

SMC with adaptive resampling: large sample asymptotics

Élise Arnaud¹

¹ Université Joseph Fourier, Laboratoire Jean Kuntzmann and INRIA Grenoble — Rhône-Alpes, France — elise.arnaud@inrialpes.fr

² INRIA Rennes — Bretagne Atlantique and IRISA, France — legland@irisa.fr

François Le Gland²

Introduction

SMC folklore, based on practical evidence

- it is a good idea to use importance weights to resample ... can anything be proved ?
- it is better not to resample at each time instant
practical use of heuristics (effective sample size, sample entropy, etc.) to decide when to resample ... can anything be proved ?

Monte Carlo approximation of Feynman-Kac distributions

Consider the unnormalized and normalized Feynman–Kac distributions defined on E by

$$\langle \gamma_n, \phi \rangle = \int_E \cdots \int_E \phi(x_n) \gamma_0(dx_0) \prod_{k=1}^n R_k(x_{k-1}, dx_k) \quad \text{and} \quad \langle \mu_n, \phi \rangle = \frac{\langle \gamma_n, \phi \rangle}{\langle \gamma_n, 1 \rangle},$$

characterized by the nonnegative measure $\gamma_0(dx)$, and the nonnegative kernels $R_k(x, dx')$. It is always possible to decompose the nonnegative measure

$$\gamma_0(dx) = W_0(x) p_0(dx),$$

in terms of a nonnegative function and a normalized probability distribution, and to decompose the nonnegative kernel

$$R_k(x, dx') = W_k(x, x') P_k(x, dx'),$$

in terms of a nonnegative function and a normalized Markov kernel, for any $k = 1, \dots, n$. With these decompositions is associated the equivalent probabilistic representation

$$\langle \gamma_n, \phi \rangle = \mathbb{E}[\phi(X_n) \prod_{k=0}^n W_k(X_{k-1}, X_k)],$$

where $\{X_k, k = 0, 1, \dots, n\}$ is a Markov chain taking values in E and characterized by its initial probability distribution $p_0(dx)$, and its transition probabilities $P_k(x, dx')$.

Starting from the probabilistic representation for the unnormalized distribution, a first type of Monte Carlo approximation can be designed as

$$\langle \gamma_n, \phi \rangle \approx \frac{1}{N} \sum_{i=1}^N \phi(\xi_n^i) \prod_{k=0}^n W_k(\xi_{k-1}^i, \xi_k^i),$$

where $\xi_{0:n}^i = (\xi_0^i, \dots, \xi_n^i) \sim p_0(dx_0) P_1(x_0, dx_1) \cdots P_n(x_{n-1}, dx_n)$. This results in a self-normalized approximation of the normalized distribution

$$\mu_n \approx \mu_n^N = \frac{\gamma_n^N}{\langle \gamma_n^N, 1 \rangle} = \sum_{i=1}^N w_n^i \delta_{\xi_n^i} \quad \text{where} \quad w_n^i \propto \prod_{k=0}^n W_k(\xi_{k-1}^i, \xi_k^i).$$

Issues

- a well know limitation of this Monte Carlo approximation is the *weight degeneracy* problem,
- a possible remedy is to use resampling, which introduces additional variance,
- the question that arises is when to resample, i.e. how to take the decision to resample,
- adaptive resampling has been proposed, which monitors the *effective sample size*, a statistics related to the variance of the weights.

The contribution of this paper is to evaluate the asymptotic variance, as the sample size N goes to infinity, of the adaptive resampling algorithm (Doucet et al., 2000) and to study the impact of the prescribed level, or threshold.

Nonlinear adaptive Markov model

Following (Le Gland, 2007), on the product set $E \times [0, \infty)$, define the probability distribution

$$\mu_0^e(dx_0, dv_0) = p_0(dx_0) \delta_{W_0(x_0)}(dv_0),$$

and the nonnegative kernels

$$R_k^e(\mu, x, v, dx', dv') = 1_{(\mu \notin D)} \underbrace{P_k(x, dx') \delta_{v W_k(x, x')}(dv')}_{P_k^{\text{imp}}(x, v, dx', dv')} + 1_{(\mu \in D)} \underbrace{v P_k(x, dx') \delta_{W_k(x, x')}(dv')}_{R_k^{\text{red}}(x, v, dx', dv')},$$

and let $e(v) = v$ for any $v \geq 0$. Notice that

$$R_k^e(\mu) (\phi \otimes e)(x, v) = \int_E \int_0^\infty R_k^e(\mu, x, v, dx', dv') \phi(x') v' = R_k \phi(x) v,$$

hence $R_k^e(\mu) (\phi \otimes e) = (R_k \phi) \otimes e$, **no matter how D is defined**. For instance, we may define

$$D = \{\mu \in \mathcal{P} : \frac{\langle \mu, 1 \otimes e^2 \rangle}{\langle \mu, 1 \otimes e \rangle^2} \geq c\},$$

where $c \geq 1$ is a threshold to be fixed.

