Rare event simulation - part II

Josselin Garnier

ETICS 2020

Interacting particle systems for the analysis of rare events

- Problem: estimation of the probability of occurence 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:

– $(\boldsymbol{X}_p)_{0 \leq p \leq M}$: a E-valued Markov chain ($E = \mathbb{R}, \mathbb{R}^d, \ldots$):

$$\mathbb{P}(\boldsymbol{X}_p \in A \,|\, \boldsymbol{X}_{p-1} = \boldsymbol{x}_{p-1}, \ldots, \boldsymbol{X}_0 = \boldsymbol{x}_0) = \mathbb{P}(\boldsymbol{X}_p \in A \,|\, \boldsymbol{X}_{p-1} = \boldsymbol{x}_{p-1})$$

– $V : E \to \mathbb{R}$: the risk function.

– $a \in \mathbb{R}$: the threshold level.

- Problem: estimation of the probability

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

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

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

- *Example:* $X_p = X_{p-1} + \theta_p$, $X_0 = 0$, where θ_p is a sequence of i.i.d. Gaussian random variables with mean zero and variance one. Here

– $E = \mathbb{R}$,

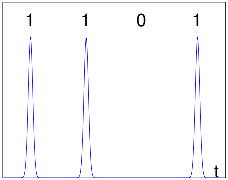
– $V(x) = x$,

– the solution is known: $X_M = V(X_M) \sim \mathcal{N}(0, M)$.

Example: Communication in transoceanic optical fibers

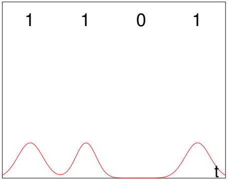
- Optical fiber transmission principle:

- 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.
$\hookrightarrow$ black box.
$\to$ Truncation of $[0,Z]$ into M segments $[z_{p-1},z_p)$, $z_p = pZ/M$, $1 \le p \le M$.
$\to$ $\boldsymbol{X}_p = u(z_p,t)_{t\in\mathbb{R}}$ is the pulse profile at distance z_p.
Here $(\boldsymbol{X}_p)_{0\le p\le M}$ 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 $V : \left| \begin{array}{l} E \to \mathbb{R} \\ V(\boldsymbol{X}) = \int t^2 |\boldsymbol{X}(t)|^2 dt \ / \int |\boldsymbol{X}(t)|^2 dt \end{array} \right.$

- Problem: estimation of the probability

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

Monte Carlo method

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

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

- Unbiased estimator:

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

- Variance:

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

- Relative error:

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

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

Importance Sampling method

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

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

Unbiased estimator:

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

Variance:

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

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

Importance Sampling method

• Optimal choice of $\mathbb{Q}$: $d\mathbb{Q} = \frac{\mathbf{1}_{V(\boldsymbol{X}_M)\geq a}}{\mathbb{P}(V(\boldsymbol{X}_M)\geq a)} d\mathbb{P}$.
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, $\lambda > 0$:

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

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

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

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

- Variance:

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

Importance Sampling driven by Large Deviations Principle

- Variance:

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

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

- For a judicious choice of λ,

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

(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 $\mathbb{Q}$:

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

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

Examples:

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

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

Original measures

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

$$\mathbb{P}(\boldsymbol{X}_p \in A \,|\, \boldsymbol{X}_{p-1} = \boldsymbol{x}_{p-1}, \ldots, \boldsymbol{X}_0 = \boldsymbol{x}_0) = \int_A K_p(\boldsymbol{x}_{p-1}, d\boldsymbol{x}_p)$$

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

- Denote the (partial) path

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

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

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

- Expression of P in terms of μ_M:

$$P = \mu_M(f)$$

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

$\rightarrow$ If one can compute/estimate μ_M, then one can compute/estimate P.

Unnormalized Feynman-Kac measures

$$\boldsymbol{Y}_p =_{\rm def.} (\boldsymbol{X}_0, \ldots, \boldsymbol{X}_p) \in E^{p+1}, \qquad p = 0, \ldots, M$$

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

$$\gamma_p(f_p) = \mathbb{E}\Big[f_p(\boldsymbol{Y}_p) \prod_{0 \le k < p} G_k(\boldsymbol{Y}_k) \Big]$$

- Expression of P in terms of γ_M:

$$P = \gamma_M(g)$$

$$g(\boldsymbol{y}_M) = g(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_M) = \mathbf{1}_{V(\boldsymbol{x}_M) \ge a} \prod_{0 \le p < M} G_p^{-1}(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_p)$$

$\rightarrow$ If one can compute/estimate γ_M, then one can compute/estimate P.

