PARTICLE METHODS FOR THE ESTIMATION OF CREDIT PORTFOLIOS LOSS DISTRIBUTION

RENÉ CARMONA AND STÉPHANE CRÉPEY

Abstract. The goal of the paper is the numerical analysis of the performance of Monte Carlo simulation based methods for the computation of credit-portfolio loss-distributions in the context of Markovian intensity models of credit risk. We concentrate on two of the most frequently touted methods of variance reduction in the case of stochastic processes: importance sampling (IS) and interacting particle systems (IPS) based algorithms. Because the subtle differences between these methods are often misunderstood, as IPS is often regarded as a mere particular case of IP, we describe in detail the two kinds of algorithms, and we highlight their fundamental differences. We then proceed to a detailed comparative case study based on benchmark numerical experiments chosen for their popularity in the quantitative finance circles.

(RC) Bendheim Center for Finance
ORFE, Princeton University
Princeton, NJ 08544 USA
(SC) Département de Mathématiques
Université d'Évry Val d'Essonne
91025 Évry Cedex, France

Contents

Date: February 2009 (This Version).
Key words and phrases. Importance Sampling, Interacting Particle Systems, Rare Events, Credit Portfolios, Loss Distribution Estimation.
The first named author (RC) was partially supported by NSF-FRG-DMS-0455982 while the research of the second named author (SC) benefited from the support of the Europlace Institute of Finance and the "Chair Risque de Crédit" (Fédération Bancaire Française).

1. **Introduction**

The purpose of this paper is to study the performance of Monte Carlo methods for the computation of credit portfolio loss distributions, in the context of Markovian intensity models of credit risk. For simplicity, we assume homogeneity of individual losses given default, so that the portfolio loss can be identified with the number of defaults in the portfolio. Since most Monte Carlo methods can produce equally sharp estimates of the main parts of these distributions, they need to be tested and compared on their ability to estimate accurately the *rare events* which typically requires variance reduction techniques. Denoting by L_T the cumulative loss (the number of defaults according to our homogeneity assumption) at time T of a credit portfolio, we are interested in the computation of the probabilities $p_\ell(T) = \mathbb{P}\{L_T = \ell\}$ for ℓ in $\{0, \cdots, n\}$ where n is the number of names of the portfolio. The naive Monte Carlo estimate

$$p_\ell^m(T) = \frac{\#\{j;\ 0 \le j \le m,\ L_T(\omega_j) = \ell\}}{m} \tag{1}$$

computed from a set of m Monte Carlo samples $\omega_1, \cdots, \omega_m$, is given by the proportion of samples giving exactly ℓ losses. It suffers from several shortcomings. First, it is very unlikely that the Monte Carlo samples will reach the loss level ℓ when $\{L_T = \ell\}$ is a rare event. Furthermore, the variance of the estimator, which is equal to

$$\sigma_\ell^{2,m}(T) = \frac{1}{m} p_\ell(T)(1 - p_\ell(T))$$

can be quite large.

In the *factor copulae models* most commonly used in practice, the conditional loss distribution can be recovered analytically by various means, and the unconditional distribution can then be obtained by quadrature (numerical integration). As for

Monte Carlo calculations, generating a direct sample of the unconditional distribution is typically too slow. However it is possible to apply various variance reduction techniques to the simulation of the conditional distribution, possibly in conjunction with a shift on the common factor (see for example Glasserman et al. [7, 9, 8]).
Note that for the specific purpose of valuing credit portfolio loss derivatives by simulation, various direct (unconditional) variance reduction techniques are available, among them, the use of the portfolio loss at maturity as control variable.

In this paper we consider *Markovian intensity models* of credit risk, such as those used for pricing purposes in Frey and Backhaus [5], Bielecki et al. [2], or Herbertsson [10]. Financial applications can be divided into two categories depending on whether one is interested in pricing or risk management. Accordingly, probability models are organized into risk neutral and statistical models. Some of the interesting features of Markovian intensity models of credit risk are:
• to provide a consistent pricing *and hedging* theory, as opposed to the abovementioned static models which are essentially pricing tools (even if they can be used for hedging in an ad-hoc manner);
• to account for contagion effects which play an important role in pricing, especially since the start of the sub-prime crisis.
We shall see in section 5 the potential interest of IPS with respect to IS in relation to this second bullet point.

