Interacting particle systems
for the analysis of rare events

Josselin Garnier (Université Paris Diderot)
http://www.proba.jussieu.fr/~garnier

Cf http://www.proba.jussieu.fr/~garnier/expo2.pdf

Problem: estimation of the probability of occurence of a rare event.

Simulation by an Interacting Particle System.

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

Rare events

- Description of the system: Let E be a measurable space.

– $(\boldsymbol{X}_p)_{0\le p\le M}$: a E-valued Markov chain:

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

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

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

- Problem: estimation of the probability

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

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

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

Rare events

- Description of the system: Let E be a measurable space.

– $(\boldsymbol{X}_p)_{0\le p\le M}$: a E-valued Markov chain:

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

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

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

- Problem: estimation of the probability

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

when a is large $\Longrightarrow P \ll 1$.
We know how to simulate the Markov chain $(\boldsymbol{X}_p)_{0\le p\le M}$.

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

– $E = \mathbb{R}$,

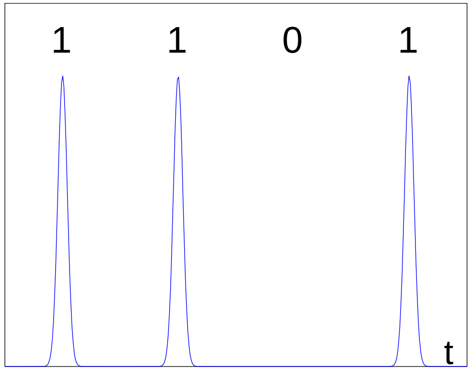
– $V(x) = x$,

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

Example: Optical communication in transoceanic optical fibers

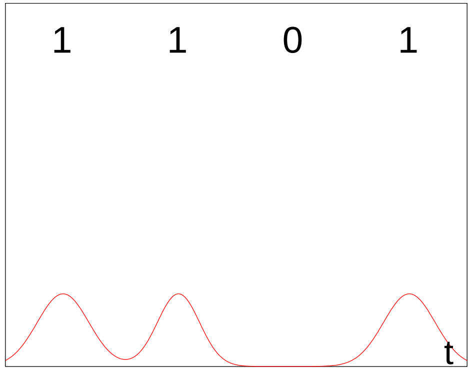
Optical fiber transmission principle:

- a binary message is encoded as a train of short light pulses.
- the pulse train propagates in a long optical fiber.
- the message is read at the output of the fiber.



Input pulse train

Transmission
$\longrightarrow$



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: Optical 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 (code OCEAN, Alcatel).

$\hookrightarrow$ 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: Optical communication in transoceanic optical fibers

Question: estimation of the probability of anomalous pulse spreading.

Rms pulse width after propagation distance z:

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

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 independent copies $((\boldsymbol{X}_0^i, \ldots, \boldsymbol{X}_M^i))_{1 \le i \le n}$ of $(\boldsymbol{X}_0, \ldots, \boldsymbol{X}_M)$ distributed with the original $\mathbb{P}$.

- Proposed estimator:

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

Unbiased estimator:

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

Variance:

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

The absolute error $\mathrm{Std}(\hat{P}_n) \simeq \sqrt{P}/\sqrt{n}$.
The 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.

Of course: P^{-1} is the minimum size of the sample required for one element to reach the rare level !

Importance Sampling method

• n independent copies $((\boldsymbol{X}_0^i, \ldots, \boldsymbol{X}_M^i))_{1\le i\le n}$ of $(\boldsymbol{X}_0, \ldots, \boldsymbol{X}_M)$ distributed with a biased distribution $\mathbb{Q}$.

• Proposed estimator:

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

Unbiased estimator:

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

Variance:

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

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

Optimal choice: $d\mathbb{Q} = \frac{\mathbf{1}_{V(\boldsymbol{X}_M)\ge a}}{\mathbb{P}(V(\boldsymbol{X}_M)\ge a)} d\mathbb{P}$. Impossible to apply ! But this result gives ideas (adaptive strategy, ...)

• Critical points: choice of the biased distribution + evaluation of the likelihood ratio + simulation of the biased dynamics (intrusive method).

Importance Sampling method driven by Large Deviations Principle

- Consider the family of twisted 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 independent copies $(\boldsymbol{X}_M^i)_{1\le i\le n}$ distributed with $\mathbb{P}^{(\lambda)}$.
- Estimator:

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

Variance:

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

where $\Lambda_M(\lambda) = \log \mathbb{E}_{\mathbb{P}}[e^{\lambda V(\boldsymbol{X}_M)}]$. For a judicious choice of λ,
$\lambda^* a - \Lambda_M(\lambda^*) = \sup_{\lambda>0}[\lambda a - \Lambda_M(\lambda)] \simeq -\ln P$ (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: $1/\sqrt{Pn}$).

Twisted Feynman-Kac path measures

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

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

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

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

Examples:

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

- 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 $K_p(\boldsymbol{x}_{p-1}, d\boldsymbol{x}_p)$:

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

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

- Recursive relation:

$$\mu_p = \Theta_p(\mu_{p-1}) =_{\text{def.}} \int_{E^p} \mu_{p-1}(d\boldsymbol{y}_{p-1})\mathcal{K}_p(\boldsymbol{y}_{p-1}, \cdot)$$

with $\mu_0 = \delta_{\boldsymbol{x}_0}$.
$\mathcal{K}_p(\boldsymbol{y}_{p-1}, d\boldsymbol{y}'_p)$: Markov transitions associated to the chain $\boldsymbol{Y}_p$:

$$\mathcal{K}_p(\boldsymbol{y}_{p-1}, d\boldsymbol{y}'_p) = \delta_{\boldsymbol{y}_{p-1}}(d\boldsymbol{y}'_{p,0}, \ldots, d\boldsymbol{y}'_{p,p-1})K_p(\boldsymbol{y}_{p-1,p-1}, d\boldsymbol{y}'_{p,p})$$

Here $\boldsymbol{y}_{p-1} = (\boldsymbol{y}_{p-1,0}, \ldots \boldsymbol{y}_{p-1,p-1}) \in E^p$, $\boldsymbol{y}'_p = (\boldsymbol{y}'_{p,0}, \ldots \boldsymbol{y}'_{p,p}) \in E^{p+1}$):

$\hookrightarrow$ Linear evolution.