Normalized Feynman-Kac measures

Introduce the normalized measure η_p:

$$\eta_p(f_p) = \gamma_p(f_p)/\gamma_p(1), \qquad p = 0, \ldots, M$$

- Expression of P in terms of η_p:

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

Proof:

$$P = \mathbb{E}\Big[g(\boldsymbol{Y}_M) \prod_{0 \le k < M} G_k(\boldsymbol{Y}_k)\Big] = \gamma_M(g) = \eta_M(g)\gamma_M(1)$$

Normalizing constant:

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

□

→ 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 μ_p

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

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

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

- Goal: simulate the normalized measures η_p

$$\eta_p(f_p) = \frac{\mathbb{E}\Big[f_p(\boldsymbol{Y}_p) \prod_{0\le k<p} G_k(\boldsymbol{Y}_k)\Big]}{\mathbb{E}\Big[\prod_{0\le k<p} G_k(\boldsymbol{Y}_k)\Big]}$$

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

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

Interacting path-particle system

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

$$d\eta_M = \frac{1}{\mathcal{Z}_M} \left\{ \prod_{p=0}^{M-1} G_p(\boldsymbol{X}_0, \ldots, \boldsymbol{X}_p) \right\} d\mathbb{P}$$

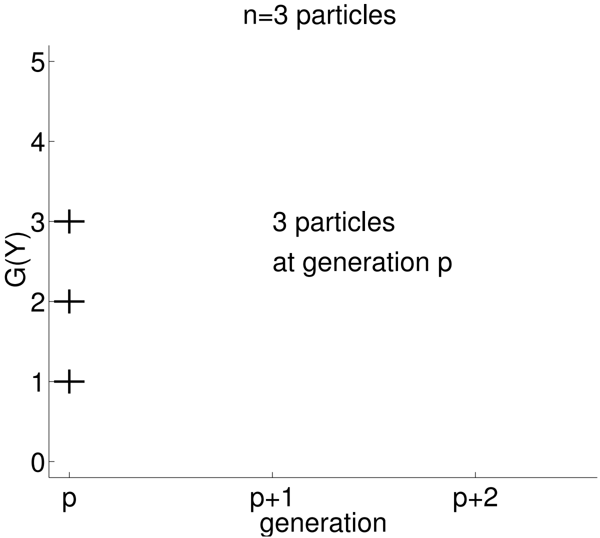
Answer: System of path-particles, whose empirical measure will be approximately $\mathbb{Q}$.

- Path-particle: $\boldsymbol{Y}_p = (\boldsymbol{X}_0, \ldots, \boldsymbol{X}_p)$ taking values in E^{p+1}, $0 \leq p \leq M$.
- System of n path-particles: $\mathbb{Y}_p = (\boldsymbol{Y}_p^{(i)})_{1 \leq i \leq n}$ taking values in $(E^{p+1})^n$.

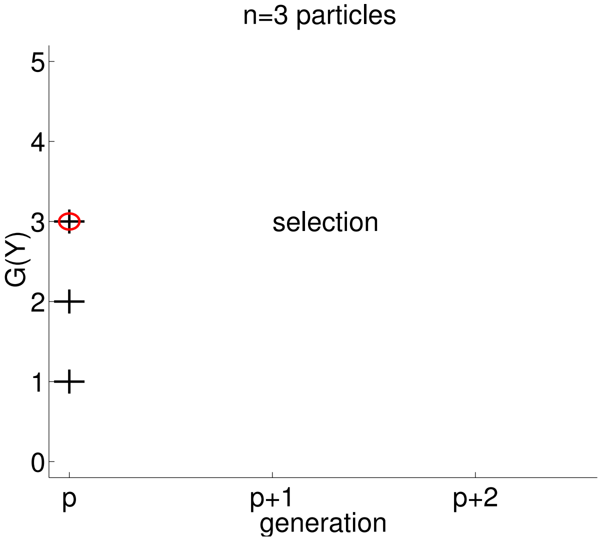
– Initialization: $p = 0$: $\boldsymbol{Y}_0^{(i)} = \boldsymbol{x}_0$ for all $i = 1, \ldots, n$.

