GENEALOGICAL PARTICLE ANALYSIS OF RARE EVENTS

PIERRE DEL MORAL* AND JOSSELIN GARNIER*

Abstract. In this paper an original interacting particle system approach is developed for studying Markov chains in rare event regimes. The proposed particle system is theoretically studied through a genealogical tree interpretation of Keynman-Kac path measures. The algorithmic implementation of the particle system is presented. An efficient estimator for the probability of ocurrence of a rare event is proposed and its variance is computed. Applications and numerical implementations are discussed. First, we apply the particle system technique to a toy model (a Gaussian random walk), which permits to illustrate the theoretical predictions. Second, we address a physically relevant problem consisting in the estimation of the outage probability due to polarization-mode dispersion in optical fibers.

Key words. Rare events, Monte Carlo Markov chains, importance sampling, interacting particle systems, genetic algorithms.

AMS subject classifications. 65C35, 65C20, 60F10, 68U20, 62P35

1. Introduction. The simulations of rare events has become an extensively studied subject in queueing and reliability models [19], in particular in telecommunication systems. The rare events of interest are long waiting times or buffer overflows in queueing systems, and system failure events in reliability models. The issue is usually the estimation of the probability of occurence of a rare event (such as the failure of a system), and we shall focus mainly on that point. But our method will be shown to be also efficient for the analysis of the cascade of events leading to such a rare event, in order to exhibit the typical physical path that the system uses to achieve the rare event.

Standard Monte Carlo (MC) simulations are usually prohibited in these situations because very few (or even zero) simulations will achieve the rare event. The general approach to speeding up such simulations is to accelerate the occurrence of the rare events by using importance sampling (IS) [19, 31]. In IS the system is simulated using a new set of input probability distributions, and unbiased estimates are recovered by multiplying the simulation output by a likelihood ratio. The tricky part in IS is to properly choose the twisted distribution. The user is expected to guess a more or less correct twisted distribution otherwise IS may completely fail. Our aim is to propose a more elaborate scheme that will select the twisted distribution in an adaptative way, without any operation of the user. The method consists in simulating an interacting particle system (IPS) with selection and mutation steps [6]. This interacting particle methodology is closely related to a class of Monte Carlo acceptance/rejection simulation techniques used in physics and biology. These methods were first designed in the fifties to estimate particle energy transmission [18], self avoiding random walks, and macromolecule evolutions [30]. The application model areas of these particle methods now have a range going from advanced signal processing, including speech recognition, tracking, and filtering, to financial mathematics and telecommunication [12]. The present paper is devoted to new applications towards rare event estimation.

The idea is the following one. Consider a E-valued Markov chain $(X_p)_{0\leq p\leq n}$ with non-homogeneous transition kernels K_p. The problem consists in estimating the

*Laboratoire de Statistique et Probabilités, Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex 4, France. E-mail: garnier@cict.fr (J. Garnier) and delmoral@cict.fr (P. Del Moral).

probability P_A of occurence of a rare event of the form $\{V(X_n) \in A\}$ where V is some function from E to $\mathbb{R}$. The IPS consists of a set of N particles $(X_p^{(i)})_{1\leq i\leq N}$ evolving from time $p=0$ to $p=n$. The initial generation at $p=0$ is a set of independent copies of X_0. The updating from the generation p to the generation $p+1$ is divided into two stages.
1) The selection stage consists in choosing randomly and independently N particles amongst $(X_p^{(i)})_{1\leq i\leq N}$ according to a weighted Boltzmann-Gibbs particle measure, with a weight function that is varying monotoneously with V. Thus, particles with low scores are killed, while particles with high scores are multiplied. Note that the total number of particles is kept constant.
2) The mutation step consists in mutating independently the particles according to the kernel K_p. Note that the true transition kernel is applied, in contrast with IS.
The description is rough in that the IPS actually acts on the path level. The mathematical tricky part consists in proposing an estimator of the probability P_A and analyzing its variance. The variance analysis will provide useful information for a proper choice of the weight function.

The paper is organized as follows. Section 2 contains all the theoretical results formulated in an abstract framework. We give a summary of the method and present a user-friendly implementation in Section 3. We consider a toy model (a Gaussian random walk) in Section 4 to illustrate the theoretical predictions on an example where all relevant quantities can be explicitly computed. Finally, in Section 5, we apply the method to a physical situation emerging in telecommunication.

2. Simulations of rare events by interacting particle systems.

2.1. Introduction. In this section, we design an original IPS approach for analyzing of Markov chains evolving in rare event regime.

In the first subsection 2.2, we use a natural large deviation perspective to exhibit natural changes of reference measures under which the underlying process is more likely to enter in a given rare level set. This technique is more or less well known. It often offers a powerful and elegant strategy for analyzing rare deviation probabilities. Loosely speaking, the twisted distributions associated to the deviated process represent the evolution of the original process in the rare event regime. In MC Markov chain literature, these changes of measure strategy is also called the importance sampling (IS) technique.

In Subsection 2.3, we present a Feynman-Kac formulation of twisted reference path distributions. We examine a pair of Gaussian models for which these changes of measures have a nice explicit formulation. In this context, we initiate a comparison of the fluctuation-error variances of the "pure" MC and the IS techniques. In general, the twisted distribution suggested by the physical model is rather complex, and its numerical analysis often requires extensive calculations. The practioners often need to resort to another "sub-optimal" reference strategy, based on a more refined analysis of the physical problem at hand. The main object of this section is to complement this IS methodology, with presenting a genetic type particle interpretation of a general and abstract class of twisted path models. Instead of hand crafting or simplified simulation models this new particle methodology provides a powerful, and very flexible way, to produce samples according to any complex twisted measures dictated by the physical properties of the model at hand. But, from the strict practical point of view, if there exists already a good specialized IS method for a specific rare event problem then our IPS methodology may not be the best tool for that application.

In Subsection 2.4, we introduce the reader to a new developing genealogical tree interpretation of Feynman-Kac path measures. For a more thorough study on this theme we refer to the monograph [6], and references therein. We connect this IPS methodology with rare event analysis. Intuitively speaking, the ancestral lines associated to these genetic evolution models represent the physical ways that the process uses to reach the desired rare level set.

In the final section 2.5, we analyze the fluctuations of rare event particle simulation models. We discuss the performance of these interpretations on a class of warm up Gaussian models. We compare the asymptotic error-variances of genealogical particle models and the more traditional noninteracting IS schemes. For Gaussian models, we show that the exponential fluctuation orders between these two particle simulation strategies are equivalent.

2.2. A large deviation perspective. Let X_n be a Markov chain, taking values at each time n, in some measurable state space $(E_n, \mathcal{E}_n)$, that may depend on the time parameter n. Suppose we want to estimate the probability $P_n(a)$ that X_n enters, at a given fixed date n, into the a-level set $V_n^{-1}([a,\infty))$ of a given energy like function V_n on E_n, for some $a \in \mathbb{R}$:

$$P_n(a) = \mathbb{P}(V_n(X_n) \geq a). \tag{2.1}$$

To avoid some unnecessary technical difficulties, we further assume that $P_n(a) > 0$, and the pair (X_n, V_n) satisfies Cramer's condition $\mathbb{E}(e^{\lambda V_n(X_n)}) < \infty$, for all $\lambda \in \mathbb{R}$. This condition ensures the exponential decrease of the probabilities $\mathbb{P}(V_n(X_n) \geq a) \downarrow 0$, as $a \uparrow \infty$. To see this claim, we simply use the exponential version of Chebychev's inequality to check that, for any $\lambda > 0$ we have

$$\mathbb{P}(V_n(X_n) \geq a) \leq e^{-\lambda(a-\lambda^{-1}\Lambda_n(\lambda))} \quad \text{with} \quad \Lambda_n(\lambda) =_{\text{def.}} \log \mathbb{E}(e^{\lambda V_n(X_n)})$$

As an aside, it is also routine to prove that the maximum of $(\lambda a - \Lambda_n(\lambda))$, with respect to the parameter $\lambda > 0$, is attained at the value $\lambda_n(a)$ determined by the equation $a = \mathbb{E}(V_n(X_n)e^{\lambda V_n(X_n)}))/\mathbb{E}(e^{\lambda V_n(X_n)}))$. The resulting inequality

$$\mathbb{P}(V_n(X_n) \geq a) \leq e^{-\Lambda_n^\star(a)} \quad \text{with} \quad \Lambda_n^\star(a) = \sup_{\lambda>0}(\lambda a - \Lambda_n(\lambda))$$

is known as large-deviation inequality. When the Laplace transforms Λ_n are explicitly known, this variational analysis often provides sharp tail estimates. We illustrate this observation on an elementary Gaussian model. This warm up example will be used in several places in the further development of this article. Suppose that X_n is given by the recursive equation

$$X_p = X_{p-1} + W_p \tag{2.2}$$

where $X_0 = 0$ and $(W_p)_{p\in\mathbb{N}^*}$ represents a sequence of independent and identically distributed (i.i.d.) Gaussian random variables, with $(\mathbb{E}(W_1), \mathbb{E}(W_1^2)) = (0,1)$. If we take $V_n(x) = x$, then we find that $\Lambda_n(\lambda) = \lambda^2 n/2$, $\lambda_n(a) = a/n$, and $\Lambda_n^\star(a) = a^2/(2n)$, from which we recover the well-known sharp exponential tails $\mathbb{P}(X_n \geq a) \leq e^{-a^2/(2n)}$.

In more general situations, the analytical expression of $\Lambda_n^\star(a)$ is out of reach, and we need to resort to judicious numerical strategies. The first rather crude MC method is to consider the estimate

$$P_n^N(a) = \frac{1}{N}\sum_{i=1}^N \mathbf{1}_{V_n(X_n^i)\geq a}$$

based on N independent copies $(X_n^i)_{1\leq i\leq N}$ of X_n. There are no difficulties to check that the resulting error-variance is given by

$$\sigma_n^2(a) = N\ \mathbb{E}\left[(P_n^N(a) - P_n(a))^2\right] = P_n(a)\ (1 - P_n(a))$$

In practice, $P_n^N(a)$ is very poor estimate mainly because the whole sample set is very unlikely to reach the rare level.

A more judicious choice of MC exploration model is dictated by the large deviation analysis presented above. To be more precise, let us suppose that $a > \lambda^{-1}\Lambda_n(\lambda)$, with $\lambda > 0$. To simplify the presentation, we also assume that the initial value $X_0 = x_0$ is fixed, and we set $V_0(x_0) = 0$. Let $\mathbb{P}_n^\lambda$ be the new reference measure on the path space $F_n =_{\text{def.}} (E_0 \times \cdots \times E_n)$ defined by the formula

$$d\mathbb{P}_n^{(\lambda)} = \frac{1}{\mathbb{E}(e^{\lambda V_n(X_n)})}\ e^{\lambda V_n(X_n)}\ d\mathbb{P}_n \tag{2.3}$$

where $\mathbb{P}_n$ is the distribution of the original and canonical path $(X_p)_{0\leq p\leq n}$. By construction, we have that

$$\begin{aligned}
\mathbb{P}(V_n(X_n) \geq a) &= \mathbb{E}_n^{(\lambda)}\left[\mathbf{1}_{V_n(X_n)\geq a}\ d\mathbb{P}_n/d\mathbb{P}_n^{(\lambda)}\right] \\
&= \mathbb{E}_n^{(\lambda)}\left[\mathbf{1}_{V_n(X_n)\geq a}\ e^{-\lambda V_n(X_n)}\right]\ \mathbb{E}\left[e^{\lambda V_n(X_n)}\right] \\
&\leq e^{-\lambda(a-\lambda^{-1}\Lambda_n(\lambda))}\ \mathbb{P}_n^{(\lambda)}(V_n(X_n) \geq a)
\end{aligned}$$

where $\mathbb{E}_n^{(\lambda)}$ represents the expectation operator with respect to the distribution $\mathbb{P}_n^{(\lambda)}$. By definition, the measure $\mathbb{P}_n^{(\lambda)}$ tends to favor random evolutions with high potential values $V_n(X_n)$. As a consequence, the random paths under $\mathbb{P}_n^{(\lambda)}$ are much more likely to enter in the rare level set. For instance, in the Gaussian example described earlier, we have that

$$d\mathbb{P}_n^{(\lambda)}/d\mathbb{P}_n = \prod_{p=1}^n e^{\lambda(X_p - X_{p-1}) - \lambda^2/2} \tag{2.4}$$

In other words, under $\mathbb{P}_n^{(\lambda)}$ the chain takes the form $X_p = X_{p-1} + \lambda + W_p$, and we have $\mathbb{P}_n^{(\lambda)}(X_n \geq a) = \mathbb{P}_n(X_n \geq a - \lambda n)$ ($= 1/2$ as soon as $a = \lambda n$).

These observations suggest to replace $P^N(a)$ by the weighted MC model

$$P_n^{N,\lambda}(a) = \frac{1}{N}\sum_{i=1}^N \frac{d\mathbb{P}_n}{d\mathbb{P}_n^{(\lambda)}}(X_0^{\lambda,i}, \ldots, X_n^{\lambda,i})\ \mathbf{1}_{V_n(X_n^{\lambda,i})\geq a}$$

associated to N independent copies $(X_n^{\lambda,i})_{1\leq i\leq N}$ of the chain under $\mathbb{P}_n^{(\lambda)}$. Observe that the corresponding error-variance is given by

$$\begin{aligned}
\sigma_n^{(\lambda)}(a)^2 &= N\ \mathbb{E}\left[(P_n^{N,\lambda}(a) - P_n(a))^2\right] \\
&= \mathbb{E}\left[\mathbf{1}_{V_n(X_n)\geq a}\ e^{-\lambda V_n(X_n)}\right]\ \mathbb{E}[e^{\lambda V_n(X_n)}] - P_n^2(a) \\
&\leq e^{-\lambda(a-\lambda^{-1}\Lambda_n(\lambda))}\ P_n(a) - P_n^2(a)
\end{aligned} \tag{2.5}$$

For judicious choices of λ, one expects the exponential large-deviation term to be proportional to the desired tail probabilities $P_n(a)$. In this case, we have $\sigma_n^{(\lambda)}(a)^2 \leq$

Cte. $P_n^2(a)$. Returning to the Gaussian situation, and using Mill's inequalities

$$\frac{1}{\lambda+1/\lambda} \leq \mathbb{P}(\mathcal{N}(0,1)\geq\lambda)\ \sqrt{2\pi}e^{\frac{\lambda^2}{2}} \leq \frac{1}{\lambda}$$

which are valid for any $\lambda > 0$, and any reduced Gaussian random variable $\mathcal{N}(0,1)$ (see for instance (6) on page 237 in [32]), we find that

$$\sigma_n^{(\lambda)}(a)^2 \leq e^{-a^2/(2n)}\ P_n(a) - P_n^2(a) \leq P_n^2(a)\ [\sqrt{2\pi}(a/\sqrt{n}+\sqrt{n}/a)-1]$$

for the optimal value of $\lambda = \lambda_n(a) = a/n$ computed above. For typical Gaussian type level indexes $a = a_0\sqrt{n}$, with large values of a_0, we find that $\lambda_n(a) = a_0/\sqrt{n}$ and

$$\sigma_n^{(\lambda)}(a)^2 \leq P_n^2(a)\ [\sqrt{2\pi}(a_0+1/a_0)-1]$$

As an aside, although we shall be using most of time upper bound estimates, Mill's inequalities ensure that most of the Gaussian exponential deviations discussed in the paper are sharp.

