Rare event simulation for molecular dynamics (Part 2)

F. Cerou

Frederic.Cerou@inria.fr

Joint work with P. Del Moral, A. Guyader, F. Malrieu

Inria Rennes Bretagne Atlantique

ICMS Computational methods

Rare event

Low but non-zero probability, say $0 < \mathbb{P}(X \in A) \leq 10^{-8}$.
It is very important to know this probability, or to generate typical realizations.
We assume that we know how to draw (pseudo) realizations of X.
Different from extreme values: here we have a model that we can simulate, extreme values refer to some dataset used with purely statistical approaches.

Why naive Monte-Carlo does not work ? Let $p = \mathbb{P}(X \in A)$.
$\hat{p} = \sum_{i=1}^{N} \mathbb{1}_A(X_i)$, and the relative mean square error is $\frac{\mathrm{var}(\hat{p})}{p^2} = \frac{1-p}{Np}$. So we need $N \simeq 10/p$ at least...

Two main frameworks: importance sampling vs. importance splitting.
Splitting can work with a blackbox for $\mathbb{1}_A$, and gives rare events drawn with the original distribution.

Some applications

- Particle transmission (Kahn and Harris 1951)
- Queueing networks
- Air traffic management
- Satellite versus debris collision
- Finance
- Food contaminant exposure
- Molecular dynamics
- Digital Rights management
- Counting finite sets (satisfiability)
- Epidemiology

Importance splitting based algorithms
X on $\mathbb{R}^d$, distribution η, we want to estimate $p = \mathbb{P}(X \in A) = \eta(A)$, or more generally $\mathbb{E}(f(X)\mathbb{1}_A(X))/p = \eta(f\mathbb{1}_A)/\eta(A)$ for any test function f.

$$A = A_n \subset A_{n-1} \subset \cdots \subset A_1 \subset A_0 = \mathbb{R}^d \ ,$$

such that $\mathbb{P}(A_{k+1} \,|\, A_k)$ is not small.
At each level, we need an ingredient to improve the diversity of the sample.
Denote $\eta_k = \mathbb{1}_{A_k}\eta/\eta(A_k)$.
Choose a Markov kernel K such that η is K invariant and
$\forall x, \forall y, \eta(dx)K(x,dy) = \eta(dy)K(y,dx)$.
Construct $M_k(x,dy) =$
$(K(x,dy)\mathbb{1}_{A_k}(y) + (1 - K(x,A))\delta_x(dy))\mathbb{1}_{A_k}(x) + (1 - \mathbb{1}_{A_k}(x))\delta_x(dy)$.
Then η_k is M_k invariant.

Link with SIS and Feynman-Kac formula.
Normalized/Un-normalized measures

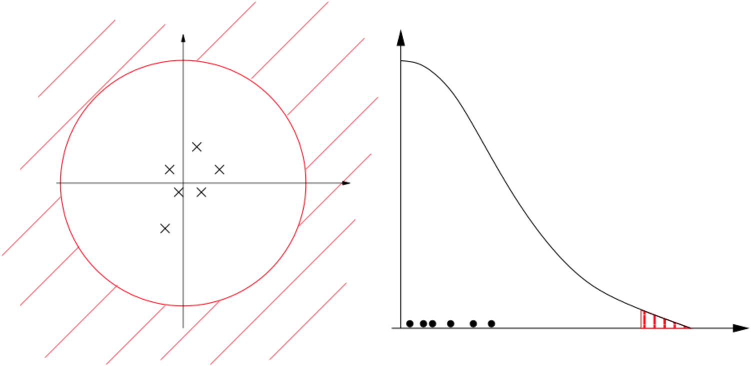
$$\eta_n(f) = \frac{\mathbb{E}[f(X_n)\prod_{k=0}^{n-1}\mathbb{1}_{A_{k+1}}(X_k)]}{\mathbb{E}[\prod_{k=0}^{n-1}\mathbb{1}_{A_{k+1}}(X_k)]} = \frac{\gamma_n(f)}{\gamma_n(1)},$$