– Dynamics: Evolution from generation p to $p + 1$ as follows:

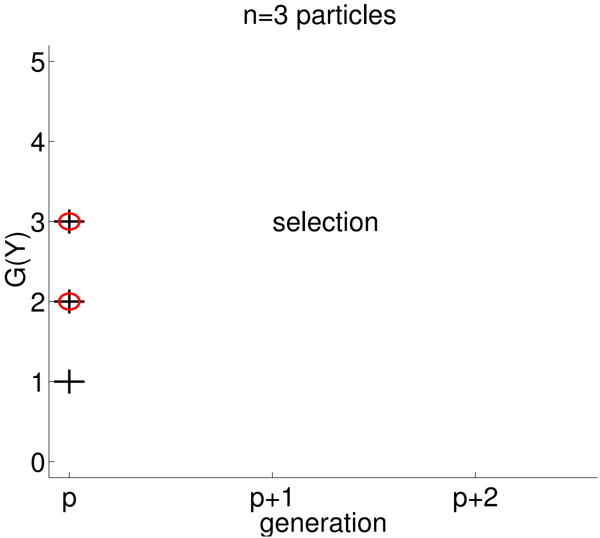
$$\mathbb{Y}_p \in (E^{p+1})^n \xrightarrow{\textit{selection}} \widehat{\mathbb{Y}}_p \in (E^{p+1})^n \xrightarrow{\textit{mutation}} \mathbb{Y}_{p+1} \in (E^{p+2})^n$$



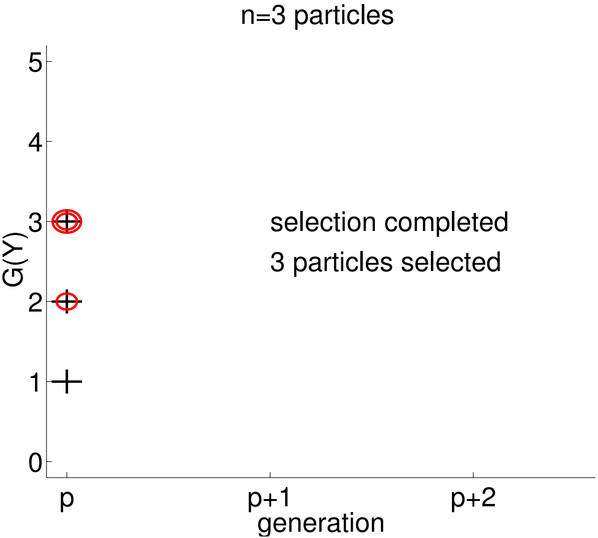
3 particles $\boldsymbol{Y}_p^{(1)}, \boldsymbol{Y}_p^{(2)}, \boldsymbol{Y}_p^{(3)}$ at generation p,
with potential weights $G(\boldsymbol{Y}_p^{(1)}) = 1$, $G(\boldsymbol{Y}_p^{(2)}) = 2$,$G(\boldsymbol{Y}_p^{(3)}) = 3$.



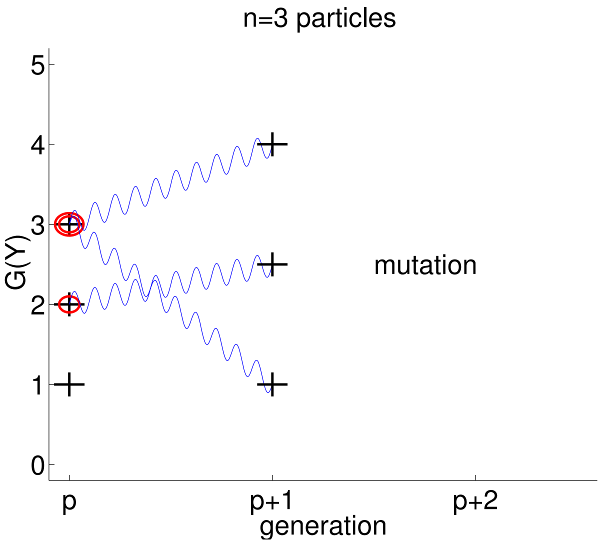
Prob. to select part. j: $\dfrac{G(\boldsymbol{Y}_p^{(j)})}{G(\boldsymbol{Y}_p^{(1)}) + G(\boldsymbol{Y}_p^{(2)}) + G(\boldsymbol{Y}_p^{(3)})} = \begin{cases} \frac{1}{6} \text{ if } j = 1 \\ \frac{1}{3} \text{ if } j = 2 \\ \frac{1}{2} \text{ if } j = 3 \end{cases}$



