP}$ .

- IS for the Gaussian walk

Consider the **biased** dynamics with no selection

$$\tilde{X}_p = \tilde{X}_{p-1} + \tilde{\theta}_p$$

where  $(\tilde{\theta}_p)_{1 \leq p \leq M}$  are i.i.d. with the distribution  $\mathcal{N}(a/M, 1)$ .

Consider the IS estimator

$$\hat{I}_n = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\tilde{X}_M^{(k)} \geq a} \exp\left(\frac{a^2}{2M} - \frac{a}{M} \tilde{X}_M^{(k)}\right), \quad \tilde{X}_M^{(k)} \text{ i.i.d. distributed as } \tilde{X}_M$$

$\hat{I}_n$  is unbiased and has the asymptotic variance

$$\sigma^2 = \mathbb{E}\left[\mathbf{1}_{X_M \geq a} \exp\left(\frac{a^2}{2M} - \frac{a}{M} X_M\right)\right] - P^2 \sim \exp\left(-\frac{a^2}{M}\right)$$

→ relative error  $\sim 1/\sqrt{n}$ .

IS requires to bias the input distribution.

- IPS for the Gaussian walk
- Consider the unbiased dynamics

$$X_p = X_{p-1} + \theta_p$$

where  $(\theta_p)_{1 \leq p \leq M}$  are i.i.d. with the distribution  $\mathcal{N}(0, 1)$ .

Selection with potential  $G_p$ .

*First choice for the potential:*

$$G_p(x_0, \dots, x_p) = \exp(\alpha x_p), \quad \text{for some } \alpha > 0$$

We find

$$\sigma^2 \simeq \sum_{p=0}^{M-1} \left[ e^{-\frac{a^2}{M}} e^{\frac{p}{M(M+p)}} [a - \alpha M(p-1)/2]^2 + \frac{1}{12} \alpha^2 (p-1)p(p+1) - p^2 \right]$$

By an approximate optimization, we take  $\alpha = 2a/[M(M-1)]$ , and we get

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

$\Leftrightarrow$  the asymptotic variance is of the order of  $P^{4/3}$

$\rightarrow$  relative error  $\sim 1/\sqrt{nP^{2/3}}$ .

- IPS for the Gaussian walk

*Second choice for the potential:*

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

We find

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

By an approximate optimization, we take  $\alpha = a/M$ , we get

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

$\Leftrightarrow$  the asymptotic variance is of the order of  $P^2$ .

$\rightarrow$  relative error  $\sim 1/\sqrt{n}$ .

By comparing with the previous case: a selection pressure depending only on the state is not efficient !

- IPS for the Gaussian walk

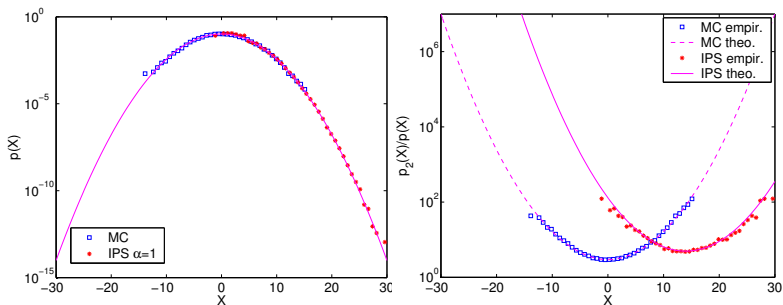
*Optimal choice for the potential:*

$$G_p^*(\mathbf{y}_p)^2 = \frac{\int_{\mathbb{R}} \left( \int_a^\infty \exp \left[ -\frac{(x'_M - x'_{p+1})^2}{2(n-p-1)} \right] dx'_M \right)^2 \exp \left[ -\frac{(x'_{p+1} - x_p)^2}{2} \right] dx'_{p+1}}{\int_{\mathbb{R}} \left( \int_a^\infty \exp \left[ -\frac{(x'_M - x'_p)^2}{2(n-p)} \right] dx'_M \right)^2 \exp \left[ -\frac{(x'_p - x_{p-1})^2}{2} \right] dx'_p}$$