where the X_k's form a Markov chain with transition kernels M_k.
We also have

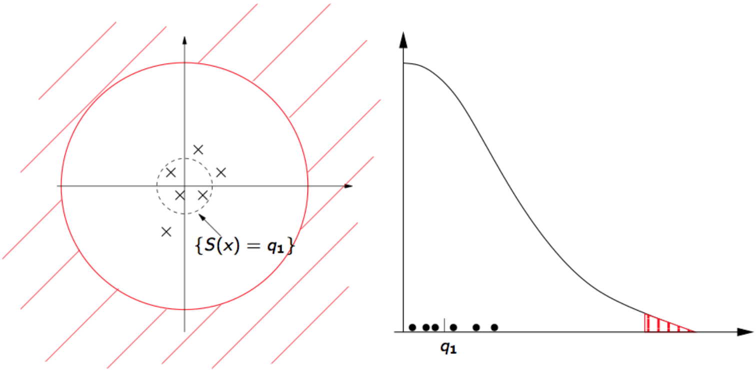
$$\eta(A) = \gamma_n(1) = \mathbb{E}[\prod_{k=0}^{n-1}\mathbb{1}_{A_{k+1}}(X_k)] = \prod_{k=0}^{n-1}\eta_k(A_{k+1}).$$

We can use SIS algorithms (interacting particles approximation) and mathematical tools (convergence).

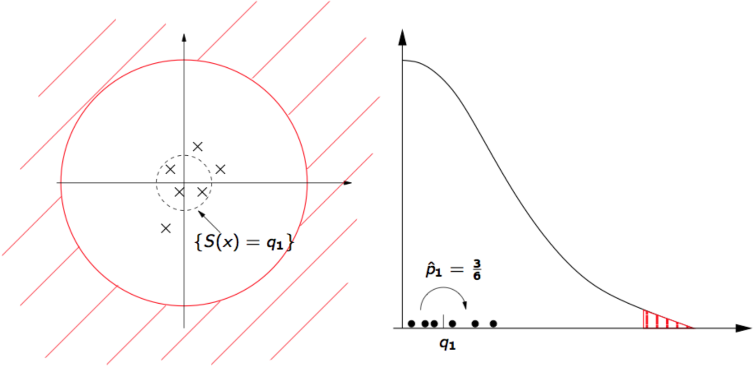
Special case $A_k = \{S(X) > q_k\}$ for some real valued S.



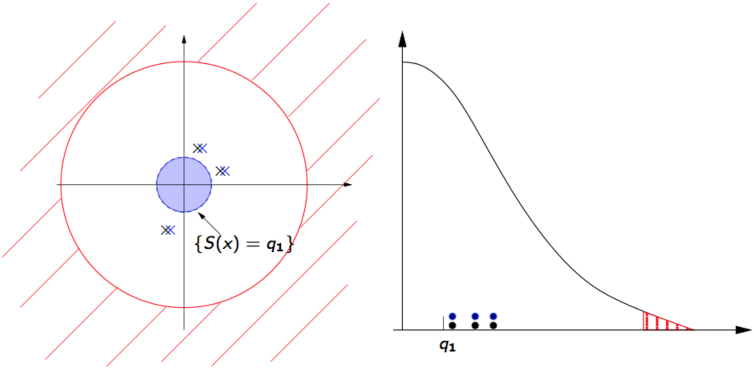
Special case $A_k = \{S(X) > q_k\}$ for some real valued S.



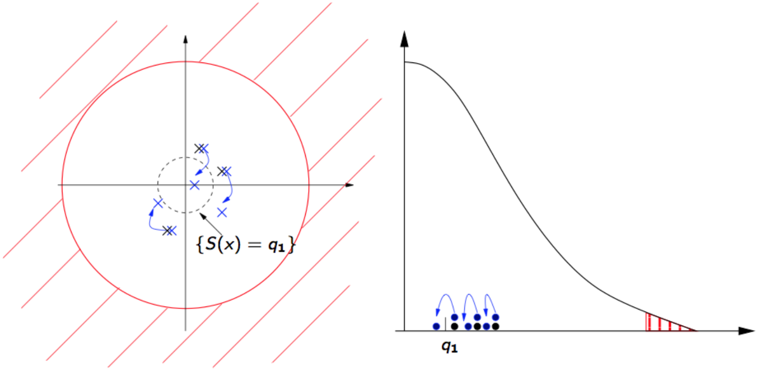
Special case $A_k = \{S(X) > q_k\}$ for some real valued S.