Proof:

$$\begin{aligned}
\mu_p(f_p) &= \mathbb{E}[f_p(\boldsymbol{Y}_{p-1}, \boldsymbol{X}_p)] \\
&= \int_{E^p} \mu_{p-1}(d\boldsymbol{y}_{p-1})\mathbb{E}\big[f_p(\boldsymbol{y}_{p-1}, \boldsymbol{X}_p)|\boldsymbol{Y}_{p-1} = \boldsymbol{y}_{p-1}\big] \\
&= \int_{E^p} \mu_{p-1}(d\boldsymbol{y}_{p-1}) \int_E K_p(\boldsymbol{y}_{p-1,p-1}, d\boldsymbol{x}_p) f_p(\boldsymbol{y}_{p-1}, \boldsymbol{x}_p) \\
&= \int_{E^p} \mu_{p-1}(d\boldsymbol{y}_{p-1}) \int_{E^{p+1}} d\boldsymbol{y}'_p \mathcal{K}_p(\boldsymbol{y}_{p-1}, d\boldsymbol{y}'_p) f_p(\boldsymbol{y}'_p) \\
&= \Theta_p(\mu_{p-1})(f_p)
\end{aligned}$$

Unnormalized measures

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

FK measure γ_p associated to the pair potentials/transitions (G_p, K_p):

$$\gamma_p(f_p) = \mathbb{E}\Big[f_p(\mathbf{Y}_p) \prod_{1 \le k < p} G_k(\mathbf{Y}_k)\Big]$$

- Expression of P in terms of γ_M:

$$P = \gamma_M(g)$$

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

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

- Recursive relation:

$$\gamma_p = \Psi_p(\gamma_{p-1}) =_{\text{def.}} \int_{E^p} \gamma_{p-1}(d\boldsymbol{y}_{p-1}) G_{p-1}(\boldsymbol{y}_{p-1}) \mathcal{K}_p(\boldsymbol{y}_{p-1}, \cdot)$$

$\mathcal{K}_p(\boldsymbol{y}_{p-1}, d\boldsymbol{y}'_p)$: Markov transitions associated to the chain $\boldsymbol{Y}_p$

$$\mathcal{K}_p(\boldsymbol{y}_{p-1}, d\boldsymbol{y}'_p) = \delta_{\boldsymbol{y}_{p-1}}(d\boldsymbol{y}'_{p,0}, \ldots, d\boldsymbol{y}'_{p,p-1}) K_p(\boldsymbol{y}_{p-1,p-1}, d\boldsymbol{y}'_{p,p})$$

$\hookrightarrow$ Linear evolution.

Normalized 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:

$$P = \eta_M(g) \prod_{1 \le p < M} \eta_p(G_p)$$

Proof:

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

Normalizing constant:

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

$\rightarrow$ If one can compute/estimate $(\eta_p)_{p=1,\ldots,M}$, then one can compute/estimate P.

- Recursive relation:

$$\eta_p = \Phi_p(\eta_{p-1}) =_{\text{def.}} \int_{E^p} \eta_{p-1}(d\boldsymbol{y}_{p-1}) G_{p-1}(\boldsymbol{y}_{p-1}) \mathcal{K}_p(\boldsymbol{y}_{p-1}, \cdot)/\eta_{p-1}(G_{p-1})$$

$\hookrightarrow$ Nonlinear evolution.

Interacting path-particle system

- Goal: simulate the original measures

$$\mu_p = \Theta_p(\mu_{p-1})$$

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

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

Interacting path-particle system

- Goal: simulate the original measures

$$\mu_p = \Theta_p(\mu_{p-1})$$

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

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

- Goal: simulate the normalized measures

$$\eta_p = \Phi_p(\eta_{p-1})$$

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

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

- Key points:

– Nonlinear $\Phi_p \to$ interacting particle system

– Simulation technique

– Fixed number of particles $(\eta_p(1) = 1)$

Interacting path-particle system

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

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

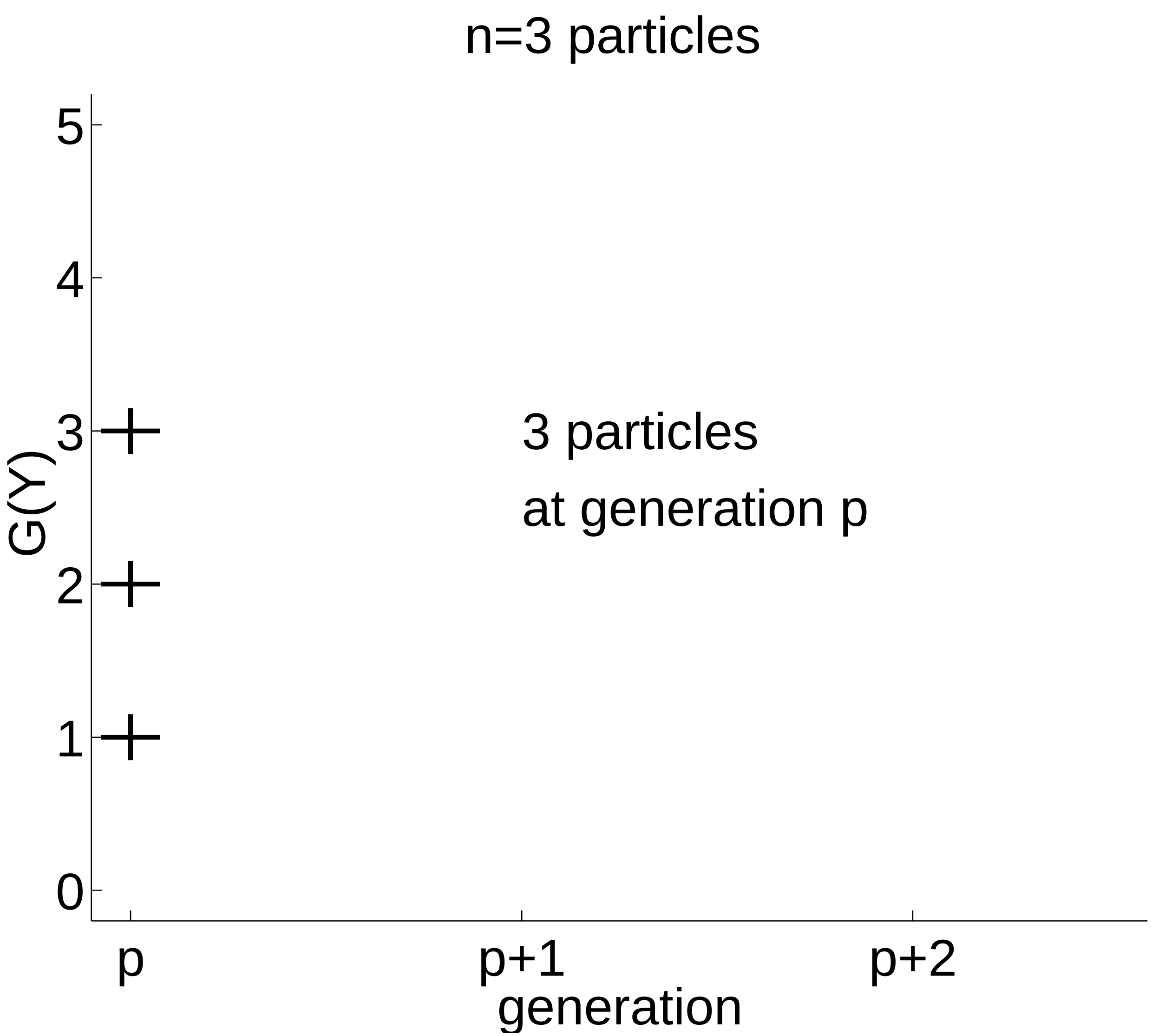
Answer: System of path-particles, whose empirical measure will be approximately $\mathbb{Q}$.