The formulation (2.5) also suggests a dual interpretation of the variance. Firstly, we note that

$$d\mathbb{P}_n/d\mathbb{P}_n^{(\lambda)} = \mathbb{E}[e^{\lambda V_n(X_n)}]\mathbb{E}[e^{-\lambda V_n(X_n)}]\ d\mathbb{P}_n^{(-\lambda)}/d\mathbb{P}_n$$

and therefore

$$\sigma_n^{(\lambda)}(a)^2 = \mathbb{P}_n^{(-\lambda)}(V_n(X_n)\geq a)\ \mathbb{E}[e^{\lambda V_n(X_n)}]\mathbb{E}[e^{-\lambda V_n(X_n)}] - P_n^2(a)$$

In contrast to $\mathbb{P}_n^{(\lambda)}$, the measure $\mathbb{P}_n^{(-\lambda)}$ now tends to favor low energy states X_n. As a consequence, we expect $\mathbb{P}_n^{(-\lambda)}(V_n(X_n)\geq a)$ to be much smaller that $P_n(a)$. For instance, in the Gaussian case we have

$$\mathbb{P}_n^{(-\lambda)}(X_n\geq a) = \mathbb{P}_n(X_n\geq a+\lambda n) \leq e^{-(a+\lambda n)^2/(2n)}$$

Since we have $\mathbb{E}[e^{\lambda X_n}] = \mathbb{E}[e^{-\lambda X_n}] = e^{\lambda^2 n/2}$, we recover the estimate obtained earlier

$$\sigma_n^{(\lambda)}(a)^2 \leq e^{-a^2/n}e^{(a-\lambda n)^2/(2n)} - P_n^2(a) = e^{-a^2/n} - P_n^2(a) \leq \text{Cte. } P_n^2(a) \tag{2.6}$$

as soon as $\lambda = a/n$.

2.3. Twisted Feynman-Kac path measures. The choice of the "twisted" measures $\mathbb{P}_n^{(\lambda)}$ introduced in (2.3) is only of pure mathematical interest. Indeed, the IS estimates described below will still require both the sampling of random paths according to $\mathbb{P}_n^{(\lambda)}$, and the computation of the normalizing constants. As we mentioned in the introduction, the key difficulty in applying IS strategies is to choose the so-called "twisted" reference measures. In the further development of Subsection 2.4, we shall present a natural genealogical tree based simulation technique of twisted Feynman-Kac path distribution of the following form

$$d\mathbb{Q}_n = \frac{1}{\mathcal{Z}_n}\left\{\prod_{p=1}^n G_p(X_0,\ldots,X_p)\right\}\ d\mathbb{P}_n \tag{2.7}$$

In the above display, $\mathcal{Z}_n > 0$ stands for a normalizing constant, and $(G_p)_{1\leq p\leq n}$ represents a given sequence of potential functions on the path spaces $(F_p)_{1\leq p\leq n}$. Note that the twisted measures defined in (2.3) corresponds to the (non-unique) choice of functions

$$G_p(X_0,\ldots,X_p) = e^{-\lambda(V_p(X_p)-V_{p-1}(X_{p-1}))} \tag{2.8}$$

As an aside, we mention that the optimal choice of twisted measure with respect to the IS criterion is the one associated to the potential functions $G_n = \mathbf{1}_{V_n^{-1}([a,\infty))}$, and $G_p = 1$, for $p < n$. In this case, we have $\mathcal{Z}_n = \mathbb{P}(V_n(X_n) \geq a)$ and $\mathbb{Q}_n$ is the distribution of the random paths ending in the desired rare level. This optimal choice is clearly infeasible, but we note that the resulting variance is null.

The rare event probability admits the following elementary Feynman-Kac formulation

$$\mathbb{P}(V_n(X_n) \geq a) = \mathbb{E}\left[g_n^{(a)}(X_0,\ldots,X_n)\ \prod_{p=1}^{n} G_p(X_0,\ldots,X_p)\right] = \mathcal{Z}_n\ \mathbb{Q}_n(g_n^{(a)})$$

with the weighted function defined by

$$g_n^{(a)}(x_0,\ldots,x_n) = \mathbf{1}_{V_n(x_n)\geq a}\ \prod_{p=1}^{n} G_p^{-1}(x_0,\ldots,x_p)$$

for any path sequence such that $\prod_{p=1}^{n} G_p(x_0,\ldots,x_p) > 0$. Otherwise, $g_n^{(a)}$ is assumed to be null.

The discussion given above already shows the improvements one might expect in changing the reference exploration measure. The central idea behind this IS methodology is to choose a twisted probability that mimics the physical behavior of the process in the rare event regime. The potential functions G_p represent the changes of probability mass, and in some sense the physical variations in the evolution of the process to the rare level set. For instance, for time homogeneous models $V_p = V$, $0 \leq p \leq n$, the potential functions defined in (2.8) will tend to favor local transitions that increases a given V-energy function. The large deviation analysis developed in Subsection 2.2 combined with the Feynman-Kac formulation (2.3) gives some indications on the way to choose the twisted potential functions $(G_p)_{1\leq p\leq n}$. Intuitively, the attractive forces induced by a particular choice of potentials are compensated by increasing normalizing constants. More formally, the error-variance of the $\mathbb{Q}_n$-importance sampling scheme is given by the formula

$$\sigma_n^{\mathbb{Q}}(a)^2 = \mathbb{Q}_n^-(V_n(X_n) \geq a)\ \mathcal{Z}_n\mathcal{Z}_n^- - P_n(a)^2 \tag{2.9}$$

where $\mathbb{Q}_n^-$ is the path Feynman-Kac measure given by

$$d\mathbb{Q}_n^- = \frac{1}{\mathcal{Z}_n^-}\left\{\prod_{p=1}^{n} G_p^{-1}(X_0,\ldots,X_p)\right\}\ d\mathbb{P}_n$$

Arguing as before, and since $\mathbb{Q}_n^-$ tends to favor random paths with low G_p energy, we expect $\mathbb{Q}_n^-(V_n(X_n) \geq a)$ to be much more smaller than the rare event probability $\mathbb{P}(V_n(X_n) \geq a)$. On the other hand, by Jensen's inequality we expect the product of normalizing constants $\mathcal{Z}_n\mathcal{Z}_n^-(\geq 1)$ to be very large. These expectations fail in the "optimal" situation examined above ($G_n = \mathbf{1}_{V_n^{-1}([a,\infty))}$, and $G_p = 1$, for $p < n$).

In this case, we simply note that $\mathbb{Q}_n = \mathbb{Q}_n^- = \mathrm{Law}(X_0, \ldots, X_n \mid V_n(X_n) \geq a)$, and $\mathbb{Q}_n^-(V_n(X_n) \geq a) = 1$, and $\mathcal{Z}_n = \mathcal{Z}_n^- = P_n(a)$.

We end this section with a brief discussion on the competition between making a rare event more attractive and controlling the normalizing constants. We return to the Gaussian example examined in (2.2), and instead of (2.4), we consider the twisted measure

$$d\mathbb{Q}_n = d\mathbb{P}_n^{(\lambda)} = \frac{1}{\mathcal{Z}_n^{(\lambda)}} \left\{ \prod_{p=1}^{n} e^{\lambda X_p} \right\} d\mathbb{P}_n \tag{2.10}$$

In this case, it is not difficult to check that for any $\lambda \in \mathbb{R}$ we have $\mathcal{Z}_n^{(\lambda)} = e^{\frac{\lambda^2}{2} \sum_{p=1}^n p^2}$. In addition, under $\mathbb{P}_n^{(\lambda)}$ the chain X_n has the form

$$X_p = X_{p-1} + \lambda\ (n - p + 1) + W_p, \quad 1 \leq p \leq n \tag{2.11}$$

When $\lambda > 0$, the rare level set is now very attractive, but the normalizing constants can become very large $\mathcal{Z}_n^{(\lambda)} = \mathcal{Z}_n^{(-\lambda)} (\geq e^{\lambda^2 n^3/12})$. Also notice that in this situation the first term in the right-hand side of (2.9) is given by

$$\begin{aligned} \mathbb{P}_n^{(-\lambda)}(V_n(X_n) \geq a)\ \mathcal{Z}_n^{(\lambda)} \mathcal{Z}_n^{(-\lambda)} &\leq e^{-\frac{1}{2n}(a+\lambda \sum_{p=1}^n p)^2 + \lambda^2 \sum_{p=1}^n p^2} \\ &\leq e^{-a^2/n}\ e^{\frac{1}{2n}(a-\lambda \sum_{p=1}^n p)^2 + \lambda^2 [\sum_{p=1}^n p^2 - (\sum_{p=1}^n p)^2/n]} \end{aligned}$$

Although we are using inequalities, we recall that these exponential estimates are sharp. Now, if we take $\lambda = 2a/[n(n+1)]$, then we find that

$$\mathbb{P}_n^{(-\lambda)}(V_n(X_n) \geq a)\ \mathcal{Z}_n^{(\lambda)} \mathcal{Z}_n^{(-\lambda)} \leq e^{-\frac{a^2}{n}\ \frac{2}{3}(1+\frac{1}{n+1})} \tag{2.12}$$

This shows that even if we adjust correctly the parameter λ, this IS estimate is less efficient than the one associated to the twisted distribution (2.4). The reader has probably noticed that the change of measure defined in (2.10) is more adapted to estimate the probability of the rare level sets $\{V_n(Y_n) \geq a\}$, with the historical chain $Y_n = (X_0, \ldots, X_n)$ and the energy function $V_n(Y_n) = \sum_{p=1}^n X_p$.

2.4. A genealogical tree based interpretation model. The probabilistic interpretation of the twisted Feynman-Kac measures (2.7) presented in this section can be interpreted as a mean field path-particle approximation of the distribution flow $(\mathbb{Q}_n)_{n \geq 1}$. We also mention that the genetic type selection/mutation evolution of the former algorithm can also be seen as a acceptance/rejection particle simulation technique. In this connection, and as we already mentioned in the introduction, we again emphasize that this IPS methodology is not useful if we already know a specialized and exact simulation technique of the desired twisted measure.

2.4.1. Rare events Feynman-Kac type distributions. To simplify the presentation, it is convenient to formulate these models in terms of the historical process

$$Y_n =_{\mathrm{def.}} (X_0, \ldots, X_n) \in F_n =_{\mathrm{def.}} (E_0 \times \ldots \times E_n)$$

We let $M_n(y_{n-1}, dy_n)$ be the Markov transitions associated to the chain Y_n. To simplify the presentation, we shall assume that the initial value $Y_0 = X_0 = x_0$ is fixed, and we also denote by $K_n(x_{n-1}, dx_n)$ the Markov transitions of X_n. We finally let $\mathcal{B}_b(E)$ be the space of all bounded measurable functions on some measurable space $(E,\mathcal{E})$, and we equip $\mathcal{B}_b(E)$ with the uniform norm.

We associate to the pair potentials/transitions (G_n, M_n) the Feynman-Kac measure defined for any test function $f_n \in \mathcal{B}_b(F_n)$ by the formula

$$\gamma_n(f_n) = \mathbb{E}\left[f_n(Y_n) \prod_{1\leq k<n} G_k(Y_k)\right]$$

We also introduce the corresponding normalized measure

$$\eta_n(f_n) = \gamma_n(f_n)/\gamma_n(1)$$

To simplify the presentation, and avoid some unnecessary technical discussions, we shall suppose that the potential functions are chosen such that

$$\sup_{(y_n,y_n')\in F_n^2} G_n(y_n)/G_n(y_n') < \infty$$

This regularity condition ensures that the normalizing constants $\gamma_n(1)$ and the measure γ_n are bounded and positive. The Feynman-Kac and the particle approximation models developed in this section, can be extended to more general situations using traditional cut-off techniques, or by considering Kato-class type of potential functions (see for instance [6, 26, 29, 33]).

In this section, we provide a Feynman-Kac formulation of rare event probabilities. The fluctuation analysis of their genealogical tree interpretations will also be described in terms of the distribution flow (γ_n^-, η_n^-), defined as (γ_n, η_n), by replacing the potential functions G_p by their inverse

$$G_p^- = 1/G_p$$

The twisted measures $\mathbb{Q}_n$ presented in (2.7) and the desired rare event probabilities have the following Feynman-Kac representation

$$\mathbb{Q}_n(f_n) = \eta_n(f_n G_n)/\eta_n(G_n) \quad \text{and} \quad \mathbb{P}(V_n(X_n) \geq a) = \gamma_n(T_n^{(a)}(1))$$

In the above displayed formulae, $T_n^{(a)}(1)$ is the weighted indicator function defined for any path $y_n = (x_0, \ldots, x_n) \in F_n$ by

$$T_n^{(a)}(1)(y_n) = T_n^{(a)}(1)(x_0,\ldots,x_n) = \mathbf{1}_{V_n(x_n)\geq a} \prod_{1\leq p<n} G_p^-(x_0,\ldots,x_p)$$

More generally, we have for any $\varphi_n \in \mathcal{B}_b(F_n)$

$$\mathbb{E}\left[\varphi_n(X_0,\ldots,X_n)\ ;\ V_n(X_n)\geq a\right] = \gamma_n(T_n^{(a)}(\varphi_n))$$

with the function $T_n^{(a)}(\varphi_n)$ given by

$$T_n^{(a)}(\varphi_n)(x_0,\ldots,x_n) = \varphi_n(x_0,\ldots,x_n)\ \mathbf{1}_{V_n(x_n)\geq a} \prod_{1\leq p<n} G_p^-(x_0,\ldots,x_p) \tag{2.13}$$

To connect the rare event probabilities with the normalized twisted measures we use the fact that

$$\gamma_{n+1}(1) = \gamma_n(G_n) = \eta_n(G_n)\ \gamma_n(1) = \prod_{p=1}^{n} \eta_p(G_p)$$

This readily implies that for any $f_n \in \mathcal{B}_b(F_n)$

$$\gamma_n(f_n) = \eta_n(f_n) \prod_{1\leq p<n} \eta_p(G_p) \tag{2.14}$$

This yields the formulae

$$\mathbb{P}(V_n(X_n) \geq a) = \eta_n(T_n^{(a)}(1)) \prod_{1\leq p<n} \eta_p(G_p)$$

$$\mathbb{E}(\varphi_n(X_0,\ldots,X_n)\ ;\ V_n(X_n) \geq a) = \eta_n(T_n^{(a)}(\varphi_n)) \prod_{1\leq p<n} \eta_p(G_p)$$

$$\mathbb{E}(\varphi_n(X_0,\ldots,X_n) \mid V_n(X_n) \geq a) = \eta_n(T_n^{(a)}(\varphi_n))/\eta_n(T_n^{(a)}(1)) \tag{2.15}$$

To take the final step, we use the Markov property to check that the twisted measures $(\eta_n)_{n\geq 1}$ satisfies the nonlinear recursive equation

$$\eta_n = \Phi_n(\eta_{n-1}) =_{\text{def.}} \int_{F_{n-1}} \eta_{n-1}(dy_{n-1})G_{n-1}(y_{n-1})M_n(y_{n-1},\cdot)/\eta_{n-1}(G_{n-1})$$

starting from $\eta_1 = M_1(x_0,\cdot)$.