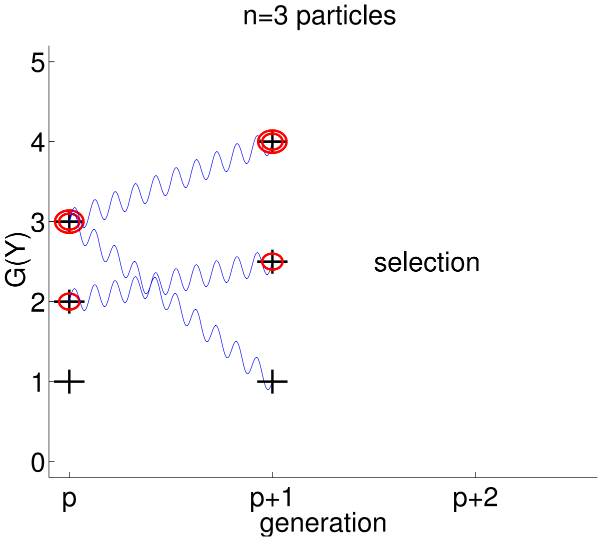
Prob. to select part. j: $\frac{G(\boldsymbol{Y}_p^{(j)})}{G(\boldsymbol{Y}_p^{(1)}) + G(\boldsymbol{Y}_p^{(2)}) + G(\boldsymbol{Y}_p^{(3)})} = \begin{cases} \frac{1}{6} \text{ if } j = 1 \\ \frac{1}{3} \text{ if } j = 2 \\ \frac{1}{2} \text{ if } j = 3 \end{cases}$



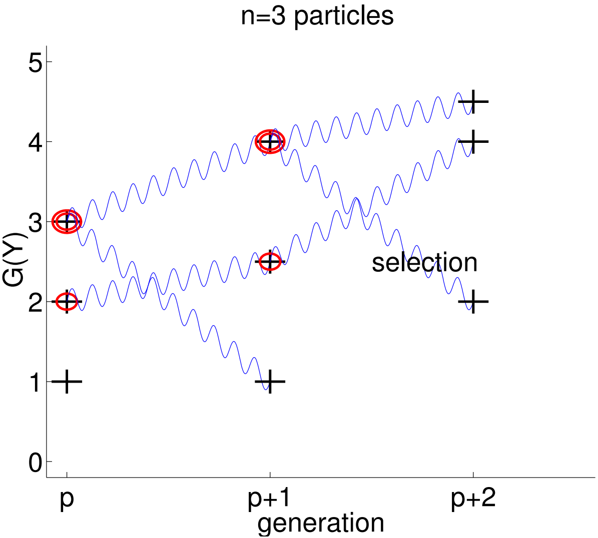
Prob. to select part. j: $\dfrac{G(\boldsymbol{Y}_p^{(j)})}{G(\boldsymbol{Y}_p^{(1)}) + G(\boldsymbol{Y}_p^{(2)}) + G(\boldsymbol{Y}_p^{(3)})} = \begin{cases} \frac{1}{6} \text{ if } j = 1 \\ \frac{1}{3} \text{ if } j = 2 \\ \frac{1}{2} \text{ if } j = 3 \end{cases}$



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



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



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

At each generation $p = 0, \ldots, M-1$:
Selection: from the system $\mathbb{Y}_p = (\boldsymbol{Y}_p^{(i)})_{1 \le i \le n}$, choose randomly and independently n path-particles

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

according to the Boltzmann-Gibbs particle measure

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

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

$$\begin{aligned} \boldsymbol{Y}_{p+1}^{(i)} &= \left(\quad (\boldsymbol{Y}_{0,p+1}^{(i)}, \ldots, \boldsymbol{Y}_{p,p+1}^{(i)}) \quad , \; \boldsymbol{Y}_{p+1,p+1}^{(i)} \right) \\ &= ((\widehat{\boldsymbol{Y}}_{0,p}^{(i)}, \ldots, \widehat{\boldsymbol{Y}}_{p,p}^{(i)}), \; \boldsymbol{Y}_{p+1,p+1}^{(i)}) \in E^{p+1} \end{aligned}$$

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

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

$$\eta_p^n =_{\rm def.} \frac{1}{n}\sum_{i=1}^n \delta_{(\boldsymbol{Y}_{0,p}^{(i)},\ldots,\boldsymbol{Y}_{p,p}^{(i)})} \stackrel{n\to\infty}{\longrightarrow} \eta_p$$

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

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

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

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

[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 $(U_1, \ldots, U_n)$ 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 $j = 1, \ldots, n$.