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

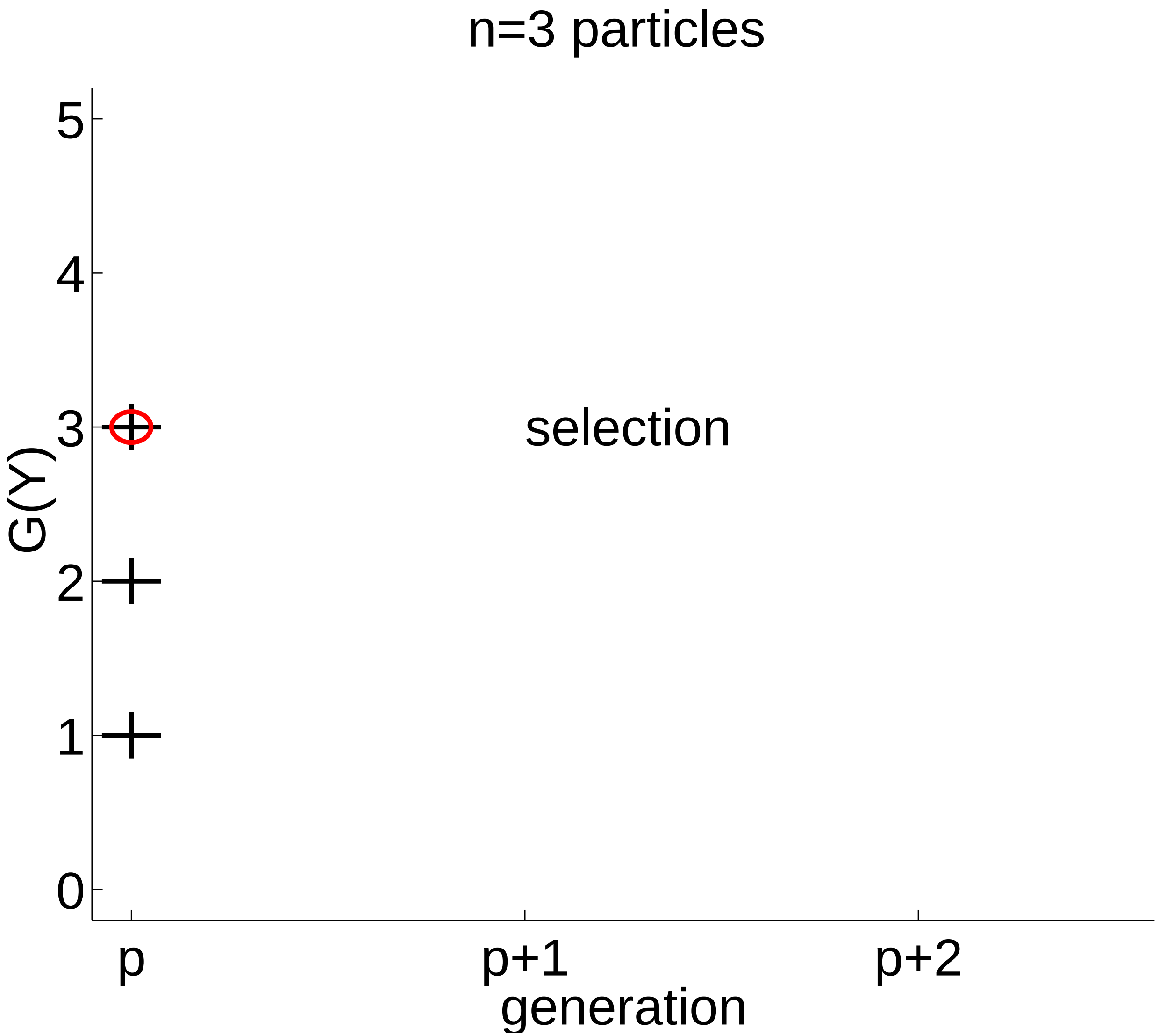
– Initialization: $p = 0$: $\boldsymbol{Y}_0^i = \boldsymbol{x}_0$ for all $i = 1, \ldots, n$.

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

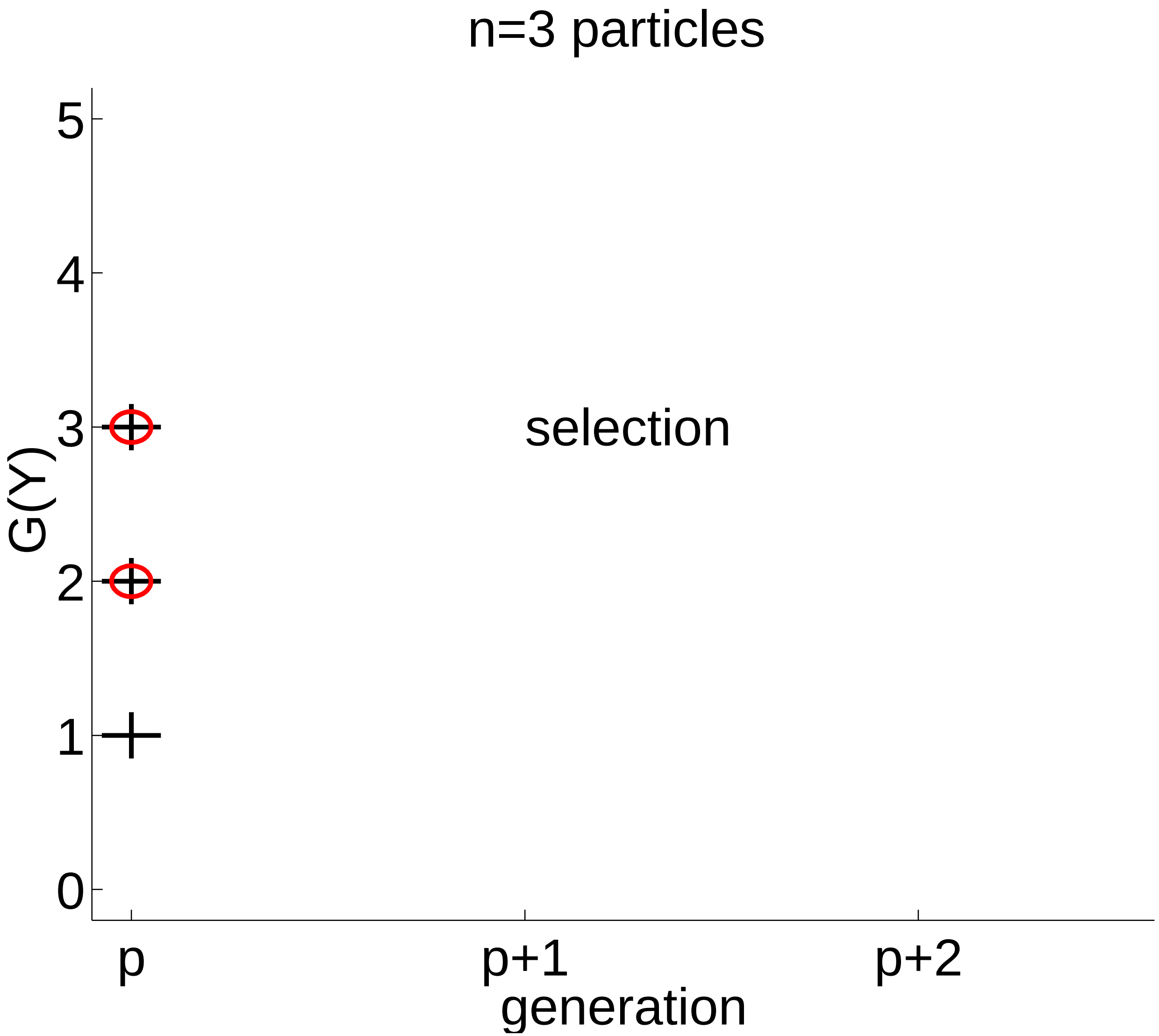
$$\mathbb{Y}_p \in (E^{p+1})^n \xrightarrow{selection} \widehat{\mathbb{Y}}_p \in (E^{p+1})^n \xrightarrow{mutation} \mathbb{Y}_{p+1} \in (E^{p+2})^n$$