2.4.2. Interacting path-particle interpretation. The mean field particle model associated with a collection of transformations Φ_n is a Markov chain $\xi_n = (\xi_n^i)_{1\leq i\leq N}$ taking values at each time $n \geq 1$ in the product spaces F_n^N. Loosely speaking, the algorithm will be conducted so that each path-particle

$$\xi_n^i = (\xi_{0,n}^i, \xi_{1,n}^i, \ldots, \xi_{n,n}^i) \in F_n = (E_0 \times \ldots \times E_n)$$

is almost sampled according to the twisted measure η_n.

The initial configuration $\xi_1 = (\xi_1^i)_{1\leq i\leq N}$ consists of N independent and identically distributed random variables with common distribution

$$\eta_1(d(y_0,y_1)) = M_1(x_0,d(y_0,y_1)) = \delta_{x_0}(dy_0)\ K_1(y_0,dy_1)$$

In other words, $\xi_1^i =_{\text{def.}} (\xi_{0,1}^i, \xi_{1,1}^i) = (x_0, \xi_{1,1}^i) \in F_1 = (E_0 \times E_1)$ can be interpreted as N independent copies $x_0 \rightsquigarrow \xi_{1,1}^i$ of the initial elementary transition $X_0 = x_0 \rightsquigarrow X_1$. The elementary transitions $\xi_{n-1} \rightsquigarrow \xi_n$ from F_{n-1}^N into F_n^N are defined by

$$\mathbb{P}(\xi_n \in d(y_n^1,\ldots,y_n^N) \mid \xi_{n-1}) = \prod_{j=1}^N \Phi_n\left(m(\xi_{n-1})\right)(dy_n^j) \tag{2.16}$$

where $m(\xi_{n-1}) =_{\text{def.}} \frac{1}{N}\sum_{i=1}^N \delta_{\xi_{n-1}^i}$, and $d(y_n^1,\ldots,y_n^N)$ is an infinitesimal neighborhood of the point $(y_n^1,\ldots,y_n^N) \in F_n^N$. By the definition of Φ_n we find that (2.16) is the overlapping of simple selection/mutation genetic transitions

$$\xi_{n-1} \in F_{n-1}^N \xrightarrow{\ selection\ } \widehat{\xi}_{n-1} \in F_{n-1}^N \xrightarrow{\ mutation\ } \xi_n \in F_n^N$$

The selection stage consists of choosing randomly and independently N path-particles

$$\widehat{\xi}_{n-1}^i = (\widehat{\xi}_{0,n-1}^i, \widehat{\xi}_{1,n-1}^i, \ldots, \widehat{\xi}_{n-1,n-1}^i) \in F_{n-1}$$

according to the Boltzmann-Gibbs particle measure

$$\sum_{j=1}^{N} \frac{G_{n-1}(\xi^j_{0,n-1},\ldots,\xi^j_{n-1,n-1})}{\sum_{j'=1}^{N} G_{n-1}(\xi^{j'}_{0,n-1},\ldots,\xi^{j'}_{n-1,n-1})}\, \delta_{(\xi^j_{0,n-1},\ldots,\xi^j_{n-1,n-1})}$$

During the mutation stage, each selected path-particle $\widehat{\xi}^i_{n-1}$ is extended by an elementary K_n-transitions. In other words, we set

$$\begin{aligned}\xi^i_n &= (\quad (\xi^i_{0,n},\ldots,\xi^i_{n-1,n}) \quad , \ \xi^i_{n,n})\\ &= ((\widehat{\xi}^i_{0,n-1},\ldots,\widehat{\xi}^i_{n-1,n-1}),\ \xi^i_{n,n}) \in F_n = F_{n-1}\times E_n\end{aligned}$$

where $\xi^i_{n,n}$ is a random variable with distribution $K_n(\widehat{\xi}^i_{n-1,n-1},\cdot)$. The mutations are performed independently.

2.4.3. Particle approximation measures. It is of course out of the scope of this article to present a full asymptotic analysis of these genealogical particle models. We rather refer the interested reader to the recent monograph [6], and the references therein. For instance, it is well known that the occupation measures of the ancestral lines converge to the desired twisted measures. That is, we have with various precision estimates the weak convergence result

$$\eta^N_n =_{\text{def.}} \frac{1}{N}\sum_{i=1}^{N}\delta_{(\xi^i_{0,n},\xi^i_{1,n},\ldots,\xi^i_{n,n})} \stackrel{N\to\infty}{\longrightarrow} \eta_n$$

In addition, several propagation-of-chaos estimates ensure that the ancestral lines $(\xi^i_{0,n},\xi^i_{1,n},\ldots,\xi^i_{n,n})$ are asymptotically independent and identically distributed with common distribution η_n. The asymptotic analysis of regular models with unbounded potential functions can be treated using traditional cut-off techniques.

Mimicking (2.14), the *un-bias* particle approximation measures γ^N_n of the unnormalized model γ_n are defined as

$$\gamma^N_n(f_n) = \eta^N_n(f_n) \prod_{1\le p<n} \eta^N_p(G_p)$$

By (2.15), we eventually get the particle approximation of the rare event probabilities $P_n(a)$ as stated in the following proposition.

Proposition 2.1. *Let*

$$P^N_n(a) = \gamma^N_n(T^{(a)}_n(1)) = \eta^N_n(T^{(a)}_n(1)) \prod_{1\le p<n}\eta^N_p(G_p) \tag{2.17}$$

$P^N_n(a)$ is an unbiased estimator of $P_n(a)$ such that

$$P^N_n(a) \stackrel{N\to\infty}{\longrightarrow} P_n(a) \qquad a.s. \tag{2.18}$$

In addition, by (2.15), the conditional distribution of the process in the rare event regime can be estimated using the weighted particle measure

$$\begin{aligned}P^N_n(a,\varphi_n) &=_{\text{def.}} \eta^N_n(T^{(a)}_n(\varphi_n))/\eta^N_n(T^{(a)}_n(1))\\ &\stackrel{N\to\infty}{\longrightarrow} P_n(a,\varphi_n) =_{\text{def.}} \mathbb{E}\left[\varphi_n(X_0,\ldots,X_n)\mid V_n(X_n)\ge a\right]\end{aligned} \tag{2.19}$$

When no particles have succeeded to reach the desired level $V_n^{-1}([a,\infty))$, at time n, we have $\eta_n^N(T_n^{(a)}(1)) = 0$, and therefore $\eta_n^N(T_n^{(a)}(\varphi_n)) = 0$, for any $\varphi_n \in \mathcal{B}_b(F_n)$. In this case, we take the convention $P_n^N(a,\varphi_n) = 0$. Also notice that $\eta_n^N(T_n^{(a)}(1)) > 0$ if, and only if, we have $P_n^N(a) > 0$. When $P_n(a) > 0$, we have the exponential decay of the probabilities $\mathbb{P}(P_n^N(a) = 0) \to 0$, as N tends to infinity.

The above asymptotic estimates are almost sure convergence results. We can also precise these convergences in various ways, including $\mathbb{L}_p$-mean-error or increasing propagations of chaos analysis, central limit theorems and Berry-Esseen type inequalities, as well as exponential estimates and large deviations principles (see [6]).

2.5. Fluctuations and variance comparisons. The fluctuations of genetic type particle models have been initiated in Ref. [7]. Under appropriate regularity conditions on the mutation transitions, this study provides a central limit theorem for the path particle model $(\xi_0^i,\ldots,\xi_n^i)_{1\leq i\leq N}$. Several extensions, including Donsker's type theorems, Berry-Esseen inequalities, and applications to nonlinear filtering problems can be found in Refs. [10, 8, 9, 11]. In this section, we design a simplified analysis essentially based on the fluctuations of random fields associated to the local sampling errors. In the first part of this subsection, we provide a brief discussion on the fluctuations analysis of the weighted particle measures introduced in Subsection 2.4. We underline several interpretations of the central limit variances in terms of twisted Feynman-Kac measures. In the final part of this section, we illustrate these general and theoretical fluctuation analysis in the warm up Gaussian situation discussed in (2.2), (2.4), and (2.10). In this context, we derive an explicit description of the error-variances, and we compare the performance of the IPS methodology with the noninteracting IS technique.

2.5.1. Central limit theorems. The fluctuations of the mean field particle models described in Subsection 2.4 are essentially based on the asymptotic analysis of the local sampling errors associated with the particle approximation sampling steps. These local errors are defined in terms of the random fields $\mathcal{W}_n^N$, given for any $f_n \in \mathcal{B}_b(F_n)$ by the formula

$$\mathcal{W}_n^N(f_n) = \sqrt{N}\ [\eta_n^N - \Phi_n(\eta_{n-1}^N)](f_n)$$

The next central limit theorem is pivotal. Its complete proof can be found in [6]. For any fixed time horizon $n \geq 1$, the sequence $(\mathcal{W}_p^N)_{1\leq p\leq n}$ converges in law, as N tends to infinity, to a sequence of n independent, Gaussian and centered random fields $(\mathcal{W}_p)_{1\leq p\leq n}$; with, for any $f_p, g_p \in \mathcal{B}_b(F_p)$, and $1 \leq p \leq n$,

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

Let $Q_{p,n}$, with $1 \leq p \leq n$, be the Feynman-Kac semi-group associated to the flow $\gamma_n = \gamma_p Q_{p,n}$. For $p = n$, we use the convention that $Q_{n,n} = Id$. Using the Markov property, it is not difficult to check that $Q_{p,n}$ has the following functional representation

$$Q_{p,n}(f_n)(y_p) = \mathbb{E}\left[f_n(Y_n)\prod_{p\leq k<n} G_k(Y_k) \mid Y_p = y_p\right]$$

for any test function $f_n \in \mathcal{B}_b(F_n)$, and any path sequence $y_p = (x_0,\ldots,x_p) \in F_p$.

To explain what we have in mind when making these definitions, we now consider the elementary telescopic decomposition

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

For $p = 1$, we recall that $\eta_0^N = \delta_{x_0}$ and $\gamma_1 = \eta_1 = M_1(x_0, \cdot)$, from which we find that $\eta_0^N Q_{0,n} = \gamma_1 Q_{1,n} = \gamma_n$. Using the fact that

$$\gamma_{p-1}^N Q_{p-1,p} = \gamma_{p-1}^N(G_{p-1}) \ \times \ \Phi_{p-1}(\eta_{p-1}^N) \quad \text{and} \quad \gamma_{p-1}^N(G_{p-1}) = \gamma_p^N(1)$$

the above decomposition readily implies that

$$\mathcal{W}_n^{\gamma,N}(f_n) =_{\text{def.}} \sqrt{N}[\gamma_n^N - \gamma_n](f_n) = \sum_{p=1}^{n} \gamma_p^N(1) \ \mathcal{W}_p^N(Q_{p,n} f_n) \tag{2.20}$$

It is now easy to check that γ_n^N is an unbiased estimate of γ_n, in the sense that $\mathbb{E}(\gamma_n^N(f_n)) = \gamma_n(f_n)$, for any $f_n \in \mathcal{B}_b(F_n)$.

To take the final step, we recall that the random sequence $(\gamma_p^N(1))_{1\leq p\leq n}$ converges in law, as N tends to infinity, to the deterministic sequence $(\gamma_p(1))_{1\leq p\leq n}$ (see for instance [6]). A simple application of Slutsky's Lemma, now implies that the random fields $\mathcal{W}_n^{\gamma,N}$ converge in law, as N tends to infinity, to the Gaussian random fields $\mathcal{W}_n^{\gamma}$ defined for any $f_n \in \mathcal{B}_b(F_n)$ by

$$\mathcal{W}_n^{\gamma}(f_n) = \sum_{p=1}^{n} \gamma_p(1) \ \mathcal{W}_p(Q_{p,n} f_n) \tag{2.21}$$

In much the same way, the sequence of random fields

$$\begin{aligned} \mathcal{W}_n^{\eta,N}(f_n) &=_{\text{def.}} \sqrt{N}[\eta_n^N - \eta_n](f_n) \\ &= \frac{\gamma_n(1)}{\gamma_n^N(1)} \ \times \ \mathcal{W}_n^{\gamma,N}\left(\frac{1}{\gamma_n(1)}(f_n - \eta_n(f_n))\right) \end{aligned} \tag{2.22}$$

converges in law, as N tends to infinity, to the Gaussian random fields $\mathcal{W}_n^{\eta}$ defined for any $f_n \in \mathcal{B}_b(F_n)$ by

$$\mathcal{W}_n^{\eta}(f_n) = \mathcal{W}_n^{\gamma}\left(\frac{1}{\gamma_n(1)}(f_n - \eta_n(f_n))\right) = \sum_{p=1}^{n} \mathcal{W}_p\left(\frac{Q_{p,n}}{\eta_p Q_{p,n}(1)}(f_n - \eta_n(f_n))\right) \tag{2.23}$$

One simple consequence of the above fluctuations is a central limit theorem for the particle rare event estimates $P_n^N(a)$ introduced in (2.17). We first notice that

$$\sqrt{N} \ [P_n^N(a) - P_n(a)] = \mathcal{W}_n^{\gamma,N}(T_n^{(a)}(1))$$

with the weighted function $T_n^{(a)}(1)$ introduced in (2.13). From previous considerations, we readily find that $\mathcal{W}_n^{\gamma,N}(T_n^{(a)}(1))$ converge in law, as N tends to infinity, to a centered Gaussian random variable $\mathcal{W}_n^{\gamma}(T_n^{(a)}(1))$ with the variance

$$\sigma_n^{\gamma}(a)^2 =_{\text{def.}} \mathbb{E}(\mathcal{W}_n^{\gamma}(T_n^{(a)}(1))^2) = \sum_{p=1}^{n} \gamma_p(1)^2 \ \eta_p([Q_{p,n}(T_n^{(a)}(1)) - \eta_p Q_{p,n}(T_n^{(a)}(1))]^2)$$

To have a more explicit description of $\sigma_n^\gamma(a)$ we notice that

$$Q_{p,n}(T_n^{(a)}(1))(x_0,\ldots,x_p) = \left\{\prod_{1\leq k<p} G_k(x_0,\ldots,x_k)^{-1}\right\} \mathbb{P}(V_n(X_n)\geq a \mid X_p = x_p)$$

By definition of η_p, we also find that

$$\eta_p(Q_{p,n}(T_n^{(a)}(1))) = \mathbb{P}(V_n(X_n)\geq a)/\gamma_p(1)$$

From these observations, we conclude that

$$\begin{aligned}\sigma_n^\gamma(a)^2 &= \sum_{p=1}^n \left\{\gamma_p(1)\ \mathbb{E}\left[\prod_{1\leq k<p} G_k^-(X_0,\ldots,X_k)\ \mathbb{E}(\mathbf{1}_{V_n(X_n)\geq a} \mid X_p)^2\right] - P_n(a)^2\right\}\\ &= \sum_{p=1}^n [\gamma_p(1)\gamma_p^-(1)\ \eta_p^-(P_{p,n}(a)^2) - P_n(a)^2]\end{aligned}$$

Our next objective is to analyze the fluctuations of the particle conditional distributions of the process in the rare event regime defined in (2.19):

$$\sqrt{N}[P_n^N(a,\varphi_n) - P_n(a,\varphi_n)] = \frac{\eta_n T_n^{(a)}(1)}{\eta_n^N T_n^{(a)}(1)} \times \mathcal{W}_n^{\eta,N}\left(\frac{T_n^{(a)}}{\eta_n T_n^{(a)}(1)}\ (\varphi_n - P_n(a,\varphi_n))\right)$$

Using the same arguments as above, and by (2.23), we have the weak convergence

$$\sqrt{N}[P_n^N(a,\varphi_n)-P_n(a,\varphi_n)]\ \mathbf{1}_{P_n^N(a)>0} \overset{N\to\infty}{\longrightarrow} \mathcal{W}_n^\eta\left(\frac{T_n^{(a)}}{\eta_n T_n^{(a)}(1)}\ (\varphi_n - P_n(a,\varphi_n))\right) \tag{2.24}$$

The limit is a centered Gaussian random variable with variance

$$\sigma_n(a,\varphi_n)^2 =_{\text{def.}} \mathbb{E}\left[\mathcal{W}_n^\eta\left(\frac{T_n^{(a)}}{\eta_n T_n^{(a)}(1)}\ (\varphi_n - P_n(a,\varphi_n))\right)^2\right]$$

Taking into account the definition of $\mathcal{W}_n^\eta$ and the identities $\eta_n T_n^{(a)}(1) = P_n(a)/\gamma_n(1)$ and $\eta_n T_n^{(a)}\ (\varphi_n - P_n(a,\varphi_n)) = 0$, we readily check that

$$\sigma_n(a,\varphi_n)^2 = P_n(a)^{-2}\sum_{p=1}^n \gamma_p(1)\ \gamma_p\left([Q_{p,n}(T_n^{(a)}\ (\varphi_n - P_n(a,\varphi_n))]^2\right)$$

We summarize the above discussion with the following theorem.