$G_p(\mathbf{y}_p)$	$mean(\hat{P})$	$\hat{\sigma}_{IPS,G}^2$
$\exp[\alpha x_p], \alpha = 0.22$	$1.05 \cdot 10^{-6}$	$2.8 \cdot 10^{-9}$
$\exp[\alpha(x_p - x_{p-1})], \alpha = 1.4$	$1.05 \cdot 10^{-6}$	$1.7 \cdot 10^{-10}$
$\exp \left[ -\frac{(x_p - a)^2}{2(n-p+1)} + \frac{(x_{p-1} - a)^2}{2(n-p+2)} \right]$	$1.05 \cdot 10^{-6}$	$1.5 \cdot 10^{-10}$
$G_p^*(\mathbf{y}_p)$	$1.05 \cdot 10^{-6}$	$1.3 \cdot 10^{-10}$

Here  $P = 1.05 \cdot 10^{-6}$ ,  $n = 2000$ ,  $M = 10$ ,  $a = 15$ .

- IPS for the Gaussian walk



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

## Example: Communication in transoceanic optical fibers

- Physical model:

$(u_0(t))_{t \in \mathbb{R}}$  = initial pulse profile.

$(u(z, t))_{t \in \mathbb{R}}$  = pulse profile after a propagation distance  $z$ .

$(u(Z, t))_{t \in \mathbb{R}}$  = output pulse profile (after a propagation distance  $Z$ ).

$\tau(z)^2 = \int |u(z, t)|^2 t^2 dt / \int |u(z, t)|^2 dt$  rms pulse width after propagation distance  $z$ .

Propagation from  $z = 0$  to  $z = Z$  governed by two coupled nonlinear Schrödinger equations with randomly  $z$ -varying coefficients.

→ Truncation of  $[0, Z]$  into  $M$  segments  $[z_{p-1}, z_p)$ ,  $z_p = pZ/M$ ,  $1 \leq p \leq M$ .

→  $\mathbf{X}_p = (u(z_p, t))_{t \in \mathbb{R}}$  is the pulse profile at distance  $z_p$ .

Here  $(\mathbf{X}_p)_{0 \leq p \leq M}$  is Markov with state space  $E = H_0^2(\mathbb{R}) \cap L_2^2(\mathbb{R})$

- Problem: estimation of the probability

$$P = \mathbb{P}(V(\mathbf{X}_M) \geq a) = \mathbb{P}(\tau(Z) \geq a)$$

The potential function is  $V : \begin{cases} E \rightarrow \mathbb{R} \\ V(\mathbf{X}) = \int t^2 |\mathbf{X}(t)|^2 dt / \int |\mathbf{X}(t)|^2 dt \end{cases}$

1) asymptotic model (separation of scales technique)

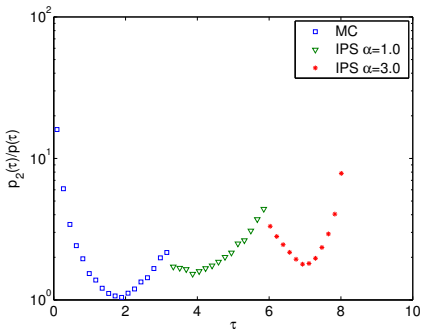
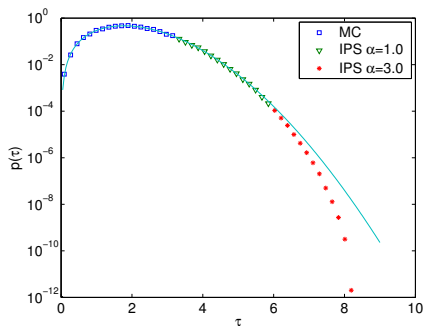
→ the rms pulse width  $\tau(z)$  is a diffusion process and its pdf is

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

2) realistic model: impossible to get a closed-form expression for the pdf of  $\tau(z)$ .

3) experimental observations: the pdf tail of the rms pulse width does not fit with the Maxwellian distribution in realistic configurations.





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

The solid line stands for the Maxwellian pdf predicted by the asymptotic model.

# Multilevel splitting

- Description of the system:
  - Let  $\mathbf{X}$  be a  $\mathbb{R}^d$ -valued random variable with pdf  $p(\mathbf{x})$ .
  - Let  $V : \mathbb{R}^d \rightarrow \mathbb{R}$  be the risk function.
  - Let  $a$  be the threshold level.
- Problem: estimation of

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

when  $a$  is large  $\implies P \ll 1$ .