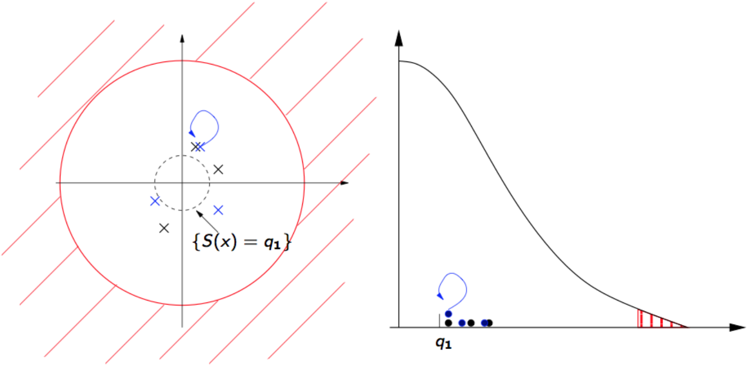
Special case $A_k = \{S(X) > q_k\}$ for some real valued S.



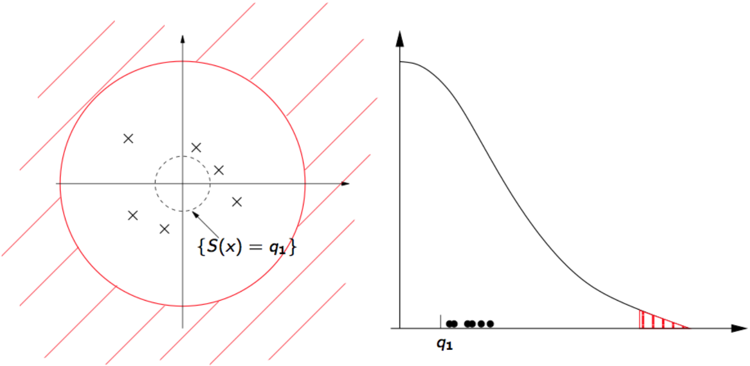
Special case $A_k = \{S(X) > q_k\}$ for some real valued S.



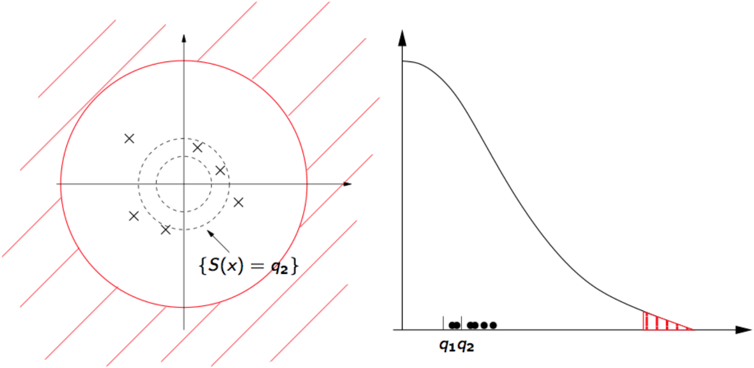
Special case $A_k = \{S(X) > q_k\}$ for some real valued S.



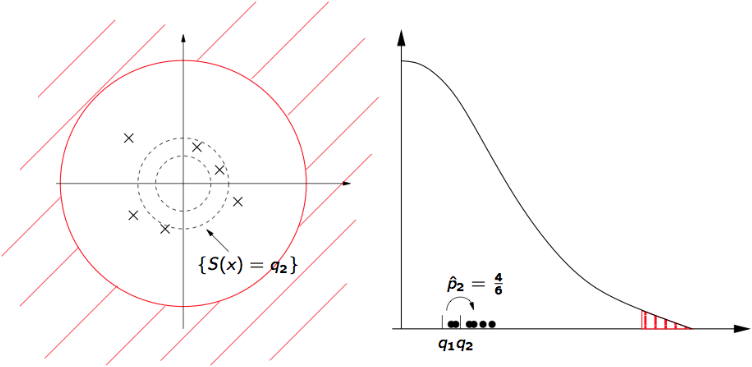
Special case $A_k = \{S(X) > q_k\}$ for some real valued S.