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

$$\hat{P}_n \overset{n\to\infty}{\longrightarrow} P \qquad \text{a.s.}$$

Central limit theorem

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

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

with the asymptotic variance

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

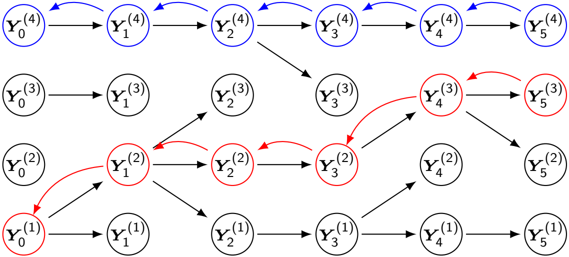
Here the functions $P^a_{p,M}$ are defined by

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

- 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

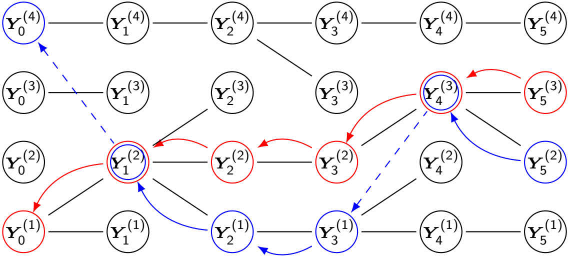


An IPS with two lineages with distinct parents $a(4) = 4$ and $a(3) = 1$.

- Unbiased and weakly consistent estimator of σ^2:

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

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



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

- Unbiased and weakly consistent estimator of σ^2:

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

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

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

[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:

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

• The potential functions minimizing σ^2 are proportional to G_p^*.

The optimal variance of the IPS method is then

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

• 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

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

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

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

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

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

$\to$ 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 $(\theta_p)_{1\le p\le M}$ 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

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

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

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

$\hookrightarrow$ the asymptotic variance is of the order of $P^{4/3}$
$\to$ relative error $\sim 1/\sqrt{nP^{2/3}}$.

• IPS for the Gaussian walk

Second choice for the potential:

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

We find

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

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

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

$\hookrightarrow$ 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

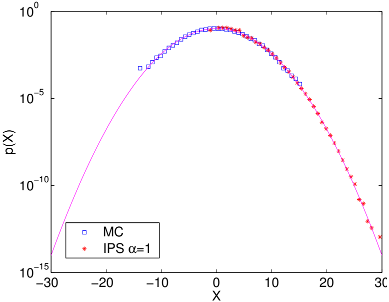
Optimal choice for the potential:

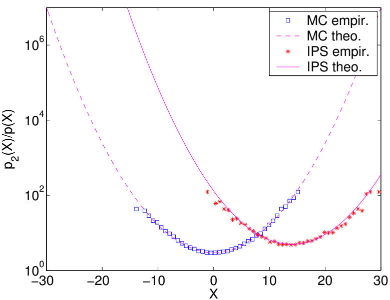
$$G_p^*(\boldsymbol{y}_p)^2 = \frac{\int_{\mathbb{R}} \left(\int_a^\infty \exp\left[-\frac{(x_M'-x_{p+1}')^2}{2(n-p-1)} \right] dx_M' \right)^2 \exp\left[-\frac{(x_{p+1}'-x_p)^2}{2} \right] dx_{p+1}'}{\int_{\mathbb{R}} \left(\int_a^\infty \exp\left[-\frac{(x_M'-x_p')^2}{2(n-p)} \right] dx_M' \right)^2 \exp\left[-\frac{(x_p'-x_{p-1})^2}{2} \right] dx_p'}$$