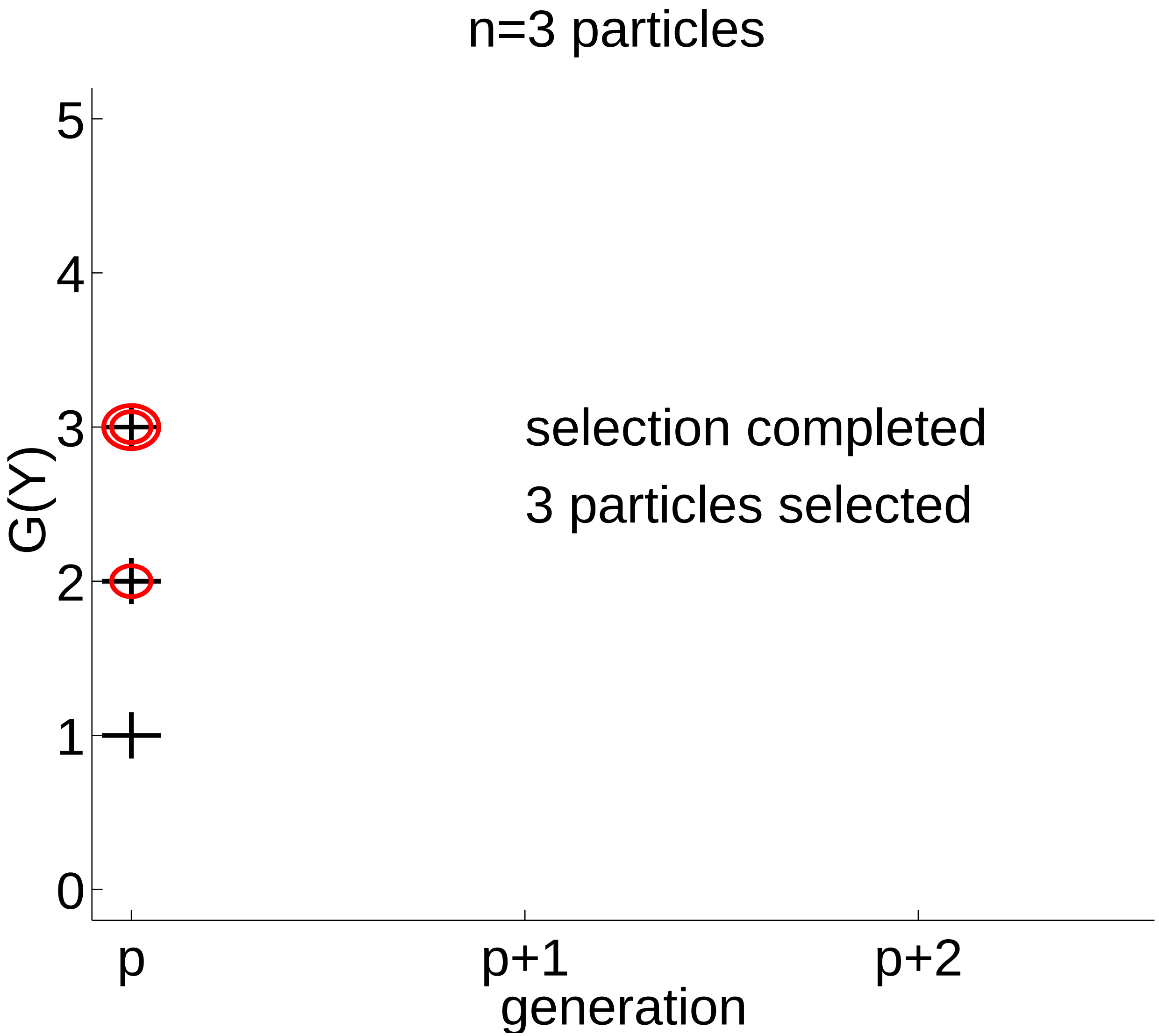
3 particles $\boldsymbol{Y}_p^1, \boldsymbol{Y}_p^2, \boldsymbol{Y}_p^3$ at generation p,
with potential weights $G(\boldsymbol{Y}_p^1) = 1$, $G(\boldsymbol{Y}_p^2) = 2$,$G(\boldsymbol{Y}_p^3) = 3$.