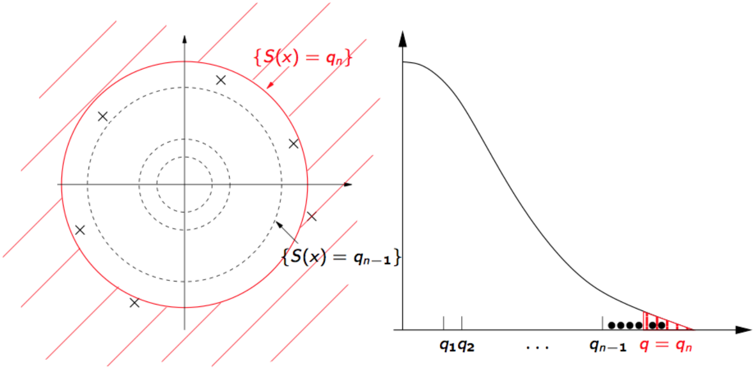
Special case $A_k = \{S(X) > q_k\}$ for some real valued S.



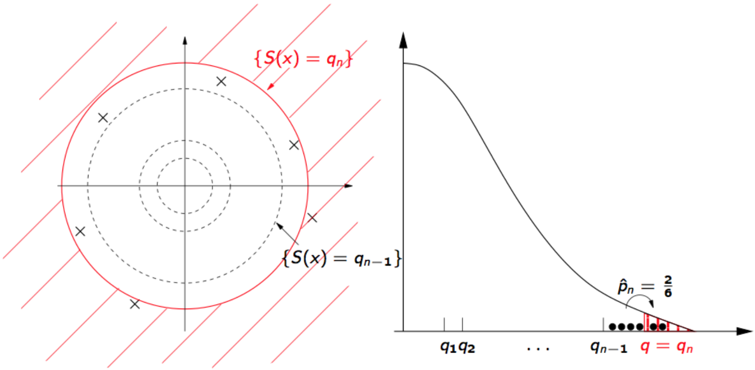
Special case $A_k = \{S(X) > q_k\}$ for some real valued S.



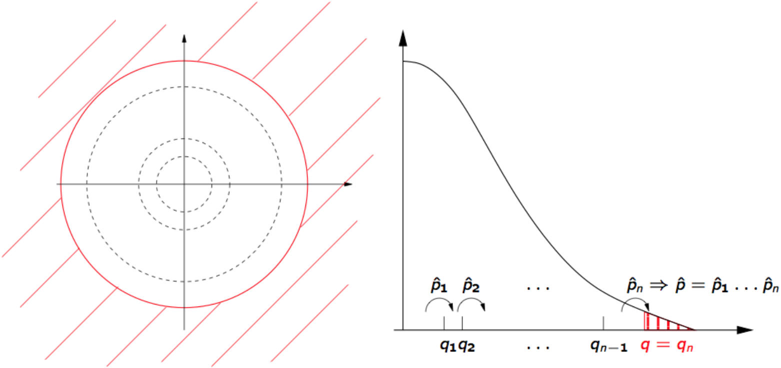
Special case $A_k = \{S(X) > q_k\}$ for some real valued S.



Special case $A_k = \{S(X) > q_k\}$ for some real valued S.



Special case $A_k = \{S(X) > q_k\}$ for some real valued S.



Particle Approximation

- At each level we have a sample $(X_k^1, \ldots, X_k^N)$ identically distributed with η_k but non independent.
- $\hat{p}_{j+1} = \eta_j^N(A_{j+1})$ the proportion of particles that reach the next level.
- Empirical normalized measures:

$$\eta_k^N(f) = \frac{1}{N} \sum_{i=1}^{N} f(X_k^i) \longrightarrow \mathbb{E}[f(X) \,|\, X \in A_k]$$

- Empirical unnormalized measures:

$$\gamma_k^N(1) = \hat{p}_1 \ldots \hat{p}_k \longrightarrow p_k = \eta(A_k)$$

$$\gamma_k^N(f) = \gamma_k^N(1)\eta_k^N(f) \longrightarrow \mathbb{E}[f(X)\mathbb{1}_{A_k}(X)].$$

Some remarks

The kernel K should be mixing, but not too much, otherwise the acceptance transition rate may be low. It can also depend on k.