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

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

- IPS for the Gaussian walk





$$M = 15,\ n = 2\ 10^4 \text{ particles},\ \alpha = 1.$$

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.
$\rightarrow$ Truncation of $[0,Z]$ into M segments $[z_{p-1}, z_p)$, $z_p = pZ/M$, $1 \leq p \leq M$.
$\rightarrow$ $\boldsymbol{X}_p = (u(z_p,t)_{t\in\mathbb{R}})$ is the pulse profile at distance z_p.
Here $(\boldsymbol{X}_p)_{0\leq p\leq M}$ is Markov with state space $E = H_0^2(\mathbb{R}) \cap L_2^2(\mathbb{R})$

• Problem: estimation of the probability

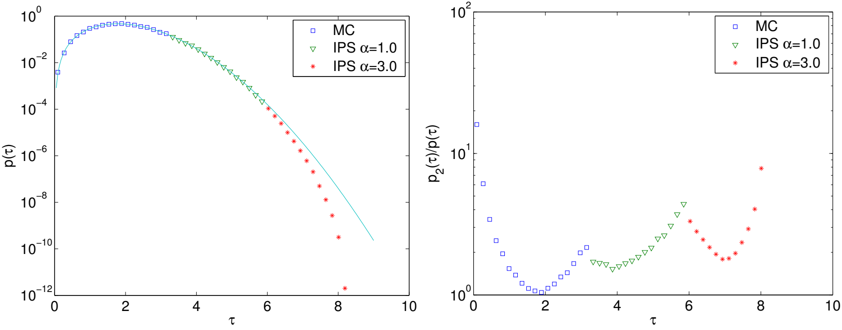
$$P = \mathbb{P}(V(\boldsymbol{X}_M) \geq a) = \mathbb{P}(\tau(Z) \geq a)$$

The potential function is $V : \left| \begin{array}{l} E \to \mathbb{R} \\ V(\boldsymbol{X}) = \int t^2 |\boldsymbol{X}(t)|^2 dt / \int |\boldsymbol{X}(t)|^2 dt \end{array} \right.$

1) asymptotic model (separation of scales technique)
$\to$ 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.



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

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

Multilevel splitting

- Description of the system:

– Let $\boldsymbol{X}$ be a $\mathbb{R}^d$-valued random variable with pdf $p(\boldsymbol{x})$.
– Let $V : \mathbb{R}^d \to \mathbb{R}$ be the risk function.
– Let a be the threshold level.

- Problem: estimation of

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

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

Multilevel splitting

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

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

- Splitting strategy:

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

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

- Estimate P_j separately.

- Two key issues:

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

Multilevel splitting

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

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

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

$\hookrightarrow$ one should take $\alpha \to 1$.

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

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

- For $j - 1 \to j$,

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

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

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

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

- 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 μ_{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 $V(\boldsymbol{X})$ 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,

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

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

Proof of (a).

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

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

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

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

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

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

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

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

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

Therefore

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

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

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

Therefore $\Lambda_j(y) = \Lambda(y) - \Lambda(a_j)$.
As above: $(\Lambda_j(V(\boldsymbol{X}_j^{(k)})))_{k=1,\dots,n}$ are i.i.d. with the distribution $\mathcal{E}(1)$.
Let $a_{j+1} = \min_{k=1,\dots,n}\{V(\boldsymbol{X}_j^{(k)})\}$. As above $\Lambda_j(a_{j+1}) \sim \frac{1}{n}E_{j+1}$, $E_j \sim \mathcal{E}(1)$.
Therefore

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

□

- Estimator:

$$\hat{P}_n = \Big(1 - \frac{1}{n}\Big)^{\hat{J}_n}$$

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

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

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

Proof:

$$\mathbb{P}\Big(\hat{P}_n = \Big(1 - \frac{1}{n}\Big)^j\Big) = \mathbb{P}(\hat{J}_n = j) = \frac{P^n(-n\log P)^j}{j!}$$

- Result 3: Denote

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

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

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

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

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

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

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

$$\rho(\boldsymbol{x}', \boldsymbol{x}) = \min\Big(\frac{p_a(\boldsymbol{x})q(\boldsymbol{x}, \boldsymbol{x}')}{p_a(\boldsymbol{x}')q(\boldsymbol{x}', \boldsymbol{x})}, 1\Big)$$

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

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

• We have (because $p_a(\boldsymbol{x}')[q(\boldsymbol{x}',\boldsymbol{x})\rho(\boldsymbol{x}',\boldsymbol{x})] = p_a(\boldsymbol{x})[q(\boldsymbol{x},\boldsymbol{x}')\rho(\boldsymbol{x},\boldsymbol{x}')]$)

$$\int d\boldsymbol{x}' p_a(\boldsymbol{x}') K(\boldsymbol{x}', d\boldsymbol{x}) = p_a(\boldsymbol{x}) d\boldsymbol{x}$$

$\hookrightarrow$ μ_a is stationary for the Markov chain.