# Multilevel splitting

- Let  $\mathbf{X}$  be a  $\mathbb{R}^d$ -valued random variable with pdf  $p(\mathbf{x})$ . Estimation of

$$P = \mathbb{P}(V(\mathbf{X}) \geq a) = \int_{\mathbb{R}^d} \mathbf{1}_{V(\mathbf{x}) \geq a} p(\mathbf{x}) d\mathbf{x}$$

- Splitting strategy:

- Write the decomposition (with  $a_M = a > \dots > a_0 = -\infty$ )

$$P = \prod_{j=1}^M P_j, \quad P_j = \mathbb{P}(V(\mathbf{X}) \geq a_j | V(\mathbf{X}) \geq a_{j-1})$$

- Estimate  $P_j$  separately.

- Two key issues:

- 1) Algorithm to evaluate each  $P_j$ ,
- 2) Selection of the levels  $a_j$ .

## Multilevel splitting

$$P = \mathbb{P}(V(\mathbf{X}) \geq a) = \prod_{j=1}^M P_j, \quad P_j = \mathbb{P}(V(\mathbf{X}) \geq a_j | V(\mathbf{X}) \geq a_{j-1})$$

- Two key issues:

- 1) Algorithm to evaluate each  $P_j$ ,
- 2) Selection of the levels  $a_j$ .

- Answers:

Answer to 1): use an interacting particle method (based on a Markov process whose invariant distribution has pdf  $p$ )  $\rightarrow \hat{P}_n$ .

Answer to 2): choose  $a_j$  such that the  $P_j$ 's are all equal to the same  $\alpha \in (0, 1)$ . Then

$$\text{Var}(\hat{P}_n) = \frac{P^2}{n} \left( \frac{(1 - \alpha) \ln P}{\alpha \ln \alpha} \right) + o(n^{-1})$$

$\Leftrightarrow$  one should take  $\alpha \rightarrow 1$ .

- Multilevel splitting strategy with “ $\alpha = 1 - 1/n$ ”:
- Generate  $n$  particles (with pdf  $p$ ) to create generation zero:

$$\hookrightarrow (\mathbf{X}_0^{(1)}, \dots, \mathbf{X}_0^{(n)}) \text{ i.i.d. with pdf } p(\mathbf{x})$$

- For  $j - 1 \rightarrow j$ ,
  - define the level  $a_j$  as the minimum of  $V(\mathbf{x})$  evaluated on the  $n$  particles:  $a_j = \min_{k=1, \dots, n} \{V(\mathbf{X}_{j-1}^{(k)})\}$ ,
  - remove the particle that achieves the minimum,
  - generate a new particle with the conditional distribution  $\mu_{a_j}$  of  $\mathbf{X}$  knowing that  $V(\mathbf{X}) \geq a_j$ :

$$\mu_{a_j}(d\mathbf{x}) = p_{a_j}(\mathbf{x})d\mathbf{x}, \quad p_{a_j}(\mathbf{x}) = \frac{\mathbf{1}_{V(\mathbf{x}) \geq a_j} p(\mathbf{x})}{\int_{\mathbb{R}^d} \mathbf{1}_{V(\mathbf{x}') \geq a_j} p(\mathbf{x}') d\mathbf{x}'}$$

(see below: use the Metropolis-Hastings algorithm).

$$\hookrightarrow (\mathbf{X}_j^{(1)}, \dots, \mathbf{X}_j^{(n)}) \text{ i.i.d. with the distribution } \mu_{a_j}$$

- Stop when  $a_j \geq a$ . Denote  $\hat{J}_n = \min\{j, a_j \geq a\} - 1$ .

- Result 1: if one knows how to generate the new particle with the distribution  $\mu_{a_j}$ ,

then  $\hat{J}_n$  follows a Poisson distribution with parameter  $-n \ln P$ :

$$\mathbb{P}(\hat{J}_n = j) = \frac{P^n (-n \log P)^j}{j!}$$

Proof:

we assume that  $V(\mathbf{X})$  has continuous cdf  $F$ .

(a) the random variables  $-\log(1 - F(a_j))$ ,  $j \geq 1$ , are distributed as the successive arrival times of a Poisson process with rate  $n$ ,

$$-\log(1 - F(a_j)) \stackrel{\text{dist.}}{=} \frac{1}{n} \sum_{i=1}^j E_i$$

where  $E_i$  are i.i.d. exponential random variables.