Several different ways of resampling:

- Multinomial on the succesful particles (not optimal)
- Multinomial only for the failed particles, the successful ones stay where they are
- Maximal spread of the failed particles among the successful ones (difficult to analyse)

Fluctuations analysis

Multinomial resampling.
For $0 \leq \ell \leq n$ we define

$$Q_{\ell,n}(f)(x) = \mathbb{E}[f(X_n) \prod_{k=\ell}^{n-1} \mathbb{1}_{A_{k+1}}(X_k) \,|\, X_\ell = x]$$

and $\bar{Q}_{\ell,n}(f) = Q_{\ell,n}(f)/Q_{\ell,n}(1)$.

Theorem (CLT, Del Moral and Jacod 2001)

When $N \to \infty$, we have the following convergences in distribution

$$\sqrt{N}(\gamma_n^N(f) - \gamma_n(f)) \longrightarrow N(0, p^2\Sigma(f)),$$

and

$$\sqrt{N}(\eta_n^N(f) - \eta_n(f)) \longrightarrow N(0, \Sigma(f)),$$

where $\Sigma(f) = \sum_{k=0}^{n} \eta_k(\bar{Q}_{k,n}(f)^2 - \eta_n(f)^2)$.

Proof

Strong law of large numbers can be done by recurrence. Martingale decomposition

$$\gamma_n^N(f) - \gamma_n(f) = M_n^N(f) = \sum_{p=0}^{n} \gamma_p^N(1)[\eta_p^N(Q_{p,n}(f)) - \Phi_p(\eta_{p-1}^N)(Q_{p,n}(f))]$$

Each term can be viewed as a propagation of a local error term.
Then apply a general CLT-type theorem (for triangular arrays of Martingales).
Normalized measures

$$\eta_n^N(f) - \eta_n(f) = \frac{\gamma_n(1)}{\gamma_n^N(1)} \gamma_n^N(\frac{1}{\gamma_n(1)}(f - \eta_n(f)).$$

Best possible variance

Estimation of p. $p_k = \mathbb{P}(X \in A_k \mid X \in A_{k-1})$ and $\hat{p}_k = \eta^N_{k-1}(A_k)$, $\hat{p} = \gamma^N_n(1) = \hat{p}_1 \dots \hat{p}_n$.

$$\sqrt{N}(\gamma^N_n(1) - \gamma_n(1)) \longrightarrow N(0, p^2 \Sigma(1)),$$

with

$$\Sigma(1) = \sum_{k=1}^{n} \frac{1-p_k}{p_k} + \sum_{k=1}^{n-1} \frac{1}{p_k} \mathbb{E}\left[\left(\frac{\mathbb{P}(X_n \in A_n | X_k)}{\mathbb{P}(X_{n-1} \in A_n | X_{k-1} \in A_k)} - 1 \right)^2 \middle| X_{k-1} \in A_k \right].$$

Best possible variance

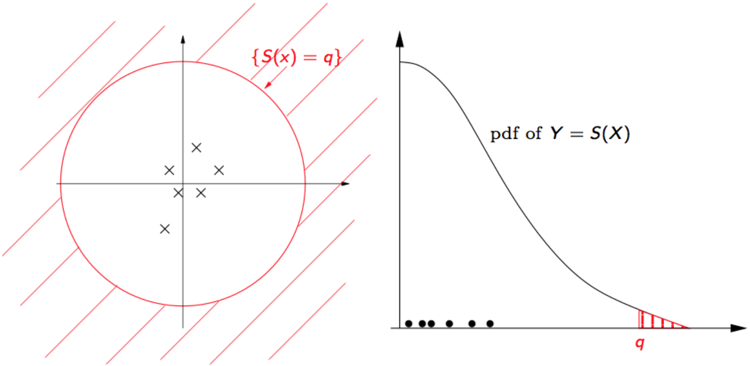
$$\Sigma(1) \geq \sum_{k=1}^{n} \frac{1-p_k}{p_k},$$

with is reached if at each step the X_k^N are i.i.d. according to η_k.
Consider the i.i.d. (idealized) case, the levels giving the best variance are solution of