But pricing is not the only application of these models. Indeed, they can also be used for the purpose of the thorny risk management issue highlighted by the credit crisis. Note that the probabilities of the rare events of interest are significantly smaller under the statistical measure than under the risk-neutral probability for which contagion effects are most prevalent. Consequently methods such as those presented in this paper are especially important in that context.

In Markovian models, one should be able to compute the loss distribution by numerical resolution of the forward Kolmogorov equations. However practical implementation of deterministic numerical schemes is precluded by the curse of dimensionality for models of dimension greater than a few units. Simulation approaches appear as the only reasonable alternative. Unfortunately, simulation methods are typically slow, and their accuracy depends strongly on the control of the variances of the estimates (see, e.g., Glasserman [6], for a general reference on these issues).

Importance sampling (IS for short) is often regarded as the method of choice when it comes to variance reduction. In this paper we show how explicit changes of measures can be implemented in Markovian models of portfolio credit risk, and we document the efficiency of these IS estimators. However, we argue that IS does not always make sense. Sometimes, it is not clear which change of measure will significantly reduce the variance, and moreover, producing Monte Carlo random samples after the change of measure can be impractical if at all possible.

We thus present an alternative variance reduction method based on the properties of twisted Feynman-Kac expectations and the approximation of their values by interacting particle systems (IPS for short), also called in this paper *implicit importance sampling* as opposed to the previous *explicit importance sampling*. The use of IPS

for the computation of the probabilities of rare events was introduced by Del Moral and Garnier in [14]. We refer the reader to Del Moral's account [13] for the general theory in a textbook form.

An important feature of IPS estimates is that they can be computed by generating Monte Carlo samples *from the original model distribution* $\mathbb{P}$ rather than from the changed measure as in the case of importance sampling. In many cases, this represents an important advantage of IPS over IS. Indeed, in many practical applications, large programs are used to generate Monte Carlo samples. These programs have been developed based on complex mathematical models and extensive historical data bases (see for example the Standard & Poor's documentation on CPDOs [15]). The level of intricacy of their implementation files precludes any change of measure and the generation of Monte Carlo samples from a different distribution. They need to be used *as is*, as a *black box*. IPS algorithms can be used without knowing what is in the black boxes. This is a major advantage of IPS over IS in this realm of applications.

Remark. A first application of IPS to the computation of credit portfolio losses and CDO tranche spreads was considered by Carmona, Fouque and Vestal in [3] in the framework of *structural models* of credit risk. In the present paper we consider the application of IS and IPS to the computation of a credit portfolio loss distributions, in a Markovian *intensity model*.

2. **Importance Sampling and Interacting Particle Systems**

2.1. **Point Process Set-Up.** Our goal is to compute the marginal probability distributions of a (continuous time) point process which we denote by $\{L_t\}_{t\geq 0}$ in analogy with the notation used for the loss process (number of defaults) of a credit portfolio in the introduction. We assume that at each time t, the value L_t of the point process can be derived from the values of a finite set of factors, say $x = (x^1, \cdots, x^d)$ whose time evolution can be described in discrete time by a Markov chain. In other words, we are interested in the evaluation of the probabilities of events relative to a (possibly time inhomogeneous) Markov chain $\{X_n\}_n$, the background factor X_n being a random element of a general measurable space $(E_n, \mathcal{E}_n)$ which can change with n.

We denote by $K_n(x_{n-1}, dx_n)$ the transition kernel of the background Markov chain at time n, and we denote by $\{Y_n\}_n$ the historical process of $\{X_n\}_n$ defined by:

$$Y_n = (X_0, \cdots, X_n) \in F_n = E_0 \times \cdots \times E_n.$$

Next, we let $M_n(y_{n-1}, dy_n)$ denote the Markov transition kernels of the inhomogeneous Markov chain $\{Y_n\}_n$. Finally, for each integer $n \geq 0$ we consider non-negative measurable functions w_n on F_n equipped with the product σ-field, and we interpret these functions as weight functions.