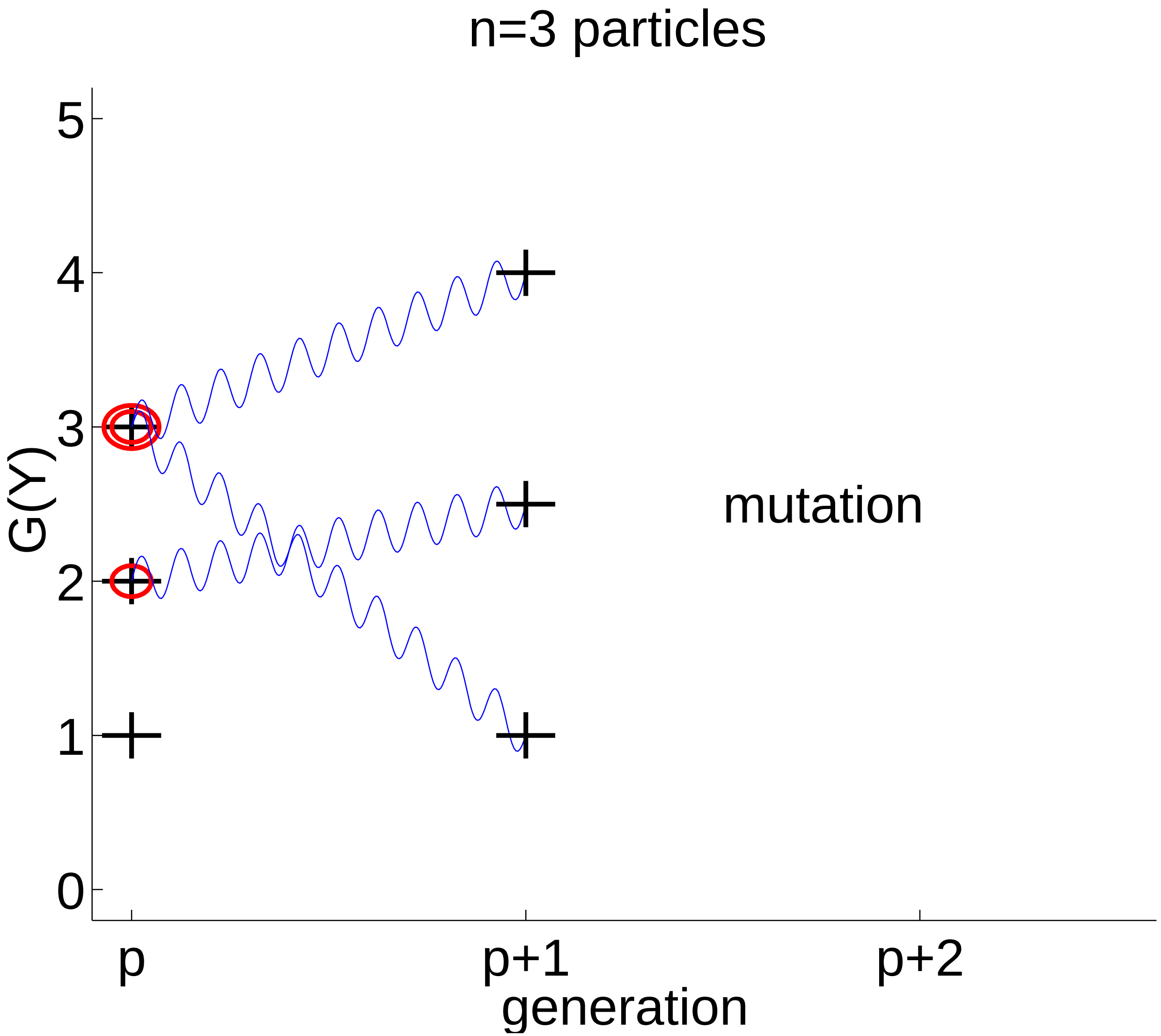
Probability to select particle j: $\dfrac{G(\boldsymbol{Y}_p^j)}{G(\boldsymbol{Y}_p^1)+G(\boldsymbol{Y}_p^2)+G(\boldsymbol{Y}_p^3)} = \begin{cases} \frac{1}{6} \text{ if } j=1 \\ \frac{1}{3} \text{ if } j=2 \\ \frac{1}{2} \text{ if } j=3 \end{cases}$



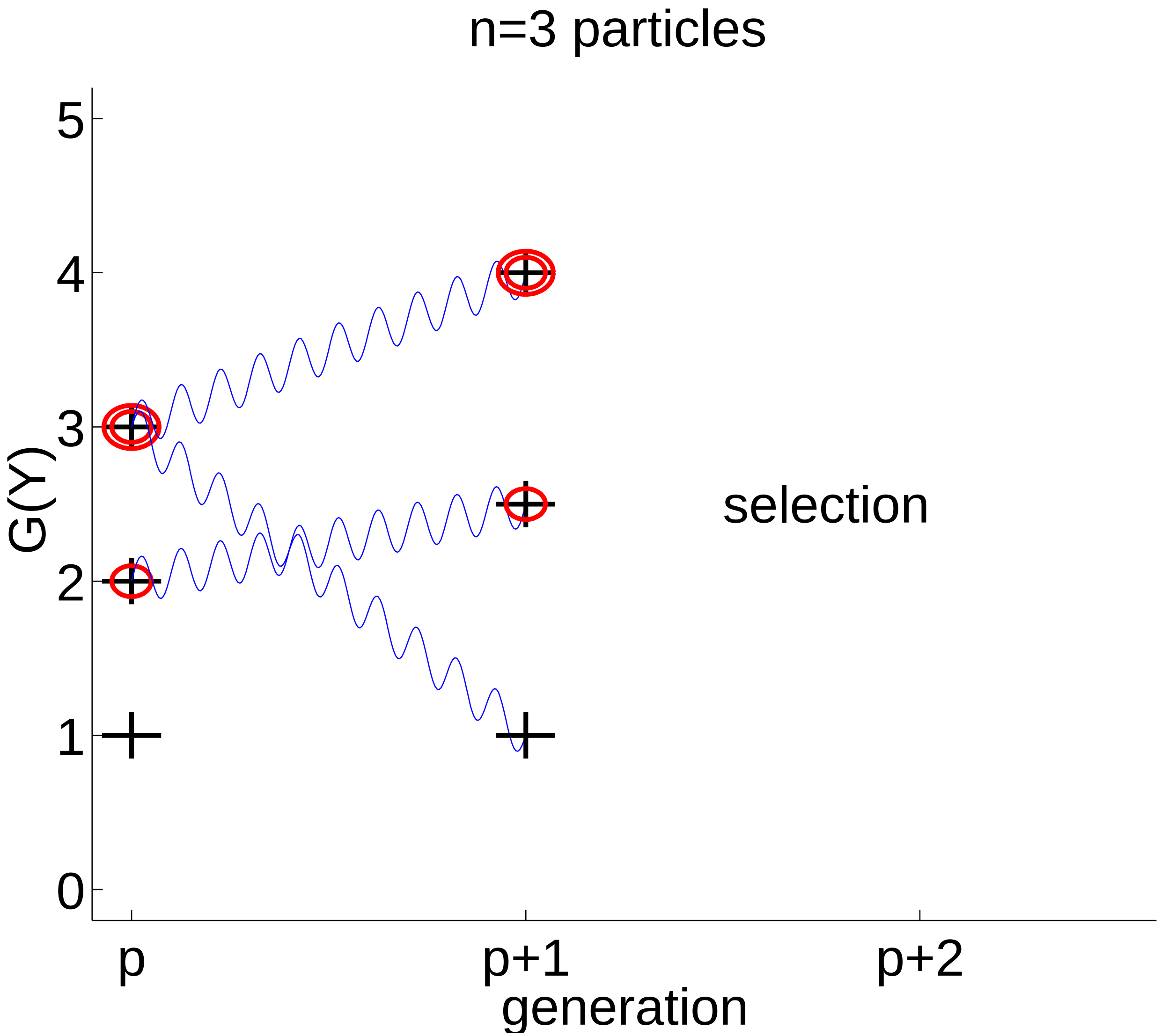
Probability to select particle j: $\dfrac{G(\mathbf{Y}_p^j)}{G(\mathbf{Y}_p^1)+G(\mathbf{Y}_p^2)+G(\mathbf{Y}_p^3)} = \begin{cases} \frac{1}{6} \text{ if } j=1 \\ \frac{1}{3} \text{ if } j=2 \\ \frac{1}{2} \text{ if } j=3 \end{cases}$