Consider next the unnormalized and normalized Feynman–Kac distributions defined on the product set $E \times [0, \infty)$ by

$$\langle \gamma_n^e, F \rangle = \int_E \int_0^\infty \cdots \int_E \int_0^\infty F(x_n, v_n) \mu_0^e(dx_0, dv_0) \prod_{k=1}^n R_k^e(\mu_{k-1}^e, x_{k-1}, v_{k-1}, dx_k, dv_k)$$

$$\text{and} \quad \langle \mu_n^e, F \rangle = \frac{\langle \gamma_n^e, F \rangle}{\langle \gamma_n^e, 1 \rangle}.$$

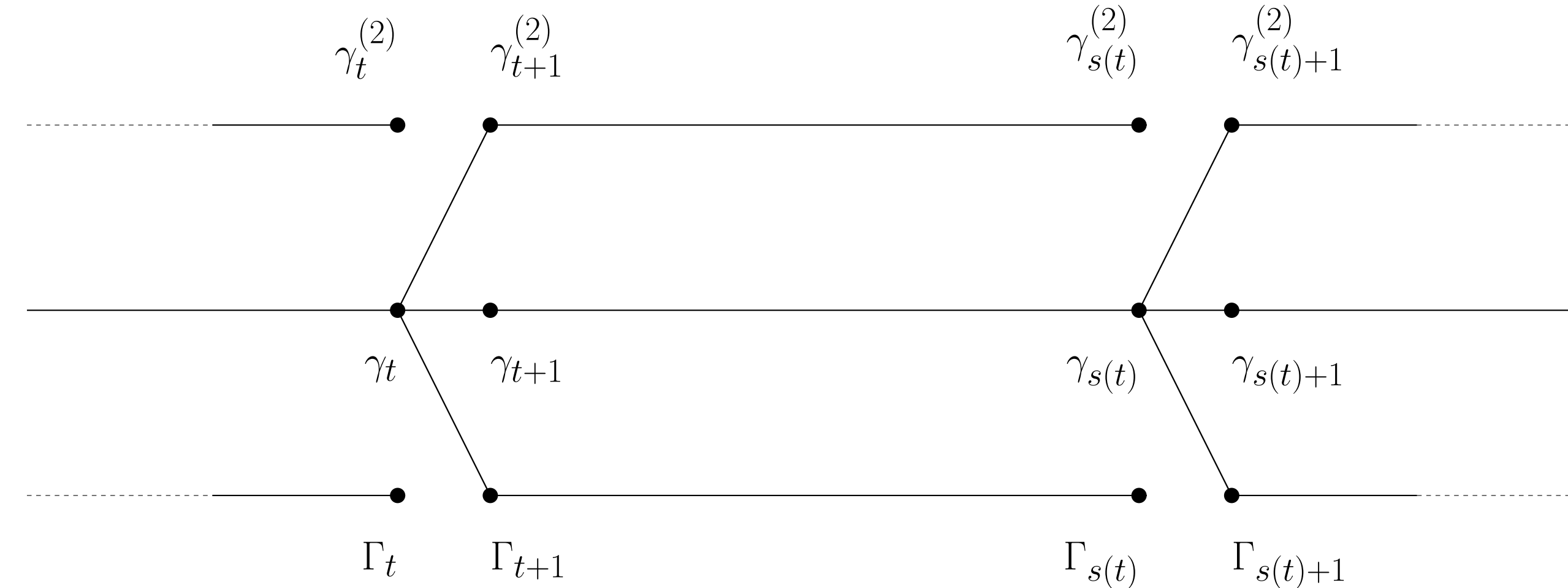
Important special cases

$$\Gamma_k = \langle \gamma_k^e, 1 \rangle, \quad \langle \gamma_k^{(1)}, \phi \rangle = \langle \gamma_k^e, \phi \otimes e \rangle \quad \text{and} \quad \langle \gamma_k^{(2)}, \phi \rangle = \langle \gamma_k^e, \phi \otimes e^2 \rangle,$$

and in particular $\gamma_k^{(1)} = \gamma_k$, so that

$$\frac{\langle \mu_k^e, \phi \otimes e \rangle}{\langle \mu_k^e, 1 \otimes e \rangle} = \langle \mu_k, \phi \rangle \quad \text{and} \quad c_k = \frac{\langle \mu_k^e, 1 \otimes e^2 \rangle}{\langle \mu_k^e, 1 \otimes e \rangle^2} = \frac{\langle \gamma_k^{(2)}, 1 \rangle \Gamma_k}{\langle \gamma_k, 1 \rangle^2}.$$