Note that for IPS it is assumed in [14, 13] that these weight functions are bounded and bounded away from zero in the sense that:

$$\sup_{(y_n, y'_n)\in F_n\times F_n} \frac{w_n(y_n)}{w_n(y'_n)} < \infty. \tag{2}$$

This restrictive boundedness assumption can be relaxed in many cases. However, we only mention it for the sake of completeness since we are not interested in mathematical proofs of convergence results. Indeed, when in need of convergence results for the IPS algorithm, we directly appeal to [14].

Both IS and IPS computations rely on changes from the original probability measure to absolutely continuous *twisted* measures, in order to compute quantities of the form

$$\mathbb{E}\{f_n(Y_n)\}$$

for suitable functions f_n (like, for instance, the indicator function of some "rare" set of interest). Yet, despite this similarity, the two methods differ in very significant ways.

2.2. **Importance Sampling for Markov Chains.** Importance Sampling is especially useful when not enough Monte Carlo samples from the original distribution contribute significantly to the computation of the desired probabilities or expectations. In this case, typical estimates are unstable and carry higher variance than desired. A simple fix could be to use random samples from a probability distribution obtained by deformation of the original measure in such a way that more Monte Carlo samples contribute significantly, and to correct the naive estimator (1) for the fact that the samples were generated from a different distribution.

Importance Sampling is well known in the context of random variables, or (see Appendix A) in the context of diffusions. Let us specialize the description of the method to the case of interest in this paper. Given a set of weight functions w_n as before, for each fixed n, we let $\widetilde{\mathbb{P}}_n$ stand for the *twisted probability measure* defined by its density with respect to the original measure $\mathbb{P}$:

$$\frac{d\widetilde{\mathbb{P}}_n}{d\mathbb{P}} = \frac{\prod_{1\le i\le n} w_i(Y_i)}{\mathbb{E}\left\{\prod_{1\le i\le n} w_i(Y_i)\right\}}. \tag{3}$$

We then have:

$$\begin{aligned}\mathbb{E}\{f_n(Y_n)\} &= \widetilde{\mathbb{E}}\left\{f_n(Y_n)\frac{d\mathbb{P}}{d\widetilde{\mathbb{P}}_n}\right\}\\ &= \widetilde{\mathbb{E}}\left\{f_n(Y_n)\prod_{1\le i\le n} w_i^-(Y_i)\right\}\mathbb{E}\left\{\prod_{1\le i\le n} w_i(Y_i)\right\},\end{aligned} \tag{4}$$

where we used the notation w_i^- for the inverse $1/w_i$ of the i-th weight function. Assuming $\mathbb{E}\{w_i(Y_i)\,|\,\mathcal{F}_{i-1}^X\} = 1$ for every $i \ge 1$, we see that the second expectation in the right hand side is equal to one and that (4) further reduces to

$$\mathbb{E}\{f_n(Y_n)\} = \widetilde{\mathbb{E}}\left\{f_n(Y_n)\prod_{1\le i\le n} w_i^-(Y_i)\right\}. \tag{5}$$

For the computation of $\mathbb{E}\{f_n(Y_n)\}$, the importance sampling algorithm based on $\widetilde{\mathbb{P}}_n$ relies on the generation of m Monte Carlo samples $\widetilde{\zeta}_n^j = (\widetilde{\xi}_0^j, \widetilde{\xi}_1^j, \cdots, \widetilde{\xi}_n^j)$ for $j = 1, 2, \cdots, m$ under the twisted distribution $\widetilde{\mathbb{P}}_n$. These samples are generated as follows. For each j, and given a fixed initial condition $\widetilde{\xi}_0$ for X_0, for each $i = 1, \cdots, n$, use the

transition kernel $\widetilde{K}_i$ of the Markov chain X at time $i-1$ under the twisted probability measure $\widetilde{\mathbb{P}}_i$ to generate a sample $\widetilde{\xi}_i^j$ from $\widetilde{\xi}_{i-1}^j$. Then for each fixed Monte Carlo sample of size m, we have the following unbiased estimate:

$$\mathbb{E}\{f_n(Y_n)\} \approx \mathbb{E}_n^m \left\{ f_n(\widetilde{\zeta}_n) \prod_{1\le i\le n} w_i^-(\widetilde{\xi}_0, \cdots, \widetilde{\xi}_i) \right\} \tag{6}$$

where $\mathbb{E}_n^m$ refers to expectation with respect to the empirical distribution of the m samples $\widetilde{\zeta}^j$'s. For the record we notice that this estimator is asymptotically consistent in the sense that it converges to $\mathbb{E}\{f_n(Y_n)\}$ as $m \to \infty$ (see for example Glasserman [6]).

Remark 1. Two conditions need to be satisfied for this change of measure to be practical. First one needs to be able to compute the new transition kernel $\widetilde{K}$ of the chain under $\widetilde{\mathbb{P}}$, and most importantly, one needs to be able to generate efficiently Monte Carlo samples from the distribution $\widetilde{K}(x_{n-1}, dx_n)$. Second, one needs to be able to compute the distorted integrand appearing in the right hand side of (6), and in particular the product of the inverses of the weight functions with a low overhead. As in the case of the applications given in this paper, this is typically done by resorting to a suitable version of the Girsanov theorem which is the time-honored method to change measure for stochastic processes.

Remark 2. In order to minimize the variance, one should choose, for each integer $i \ge 1$, a weight function w_i such that $\prod_{1\le i\le n} w_i$ is proportional to $|f_n|$. But of course, in this case, the second expectation in the right hand side of (4) is typically unknown (not equal to 1), since it is the quantity that we aim at computing (at least for $f_n \ge 0, f_n = |f_n| = \prod_{1\le i\le n} w_i$).

2.3. **Twisted Feynman-Kac Expectations.** The purpose of this subsection is to give a crash course on Feynman-Kac path measures. We describe their subsequent approximations by interacting particle system in the next subsection. The basic material is borrowed from [13] and the actual application to the Monte Carlo computation of probabilities of rare events from [14]. Technical details are included for the sake of completeness and to ease the introduction of a specific set of notations.

For any bounded measurable function f_n on F_n, we define the Feynman-Kac expectation $\gamma_n(f_n)$ by:

$$\gamma_n(f_n) = \mathbb{E}\left\{ f_n(Y_n) \prod_{1\le i<n} w_i(Y_i) \right\}. \tag{7}$$

Note that as a non-negative linear form on a cone of non-negative functions, γ_n can be viewed as a measure. We shall denote by η_n the corresponding normalized measure which is naturally defined as

$$\eta_n(f_n) = \frac{\mathbb{E}\left\{ f_n(Y_n) \prod_{1\le i<n} w_i(Y_i) \right\}}{\mathbb{E}\left\{ \prod_{1\le i<n} w_i(Y_i) \right\}} = \frac{\gamma_n(f_n)}{\gamma_n(1)}. \tag{8}$$

Notice that

$$\gamma_{n+1}(1) = \gamma_n(w_n) = \eta_n(w_n)\gamma_n(1) = \prod_{i=1}^{n} \eta_i(w_i). \tag{9}$$

This seemingly innocent remark will play a crucial role in the following. Consequently, for any given bounded measurable function f_n, we have

$$\gamma_n(f_n) = \eta_n(f_n) \prod_{1 \le i < n} \eta_i(w_i). \tag{10}$$

The above relationship has the merit of relating the un-normalized expectations in the left hand side to normalized *twisted* expectations in the right hand side. Using the notation w_i^- for the inverse of the weight function w_i introduced earlier, we have (to be compared with (4) in the case of IS):

$$\begin{aligned} \mathbb{E}\{f_n(Y_n)\} &= \mathbb{E}\left\{ f_n(Y_n) \prod_{1 \le i < n} w_i^-(Y_i) \prod_{1 \le i < n} w_i(Y_i) \right\} \\ &= \gamma_n\left(f_n \prod_{1 \le i < n} w_i^- \right) \\ &= \eta_n\left(f_n \prod_{1 \le i < n} w_i^- \right) \prod_{1 \le i < n} \eta_i(w_i). \end{aligned} \tag{11}$$