(b)  $\mathbb{P}(\hat{J}_n = j) = \mathbb{P}(a_j \leq a, a_{j+1} > a) = \mathbb{P}(\sum_{i=1}^j E_i \leq -n \ln P < \sum_{i=1}^{j+1} E_i)$ .

Proof of (a).

Let  $\Lambda(y) = -\log(1 - F(y))$ .  $\Lambda : \mathbb{R} \rightarrow (0, \infty)$  is continuous and increasing.

• Generation 0:  $(\Lambda(V(\mathbf{X}_0^{(k)})))_{k=1, \dots, n}$  are i.i.d.

$F$  is the cdf of  $V(\mathbf{X})$ , so  $F(V(\mathbf{X})) \sim \mathcal{U}(0, 1)$

Therefore  $\Lambda(V(\mathbf{X})) = -\log(1 - F(V(\mathbf{X}))) \sim \mathcal{E}(1)$ :

$$\mathbb{P}(\Lambda(V(\mathbf{X})) \geq \lambda) = e^{-\lambda}$$

Therefore  $(\Lambda(V(\mathbf{X}_0^{(k)})))_{k=1, \dots, n}$  are i.i.d. with the distribution  $\mathcal{E}(1)$ .

Let  $a_1 = \min_{k=1, \dots, n} \{V(\mathbf{X}_0^{(k)})\}$ . We have

$$\Lambda(a_1) = \min_{k=1, \dots, n} \{\Lambda(V(\mathbf{X}_0^{(k)}))\}$$

$$\mathbb{P}(\Lambda(a_1) \geq \lambda) = \mathbb{P}(\Lambda(V(\mathbf{X})) \geq \lambda)^n = e^{-n\lambda}$$

Therefore

$$\Lambda(a_1) \sim \frac{1}{n} E_1, \quad E_1 \sim \mathcal{E}(1)$$

- Generation  $j$ . Let  $\Lambda_j(y) = -\log(1 - F_j(y))$  where  $F_j$  is the cdf of  $V(\mathbf{X})$  given  $V(\mathbf{X}) \geq a_j$ :

$$F_j(y) = \mathbb{P}(V(\mathbf{X}) \leq y | V(\mathbf{X}) \geq a_j) = \frac{\mathbb{P}(a_j \leq V(\mathbf{X}) \leq y)}{\mathbb{P}(V(\mathbf{X}) \geq a_j)} = \frac{F(y) - F(a_j)}{1 - F(a_j)}$$

Therefore  $\Lambda_j(y) = \Lambda(y) - \Lambda(a_j)$ .

As above:  $(\Lambda_j(V(\mathbf{X}_j^{(k)})))_{k=1, \dots, n}$  are i.i.d. with the distribution  $\mathcal{E}(1)$ .

Let  $a_{j+1} = \min_{k=1, \dots, n} \{V(\mathbf{X}_j^{(k)})\}$ . As above  $\Lambda_j(a_{j+1}) \sim \frac{1}{n} E_{j+1}$ ,  $E_j \sim \mathcal{E}(1)$ .  
Therefore

$$\Lambda(a_{j+1}) = \Lambda(a_j) + \Lambda_j(a_{j+1}) \sim \frac{1}{n} \sum_{i=1}^{j+1} E_i, \quad E_i \sim \mathcal{E}(1)$$





- Estimator:

$$\hat{P}_n = \left(1 - \frac{1}{n}\right)^{J_n}$$

- Result 2: if one knows how to generate the new particle with the distribution  $\mu_{a_j}$ ,

then  $\hat{P}_n$  is an unbiased estimator of  $P$  with variance

$$\text{Var}(\hat{P}_n) = P^2(P^{-1/n} - 1) \simeq \frac{-P^2 \ln P}{n}$$

Proof:

$$\mathbb{P}\left(\hat{P}_n = \left(1 - \frac{1}{n}\right)^j\right) = \mathbb{P}(J_n = j) = \frac{P^n (-n \log P)^j}{j!}$$

- Result 3: Denote

$$\hat{P}_{n,\pm} = \hat{P}_n \exp\left(\pm \frac{z_{1-\alpha/2}}{\sqrt{n}} \sqrt{-\log \hat{P}_n}\right)$$

where  $z_{1-\alpha/2}$  is the  $1 - \alpha/2$ -quantile of the standard normal distribution.