Let $\mathbb{T} = \{t_1, \dots, t_p\} \subset \{0, 1, \dots, n-1\}$ denote the set of redistribution times, i.e. $t \in \mathbb{T}$ iff $\mu_t^e \in D$. By definition, no redistribution occurs before the first redistribution time t_{\min} , between two successive redistribution times t and $s(t)$ where $s(t)$ denotes the next redistribution time after t , or after the last redistribution time t_{\max} .



Evolution of $\gamma_k^{(2)}$ and Γ_k between redistribution times, and resetting of $\gamma_{t+1}^{(2)}$ and Γ_{t+1} after the redistribution time t .

Particle approximation

Introducing a particle approximation of the form

$$\mu_{k-1}^e \approx \mu_{k-1}^{e,N} = \frac{1}{N} \sum_{i=1}^N \delta_{(\xi_{k-1}^i, v_{k-1}^i)},$$

yields : if $\mu_{k-1}^{e,N} \notin D$, then

$$\mu_k^{e,N} = \frac{1}{N} \sum_{i=1}^N \delta_{(\xi_k^i, v_k^i)},$$

where $\xi_k^i \sim P_k(\xi_{k-1}^i, dx')$ and $v_k^i = v_{k-1}^i W_k(\xi_{k-1}^i, \xi_k^i)$, and if $\mu_{k-1}^{e,N} \in D$, then

$$\mu_k^{e,N} = \frac{1}{N} \sum_{i=1}^N \delta_{(\xi_k^i, v_k^i)},$$

where $\xi_k^i \sim P_k(\xi_{k-1}^i, dx')$ and $v_k^i = W_k(\xi_{k-1}^i, \xi_k^i)$, and where the index $\tau_k^i \in \{1, \dots, N\}$ is selected according to the weights $(w_{k-1}^1, \dots, w_{k-1}^N)$.

Notice that

$$\frac{\langle \mu_k^{e,N}, \phi \otimes e \rangle}{\langle \mu_k^{e,N}, 1 \otimes e \rangle} = \frac{\sum_{i=1}^N v_k^i \phi(\xi_k^i)}{\sum_{i=1}^N v_k^i} = \langle \mu_k^N, \phi \rangle,$$

which defines μ_k^N as the corresponding particle approximation for μ_k .

This particle approximation for the nonlinear adaptive Markov model coincides with the classical particle approximation with adaptive resampling, and it satisfies a CLT that can be proved using standard techniques. Notice also that the *effective sample size* N_{eff} can be expressed in terms of

$$\frac{\langle \mu_{k-1}^{e,N}, 1 \otimes e^2 \rangle}{\langle \mu_{k-1}^{e,N}, 1 \otimes e \rangle^2} = \frac{\frac{1}{N} \sum_{i=1}^N |v_{k-1}^i|^2}{\left| \frac{1}{N} \sum_{i=1}^N v_{k-1}^i \right|^2} = N \sum_{i=1}^N |w_{k-1}^i|^2 = \frac{N}{N_{\text{eff}}} \geq 1.$$

Central limit theorem

Recall that $\mathbb{T} = \{t_1, \dots, t_p\} \subset \{0, 1, \dots, n-1\}$ denotes the set of redistribution times, i.e. $t \in \mathbb{T}$ iff $\mu_t^e \in D$, and let $\mathbb{T}^+ = \mathbb{T} \cup \{n\} = \{t_1, \dots, t_p, t_{p+1}\}$ with $t_{p+1} = n$ by convention.