This shows that expectations over the original process can be computed if one can compute normalized twisted expectations. This is in fact possible, in a dynamic way because, like in classical filtering theory, it is easily checked that the sequence of normalized twisted probability measures form a well defined *dynamical system* in the space of probability measures. Indeed we have

$$\eta_n = \Phi_n(\eta_{n-1}), \qquad \eta_1 = M_1(X_0, \cdot) \tag{12}$$

where the nonlinear operators Φ_n giving the dynamics are defined as

$$\Phi_n(\eta) = \frac{1}{\eta(w_{n-1})} \int_{F_{n-1}} \eta(dy_{n-1}) w_{n-1}(y_{n-1}) M_n(y_{n-1}, \cdot). \tag{13}$$

2.4. **Interacting Particle Systems for the Computation of Rare Events.** The IPS method is based on the deformation of the Markov chain successive transitions by way of mutations and selections in order to force the chain into the rare events of interest. Because the deformations of the chain can be understood mathematically, at least locally in time, as changes of measures, this strategy is reminiscent of classical importance sampling as described earlier. However, as seen in the context of diffusion processes (see Appendix A), IPS involves Feynman-Kac changes of measures as of section 2.3, whereas Girsanov-like changes of measure underlie IS. As we already emphasized in the introduction, this implies the important practical difference that while the Monte Carlo samples of an importance sampling computation are generated from the *twisted* distribution, the Monte Carlo samples used in an IPS Monte Carlo computation are generated under the original distribution of the chain. As we shall see in detail now, all we need to have in order to implement the IPS Monte Carlo

computations is a *black box* capable of generating Monte Carlo samples from the distribution of the chain: no need to open the box to perform the changes necessary to generate samples from the twisted distribution!

2.4.1. ***Model Simulation***. For the purpose of numerical computations and Monte Carlo estimation of expectations of the form (11), we introduce approximations of the above probability distributions by convex combinations of Dirac measures, and we show that the time evolution of the measures η_n given by the dynamical system (12) and (13) implies a natural time evolution for the point masses of the convex combinations of Dirac measures, hence the interpretation of these approximations as an interacting particles system.

We choose a ('large') integer m which we shall interpret as the number of particles. A particle at time n is an element

$$\zeta_n^j = (\xi_{0,n}^j, \xi_{1,n}^j, \cdots, \xi_{n,n}^j) \in F_n = E_0 \times E_1 \times \cdots \times E_n.$$

where the superscript j of the particle ranges from 1 to m. We start with an initial configuration $\zeta_1 = (\zeta_1^j)_{1 \le j \le m}$ that consists of m independent and identically distributed random samples from the distribution:

$$\eta_1(d(x_0, x_1)) = M_1(X_0, d(x_0, x_1)) = \delta_{X_0}(dx_0) K_1(x_0, dx_1)$$

where we use the notation δ_x for the Dirac measure at the point x. In other words, the $\zeta_1^j = (\xi_{0,1}^j, \xi_{1,1}^j) = (X_0, \xi_{1,1}^j) \in F_1 = E_0 \times E_1$ are independent and in such a way that all the $\xi_{1,1}^j$ are all independent with the same distribution $K_1(X_0, \cdot)$. Based on the transition given by the dynamic equation (13), we define the transition for the particles which are providing the approximation of η_2 by $\zeta_{n-1} \to \zeta_n$ from F_{n-1}^m into F_n^m according to the transition probability

$$\mathbb{P}\{\zeta_n \in d(y_n^1, \cdots, y_n^m) \,|\, \zeta_{n-1}\} = \prod_{j=1}^m \Phi_n(\rho(\zeta_{n-1}))(dy_n^j), \tag{14}$$

where $\rho(\zeta_{n-1})$ is the empirical measure defined by

$$\rho(\zeta_{n-1}) = \frac{1}{m} \sum_{j=1}^m \delta_{\zeta_{n-1}^j}$$

and $d(y_n^1, \cdots, y_n^m)$ represents an infinitesimal neighborhood of the point $(y_n^1, \cdots, y_n^m)$ in F_n^m. Recalling the definition of the operators Φ_n giving the dynamics of the η_n, one sees that (14) is the superposition of two clearly identifiable elementary transitions, a *selection* followed by a *mutation*. In other words:

$$F_{n-1}^m \ni \zeta_{n-1} \overset{\text{selection}}{\longrightarrow} \hat{\zeta}_{n-1} \in F_{n-1}^m \overset{\text{mutation}}{\longrightarrow} \zeta_n \in F_n^m$$

as follows. The selection stage is performed by resampling with replacement (i.e. choosing independently) m path-particles

$$\hat{\zeta}_{n-1}^j = (\hat{\xi}_{0,n-1}^j, \hat{\xi}_{1,n-1}^j, \cdots, \hat{\xi}_{n-1,n-1}^j) \in F_{n-1},$$

with possible repetitions according to the Gibbs measure

$$\sum_{j=1}^{m} \frac{w_{n-1}(\zeta_{n-1}^j)}{\sum_{l=1}^m w_{n-1}(\zeta_{n-1}^l)} \delta_{\zeta_{n-1}^j}. \tag{15}$$

Then, the mutation stage is performed by extending each selected path-particle $\hat{\zeta}_{n-1}^j$ into a path-particle $\zeta_n^j \in F_n = F_{n-1} \times E_n$ of the form

$$\begin{aligned} \zeta_n^j &= (\xi_{0,n}^j, \xi_{1,n}^j, \cdots, \xi_{n,n}^j) \\ &= (\hat{\xi}_{0,n-1}^j, \xi_{1,n-1}^j, \cdots, \hat{\xi}_{n-1,n-1}^j, \xi_{n,n}^j) \end{aligned}$$

where the m samples $\xi_{n,n}^j$ are independently drawn from the distributions $K_n(\hat{\xi}_{n-1,n-1}^j, \cdot)$.

2.4.2. ***Convergence Results***. We are now in a position to quote the theoretical result on which the Monte Carlo approximations are based. See for example [13] or [14] for details. For each fixed n we have

$$\lim_{m\to\infty} \eta_n^m = \eta_n$$

in distribution, where the empirical measures η_n^m are defined by:

$$\eta_n^m = \frac{1}{m} \sum_{j=1}^{m} \delta_{\zeta_n^j}.$$

This result is screaming for the introduction of the particle approximation

$$\gamma_n^m(f_n) = \eta_n^m(f_n) \prod_{1\le i<n} \eta_n^m(w_i)$$

for $\gamma_n(f_n)$. The main result of [14] which we use below states that γ_n^m is an unbiased estimator of γ_n in the sense that for any integer $p \ge 1$ and any bounded measurable function f_n on F_n with $\sup|f_n| \le 1$, we have

$$\mathbb{E}\{\gamma_n^m(f_n)\} = \gamma_n(f_n),$$

and in addition

$$\sup_{m\ge 1} \sqrt{m}\mathbb{E}\{|\gamma_n^m(f_n) - \gamma_n(f_n)|^p\}^{1/p} \le c_p(n),$$

for some positive constant $c_p(n) < \infty$ whose value does not depend upon the particular choice of the function f_n.

In view of (11), we thus get the following unbiased (for fixed m) and asymptotically convergent (as $m \to \infty$) estimate of $\mathbb{E}\{f_n(Y_n)\}$:

$$\begin{aligned} \gamma_n^m(f_n \prod_{1\le i<n} w_i^-) &= \eta_n^m(f_n \prod_{1\le i<n} w_i^-) \prod_{1\le i<n} \eta_i^m(w_i) \\ &= \mathbb{E}_n^m \left\{ f_n(\zeta_n) \prod_{1\le i<n} w_i^-(\xi_{0,n}, \cdots, \xi_{i,n}) \right\} \prod_{1\le i<n} \mathbb{E}_i^m\{w_i(\zeta_i)\} \end{aligned} \tag{16}$$

where as before, for every i the notation $\mathbb{E}_i^m$ refers to expectation under the empirical distribution defined by the ζ_i^j's. More importantly, the variance of the estimator can

be analyzed. In case the expectation of interest is tantamount to the probability of a "rare" event

$$\mathbb{E}\{f_n(Y_n)\} = \mathbb{E}\{\mathbf{1}_A(V(Y_n))\} = \mathbb{P}\{V(Y_n) \in A\}, \tag{17}$$

where V is a function from E_n to $\mathbb{R}$ the conclusion is that in order to minimize this variance, one should use weight functions w favoring the occurrence of the rare event without involving too large normalizing constants. Moreover, the choice of w should give rise to an algorithm that can be easily implemented.