Theorem 2.2. *The estimator $P_n^N(a)$ given by (2.17) is unbiased, and it satisfies the central limit theorem*

$$\sqrt{N}\left[P_n^N(a) - P_n(a)\right] \overset{N\to\infty}{\longrightarrow} \mathcal{N}(0,\sigma_n^\gamma(a)^2) \tag{2.25}$$

with the asymptotic variance

$$\sigma_n^\gamma(a)^2 = \sum_{p=1}^n [\gamma_p(1)\gamma_p^-(1)\ \eta_p^-(P_{p,n}(a)^2) - P_n(a)^2] \tag{2.26}$$

and the collection of functions $P_{p,n}(a)$ *defined by*

$$x_p \in E_p \mapsto P_{p,n}(a)(x_p) = \mathbb{E}[\mathbf{1}_{V_n(X_n)\geq a} \mid X_p = x_p] \in [0,1] \tag{2.27}$$

In addition, for any $\varphi_n \in \mathcal{B}_b(F_n)$*, the estimator* $P_n^N(a,\varphi_n)$ *given by (2.19) satisfies the central limit theorem*

$$\sqrt{N}\left[P_n^N(a,\varphi_n) - P_n(a,\varphi_n)\right] \stackrel{N\to\infty}{\longrightarrow} \mathcal{N}(0,\sigma_n(a,\varphi_n)^2) \tag{2.28}$$

with the asymptotic variance

$$\sigma_n(a,\varphi_n)^2 = P_n(a)^{-2}\sum_{p=1}^{n} \gamma_p(1)\gamma_p^-(1)\ \eta_p^-(P_{p,n}(a,\varphi_n)^2)$$

and the collection of functions $P_{p,n}(a,\varphi_n) \in \mathcal{B}_b(F_p)$ *defined by*

$$\begin{aligned} &P_{p,n}(a,\varphi_n)(x_0,\ldots,x_p) \\ &= \mathbb{E}\left[(\varphi_n(X_0,\ldots,X_n) - P_n(a,\varphi_n))\ \mathbf{1}_{V_n(X_n)\geq a} \mid (X_0,\ldots,X_p) = (x_0,\ldots,x_p)\right] \end{aligned} \tag{2.29}$$

Arguing as in the end of Section 2.2, we note that the measures η_p^- tend to favor random paths with low $(G_k)_{1\leq k<p}$-potential values. Recalling that these potentials are chosen so as to represent the process evolution in the rare level set, we expect the quantities $\eta_p^-(P_{p,n}(a)^2)$ to be much smaller than $P_n(a)$. In the reverse angle, by Jensen's inequality we expect the normalizing constants products $\gamma_p(1)\gamma_p^-(1)$ to be rather large. We shall make precise these intuitive comments in the next section, with explicit calculations for the pair Gaussian models introduced in (2.4) and (2.10). We end the section by noting that

$$\sigma_n(a,\varphi_n)^2 \leq P_n(a)^{-2}\sum_{p=1}^{n} \gamma_p(1)\gamma_p^-(1)\ \eta_p^-(P_{p,n}(a)^2)$$

for any test function φ_n, with $\sup_{(y_n,y_n')\in F_p^2} |\varphi_n(y_n) - \varphi_n(y_n')| \leq 1$.

2.6. On the weak negligible bias of genealogical models. In this subsection, we complete the fluctuation analysis developed in Subsection 2.5.1, with the study of the bias of the genealogical tree occupation measures η_n^N, and the corresponding weighted measures $P_n^N(a,\varphi_n)$ defined by (2.19). The forthcoming analysis also provide sharp estimates, and a precise asymptotic description of the law of a given particle ancestral line. In this sense, this study also completes the propagation-of-chaos analysis developed in [6].

Before getting into further details, we recall from [6] that for $\nu \in \{\gamma,\eta\}$, and for any $f_n \in \mathcal{B}_b(F_n)$, and $p\geq 1$, we have the $\mathbb{L}_p$-mean error estimates

$$\sup_{N\geq 1} \mathbb{E}\left[|\mathcal{W}_n^{\nu,N}(f_n)|^p\right]^{1/p} < \infty$$

with the random fields $(\mathcal{W}_n^{\gamma,N},\mathcal{W}_n^{\eta,N})$ defined in (2.20) and (2.22). By the Borel-Cantelli lemma this property ensures that $(\gamma_n^N(f_n),\eta_n^N(f_n))$ converges almost surely to $(\gamma_n(f_n),\eta_n(f_n))$, as N tends to infinity. Using Hölder's inequality, we can also prove that any polynomial function of terms $\mathcal{W}_n^{\nu,N}(f_n)$, $\nu \in \{\gamma,\eta\}$, $f_n \in \mathcal{B}_b(F_n)$, forms a uniformly integrable collection of random variables (indexed by the size and precision parameter $N \geq 1$). This property, combined with the continuous mapping theorem, and Skorohod embedding theorem, allows us to state the following technical lemma.

LEMMA 2.3. *For any $n, d \geq 1$, any collection of functions $(f_n^i)_{1\leq i\leq d} \in \mathcal{B}_b(F_n)^d$, and any sequence $(\nu^i)_{1\leq i\leq d} \in \{\gamma,\eta\}^d$, the random products $\prod_{i=1}^d \mathcal{W}_n^{\nu^i,N}(f_n^i)$ converge in law, as N tends to infinity, to the Gaussian products $\prod_{i=1}^d \mathcal{W}_n^{\nu^i}(f_n^i)$. In addition, we have*

$$\lim_{N\to\infty} \mathbb{E}\left[\prod_{i=1}^d \mathcal{W}_n^{\nu^i,N}(f_n^i)\right] = \mathbb{E}\left[\prod_{i=1}^d \mathcal{W}_n^{\nu^i}(f_n^i)\right]$$

This result is pivotal in our way to analyze the bias of the path-particle models. To illustrate our approach, we already present an elementary consequence of Lemma 2.3. We first rewrite (2.22) as follows

$$\begin{aligned}
&\mathcal{W}_n^{\eta,N}(f_n)\\
&= \mathcal{W}_n^{\gamma,N}\left(\tfrac{1}{\gamma_n(1)}(f_n-\eta_n(f_n))\right) + \left(\tfrac{\gamma_n(1)}{\gamma_n^N(1)}-1\right) \;\times\; \mathcal{W}_n^{\gamma,N}\left(\tfrac{1}{\gamma_n(1)}(f_n-\eta_n(f_n))\right)\\
&= \mathcal{W}_n^{\gamma,N}(\tilde{f}_n) - \tfrac{1}{\sqrt{N}}\;\tfrac{\gamma_n(1)}{\gamma_n^N(1)}\;\mathcal{W}_n^{\gamma,N}(\tilde{f}_n)\mathcal{W}_n^{\gamma,N}(\tilde{g}_n)
\end{aligned}$$

with the pair of functions $(\tilde{f}_n, \tilde{g}_n)$ defined by

$$\tilde{f}_n = \frac{1}{\gamma_n(1)}(f_n - \eta_n(f_n)) \quad \text{and} \quad \tilde{g}_n = \frac{1}{\gamma_n(1)}$$

This readily yields that

$$N\;\mathbb{E}\left[\eta_n^N(f_n) - \eta_n(f_n)\right] = -\mathbb{E}\left[\frac{\gamma_n(1)}{\gamma_n^N(1)}\;\mathcal{W}_n^{\gamma,N}(\tilde{f}_n)\mathcal{W}_n^{\gamma,N}(\tilde{g}_n)\right]$$

Since the sequence of random variables $(\gamma_n(1)/\gamma_n^N(1))_{N\geq 1}$ is uniformly bounded, and it converges in law to 1, as N tends to infinity, by Lemma 2.3 we conclude that

$$\begin{aligned}
\lim_{N\to\infty} N\;\mathbb{E}[\eta_n^N(f_n) - \eta_n(f_n)] &= -\mathbb{E}\left[\mathcal{W}_n^{\gamma}(\tilde{f}_n)\mathcal{W}_n^{\gamma}(\tilde{g}_n)\right]\\
&= -\sum_{p=1}^n \eta_p\left(\overline{Q}_{p,n}(1)\;\overline{Q}_{p,n}(f_n-\eta_n(f_n))\right)
\end{aligned} \tag{2.30}$$

with the re-normalized semi-group $\overline{Q}_{p,n}$ defined by

$$\overline{Q}_{p,n}(f_n) = \frac{Q_{p,n}(f_n)}{\eta_p Q_{p,n}(1)} = \frac{\gamma_p(1)}{\gamma_n(1)}\;Q_{p,n}(f_n)$$

We are now in position to state and prove the main result of this subsection.

THEOREM 2.4. *For any $n \geq 1$ and $\varphi_n \in \mathcal{B}_b(F_n)$, we have*

$$\begin{aligned}
&N\;\mathbb{E}\left[(P_n^N(a,\varphi_n) - P_n(a,\varphi_n))\;\mathbf{1}_{P_n^N(a)>0}\right]\\
&\stackrel{N\to\infty}{\longrightarrow} -P_n(a)^{-2}\;\sum_{p=1}^n \gamma_p(1)\gamma_p^-(1)\;\eta_p^-\;[P_{p,n}(a)P_{p,n}(a,\varphi_n)]
\end{aligned}$$

with the collection of functions $P_{p,n}(a)$, $P_{p,n}(a,\varphi_n) \in \mathcal{B}_b(F_p)$ defined respectively in (2.27) and (2.29).

Proof:

The proof is essentially based on a judicious way to rewrite (2.24). If we define

$$f_n^{(a)} = \frac{T_n^{(a)}}{\eta_n T_n^{(a)}(1)} \; (\varphi_n - P_n(a,\varphi_n)) \quad \text{and} \quad g_n^{(a)} = \frac{T_n^{(a)}(1)}{\eta_n T_n^{(a)}(1)}$$

then, on the event $\{P_n^N(a) > 0\}$, it is easy to check that

$$\begin{aligned} & N[P_n^N(a,\varphi_n) - P_n(a,\varphi_n)] \\ & = N[\eta_n^N(f_n^{(a)}) - \eta_n(f_n^{(a)})] - \tfrac{1}{\eta_n^N(g_n^{(a)})} \; \mathcal{W}_n^{\eta,N}(f_n^{(a)})\mathcal{W}_n^{\eta,N}(g_n^{(a)}) \end{aligned}$$

By Lemma 2.3 and (2.30) we conclude that

$$\begin{aligned} & N \; \mathbb{E}\left[(P_n^N(a,\varphi_n) - P_n(a,\varphi_n)) \; \mathbf{1}_{P_n^N(a)>0}\right] \\ & \stackrel{N\to\infty}{\longrightarrow} -\mathbb{E}\left[\mathcal{W}_n^{\eta}(f_n^{(a)})\mathcal{W}_n^{\eta}(g_n^{(a)})\right] - \mathbb{E}\left[\mathcal{W}_n^{\gamma}(\frac{f_n^{(a)}}{\gamma_n(1)})\mathcal{W}_n^{\gamma}(\frac{1}{\gamma_n(1)})\right] \end{aligned}$$

On the other hand, using (2.23) we find that

$$\begin{aligned} \mathbb{E}\left[\mathcal{W}_n^{\eta}(f_n^{(a)})\mathcal{W}_n^{\eta}(g_n^{(a)})\right] &= \sum_{p=1}^{n} (\gamma_p(1)/\gamma_n(1))^2 \; \mathbb{E}\left[\mathcal{W}_p\left(Q_{p,n}(f_n^{(a)})\right)\mathcal{W}_p\left(Q_{p,n}(g_n^{(a)}-1)\right)\right] \\ &= \sum_{p=1}^{n} (\gamma_p(1)/\gamma_n(1))^2 \; \eta_p\left(Q_{p,n}(f_n^{(a)})Q_{p,n}(g_n^{(a)}-1)\right) \end{aligned}$$

Similarly, by (2.21) we have

$$\mathbb{E}\left[\mathcal{W}_n^{\gamma}(\frac{f_n^{(a)}}{\gamma_n(1)})\mathcal{W}_n^{\gamma}(\frac{1}{\gamma_n(1)})\right] = \sum_{p=1}^{n} \gamma_p(1)^2 \; \mathbb{E}\left[\mathcal{W}_p(Q_{p,n}\frac{f_n^{(a)}}{\gamma_n(1)})\mathcal{W}_p(Q_{p,n}\frac{1}{\gamma_n(1)})\right]$$

It is now not difficult to check that

$$\mathbb{E}\left[\mathcal{W}_n^{\gamma}(\frac{f_n^{(a)}}{\gamma_n(1)})\mathcal{W}_n^{\gamma}(\frac{1}{\gamma_n(1)})\right] = \sum_{p=1}^{n} (\gamma_p(1)/\gamma_n(1))^2 \, \eta_p\left(Q_{p,n}(1)Q_{p,n}(f_n^{(a)})\right)$$

from which we conclude that

$$\begin{aligned} & N \; \mathbb{E}([P_n^N(a,\varphi_n) - P_n(a,\varphi_n)] \; \mathbf{1}_{P_n^N(a)>0}) \\ & \stackrel{N\to\infty}{\longrightarrow} \sum_{p=1}^{n} (\gamma_p(1)/\gamma_n(1))^2 \; \eta_p\left(Q_{p,n}(f_n^{(a)})Q_{p,n}(g_n^{(a)})\right) \end{aligned}$$