Theorem For the normalization constant and for the normalized distribution, it holds

$$\sqrt{N} [\frac{\langle \gamma_n^N, 1 \rangle}{\langle \gamma_n, 1 \rangle} - 1] \implies \mathcal{N}(0, V_n),$$

and

$$\sqrt{N} \langle \mu_n^N - \mu_n, \phi \rangle \implies \mathcal{N}(0, v_n(\phi)),$$

in distribution as $N \uparrow \infty$, for any bounded measurable function ϕ , with the following expression for the asymptotic variances

$$V_n = \sum_{t \in \mathbb{T}^+} \left[\frac{\langle \mu_t^{(2)}, |R_{t+1:n} 1|^2 \rangle}{\langle \mu_t, R_{t+1:n} 1 \rangle^2} c_t - 1 \right],$$

and

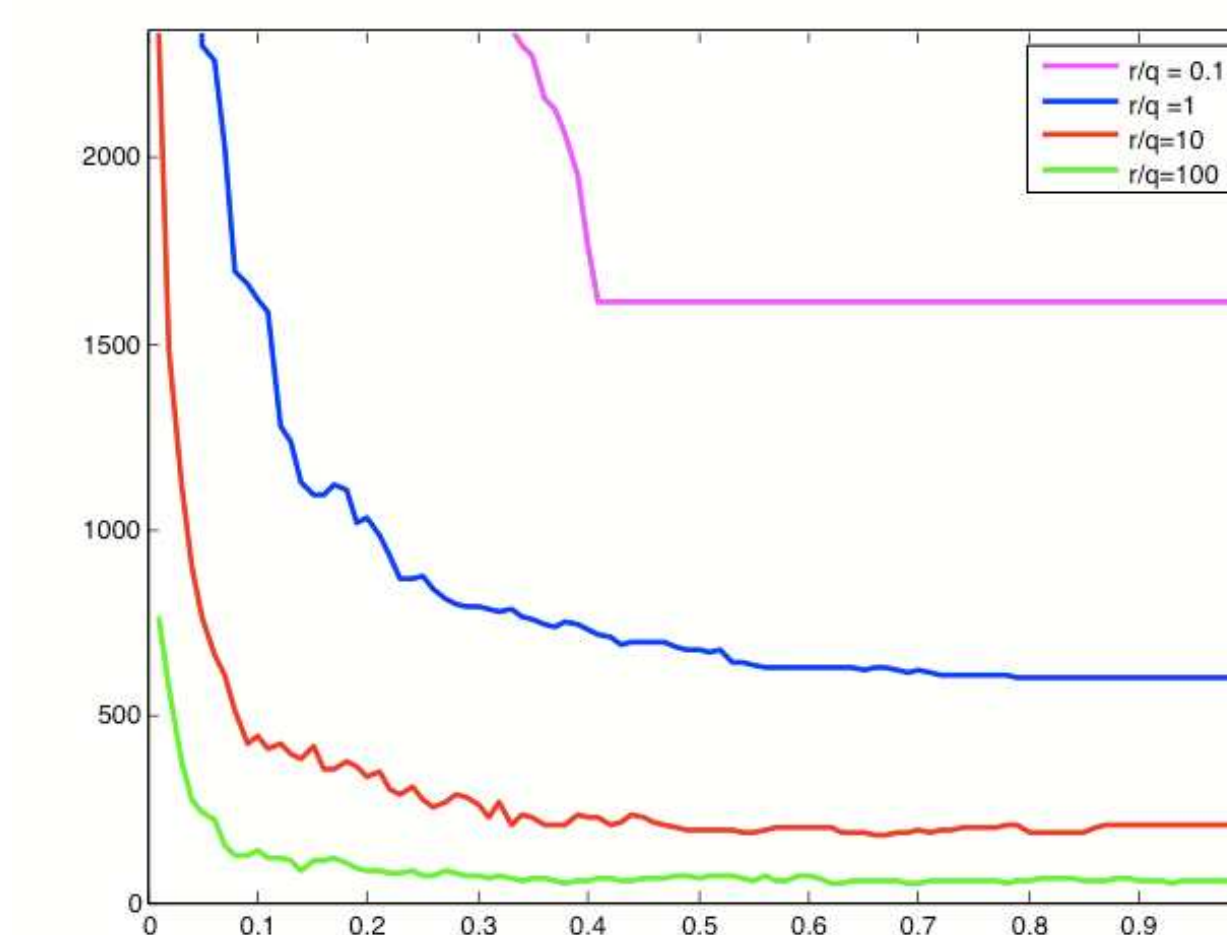
$$v_n(\phi) = \sum_{t \in \mathbb{T}^+} \frac{\langle \mu_t^{(2)}, |R_{t+1:n}(\phi - \langle \mu_n, \phi \rangle)|^2 \rangle}{\langle \mu_t, R_{t+1:n} 1 \rangle^2} c_t,$$