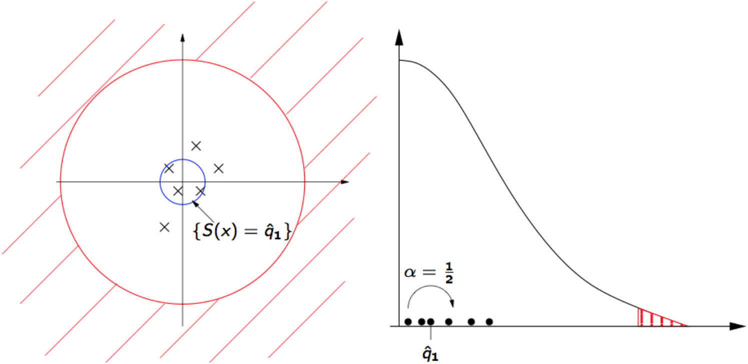
$$\arg\min \sum_{k=1}^{n} \frac{1-p_k}{p_k} \text{ subject to } \prod_{k=1}^{n} p_k = p.$$

Solution: take the p_k's all the same, $p_k = p^{\frac{1}{n}}$.
This suggest to adapt the algorithm in order to have evenly spaces levels (in terms of probability of success).
The more levels, the lower the variance.

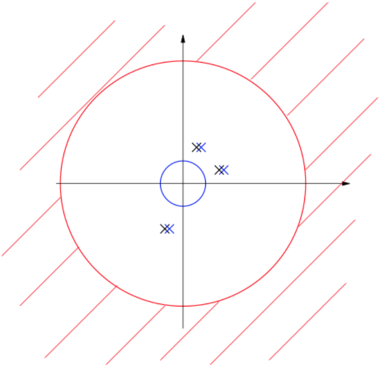
$A = \{S(X) > q\}$ for some real valued S and $\alpha = 1/2$.

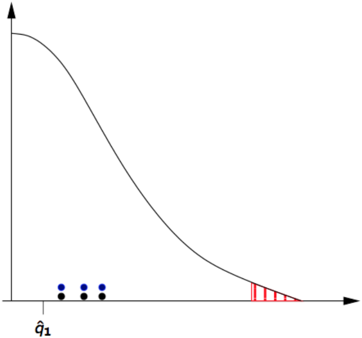


$A = \{S(X) > q\}$ for some real valued S and $\alpha = 1/2$.

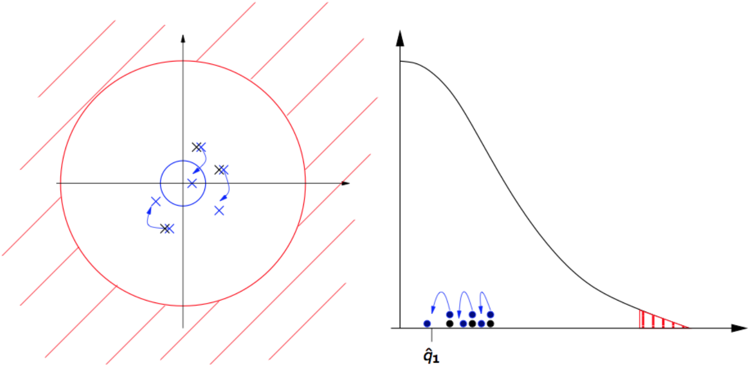


$A = \{S(X) > q\}$ for some real valued S and $\alpha = 1/2$.