By the definition of the function $T_n^{(a)}(\varphi_n)$ we have $\eta_n T_n^{(a)}(1) = P_n(a)/\gamma_n(1)$ and for any $y_p = (x_0,\ldots,x_p) \in F_p$

$$\begin{aligned} Q_{p,n}(T_n^{(a)}(\varphi_n))(x_0,\ldots,x_p) &= [\textstyle\prod_{1\le k<p} G_k^-(x_0,\ldots,x_k)] \\ & \quad \times\mathbb{E}\left[\varphi_n(X_0,\ldots,X_n) \; \mathbf{1}_{V_n(X_n)\ge a} \mid (X_0,\ldots,X_p) = (x_0,\ldots,x_p)\right] \end{aligned}$$

By the definition of the pair of functions $(f_n^{(a)}, g_n^{(a)})$, these observations yield

$$Q_{p,n}(f_n^{(a)})(x_0,\ldots,x_p) = \frac{\gamma_n(1)}{P_n(a)}[\prod_{1\leq k<p} G_k^-(x_0,\ldots,x_k)]\ P_{p,n}(a,\varphi_n)(x_0,\ldots,x_p)$$

$$Q_{p,n}(g_n^{(a)})(x_0,\ldots,x_p) = \frac{\gamma_n(1)}{P_n(a)}[\prod_{1\leq k<p} G_k^-(x_0,\ldots,x_k)]\ P_{p,n}(a)(x_0,\ldots,x_p)$$

The end of the proof is now straightforward. ∎

2.7. Variance comparisons for Gaussian particle models. Let $(X_p)_{1\leq p\leq n}$ be the Gaussian sequence defined in (2.2). We consider the elementary energy like function $V_n(x) = x$, and the Feynman-Kac twisted models associated to the potential functions

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

Arguing as in (2.4), we prove that the Feynman-Kac distribution η_p^- is the path distribution of the chain defined by the recursion

$$X_p^- = X_{p-1}^- + W_p \quad \text{and} \quad X_k^- = X_{k-1}^- - \lambda + W_k, \quad 1\leq k<p$$

where $X_0 = 0$, and where $(W_k)_{1\leq k\leq p}$ represents a sequence of independent and identically distributed Gaussian random variables, with $(\mathbb{E}(W_1), \mathbb{E}(W_1^2)) = (0,1)$. We also observe that in this case we have

$$\gamma_p(1)\gamma_p^-(1) = \mathbb{E}[e^{\lambda X_{p-1}}]^2 = e^{\lambda^2(p-1)} \tag{2.31}$$

The next lemma is instrumental for estimating the quantities $\eta_p^-(P_{p,n}(a)^2)$ introduced in (2.26).

Lemma 2.5. *Let (W_1, W_2) be a pair of independent Gaussian random variables, with $(\mathbb{E}(W_i), \mathbb{E}(W_i^2)) = (0,\sigma_i^2)$, with $\sigma_i > 0$ and $i = 1,2$. Then, for any $a > 0$, we have the exponential estimate*

$$C(a,\sigma_1,\sigma_2) \leq \mathbb{E}\left[\mathbb{P}(W_1 + W_2 \geq a \mid W_1)^2\right] \exp\left(\frac{a^2}{2\sigma_1^2+\sigma_2^2}\right) \leq 1$$

where

$$C(a,\sigma_1,\sigma_2) = (2\pi)^{3/2}\left(\frac{\sigma_2 a}{2\sigma_1^2+\sigma_2^2} + \frac{2\sigma_1^2+\sigma_2^2}{\sigma_2 a}\right)^{-2}\left(\frac{2\sigma_1 a}{2\sigma_1^2+\sigma_2^2} + \frac{2\sigma_1^2+\sigma_2^2}{2\sigma_1 a}\right)^{-1}$$

Proof:
Using exponential version of Chebychev's inequality we first check that, for any $\lambda > 0$, we have

$$\mathbb{P}(W_1 + W_2 \geq a \mid W_1) \leq e^{\lambda(W_1-a)}\ \mathbb{E}(e^{\lambda W_2}) = e^{\lambda(W_1-a)+\lambda^2\sigma_2^2/2}$$

Integrating the random variable W_1, and choosing $\lambda = a/(2\sigma_1^2+\sigma_2^2)$ we establish the upper bound

$$\mathbb{E}\left[\mathbb{P}(W_1 + W_2 \geq a \mid W_1)^2\right] \leq e^{-2\lambda a+\lambda^2(2\sigma_1^2+\sigma_2^2)} = e^{-\frac{a^2}{2\sigma_1^2+\sigma_2^2}}$$

For any $\epsilon \in (0,1)$, we have

$$\mathbb{E}\left[\mathbb{P}(W_1+W_2 \geq a \mid W_1)^2\right] \geq \mathbb{P}(W_2 \geq \epsilon a)^2 \mathbb{P}(W_1 \geq (1-\epsilon)a)$$

Applying Mill's inequality yields

$$\mathbb{E}\left[\mathbb{P}(W_1+W_2 \geq a \mid W_1)^2\right] \geq \frac{(2\pi)^{3/2}}{\left(\frac{\epsilon a}{\sigma_2}+\frac{\sigma_2}{\epsilon a}\right)^2\left(\frac{(1-\epsilon)a}{\sigma_1}+\frac{\sigma_1}{(1-\epsilon)a}\right)} e^{-a^2\left(\frac{\epsilon^2}{\sigma_2^2}+\frac{(1-\epsilon)^2}{2\sigma_1^2}\right)}$$

Choosing $\epsilon = \sigma_2^2/(2\sigma_1^2+\sigma_2^2)$ establishes the lower bound. ■

From previous considerations, we notice that

$$\eta_p^-(P_{p,n}(a)^2) = \mathbb{E}\left[\mathbb{P}(W_1+W_2 \geq (a+\lambda(p-1)) \mid W_1)^2\right]$$

where (W_1, W_2) is a pair of independent and centered Gaussian random variables, with $(\mathbb{E}(W_1^2), \mathbb{E}(W_2^2)) = (p, n-p)$. Lemma 2.5 now implies that

$$\eta_p^-(P_{p,n}(a)^2) \leq \exp\left[-(a+\lambda(p-1))^2/(n+p)\right] \tag{2.32}$$

Using the estimates (2.31) and (2.32), we find that

$$\begin{aligned}
\sigma_n^\gamma(a)^2 &\leq \sum_{p=1}^n [e^{\lambda^2(p-1)-\frac{(a+\lambda(p-1))^2}{n+p}} - P_n(a)^2] \\
&= \sum_{0\leq p<n} [e^{-\frac{a^2}{n}}\ e^{\frac{p+1}{n(n+p+1)}\left[a-\lambda\frac{np}{p+1}\right]^2+\lambda^2\frac{p}{p+1}} - P_n(a)^2]
\end{aligned}$$

For $\lambda = a/n$, this yields that

$$\sigma_n^\gamma(a)^2 \leq \sum_{0\leq p<n} [e^{-\frac{a^2}{n}}\ e^{\frac{a^2}{n^2}\left(1-\frac{1}{n+p+1}\right))} - P_n(a)^2] \leq n\left(e^{-\frac{a^2}{n}\left(1-\frac{1}{n}\right)} - P_n(a)^2\right)$$

We find that this estimate has the same exponential decay as the one obtained in (2.6) for the corresponding noninteracting IS model. The only difference between these two asymptotic variances comes from the multiplication parameter n. This additional term can be interpreted as the number of interactions used in the construction of the genealogical tree simulation model.

Now, we consider the Feynman-Kac twisted models associated to the potential functions

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

Arguing as in (2.11), we prove that η_p^- is the distribution of the Markov chain

$$X_p^- = X_{p-1}^- + W_p \quad \text{and} \quad X_k^- = X_{k-1}^- - \lambda\,(p-k) + W_k, \quad 1 \leq k < p$$

where $X_0 = 0$, and where $(W_k)_{1\leq k\leq p}$ represents a sequence of independent and identically distributed Gaussian random variables, with $(\mathbb{E}(W_1), \mathbb{E}(W_1^2)) = (0,1)$. We also notice that

$$\gamma_p(1)\gamma_p^-(1) = \mathbb{E}\left[e^{\lambda\sum_{1\leq k<p} X_k}\right]^2 = e^{\lambda^2\sum_{1\leq k<p} k^2} \tag{2.33}$$

In this situation, we observe that

$$\eta_p^-(P_{p,n}(a)^2) = \mathbb{E}\left[\mathbb{P}\left(W_1 + W_2 \geq a + \lambda \sum_{1\leq k<p} k \mid W_1\right)^2\right]$$

where (W_1, W_2) is a pair of independent and centered Gaussian random variables, with $(\mathbb{E}(W_1^2), \mathbb{E}(W_2^2)) = (p, n-p)$. As before, Lemma 2.5 now implies that

$$\eta_p^-(P_{p,n}(a)^2) \leq \exp\left[-\frac{1}{n+p}\left(a + \lambda\frac{p(p-1)}{2}\right)^2\right] \tag{2.34}$$

Using the estimates (2.33) and (2.34), and recalling that $\sum_{1\leq k\leq n} k^2 = n(n+1)(2n+1)/6$, we conclude that

$$\begin{aligned}\sigma_n^\gamma(a)^2 &\leq \sum_{p=1}^n [e^{\frac{1}{6}\lambda^2(p-1)p(2p-1) - \frac{(a+\lambda p(p-1)/2)^2}{n+p}} - P_n(a)^2] \\ &= \sum_{p=1}^n [e^{-\frac{a^2}{n}}\ e^{\frac{p}{n(n+p)}[a-\lambda n(p-1)/2]^2 + \frac{1}{12}\lambda^2(p-1)p(p+1)} - P_n(a)^2]\end{aligned}$$

If we take $\lambda = 2a/[n(n-1)]$, then we get

$$\begin{aligned}\sigma_n^\gamma(a)^2 &\leq \sum_{p=1}^n [e^{-\frac{a^2}{n}}\ e^{\frac{a^2}{n^2(n-1)^2}\left[\frac{np}{n+p}(n-p)^2 + \frac{(p-1)p(p+1)}{3}\right]} - P_n(a)^2] \\ &= \sum_{p=1}^n [e^{-\frac{a^2}{n}}\ e^{\frac{a^2}{n}\frac{n^2}{(n-1)^2}\left[\theta(\frac{p}{n}) - \frac{p}{3n^3}\right]} - P_n(a)^2]\end{aligned}$$

with the increasing function $\theta : \epsilon \in [0,1] \longrightarrow \theta(\epsilon) = \epsilon\frac{(1-\epsilon)^2}{(1+\epsilon)} + \frac{\epsilon^3}{3} \in [0, 1/3]$. From these observations, we deduce the estimate

$$\sigma_n^\gamma(a)^2 \leq n\ \left[e^{-\frac{a^2}{n}\ \frac{2}{3}\left(1-\frac{1}{n-1}\right)} - P_n(a)^2\right]$$

Note that the inequalities are sharp in the exponential sense by the lower bound obtained in Lemma 2.5. Accordingly we get that the asymptotic variance is not of the order of $P_n(a)^2$, but rather $P_n(a)^{4/3}$. As in the first Gaussian example, we observe that this estimate has the same exponential decays as the one obtained in (2.12) for the corresponding IS algorithm.

3. Estimation of the tail of a probability density function. We collect and sum-up the general results presented in Section 2 and we apply them to propose an efficient estimator for the tail of the probability density function (pdf) of a real-valued function of a Markov chain. We consider a $(E, \mathcal{E})$-valued Markov chain $(X_p)_{0\leq p\leq n}$ with non-homogeneous transition kernels K_p. In a first time, we show how the results obtained in the previous section allow us to estimate the probability of a rare event of the form $\{V(X_n) \in A\}$

$$P_A = \mathbb{P}(V(X_n) \in A) = \mathbb{E}[\mathbf{1}_A(V(X_n))] \tag{3.1}$$

where V is some function from E to $\mathbb{R}$. We shall construct an estimator based on an IPS. As pointed out in the previous section, the quality of the estimator depends on the choice on the weight function. The weight function should fulfill two conditions. First, it should favor the occurrence of the rare event without involving too large normalizing constants. Second, it should give rise to an algorithm that can be easily implemented. Indeed the implementation of the IPS with an arbitrary weight function requires to record the complete set of path-particles. If N particles are generated and time runs from 0 to n, this set has size $(n+1) \times N \times \dim(E)$ which may exceed the memory capacity of the computer. The weight function should be chosen so that only a smaller set needs to be recorded to compute the estimator of the probability of occurrence of the rare event. We shall examine two weight functions and the two corresponding algorithms that fulfill both conditions.

Algorithm 1. Let us fix some $\beta > 0$. The first algorithm is built with the weight function

$$G_p^\beta(x) = \exp\left[\beta V(x_p)\right] \tag{3.2}$$

The practical implementation of the IPS reads as follows.

`Initialization.` We start with a set of N i.i.d. initial conditions $\hat{X}_0^{(i)}$, $1 \le i \le N$, chosen according to the initial distribution X_0. This set is complemented with a set of weights $\hat{Y}_0^{(i)} = 1$, $1 \le i \le N$. This forms a set of N particles: $(\hat{X}_0^{(i)}, \hat{Y}_0^{(i)})$, $1 \le i \le N$, where a particle is a pair $(\hat{X}, \hat{Y}) \in E \times \mathbb{R}^+$.

Now, assume that we have a set of N particles at time p denoted by $(\hat{X}_p^{(i)}, \hat{Y}_p^{(i)})$, $1 \le i \le N$.

`Selection.` We first compute the normalizing constant

$$\hat{\eta}_p^N = \frac{1}{N}\sum_{i=1}^N \exp\left[\beta V(\hat{X}_p^{(i)})\right] \tag{3.3}$$

We choose independently N particles according to the empirical distribution

$$\mu_p^N(d\check{X}, d\check{Y}) = \frac{1}{N\hat{\eta}_p^N} \times \sum_{i=1}^N \exp\left[\beta V(\hat{X}_p^{(i)})\right] \delta_{(\hat{X}_p^{(i)}, \hat{Y}_p^{(i)})}(d\check{X}, d\check{Y}) \tag{3.4}$$

The new particles are denoted by $(\check{X}_p^{(i)}, \check{Y}_p^{(i)})$, $1 \le i \le N$.