respectively, where

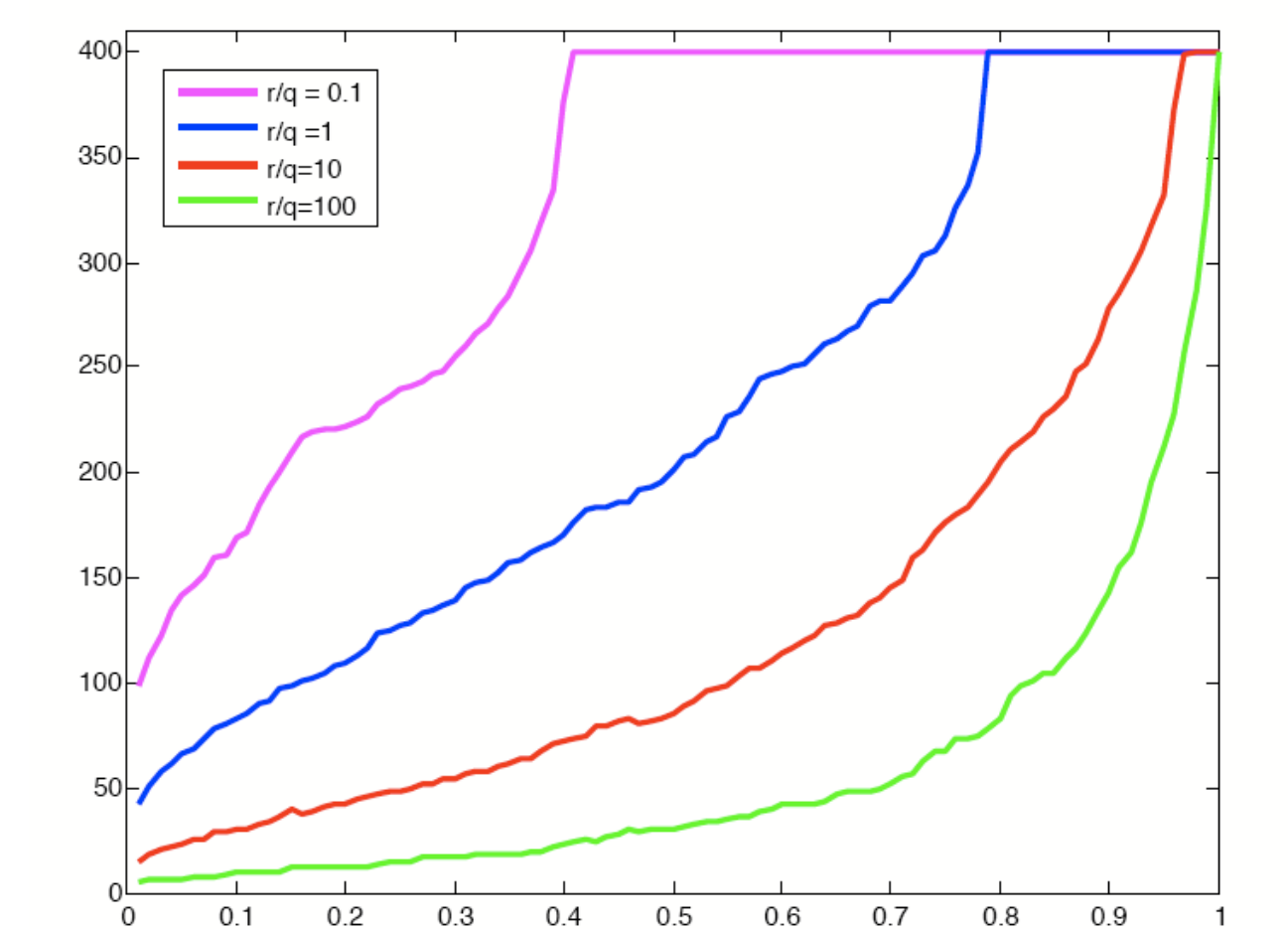
$$R_{k:n} \phi(x) = \mathbb{E}[\phi(X_n) \prod_{p=k}^n W_p(X_{p-1}, X_p) \mid X_{k-1} = x],$$

for any $k = 1, \dots, n+1$, with the convention $R_{n+1:n} \phi(x) = \phi(x)$ for any $x \in E$.

This result is of theoretical more than practical interest, since computing the asymptotic variances V_n or $v_n(\phi)$ is as complicated as computing the desired object itself. Still, it is possible in some simple cases to compute the exact expression of the asymptotic variance and to compute the exact expression of the limiting number of redistributions, as a function of the threshold c (for instance, for a one dimensional linear filtering problem)



Asymptotic variance as a function of the threshold c



Limiting number of redistributions as a function of the threshold c