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

$$\sup_{A\in\mathcal{B}(\mathbb{R}^d)} \Big| \mathbb{P}(\boldsymbol{X}_t \in A) - \mu_a(A) \Big| \stackrel{t\to\infty}{\longrightarrow} 0$$

• In practice:

- after a burn-in phase with some length t_0, the sequence $(\boldsymbol{X}_t)_{t\geq t_0}$ is stationary with distribution μ_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 $\boldsymbol{X}_0 \sim \mu_a$, then the chain is stationary. After a few accepted mutations, $\boldsymbol{X}_t \sim \mu_a$ and is quasi-independent from $\boldsymbol{X}_0$.

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

Version 1:

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

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

• Algorithm:

- $a_j =$ minimal value of the n particles.
- pick a particle $\boldsymbol{X}_{(1)}$ amongst the $n-1$ largest particles (larger than a_j).
- for $t = 1, \ldots, T$, draw a new particle $\boldsymbol{X}^*$ with the pdf $q(\boldsymbol{X}_{(1)}, \cdot)$;
if $V(\boldsymbol{X}^*) > a_j$, then $\boldsymbol{X}_{(1)} = \boldsymbol{X}^*$ with probability $\min(p(\boldsymbol{X}^*)/p(\boldsymbol{X}_{(1)}), 1)$;
otherwise keep $\boldsymbol{X}_{(1)}$.
- replace the smallest particle by $\boldsymbol{X}_{(1)}$.

• Result 3: the distribution of $\boldsymbol{X}_{(1)}$ is the distribution μ_{a_j}. As $T \to \infty$, the distribution of $\boldsymbol{X}_{(1)}$ becomes independent of the other particles.

• Problem: how to generate the new particle with the distribution μ_{a_j} (of $\boldsymbol{X}$ knowing that $V(\boldsymbol{X}) > a_j$) ?
Version 2:
• Consider a transition kernel $q(\boldsymbol{x}', \boldsymbol{x})$ such that

$$p(\boldsymbol{x}')q(\boldsymbol{x}', \boldsymbol{x}) = p(\boldsymbol{x})q(\boldsymbol{x}, \boldsymbol{x}')$$

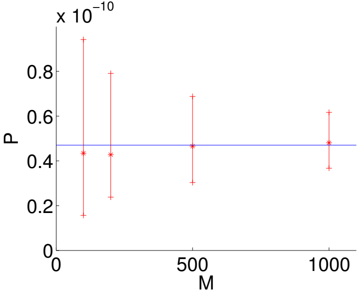
• Algorithm:
- $a_j =$ minimal value of the n particles.
- pick a particle $\boldsymbol{X}_{(1)}$ amongst the $n-1$ largest particles (larger than a_j).
- for $t = 1, \ldots, T$, draw a new particle $\boldsymbol{X}^*$ with the pdf $q(\boldsymbol{X}_{(1)}, \cdot)$; if $V(\boldsymbol{X}^*) > a_j$, then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

- Example:

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

with $\boldsymbol{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$, $d = 20$, $a = 0.95$, $V(\boldsymbol{x}) = |x_1|/|\boldsymbol{x}|$: $P = 4.704\,10^{-11}$.
Kernel $q : \boldsymbol{x}' \to \mathcal{N}\big(\frac{\boldsymbol{x}'}{\sqrt{1+\sigma^2}}, \frac{\sigma^2}{1+\sigma^2}\mathbf{I}_d\big)$, $\sigma = 0.3$, $T = 20$, ie

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



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

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

Conclusions

• Importance sampling: bias the input.
$\hookrightarrow$ Intrusive method.

• Interacting particle system: select the particles based on the output.
$\hookrightarrow$ No physical insight is required to guess the suitable biased input distribution.
But: need $V(\boldsymbol{X})$.
$\hookrightarrow$ 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.