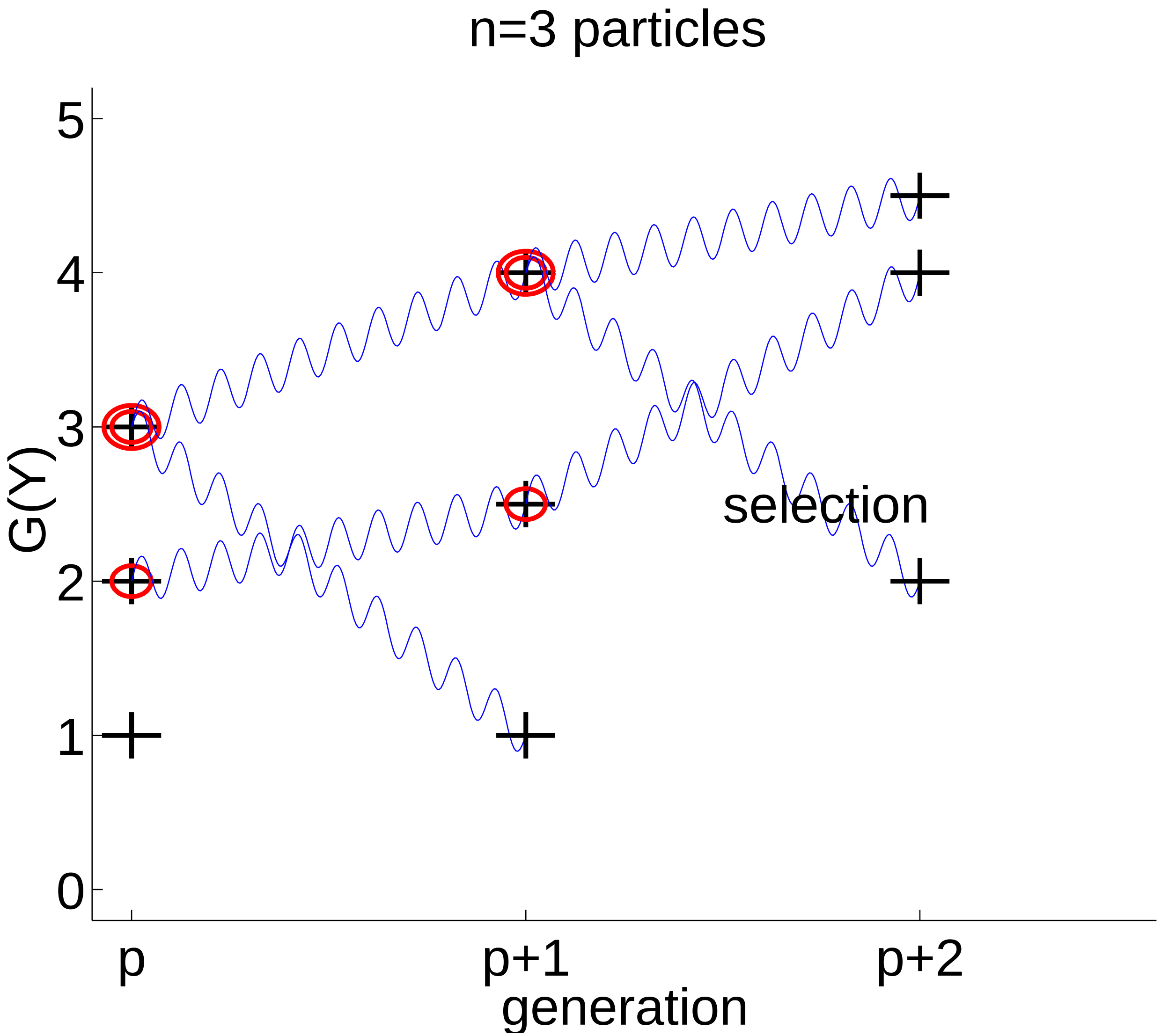
Probability to select particle j: $\dfrac{G(\boldsymbol{Y}_p^j)}{G(\boldsymbol{Y}_p^1)+G(\boldsymbol{Y}_p^2)+G(\boldsymbol{Y}_p^3)} = \begin{cases} \frac{1}{6} \text{ if } j=1 \\ \frac{1}{3} \text{ if } j=2 \\ \frac{1}{2} \text{ if } j=3 \end{cases}$



Each particle evolve 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 $\mathbb{Y}_p = (\boldsymbol{Y}_p^i)_{1 \leq i \leq n}$, 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 $\widehat{Y}_p^i$ is extended by an elementary unbiased K_p-transition:

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

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 twisted measures:

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

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

$$\hat{P}_n = \eta_M^n(g) \prod_{1\leq p<M} \eta_p^n(G_p)$$

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

Proof. asymptotic analysis of genealogical particle models.

cf P. Del Moral, *Feynman-Kac formulae, genealogical and interacting particle systems with applications,* Springer, New York, 2004.

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

Estimator of the probability of the rare event

Let

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

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

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

such that

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

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

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

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

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

- Useful for

1) the choice of "good" functions G_p (variance reduction)

2) the design of an estimator of the asymptotic variance.

Sketch of proof

Local errors: introduce the random field $\mathcal{W}_p^n$ given by

$$\mathcal{W}_p^n(f_p) = \sqrt{n}\,[\eta_p^n - \Phi_p(\eta_{p-1}^n)](f_p), \qquad \text{for } f_p \in L^\infty(E)$$

Central limit theorem: The sequence $(\mathcal{W}_p^n)_{1\le p\le M}$ converges in law, as $n \to \infty$, to a sequence of M independent, Gaussian and centered random fields $(\mathcal{W}_p)_{1\le p\le M}$

$$\mathbb{E}\,[\mathcal{W}_p(f_p)\mathcal{W}_p(g_p)] = \eta_p\big([f_p - \eta_p(f_p)][g_p - \eta_p(g_p)]\big)$$

Global error: Let $Q_{p,M}$, with $1 \le p \le M$, be the FK semi-group associated to the flow $\gamma_M = \gamma_p Q_{p,M}$. Using the Markov property,

$$Q_{p,M}(f_M)(\boldsymbol{y}_p) = \mathbb{E}\left[f_M(\boldsymbol{Y}_M) \prod_{p\le k<M} G_k(\boldsymbol{Y}_k) \mid \boldsymbol{Y}_p = \boldsymbol{y}_p\right]$$