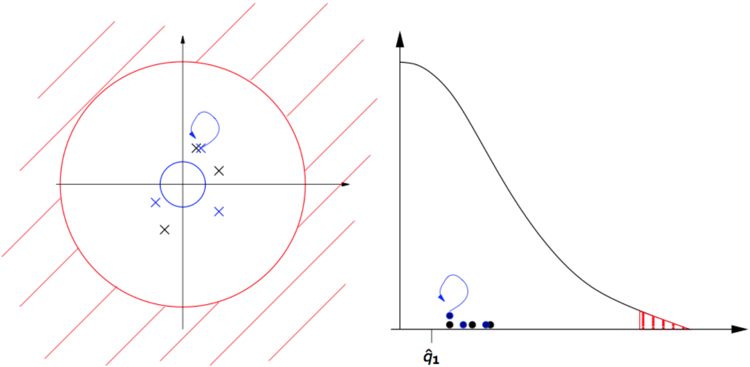




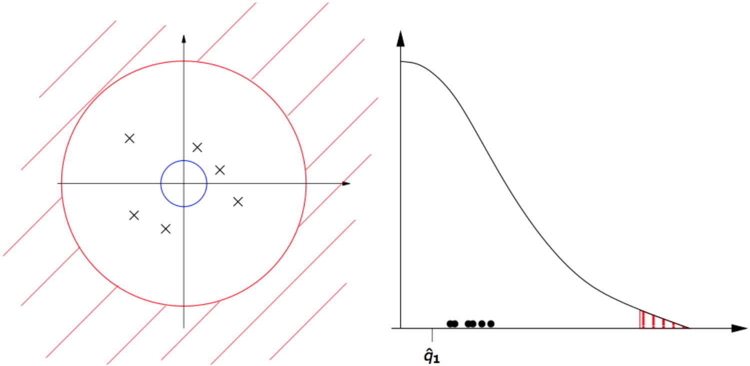
$A = \{S(X) > q\}$ for some real valued S and $\alpha = 1/2$.



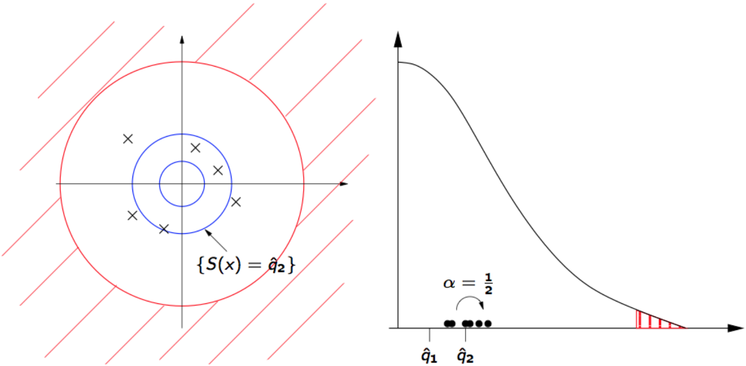
$A = \{S(X) > q\}$ for some real valued S and $\alpha = 1/2$.



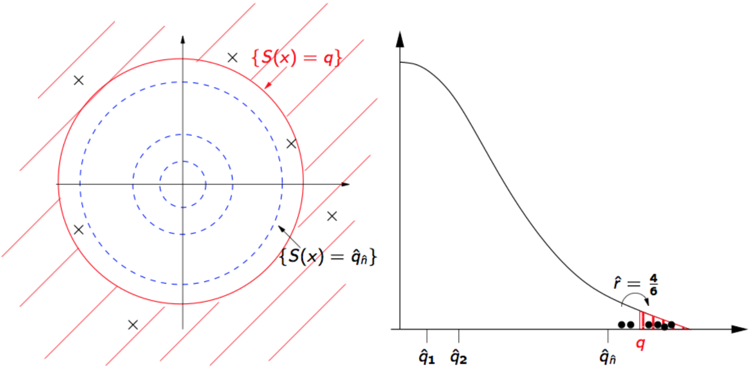
$A = \{S(X) > q\}$ for some real valued S and $\alpha = 1/2$.



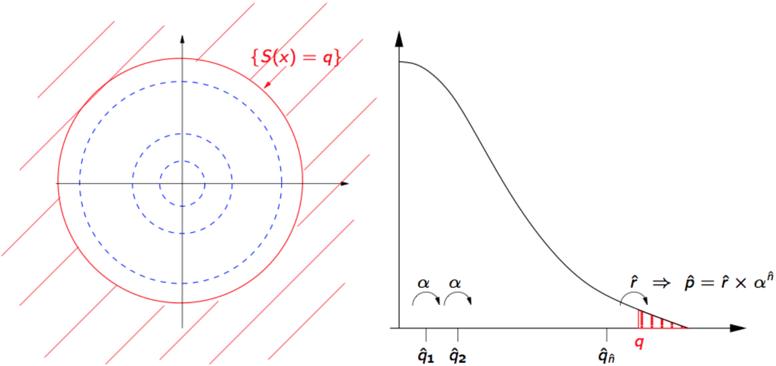
$A = \{S(X) > q\}$ for some real valued S and $\alpha = 1/2$.



$A = \{S(X) > q\}$ for some real valued S and $\alpha = 1/2$.



$A = \{S(X) > q\}$ for some real valued S and $\alpha = 1/2$.