We have

$$\mathbb{P}(P \in [\hat{P}_{n,-}, \hat{P}_{n,+}]) \approx 1 - \alpha.$$

If  $\alpha = 0.05$ , then  $z_{1-\alpha/2} \approx 2$ .

- Aparté: Metropolis-Hastings algorithm.
- Let  $\mu_a$  be a probability distribution on  $\mathbb{R}^d$  with pdf  $p_a(\mathbf{x})$  (known up to a multiplicative constant). We want to simulate an ergodic Markov chain  $(\mathbf{X}_t)_{t \geq 0}$  whose invariant distribution is  $\mu_a$ .
- Preliminary step: choose an instrumental transition density  $q$  on  $\mathbb{R}^d$ , i.e., for any fixed  $\mathbf{x}' \in \mathbb{R}^d$ ,  $\mathbf{x} \rightarrow q(\mathbf{x}', \mathbf{x})$  is a pdf and we know how to generate a random variable  $\mathbf{X}$  with this pdf.

- Algorithm:

Step 0: Choose  $\mathbf{X}_0$  arbitrarily.

Step  $t + 1$ : Choose a candidate  $\tilde{\mathbf{X}}_{t+1}$  with the distribution with pdf  $q(\mathbf{X}_t, \mathbf{x})$ . Set  $\mathbf{X}_{t+1} = \mathbf{X}_t$  with probability  $1 - \rho(\mathbf{X}_t, \tilde{\mathbf{X}}_{t+1})$  (reject) and  $\mathbf{X}_{t+1} = \tilde{\mathbf{X}}_{t+1}$  with probability  $\rho(\mathbf{X}_t, \tilde{\mathbf{X}}_{t+1})$  (accept). Here

$$\rho(\mathbf{x}', \mathbf{x}) = \min \left( \frac{p_a(\mathbf{x})q(\mathbf{x}, \mathbf{x}')}{p_a(\mathbf{x}')q(\mathbf{x}', \mathbf{x})}, 1 \right)$$

- $(\mathbf{X}_t)_{t \geq 0}$  is a Markov chain with transition

$$K(\mathbf{x}', d\mathbf{x}) = q(\mathbf{x}', \mathbf{x})\rho(\mathbf{x}', \mathbf{x})d\mathbf{x} + \left( 1 - \int q(\mathbf{x}', \mathbf{y})\rho(\mathbf{x}', \mathbf{y})d\mathbf{y} \right) \delta_{\mathbf{x}'}(d\mathbf{x})$$

- We have (because  $p_a(x')[q(x', x)\rho(x', x)] = p_a(x)[q(x, x')\rho(x, x')]$ )

$$\int dx' p_a(x') K(x', dx) = p_a(x) dx$$

$\Leftrightarrow \mu_a$  is stationary for the Markov chain.

- Under mild conditions (for instance, if  $q$  is positive), the chain  $(\mathbf{X}_t)_{t \geq 0}$  is ergodic with stationary distribution  $\mu_a$ :

$$\sup_{A \in \mathcal{B}(\mathbb{R}^d)} |\mathbb{P}(\mathbf{X}_t \in A) - \mu_a(A)| \xrightarrow{t \rightarrow \infty} 0$$

- In practice:

- after a burn-in phase with some length  $t_0$ , the sequence  $(\mathbf{X}_t)_{t \geq t_0}$  is stationary with distribution  $\mu_a$  (but not independent).

- the choice of the instrumental transition density is important to get fast convergence. Ideally the rejection rate should be around 50%.

- If  $\mathbf{X}_0 \sim \mu_a$ , then the chain is stationary. After a few accepted mutations,  $\mathbf{X}_t \sim \mu_a$  and is quasi-independent from  $\mathbf{X}_0$ .

- Problem: how to generate the new particle with the distribution  $\mu_{a_j}$  (of  $\mathbf{X}$  knowing that  $V(\mathbf{X}) > a_j$ ) ?

Version 1:

- Consider a symmetric transition kernel  $q(\mathbf{x}', \mathbf{x})$  such that

$$q(\mathbf{x}', \mathbf{x}) = q(\mathbf{x}, \mathbf{x}')$$

- Algorithm:

-  $a_j =$  minimal value of the  $n$  particles.

- pick a particle  $\mathbf{X}_{(1)}$  amongst the  $n - 1$  largest particles (larger than  $a_j$ ).