Telescopic decomposition

$$\gamma_M^n - \gamma_M = \sum_{p=1}^{M} [\gamma_p^n Q_{p,M} - \gamma_{p-1}^n Q_{p-1,M}] = \sum_{p=1}^{M} [\gamma_p^n - \gamma_{p-1}^n Q_{p-1,p}] Q_{p,M}$$

Use $\gamma_{p-1}^n Q_{p-1,p} = \gamma_{p-1}^n(G_{p-1})\Phi_{p-1}(\eta_{p-1}^n)$ and $\gamma_{p-1}^n(G_{p-1}) = \gamma_p^n(1)$.

$$\gamma_M^n - \gamma_M = \sum_{p=1}^{M} \gamma_p^n(1)[\eta_p^n - \Phi_{p-1}(\eta_{p-1}^n)]Q_{p,M}$$

As a result:

$$\mathcal{W}_M^{\gamma,n}(f_M) =_{\mathrm{def.}} \sqrt{n}[\gamma_M^n - \gamma_M](f_M) = \sum_{p=1}^{M} \gamma_p^n(1)\ \mathcal{W}_p^n(Q_{p,M} f_M)$$

Consider

$$\sqrt{n}\ [\hat{P}_n - P] = \mathcal{W}_M^{\gamma,n}(g)$$

Thus $\mathcal{W}_M^{\gamma,n}(g)$ converge in law, as $n \to \infty$, to a centered Gaussian random variable $\mathcal{W}_M^{\gamma}(g)$ with the variance

$$\sigma_M^2 =_{\mathrm{def.}} \mathbb{E}(\mathcal{W}_M^{\gamma}(g)^2) = \sum_{p=1}^{M} \gamma_p(1)^2\ \eta_p\big([Q_{p,M}(g) - \eta_p Q_{p,M}(g)]^2\big)$$

Variance comparisons for the Gaussian model $X_p = X_{p-1} + \theta_p$

where $(\theta_p)_{1 \leq p \leq M}$ independent, Gaussian, zero-mean, variance one,

$V(x) = x$.

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 $P \ll 1$.

First choice for the potential:

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

Calculations show

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

By optimizing, 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 $\sim 1/\sqrt{nP^{2/3}}$.

Consider the same model.

Second choice for the potential:

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

We obtain:

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

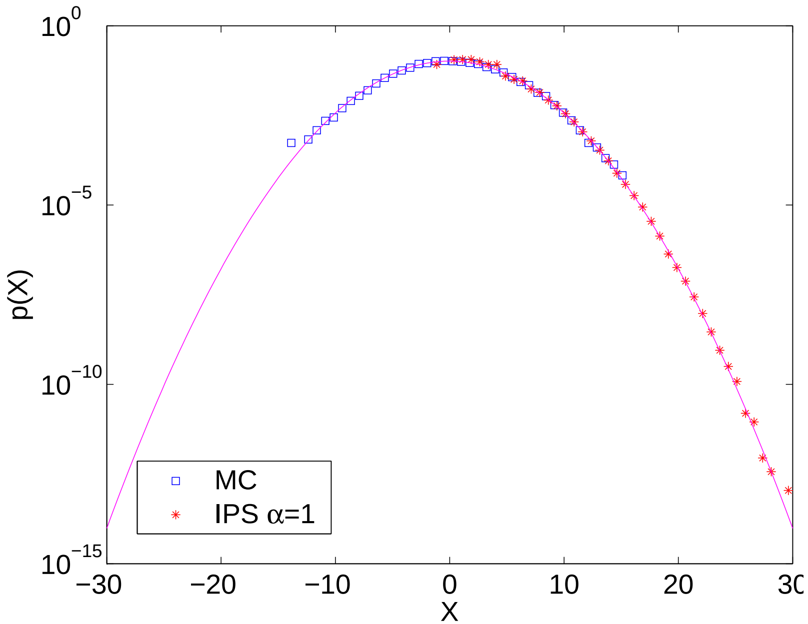
By optimizing, $\alpha = a/M$, we get

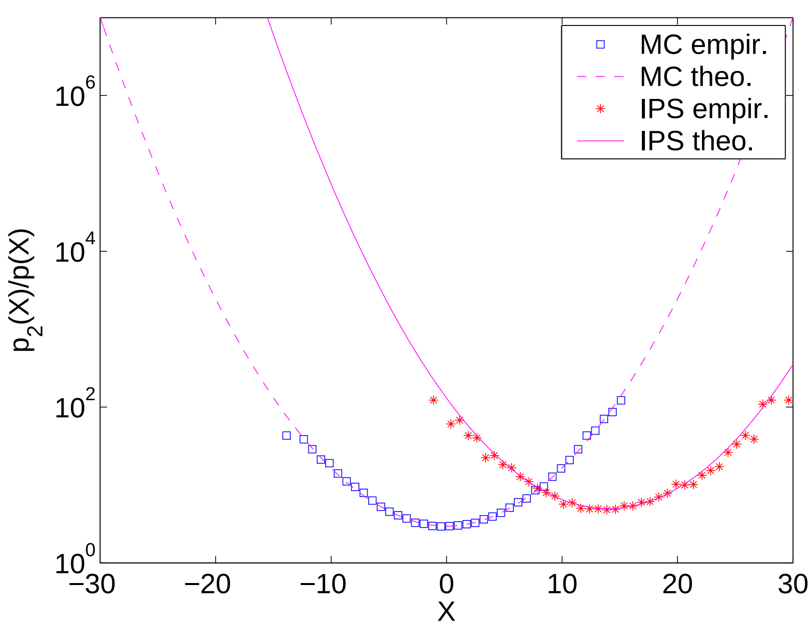
$$\sigma^2 \sim e^{-\frac{a^2}{M}\left(1 - \frac{1}{M}\right)}$$

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

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

Numerical simulations with the Gaussian model





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

Optical 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 $E = H_0^2(\mathbb{R}) \cap L_2^2(\mathbb{R})$

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

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

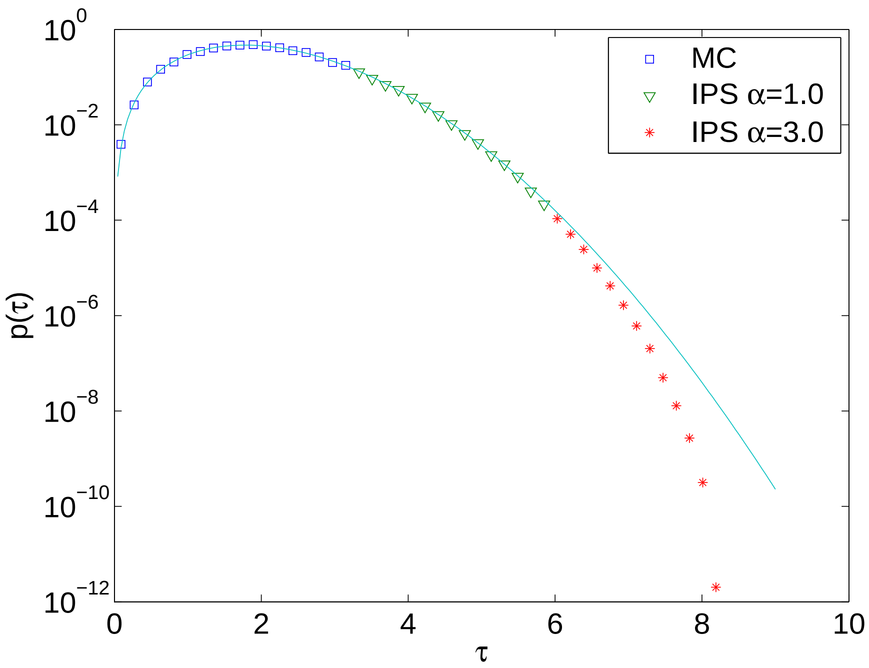
1) asymptotic model (separation of scales technique)
$\to$ the rms pulse width $\tau(z)$ is a diffusion process and its pdf is