Particle Approximation

- At each level we have a sample $(X_k^1, \dots, X_k^N)$ identically distributed with η restricted on $\{x,\, S(x) > \hat{q}_k\}$, but non independent.
- Last step $\hat{n}$ when $\hat{q}_{\hat{n}+1} \geq q$, $\hat{p} = \hat{r}\,\alpha^{\hat{n}}$. We have a.s. that, for N large enough, $\hat{n} = n = \lfloor \frac{\log p}{\log \alpha} \rfloor$.
- Empirical normalized measures: Consider $f = f\,\mathbb{1}_{S(.)>q}$,

$$\eta_{\hat{n}}^N(f) = \frac{1}{N}\sum_{i=1}^{N} f(X_{\hat{n}}^i) \longrightarrow \mathbb{E}[f(X)\,|\,S(X) > q]$$

- Empirical unnormalized measure:

$$\gamma_{\hat{n}}^N(1) = \alpha^{\hat{n}} \longrightarrow p = \eta(\mathbb{1}_{S(.)>q})$$

$$\gamma_{\hat{n}}^N(f) = \gamma_{\hat{n}}^N(1)\eta_{\hat{n}}^N(f) \longrightarrow \mathbb{E}[f(X)\mathbb{1}_{S(X)>q}].$$

Asymptotics, adaptive case

Theorem

Under some mild regularity assuptions on S, η and the kernel K, both normalized and unnormalized estimates converge a.s.

Theorem

Under some mild regularity assuptions on S, η and the kernel K, one has a CLT for both normalized and unnormalized measures, with the same asymptotic variance as in the fixed levels case, with evenly spaced levels.

Proof (adaptive CLT)

A key decomposition

$$\eta_p^N - \eta_p = \sum_{q=0}^{p} \alpha^{q-p} \left\{ T_{q,p}(\eta_q^N) - \alpha^{-1} \; T_{q-1,p}(\eta_{q-1}^N) \right\}$$

But this is not a martingale.

$$
\begin{aligned}
&[\eta_p^N - \eta_p](f) \\
&= \sum_{q=0}^{p} \alpha^{q-p} \left\{ T_{q,p}(\eta_q^N)(f) - \mathbb{E}\left[T_{q,p}(\eta_q^N)(f) \middle| \mathcal{G}_{q-1}^N \right] \right\} \\
&+ \sum_{q=0}^{p} \alpha^{q-p} \left\{ \epsilon_q^N \, \mathbb{E}\left[T_{q,p}(\eta_q^N)(f) \middle| \mathcal{G}_{q-1}^N \right] \right. \\
&\qquad\qquad - \mathbb{E}\left[\epsilon_q^N \, \mathbb{E}\left[T_{q,p}(\eta_q^N)(f) \middle| \mathcal{G}_{q-1}^N \right] \middle| \mathcal{F}_{q-1}^N \right] \\
&\qquad + \mathbb{E}\left[\epsilon_q^N \, \mathbb{E}\left[T_{q,p}(\eta_q^N)(f) \middle| \mathcal{G}_{q-1}^N \right] \middle| \mathcal{F}_{q-1}^N \right] \\
&\qquad \left. + \left((1 - \epsilon_q^N) \mathbb{E}\left[T_{q,p}(\eta_q^N)(f) \middle| \mathcal{G}_{q-1}^N \right] - \rho_N \; T_{q,p}\left(\Phi_q \left(\eta_{q-1}^N \right) \right)(f) \right) \right\}
\end{aligned}
$$

$$T_{q,p}(\mu) = \mu Q_{q,p,\mu} \quad \text{and} \quad \Phi_{q,p}(\mu)(f) = T_{q,p}(\mu)(f)/T_{q,p}(\mu)(1)$$

Approximation lemma

$$Q_{q,p}^{\delta,-}(x, dy) \leq Q_{q,p,\mu}(x, dy) \leq Q_{q,p}^{\delta,+}(x, dy),$$

and for any bounded measurable function f on $\mathbb{R}^d$,

$$\lim_{\delta \to 0} \eta_q \left(\left| \left[Q_{q,p}^{\delta,+} - Q_{q,p}^{\delta,-} \right] (f) \right| \right) = 0,$$

In addition, there exists some $N_0 = N_0(\omega)$ such that, for any $N \geq N_0$,

$$\left| \left[Q_{q,p,\eta_q^N} - Q_{q,p,\eta_q} \right] (f) \right| \leq \left| \left[Q_{q,p}^{\delta,+} - Q_{q,p}^{\delta,-} \right] (f) \right|.$$