- for  $t = 1, \dots, T$ , draw a new particle  $\mathbf{X}^*$  with the pdf  $q(\mathbf{X}_{(1)}, \cdot)$ ;

if  $V(\mathbf{X}^*) > a_j$ , then  $\mathbf{X}_{(1)} = \mathbf{X}^*$  with probability  $\min(\rho(\mathbf{X}^*)/\rho(\mathbf{X}_{(1)}), 1)$ ;

otherwise keep  $\mathbf{X}_{(1)}$ .

- replace the smallest particle by  $\mathbf{X}_{(1)}$ .

- Result 3: the distribution of  $\mathbf{X}_{(1)}$  is the distribution  $\mu_{a_j}$ . As  $T \rightarrow \infty$ , the distribution of  $\mathbf{X}_{(1)}$  becomes independent of the other particles.

- Problem: how to generate the new particle with the distribution  $\mu_{a_j}$  (of  $\mathbf{X}$  knowing that  $V(\mathbf{X}) > a_j$ ) ?

Version 2:

- Consider a transition kernel  $q(\mathbf{x}', \mathbf{x})$  such that

$$\rho(\mathbf{x}')q(\mathbf{x}', \mathbf{x}) = \rho(\mathbf{x})q(\mathbf{x}, \mathbf{x}')$$

- Algorithm:

- $a_j$  = minimal value of the  $n$  particles.
- pick a particle  $\mathbf{X}_{(1)}$  amongst the  $n - 1$  largest particles (larger than  $a_j$ ).
- for  $t = 1, \dots, T$ , draw a new particle  $\mathbf{X}^*$  with the pdf  $q(\mathbf{X}_{(1)}, \cdot)$ ; if  $V(\mathbf{X}^*) > a_j$ , then  $\mathbf{X}_{(1)} = \mathbf{X}^*$ ; otherwise keep  $\mathbf{X}_{(1)}$ .
- replace the smallest particle by  $\mathbf{X}_{(1)}$ .
- Result 3: the distribution of  $\mathbf{X}_{(1)}$  is the distribution  $\mu_{a_j}$ . As  $T \rightarrow \infty$ , the distribution of  $\mathbf{X}_{(1)}$  becomes independent of the other particles.  
In practice:  $T =$  a few tens.

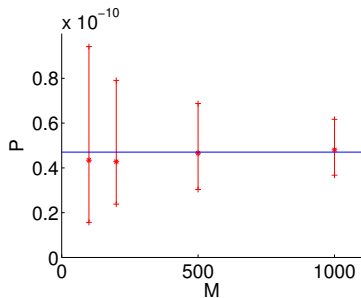
- Example:

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

with  $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ ,  $d = 20$ ,  $a = 0.95$ ,  $V(\mathbf{x}) = |x_1|/|\mathbf{x}|$ :  $P = 4.704 \cdot 10^{-11}$ .

Kernel  $q : \mathbf{x}' \rightarrow \mathcal{N}\left(\frac{\mathbf{x}'}{\sqrt{1+\sigma^2}}, \frac{\sigma^2}{1+\sigma^2} \mathbf{I}_d\right)$ ,  $\sigma = 0.3$ ,  $T = 20$ , ie

$$q(\mathbf{x}', \mathbf{x}) = \frac{(1 + \sigma^2)^{d/2}}{(2\pi\sigma^2)^{d/2}} \exp\left(-\frac{|\sqrt{1 + \sigma^2}\mathbf{x} - \mathbf{x}'|^2}{2\sigma^2}\right)$$



$n \in [100, 200, 500, 1000]$  particles.

[A. Guyader, et al., Appl. Math. Optim. **64** (2011), 171–196]

# Conclusions

- Importance sampling: bias the **input**.  
↔ Intrusive method.
- Interacting particle system: select the particles based on the **output**.  
↔ No physical insight is required to guess the suitable biased input distribution.

But: need  $V(\mathbf{X})$ .

- ↔ **Non-intrusive method**: no need to change the numerical code.
- Number of particles fixed, computational cost (almost) fixed.
- The simulation code is used with the original distribution.
- Empirical estimator of the variance of the estimator and confidence intervals can be built.
- It is possible to make the algorithm partially parallel (not fully parallel as Monte Carlo).
- Also: conditional distributions. The method is efficient for the computation of conditional expectations and for the analysis of the cascade of events leading to a rare event.