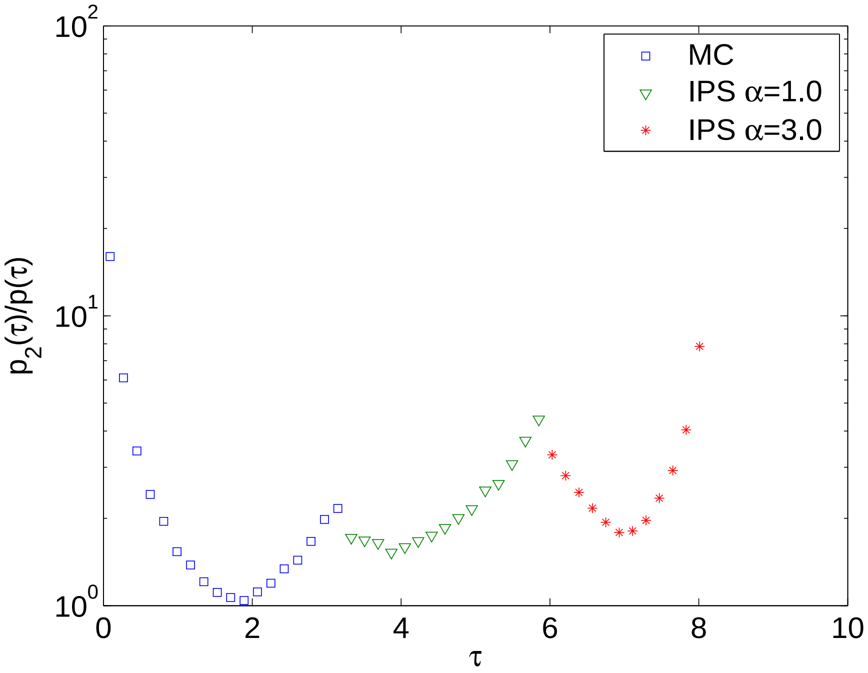
$$p_z(\tau) = \frac{\tau^{1/2}}{\sqrt{2\pi}(4\sigma^2 z)^{3/2}} \exp\left(-\frac{\tau}{8\sigma^2 z}\right) \mathbf{1}_{[0,\infty)}(\tau)$$

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.

Numerical simulations with the PMD model





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

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

- Splitting strategy:

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

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

- Estimate P_j separately.

- Two key issues:

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

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 $\alpha \to 1$.

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

$\hookrightarrow$ $(\boldsymbol{X}_0^1, \ldots, \boldsymbol{X}_0^n)$ independent and identically distributed with the distribution 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:

$a_j = \min_{i=1,\ldots,n}\{(V(\boldsymbol{X}_{j-1}^i)\}$,

  - remove the particle that achieves the minimum,
  - generate a new particle with the conditional distribution μ_{a_j} of $\boldsymbol{X}$ knowing that $V(\boldsymbol{X}) > a_j$:

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

  (use the Metropolis-Hastings algorithm).

$\hookrightarrow$ $(\boldsymbol{X}_j^1, \ldots, \boldsymbol{X}_j^n)$ independent and identically distributed with the distribution μ_{a_j}

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

• 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 $-n \ln P$.

Proof:

- if $V(\boldsymbol{X})$ has continuous cumulative distribution function F, then $F(V(\boldsymbol{X}))$ is a uniform random variable and $-\log(1 - F(V(\boldsymbol{X})))$ is an exponential random variable.

- 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,

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

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

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

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

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

$$\mathbb{P}\big(\Lambda(V(\boldsymbol{X})) \geq \lambda\big) = \mathbb{P}\big(1 - F(V(X_0)) \leq 1 - e^{-\lambda}\big) = e^{-\lambda}$$

Therefore $(\Lambda(V(\boldsymbol{X}_0^i)))_{i=1,\dots,n}$ are i.i.d. with the distribution $\mathcal{E}(1)$.
Let $a_1 = \min_{i=1,\dots,n}\{V(\boldsymbol{X}_0^i)\}$. We have $\Lambda(a_1) = \min_{i=1,\dots,n}\{\Lambda(V(\boldsymbol{X}_0^i))\}$.

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

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

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

- 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

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

In fact

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

Moreover, denoting

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

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 can check (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:

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

• 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 $t = 1, \dots, T$, 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.

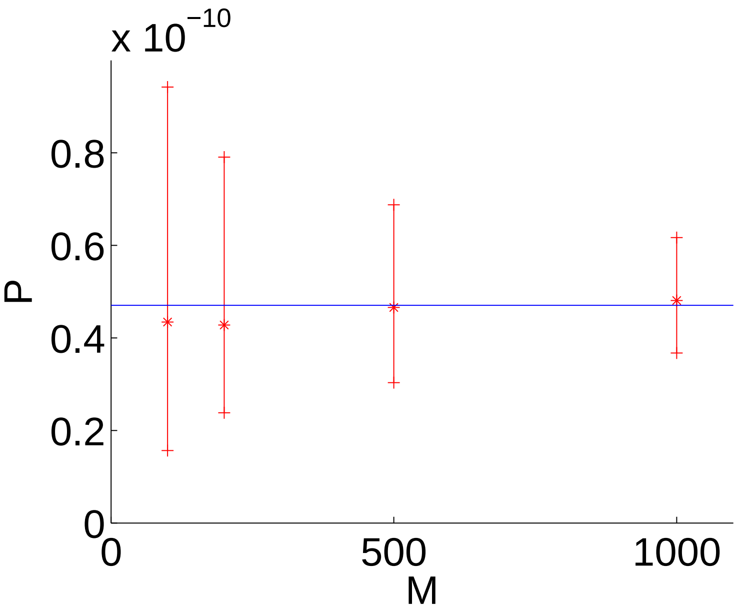
Example:

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

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

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.

Cf: F. Cerou, A. Guyader (Rennes), P. Glasserman, R. Rubinstein.

Conclusions

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

• The real distribution is used, not a twisted one.
↪ Non-intrusive method: no need to change the numerical code.

• Number of particles fixed, computational cost (almost) fixed.

• 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.