`Mutation.` For every $1 \le i \le N$, the particle $(\check{X}_p^{(i)}, \check{Y}_p^{(i)})$ is transformed into $(\hat{X}_{p+1}^{(i)}, \hat{Y}_{p+1}^{(i)})$ by the mutation procedure

$$\check{X}_p^{(i)} \xrightarrow{K_{p+1}} \hat{X}_{p+1}^{(i)}, \tag{3.5}$$

where the mutations are performed independently, and

$$\hat{Y}_{p+1}^{(i)} = \check{Y}_p^{(i)} \exp\left[-\beta V(\check{X}_p^{(i)})\right]. \tag{3.6}$$

The memory required by the algorithm is $N\dim(E) + N + n$, where $N\dim(E)$ is the memory required by the record of the set of particles, N is the memory required by the record of the set of weights, and n is the memory required by the record of the normalizing constants $\hat{\eta}_p^N$, $0 \le p \le n-1$. The estimator of the probability P_A is then

$$P_A^N = \left[\frac{1}{N}\sum_{i=1}^N \mathbf{1}_A(V(\hat{X}_n^{(i)}))\hat{Y}_n^{(i)}\right] \times \prod_{k=0}^{n-1} \hat{\eta}_p^N \tag{3.7}$$

This estimator is unbiased in the sense that $\mathbb{E}[P_A^N] = P_A$. The central limit theorem for the estimator states that

$$\sqrt{N}\left(P_A^N - P_A\right) \stackrel{N\to\infty}{\longrightarrow} \mathcal{N}(0, Q_A) \tag{3.8}$$

where the variance is

$$Q_A = \sum_{p=1}^{n} \mathbb{E}\left[\mathbb{E}_{X_p}[\mathbf{1}_A(V(X_n))]^2 \prod_{k=0}^{p-1} G_k^{-1}(X)\right] \mathbb{E}\left[\prod_{k=0}^{p-1} G_k(X)\right] - \mathbb{E}\left[\mathbf{1}_A(X_n)\right]^2 \tag{3.9}$$

Algorithm 2. Let us fix some $\alpha > 0$. The second algorithm is built with the weight function

$$G_p^\alpha(x) = \exp\left[\alpha(V(x_p) - V(x_{p-1}))\right] \tag{3.10}$$

`Initialization.` We start with a set of N i.i.d. initial conditions $\hat{X}_0^{(i)}$, $1 \le i \le N$, chosen according to the initial distribution X_0. This set is complemented with a set of parents $\hat{W}_0^{(i)} = x_0$, $1 \le i \le N$, where x_0 is an arbitrary point of E with $V(x_0) = V_0$. This forms a set of N particles: $(\hat{W}_0^{(i)}, \hat{X}_0^{(i)})$, $1 \le i \le N$, where a particle is a pair $(\hat{W}, \hat{X}) \in E \times E$.
Now, assume that we have a set of N particles at time p denoted by $(\hat{W}_p^{(i)}, \hat{X}_p^{(i)})$, $1 \le i \le N$.
`Selection.` We first compute the normalizing constant

$$\hat{\eta}_p^N = \frac{1}{N} \sum_{i=1}^{N} \exp\left[\alpha(V(\hat{X}_p^{(i)}) - V(\hat{W}_p^{(i)}))\right] \tag{3.11}$$

We choose independently N particles according to the empirical distribution

$$\mu_p^N(d\check{W}, d\check{X}) = \frac{1}{N\hat{\eta}_p^N} \sum_{i=1}^{N} \exp\left[\alpha(V(\hat{X}_p^{(i)}) - V(\hat{W}_p^{(i)}))\right] \delta_{(\hat{W}_p^{(i)}, \hat{X}_p^{(i)})}(d\check{W}, d\check{X}) \tag{3.12}$$

The new particles are denoted by $(\check{W}_p^{(i)}, \check{X}_p^{(i)})$, $1 \le i \le N$.
`Mutation.` For every $1 \le i \le N$, the particle $(\check{W}_p^{(i)}, \check{X}_p^{(i)})$ is transformed into $(\hat{W}_{p+1}^{(i)}, \hat{X}_{p+1}^{(i)})$ by the mutation procedure $\check{X}_p^{(i)} \stackrel{K_{p+1}}{\longrightarrow} \hat{X}_{p+1}^{(i)}$ where the mutations are performed independently, and $\hat{W}_{p+1}^{(i)} = \check{X}_p^{(i)}$.

The memory required by the algorithm is $2N\dim(E) + n$. The estimator of the probability P_A is then

$$P_A^N = \left[\frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_A(V(\hat{X}_n^{(i)})) \exp\left(\alpha(V(\hat{W}_n^{(i)}) - V_0)\right)\right] \times \left[\prod_{k=0}^{n-1} \hat{\eta}_p^N\right] \tag{3.13}$$

This estimator is unbiased and satisfies the central limit theorem (3.8).

Let us now focus our attention to the estimation of the pdf tail of $V(X_n)$. The rare event is then of the form $\{V(X_n) \in [a, a + \delta a)\}$ with a large a and an evanescent δa. We assume that the pdf of $V(X_n)$ is continuous so that the pdf can be seen as

$$p(a) = \lim_{\delta\to 0} \frac{1}{\delta a} \mathbb{P}(V(X_n) \in [a, a + \delta a))$$

We propose to use the estimator

$$p^N(a) = \frac{1}{\delta a} \times P^N_{[a,a+\delta a)} \tag{3.14}$$

with a small δa. The central limit theorem for the pdf estimator takes the form

$$\sqrt{N}\left(p^N(a) - p(a)\right) \stackrel{N\to\infty}{\longrightarrow} \mathcal{N}(0, p_2^2(a)) \tag{3.15}$$

where the variance is

$$p_2^2(a) = \lim_{\delta a \to 0} \frac{1}{\delta a} \mathbb{E}\left[\mathbf{1}_{[a,a+\delta a)}(V(X_n)) \prod_{k=0}^{n-1} G_k^{-1}(X)\right] \mathbb{E}\left[\prod_{k=0}^{n-1} G_k(X)\right] \tag{3.16}$$

Note that all other terms in the sum (3.9) are of order δa^2 and are therefore negligible. This is true as soon as the distribution of $V(X_n)$ given X_p for $p < n$ admits a density with respect to the Lebesgue measure. Accordingly, the variance $p_2^2(a)$ can be estimated by $(\delta a)^{-1} Q^N_{[a,a+\delta a)}$, where Q^N_A is given by

$$Q_A^N = \left[\frac{1}{N}\sum_{i=1}^N \mathbf{1}_A(V(\hat{X}_n^{(i)}))(\hat{Y}_n^{(i)})^2\right] \times \left[\prod_{k=0}^{n-1} \hat{\eta}_p^N\right]^2 \tag{3.17}$$

for the algorithm 1, and by

$$Q_A^N = \left[\frac{1}{N}\sum_{i=1}^N \mathbf{1}_A(V(\hat{X}_n^{(i)})) \exp\left(2\alpha(V(\hat{W}_n^{(i)}) - V_0)\right)\right] \times \left[\prod_{k=0}^{n-1} \hat{\eta}_p^N\right]^2 \tag{3.18}$$

for the algorithm 2. The estimators of the variances are important because confidence intervals can then be obtained.

4. A toy model. In this section we apply the IPS method to compute the probabilities of rare events for a very simple system for which we know explicit formulas. The system under consideration is the Gaussian random walk $X_{p+1} = X_p + W_{p+1}$, $X_0 = 0$, where the $(W_p)_{p=1,\dots,n}$ are i.i.d. Gaussian random variables with zero-mean and variance one. Let n be some positive integer. The goal is to compute the pdf of X_n, and in particular the tail corresponding to large positive values.

We choose the weight function

$$G_p^\alpha(x) = \exp[\alpha(x_p - x_{p-1})] \tag{4.1}$$

The theoretical pdf is such that

$$p(a)\delta a = \mathbb{E}\left[\mathbf{1}_{[a,a+\delta a)}(X_n)\right] + O(\delta a^2) \tag{4.2}$$

It is a Gaussian pdf with variance n

$$p(a) = \frac{1}{\sqrt{2\pi n}} \exp\left(-\frac{a^2}{2n}\right) \tag{4.3}$$

The theoretical variance of the pdf estimator is such that

$$p_2^2(a)\delta a = \mathbb{E}\left[\mathbf{1}_{[a,a+\delta a)}(X_n) \prod_{j=1}^{n-1} (G_p^\alpha)^{-1}(X)\right] \mathbb{E}\left[\prod_{j=1}^{n-1} G_p^\alpha(X)\right] + O(\delta a^2) \tag{4.4}$$

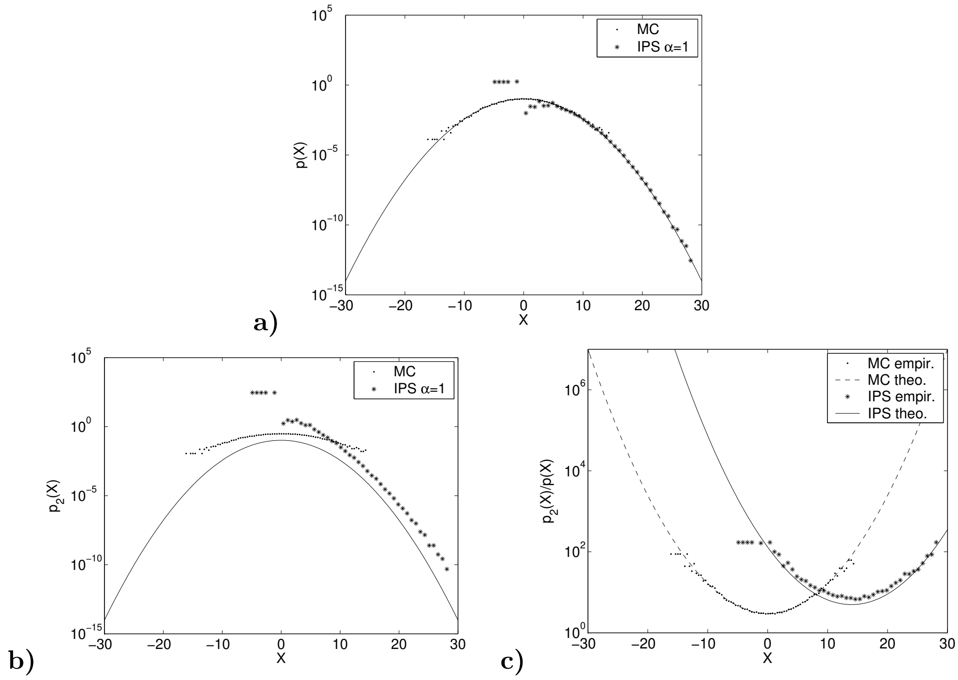

FIG. 4.1. *Picture a: Pdf estimations obtained by the usual MC technique (dots) and by the IPS with the weight function (4.1) with $\alpha = 1$ (stars). The solid line stands for the theoretical Gaussian distribution. Picture b: standard deviations of the estimators of the pdf. Picture c: empirical and theoretical ratios p_2/p.*

and it is given by

$$p_2^2(a) = p^2(a) \times \sqrt{2\pi n} \exp\left(\alpha^2 \frac{n-1}{n} + \frac{(a - \alpha(n-1))^2}{2n}\right) \tag{4.5}$$

When $\alpha = 0$, we have $p_2^2(a) = p(a)$, which is the result of standard MC. For $\alpha \neq 0$, the ratio $p_2(a)/p(a)$ is minimal when $a = \alpha(n-1)$ and then $p_2(a) \simeq p(a)\sqrt[4]{2\pi n}\exp\left(\alpha^2(n-1)/(2n)\right)$. This means that the IPS with some given α is especially relevant for estimating the pdf tail around $a = \alpha(n-1)$.

Let us assume that $n \gg 1$. Typically we look for the pdf tail for $a \simeq a_0\sqrt{n}$ with $a_0 > 1$ because $\sqrt{n}$ is the typical value of X_n. Thus we must take $\alpha = a_0/\sqrt{n}$ and then the relative error is $p_2(a)/p(a) \simeq \sqrt[4]{2\pi n}$.

In figure 4.1 we compare the results from MC simulations, IPS simulations and theoretical formulas with the weight function (4.1). We use a set of 2 10^4 particles to estimate the pdf tail of X_n with $n = 15$. The agreement shows that we can be confident with the results given by the IPS for predicting rare events with probabilities 10^{-12}.

We now choose the weight function

$$G_p^{\beta}(x) = \exp(\beta x_p) \tag{4.6}$$

We get the same results, but the explicit expression for the theoretical variance of the

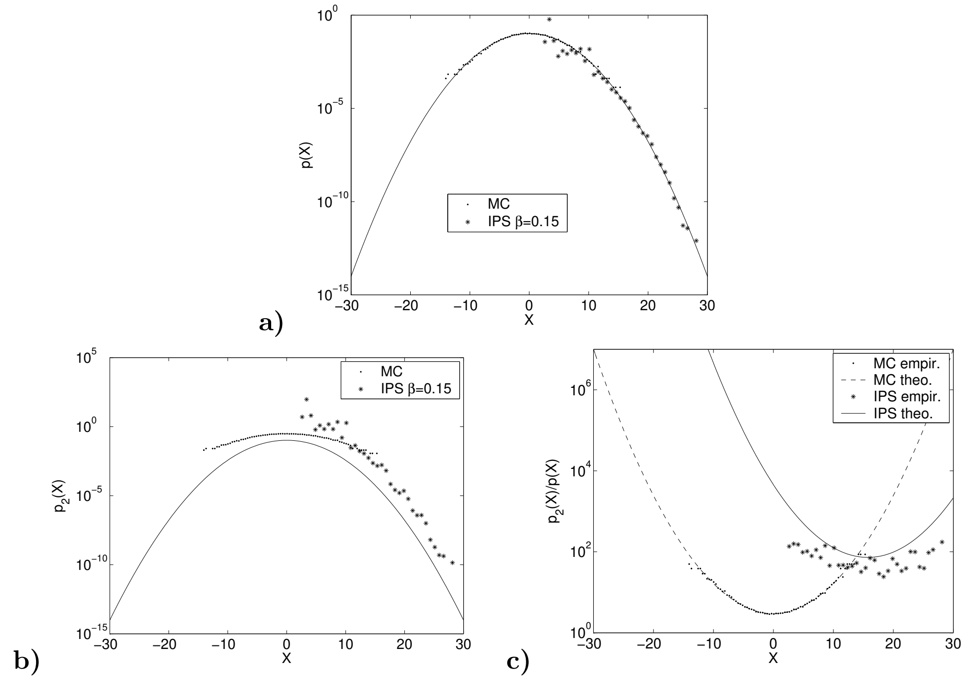

FIG. 4.2. *Picture a: Pdf estimations obtained by the usual MC technique (dots) and by the IPS with the weight function (4.6) with* $\beta = 0.15$ *(stars). The solid line stands for the theoretical Gaussian distribution. Picture b: standard deviations of the estimators of the pdf. Picture c: empirical and theoretical ratios* p_2/p.

pdf estimator is

$$p_2^2(a) = p^2(a) \times \sqrt{2\pi n} \exp\left(\beta^2 \frac{n(n^2-1)}{12} + \frac{\left(a - \frac{\beta n(n-1)}{2}\right)^2}{2n} \right) \tag{4.7}$$

When $\beta = 0$, we have $p_2^2(a) = p(a)$, which is the result of standard MC. For $\beta \neq 0$, the ratio $p_2(a)/p(a)$ is minimal when $a = \beta n(n-1)/2$ and then $p_2(a) = p(a)\sqrt[4]{2\pi n}\exp\left(\beta^2 n(n^2-1)/24\right)$. This means that the IPS with some given β is especially relevant for estimating the pdf tail around $a = \beta n(n-1)/2$.

Let us assume that $n \gg 1$. Typically we look for the pdf tail for $a \simeq a_0\sqrt{n}$ with $a_0 > 1$. Thus we must take $\beta = 2a_0/n^{3/2}$ and then the relative error is $p_2(a)/p(a) \simeq (2\pi n)^{1/4}\exp(a_0^2/6) = (2\pi n)^{-1/12} p(a)^{-1/3}$. The relative error is larger than the one we get with the weight function (4.1). In figure 4.2 we compare the results from MC simulations, IPS simulations and theoretical formulas with the weight function (4.6). This confirms that the weight function (4.6) is less efficient than (4.1).

5. Polarization mode dispersion in optical fibers.

5.1. Introduction. The study of pulse propagation in a fiber with random birefringence has become of great interest for telecommunication applications. Recent experiments have shown that Polarization Mode Dispersion (PMD) is one of the main limitations on fiber transmission links [1]. PMD has its origin in the birefringence [34], i.e. the fact that the electric field is a vector field and the index of refraction of the medium depends on the polarization state (i.e. the unit vector pointing in the direction of the electric vector field). For a fixed position in the fiber, there are two orthogonal polarization eigenstates which correspond to the maximum and the minimum of the index of refraction. These two polarization states are parameterized by an angle with respect to a fixed pair of axes that is called the birefringence angle. The difference between the maximum and the minimum of the index of refraction is the birefringence strength. If the birefringence angle and strength were constant along the fiber, then a pulse polarized along one of the eigenstates would travel at constant velocity. However the birefringence angle is randomly varying which involves coupling between the two polarized modes. The modes travel with different velocities, which involves pulse spreading. Random birefringence results from variations of the fiber parameters such as the core radius or geometry. There exist various physical reasons for the fluctuations of the fiber parameters. They may be induced by mechanical distortions on fibers in practical use, such as point-like pressures or twists [28]. They may also result from variations of ambient temperature or other environmental parameters [3].

The difficulty is that PMD is a random phenomenon. Designers want to ensure that some exceptional but very annoying event occurs only a very small fraction of time. This critical event corresponds to a pulse spreading beyond a threshold value. For example, a designer might require that such an event occurs less than 1 minute per year [5]. PMD in an optical fiber varies with time due to vibrations and variations of environmental parameters. The usual assumption is that the fiber passes ergodically through all possible realizations. Accordingly requiring that an event occurs a fraction of time p is equivalent to require that the probability of this event is p. The problem is then reduced to the estimation of the probability of a rare event. Typically the probability is 10^{-6} or less [5]. It is extremely difficult to use either laboratory experiments or MC simulations to obtain a reliable estimate of such a low probability because the number of configurations that must be explored is very large. Recently IS has been applied to numerical simulations of PMD [3]. This method gives good results, however it requires very good physical insight into the problem because it is necessary for the user to know how to produce artificially large pulse widths. We would like to revisit this work by applying the IPS strategy. The main advantage is that we do not need to specify how to produce artificially large pulse widths, as the IPS will automatically select those "particles" with large widths. As a byproduct, we shall also compute variances that allow us to give confidence intervals for our estimations of the outage probabilities, and we shall also be able to describe the typical cascade of elementary events giving rise to anomalously large pulse spreading.

5.2. Propagation of short pulses in optical fibers.

5.2.1. PMD driven by random birefringence. The evolution of polarized fields in randomly birefringent fibers is governed by the coupled Schrödinger equations with random PMD between two modes (polarizations) [24]

$$i\mathbf{A}_z + K_0\mathbf{A} + iK_1\mathbf{A}_t = 0 \tag{5.1}$$

where subscripts stand for partial differentiation with respect to corresponding variables and $\mathbf{A} = (A_x, A_y)^T$ is the column vector that denotes the envelopes of the electric field in the two eigenmodes. The z-dependent 2×2 matrices K_0 and K_1 describe random fiber birefringence. We can eliminate the fast random birefringence variations that appear in Eq. (5.1) by means of a change of variables, that leads to the new vector equation

$$iU_z = iRU_t \tag{5.2}$$

where $\mathbf{U} \equiv M^{-1}\mathbf{A}$, $\mathbf{U} = (u, v)^T$ represents the slow evolution of the field envelopes in the reference frame of the local polarization eigenmodes, and the matrix M obeys the equation $iM_z + K_0 M = 0$. $R = -M^{-1}K_1 M$ is a z-dependent matrix that involves high order PMD. In absence of losses M is unitary and R is a combination of three Pauli matrices

$$R(z) = m_1(z)\Sigma_1 + m_2(z)\Sigma_2 + m_3(z)\Sigma_3, \tag{5.3}$$

where

$$\Sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \Sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and m_j are real-valued random processes. R is associated with the coupling between the modes, as well as an accumulation of a phase mismatch. The important quantity is the pulse width defined by

$$T_w^2(z) = \frac{\int t^2(|u|^2 + |v|^2)(t,z)dt}{\int (|u|^2 + |v|^2)(t,z)dt} \tag{5.4}$$

Note that the propagation equation preserves the energy (i.e. the L^2-norm) of the optical pulse, so that T_w can be interpreted as the root mean square of the energy distribution.

5.2.2. The Stokes vector. The Fourier components $\hat{\mathbf{U}} := (\hat{u}, \hat{v})^T$ of the field defined by

$$\hat{u}(\omega, z) = \int u(t,z)\exp(i\omega t)dt, \qquad \hat{v}(\omega, z) = \int v(t,z)\exp(i\omega t)dt,$$

obey a system of ordinary differential equations:

$$\hat{\mathbf{U}}_z = i\omega R(z)\hat{\mathbf{U}} \tag{5.5}$$

There exist simple and exact analytical identities between the amount of broadening and Fourier components [4, 20]. These formulas are in fact nothing more than the usual Parseval formula applied to well chosen quantities. We first define the spectral intensity

$$\hat{E}_0(\omega) = |\hat{u}|^2(\omega, z) + |\hat{v}|^2(\omega, z) \tag{5.6}$$

which is a preserved quantity imposed by the initial condition $(u_0, v_0)^T$ at $z = 0$. A convenient representation of the polarization evolution can be obtained in terms of the Stokes vector $\hat{\mathbf{s}}(\omega, z)$ associated to the Fourier components of the field

$$\begin{aligned} \hat{s}_1 &= \left(|\hat{u}|^2 - |\hat{v}|^2\right) \,/\, \hat{E}_0, \\ \hat{s}_2 &= 2\mathrm{Re}\,(\hat{u}^*\hat{v})\,(\omega) \,/\, \hat{E}_0, \\ \hat{s}_3 &= 2\mathrm{Im}\,(\hat{u}^*\hat{v})\,(\omega) \,/\, \hat{E}_0, \end{aligned} \tag{5.7}$$

whose modulus $\sqrt{\hat{s}_1^2+\hat{s}_2^2+\hat{s}_3^2} = (|\hat{u}|^2+|\hat{v}|^2)/\hat{E}_0$ is 1. The Stokes vector thus belongs to the so-called Poincaré sphere. In terms of the Stokes vector the dynamics driven by PMD is simple

$$\hat{\mathbf{s}}_z = \omega\Omega(z)\times\hat{\mathbf{s}} \tag{5.8}$$

where $\Omega(z)$ is the column vector $2(m_3,m_1,m_2)^T(z)$. Thus the m_j appear as elementary infinitesimal generators of random rotations of the Stokes vector over the Poincaré sphere.

5.2.3. Pulse width and PMD vector. In terms of the Fourier components the pulse width T_w reads

$$T_w^2(z) = \frac{\int \hat{R}(\omega,z)\hat{E}_0(\omega)d\omega}{\int \hat{E}_0(\omega)d\omega}, \text{ with } \hat{R} := \frac{|\hat{u}'|^2+|\hat{v}'|^2}{|\hat{u}|^2+|\hat{v}|^2}. \tag{5.9}$$

Here the primes stand for partial derivatives with respect to the pulsation ω. The process $\hat{R}(\omega,z)$ obeys the differential equation $\hat{R}_z = \frac{1}{2}\Omega\cdot\hat{\mathbf{r}}$ where the vector $\hat{\mathbf{r}}(\omega,z)$

$$\begin{aligned}
\hat{r}_1 &= 2\mathrm{Im}\,(\hat{u}'\hat{u}^* - \hat{v}'\hat{v}^*)\ /\ \hat{E}_0\\
\hat{r}_2 &= 2\mathrm{Im}\,(\hat{u}'\hat{v}^* + \hat{v}'\hat{u}^*)\ /\ \hat{E}_0\\
\hat{r}_3 &= 2\mathrm{Re}\,(\hat{u}'\hat{v}^* - \hat{v}'\hat{u}^*)\ /\ \hat{E}_0
\end{aligned} \tag{5.10}$$

is solution of

$$\hat{\mathbf{r}}_z = \omega\Omega(z)\times\hat{\mathbf{r}} + \Omega(z) \tag{5.11}$$

The vector $\hat{\mathbf{r}}$ is the so-called PMD vector. Let us denote $\tau(\omega,z) = 4(\hat{R}(\omega,z)-\hat{R}(\omega,0))$. Differentiating $|\hat{\mathbf{r}}(\omega,z)|^2 - \tau(\omega,z)$ with respect to z establishes that it is constant. $\tau = |\hat{\mathbf{r}}|^2$ is the so-called square Differential Group Delay (DGD), it characterizes the pulse spreading.

5.3. Review of PMD models.

5.3.1. The white noise model. Simplified analytical models have been studied. In the standard model [27, 34, 16, 21] it is assumed that the real-valued processes m_j are random white noises with autocorrelation function $\mathbb{E}[m_i(z')m_j(z)] = \sigma^2\delta_{ij}\delta(z'-z)$. In such a case the differential equation (5.11) must be interpreted as a stochastic differential equation

$$d\hat{r}_1 = \sigma\omega\hat{r}_3\circ dW_z^2 - \sigma\omega\hat{r}_2\circ dW_z^3 + \sigma dW_z^1 \tag{5.12}$$
$$d\hat{r}_2 = \sigma\omega\hat{r}_1\circ dW_z^3 - \sigma\omega\hat{r}_3\circ dW_z^1 + \sigma dW_z^2 \tag{5.13}$$
$$d\hat{r}_3 = \sigma\omega\hat{r}_2\circ dW_z^1 - \sigma\omega\hat{r}_1\circ dW_z^2 + \sigma dW_z^3 \tag{5.14}$$

where $\circ$ stands for the Stratonovich integral. It is then easy to establish [15] that the DGD τ is a diffusion process with infinitesimal generator

$$\mathcal{L} = 8\sigma^2\tau\frac{\partial^2}{\partial\tau^2} + 12\sigma^2\frac{\partial}{\partial\tau} \tag{5.15}$$

which implies that $\tau(\omega,z)$ obeys a χ^2 distribution with three degrees of freedom also known as Maxwellian distribution. In other words the pdf of $\tau(\omega,z)$ is:

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

5.3.2. Realistic models. The white noise model gives an analytical formula for the pdf of the DGD, which in turns allows us to compute exactly the probability that the DGD exceeds a given threshold value. However it has been pointed out that the pdf tail of the DGD does not fit with the Maxwellian distribution in realistic configurations [2]. Various numerical and experimental PMD generation techniques involve the concatenation of birefringent elements with piecewise constant vectors Ω [23]. Eq. (5.11) can be solved over each segment, and continuity conditions on the segments junctions give a discrete model for the PMD vector $\hat{\mathbf{r}}$. The total PMD vector after the $n+1$-th section can then be obtained from the concatenation equation [17]

$$\hat{\mathbf{r}}_{n+1} = R_{n+1}\hat{\mathbf{r}}_n + \sigma\Omega_{n+1} \tag{5.16}$$

where σ is the DGD per section. Assuming linearly birefringent fibers, Ω_n lies in the equatorial plane of the Poincaré sphere [14]: $\Omega_n = \Omega(\theta_n)$ with

$$\Omega(\theta) = (\cos(\theta), \sin(\theta), 0)^T$$

R_n is a matrix corresponding to a rotation through an angle ϕ_n about the axis Ω_n [17]. Explicitly $R_n = R(\theta_n, \phi_n)$ with

$$R(\theta,\phi) = \begin{pmatrix} \cos^2(\theta)+\sin^2(\theta)\cos(\phi) & \sin(\theta)\cos(\theta)(1-\cos(\phi)) & \sin(\theta)\sin(\phi) \\ \sin(\theta)\cos(\theta)(1-\cos(\phi)) & \sin^2(\theta)+\cos^2(\theta)\cos(\phi) & -\cos(\theta)\sin(\phi) \\ -\sin(\theta)\sin(\phi) & \cos(\theta)\sin(\phi) & \cos(\phi) \end{pmatrix}$$

From the probabilistic point of view, the angles ϕ_n are i.i.d. random variables uniformly distributed in $(0, 2\pi)$. The angles θ_n are i.i.d. random variables such that $\cos(\theta_n)$ are uniformly distributed in $(-1, 1)$ [3]. Accordingly, $(\hat{\mathbf{r}}_n)_{n\in\mathbb{N}}$ is a Markov chain. Let us assume that the fiber is modeled as the concatenation of n segments and that the outage event is of the form $|\hat{\mathbf{r}}_n| > a$ for some fixed threshold value a. In the case where a is much larger than the expected value of the final DGD $|\hat{\mathbf{r}}_n|$, the outage probability is very small, and this is the quantity that we want to estimate.

5.4. Estimations of outage probabilities.

5.4.1. Importance sampling. In Ref. [3] IS is used to accurately calculate outage probabilities due to PMD. The outage event can be represented as a set A of particular realizations of the random process $x = (x_p)_{1\le p\le n} = (\theta_p, \phi_p)_{1\le p\le n}$, whose probability is denoted by P_A. The idea is to bias the distribution of x so as to cause large DGD events to occur more frequently. Let us denote by p the standard distribution of x (a direct product of uniform distributions as described in Subsection 5.3.2) and by p^* a twisted distribution chosen by the user. We can carry out a set of N MC simulations with the twisted distribution p^* generating a set of i.i.d. $(x^{(i)})_{1\le i\le N}$. An estimator of the probability P_A is

$$P_A^N = \frac{1}{N}\sum_{i=1}^N \mathbf{1}_A(x^{(i)})\frac{p(x^{(i)})}{p^*(x^{(i)})}$$

The key difficulty in applying IS is to properly choose p^*. The papers [3, 22, 13] present different twisted distributions and the physical explanations why such distributions are likely to produce large DGDs. As a result the authors obtain with 10^5 MC simulations good approximations of the pdf tail even for probabilities of the order 10^{-12}. The main reported physical result is that the probability tail is significantly smaller than the Maxwellian tail predicted by the white noise model.

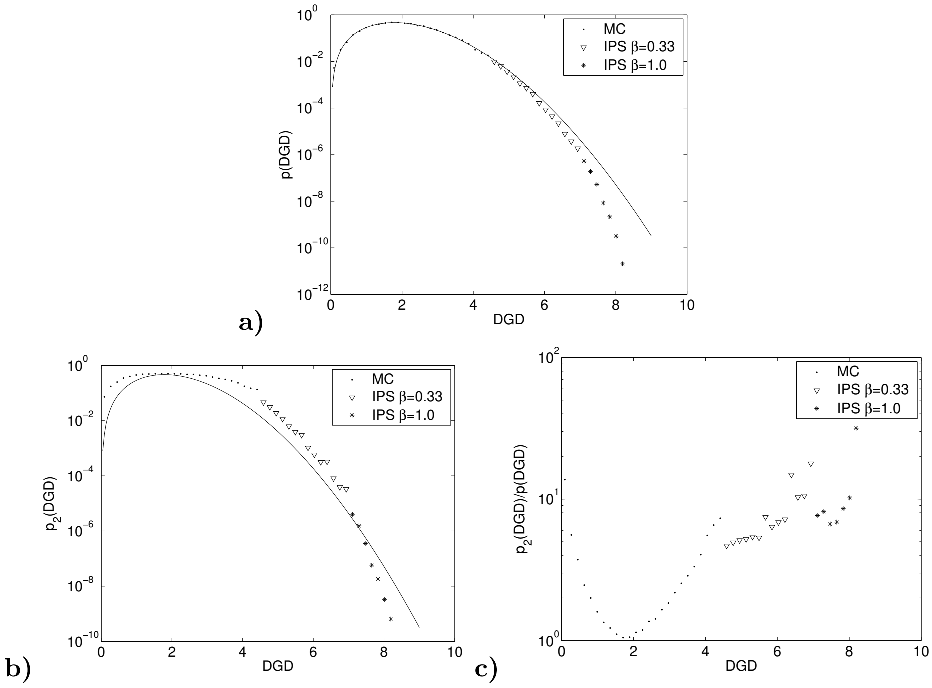

FIG. 5.1. *Picture a: Segments of the DGD pdf obtained by the usual MC technique (dots) and by the IPS with the weight function G_n^β with $\beta = 0.33$ (triangles) and $\beta = 1$ (stars). The solid line stands for the Maxwellian distribution obtained in the white noise model. The Maxwellian distribution fails to describe accurately the pdf tail. Picture b: Standard deviations p_2 of the estimators of the DGD pdf. In the MC case, the standard deviation is $p_2(a) = [p(a)(1-p(a))]^{1/2}$. In the IPS cases, the standard deviations are estimated via the formula (3.17). Picture c: Ratios p_2/p.*

5.4.2. Interacting particle systems. In this subsection we apply our IPS method and compare the results with those obtained by MC and IS. To get a reliable estimate of the outage probability of the event, it is necessary to generate realizations producing large DGDs. The main advantage of the IPS approach is that is proposes a "blink" method that does not require any physical insight. Such a method could thus be generalized to more complicated situations. Here the Markov process is the PMD vector $(\hat{\mathbf{r}}_n)_{n\in\mathbb{N}}$ at the output of the n-th fiber section. The state space is $\mathbb{R}^3$, the initial PMD vector is $\hat{\mathbf{r}}_0 = (1,0,0)^T$, the Markov transitions are described by Eq. (5.16), and the energy-like function is $V(\hat{\mathbf{r}}) = |\hat{\mathbf{r}}|$. We estimate the pdf $p(a)$ of $|\hat{\mathbf{r}}_n|$ by implementing the IPS with the two weight functions

$$G_p^\beta(\hat{\mathbf{r}}) = \exp\left(\beta|\hat{\mathbf{r}}_p|\right) \tag{5.17}$$

parameterized by $\beta \geq 0$, and

$$G_p^\alpha(\hat{\mathbf{r}}) = \exp\left[\alpha(|\hat{\mathbf{r}}_p| - |\hat{\mathbf{r}}_{p-1}|)\right] \tag{5.18}$$

parameterized by $\alpha \geq 0$. We have implemented algorithms 1 and 2 as described in Section 3.

In figure 5.1a we plot the estimation of the DGD pdf obtained by the IPS method with the weight function G_n^β defined by (5.17). The fiber consists in the concatenation of $n = 15$ segments. The DGD per section is $\sigma = 0.5$. We use a set of $N =$

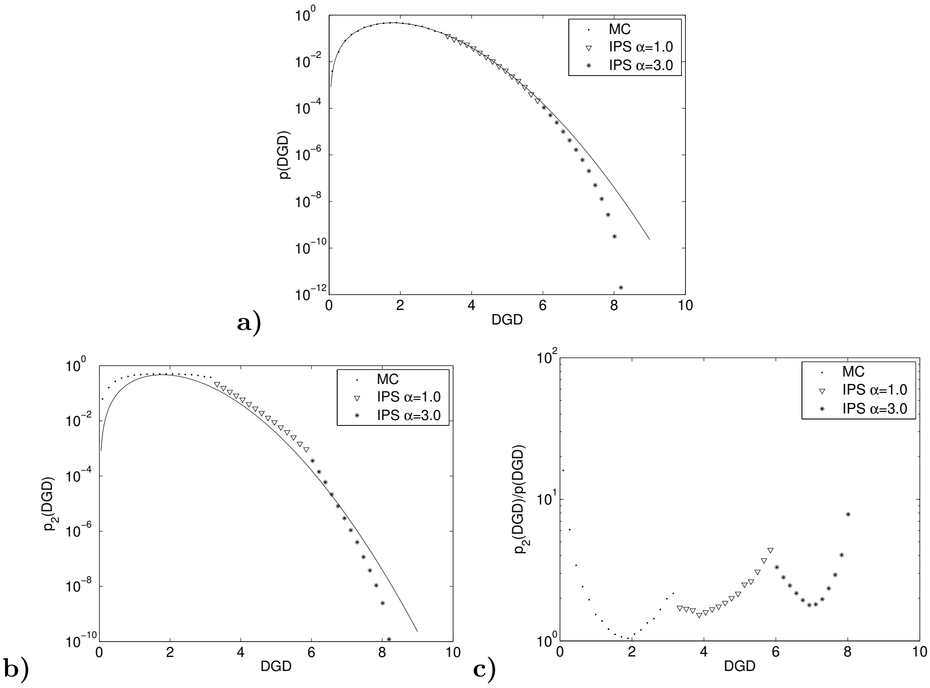

Fig. 5.2. *Picture a: Segments of the DGD pdf obtained by the usual MC technique (dots) and by the IPS with the weight function* G_p^α *with* $\alpha = 1.0$ *(triangles) and* $\alpha = 3.0$ *(stars). The solid line stands for the Maxwellian distribution obtained in the white noise model. Picture b: Standard deviations* p_2 *of the estimators of the DGD pdf. In the MC case, the standard deviation is* $p_2(a) = [p(a)(1-p(a))]^{1/2}$*. In the IPS cases, the standard deviations are estimated via the formula (3.18). Picture c: Ratios* p_2/p*.*

2 10^4 interacting particles. This result can be compared with the one obtained in Ref. [3], which shows excellent agreement. The difference is that our procedure is fully adaptative and does not require any guess from the user. The variance p_2^2 of the estimator of the DGD pdf is plotted in Figure 5.1b. This figure is actually used to determine the best estimator of the DGD pdf. Indeed the IPS and the corresponding estimator depends on the parameter β. We have actually simulated three sets of particle systems, the first one being the usual MC method, the two other ones being IPSs with two different parameters β. For each set of particle systems we have computed the empirical variances $p_2^2(a)$, and we have detected for each value of a which set gives rise to the smallest variance. Then we report in figure 5.1b-c the value of this variance, and in figure 5.1a we report the estimation $p(a)$ obtained with the corresponding particle system.

In figure 5.2a we plot the estimation of the DGD pdf obtained by the IPS method with the weight function G_p^α defined by (5.18). It turns out that the estimated variance of the estimator is smaller with the weight function G_p^α than with the weight function G_p^β (compare figures 5.2c and 5.1c). This observation confirms the theoretical predictions derived from the toy model in Section 4.

The IPS approach is also powerful to compute conditional probabilities or expectations given the occurrence of some rare event. For instance, we can be interested in the moments of the intermediate DGDs given that the final DGD lies in the rare set

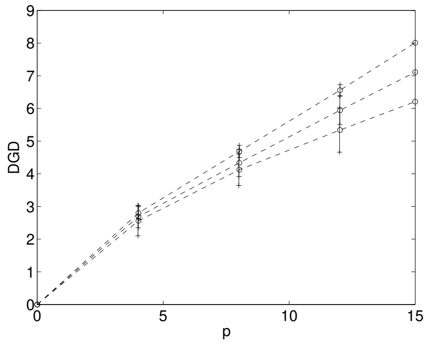

FIG. 5.3. *Conditional expectations $D^1_{a,a+\delta a}(p)$ of the intermediate DGD at $p = 4$, 8, 12, given that the final DGD lies in the interval $(a, a+\delta)$ with $\delta a = 0.18$ and (from top to bottom) $a = 8$, $a = 7.1$, $a = 6.2$. The error bars are obtained from the estimations of the conditional variances.*

$(a, a + \delta a)$:

$$D^q_a(p) = \mathbb{E}\left[|\hat{\mathbf{r}}_p|^q \ \mid\ |\hat{\mathbf{r}}_n| \in [a, a+\delta a)\right]$$

This information gives us the typical behaviors of the PMD vectors along the fiber that give rise to a large final DGD. We use the estimator (2.19) based on the IPS with the weight function (5.18). As shown by figure 5.3, the typical conditional trajectory of the DGD is a linear increase with a constant rate given by the ratio of the final DGD over the length of the fiber. The conditional variances are found to be small, which shows that fluctuations are relatively small around this average behavior.

REFERENCES

[1] B. Bahkshi, J. Handryd, P. A. Andrekson, J. Brentel, E. Kollveit, B. E. Olsson, and M. Karlsson, Experimental observation of soliton robustness to polarization dispersion pulse broadening, Electron. Lett. **35**, 65–66 (1999).

[2] G. Biondini, W. L. Kath, and C. R. Menyuk, Non-Maxwellian DGD distribution of PMD emulators, in Proc. OFC 2001, Vol. ThA5, 2001, pp. 1-3.

[3] G. Biondini, W. L. Kath, and C. R. Menyuk, Importance sampling for polarization-mode dispersion, IEEE Photon. Technol. Lett. **14**, 310–312 (2002).

[4] F. Bruyere, Impact of first- and second-order PMD in optical digital transmission systems, Opt. Fiber Technol. **2**, 269–280 (1996).

[5] H. Bülow, Limitation of optical first-order PMD compensation, in Proc. ECOC 1999, Vol. WE1, 1999, pp. 74-76.

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

[7] P. Del Moral and A. Guionnet, A central limit theorem for nonlinear filtering using interacting particle systems, Ann. Appl. Probab. **9**, 275–297 (1999).

[8] P. Del Moral and J. Jacod, Interacting particle filtering with discrete observations, In N. J. Gordon, A. Doucet, and J. F. G. de Freitas, editors, *Sequential Monte-Carlo Methods in Practice*, Springer-Verlag, New York, 2001.

[9] P. Del Moral and J. Jacod, The Monte-Carlo method for filtering with discrete-time observations: Central limit theorems, In *Numerical Methods and stochastics* (Toronto, ON, 1999), volume 34 of Fields Inst. Commun., pages 29–53, American Mathematical Society, Providence, RI, 2002.

[10] P. Del Moral and M. Ledoux, On the Convergence and the Applications of Empirical Processes for Interacting Particle Systems and Nonlinear Filtering, Journal of Theoret. Probability **13**, 225–257 (2000).

[11] P. Del Moral and S. Tindel, A Berry-Esseen theorem for Feynman-Kac and interacting particle models, Publications de l'Institut Elie Cartan, no. 44, Univ. Henri Poincaré, Nancy, 2003.

[12] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Pratice,* Statistics for engineering and Information Science, Springer, New York, 2001.
[13] S. L. Fogal, G. Biondini, and W. L. Kath, Multiple importance sampling for first- and second-order polarization-mode dispersion, IEEE Photon. Technol. Lett. **14**, 1273–1275 (2002).
[14] A. Galtarossa, G. Gianello, C. G. Someda, and M. Schiano, In-field comparison among polarization mode dispersion measurement techniques, J. Lightwave Technol. **14**, 42–49 (1996).
[15] J. Garnier, J. Fatome, and G. Le Meur, Statistical analysis of pulse propagation driven by polarization-mode dispersion, J. Opt. Soc. Am. B **19**, 1968–1977 (2002).
[16] N. Gisin, J. P. Pellaux, and J. P. Von der Weid, Polarization mode dispersion for short and long single-mode fibers, J. Lightwave Technol. **9**, 821–827 (1991).
[17] J. P. Gordon and H. Kogelnik, PMD fundamentals: Polarization-mode dispersion in optical fibers, in Proc. Nat. Acad. Sci. **97**, 4541 (2000).
[18] T. E. Harris and H. Kahn, Estimation of particle transmission by random sampling, Natl. Bur. Stand. Appl. Math. Ser. **12**, 27–30 (1951).
[19] P. Heidelberg, Fast simulation of rare events in queueing and reliability models, ACM Transactions on Modeling and Computer Simulation **5**, 43–85 (1995).
[20] M. Karlsson, Polarization mode dispersion-induced pulse broadening in optical fibers, Opt. Lett. **23**, 688–690 (1998).
[21] T .I. Lakoba and D. J. Kaup, Perturbation theory for the Manakov soliton and its application to pulse propagation in randomly birefringent fibers, Phys. Rev. E **56**, 6147–6165 (1997).
[22] I. T. Lima, Jr., G. Biondini, B. S. Marks, W. L. Kath, and C. R. Menyuk, Analysis of PMD compensators with fixed DGD using importance sampling, IEEE Photon. Technol. Lett. **14**, 627–629 (2002).
[23] D. Marcuse, C. R. Menyuk, and P. K. A. Wai, Application of the Manakov-PMD equation to studies of signal propagation in optical fibers with randomly varying birefringence, J. Lightwave Technol. **15**, 1735–1746 (1997).
[24] C. R. Menyuk, Pulse propagation in an elliptically birefringent Kerr medium, IEEE J. Quantum Electron. **25**, 2674–2682 (1989).
[25] R. O. Moore, G. Biondini, and W. L. Kath, Importance sampling for noise-induced amplitude and timing jitter in soliton transmission system, Opt. Lett. **28**, 105–107 (2003).
[26] M. Nagasawa, Stochastic Processes in Quantum Physics, Monographs in Mathematics, Vol. 94, Birkhäuser-Verlag, Boston, 1991.
[27] C. D. Poole and R. E. Wagner, Phenomenological approach to polarization dispersion in long single-mode fibers, Electron. Lett. **22**, 1029–1030 (1986).
[28] S. C. Rashleigh, Origins and control of polarization effects in single-mode fibers, J. Lightwave Technol. **1**, 312–331 (1983).
[29] M. Reed and B. Simon, *Methods of Modern Mathematical Physics, II, Fourier Analysis, Self Adjointness,* Academic Press, New York, 1975.
[30] M. N. Rosenbluth and A. W. Rosenbluth, Monte-carlo calculations of the average extension of macromolecular chains, J. Chem. Phys. **23** 356–359 (1955).
[31] R. Y. Rubinstein, *Simulation and the Monte Carlo method,* Wiley, New York, 1981.
[32] G. R. Shorack, *Probability for Statisticians,* Springer Texts in Statistics, Springer-Verlag, New York, 2000.
[33] A. S. Sznitman. *Brownian Motion Obstacles and Random Media,* Springer-Verlag, Monographs in Mathematics, New York, 1998.
[34] P. K. A. Wai and C. R. Menyuk, Polarization mode dispersion, decorrelation, and diffusion in optical fibers with randomly varying birefringence, J. Lightwave Technol. **14**, 148–157 (1996).