We use the guidelines of this and the previous subsection for rare events of the form $\{V(X_n) = \ell\}$ (since our aim is to estimate portfolio loss distributions at fixed points in time; it is interesting to note however that IPS estimates are also available for more general 'path-dependent' events as of (17)).

3. **Benchmark Models**

We now introduce the family of Markovian intensity models of credit risk which we use to illustrate and compare explicit implementations of IS and IPS. Even though the basic principle of IS does not require the Markov property, we saw that IPS algorithms are based on a background Markov chain. For this reason, we described IS in section 2.2 with a definite Markovian bent, and the choice of the models used as implementation test-beds is based on the specification of suitable Markovian factors.

3.1. **Local Intensity Model.** Within the context of the *Top-Down* approach to credit risk, we assume that the cumulative loss process $L = \{L_t;\, t \geq 0\}$ of a credit portfolio of n names (we can think of $n = 125$) is modeled as a Markov point process. More specifically, we assume that $L_0 = 0$ and that L is a pure birth process with local intensity $\lambda(t, L_t)$ given by a deterministic function $\{\lambda(t,i)\}_{t\geq 0, i\geq 0}$ satisfying $\lambda(t,i) = 0$ for $i \geq n$ (see, for instance, [12, 2, 4]). This last condition guarantees that the process L is actually stopped at the level n, as there are only n names in the pool. In any case, conditionally on the information $\mathcal{F}_t = \mathcal{F}_t^L$ available at time t, up to first order in dt, the probability of a jump in the infinitesimal time interval $(t, t+dt)$ is given by $\lambda(t, L_t)dt$.

The infinitesimal generator $\mathcal{G}_t$ of the process is given by the $(n+1)\times(n+1)$ matrices

$$\mathcal{G}_t = \begin{pmatrix} -\lambda(t,0) & \lambda(t,0) & 0 & 0 & 0 \\ 0 & -\lambda(t,1) & \lambda(t,1) & 0 & 0 \\ & & \cdots & & \\ 0 & 0 & 0 & -\lambda(t,n-1) & \lambda(t,n-1) \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

For each time t, we denote by $p(t)$ the $(n+1)$-dimensional vector $p(t) = [p_i(t)]_{i=0,1,\cdots,n}$ of loss probabilities $p_i(t) = \mathbb{P}\{L_t = i\}$. They satisfy the forward Kolmogorov equation:

$$(\partial_t - \mathcal{G}_t^*)p = 0 \text{ on } (0, +\infty)\,, \qquad p(0) = \delta_0\,, \tag{18}$$

which is in fact a system of ordinary differential equations. Here and throughout the paper, $\mathcal{G}^*$ denotes the adjoint (or transpose) of a generic matrix $\mathcal{G}$. In the time-homogeneous case (when the intensities $\lambda(t,i)$ do not depend upon t, in which case

the infinitesimal generator matrices $\mathcal{G}_t$ do not depend upon t either), the solution of the Kolmogorov equation is given by a matrix exponential of the form

$$p(t) = \exp(t\mathcal{G}^*)\delta_0, \qquad t \geq 0.$$

In the general case, the exponential needs to be replaced by the propagator of equation (18) which has the intuitive interpretation of a *time-ordered exponential*.

The sample paths of the loss process L are piecewise constant. Let $\widetilde{t}_i$ denote the i^{th} ordered jump time of L, for each $i = 1, \cdots, n$, (or $\widetilde{t}_i = +\infty$, in case there are less than i jumps on a given trajectory). We also set $\widetilde{t}_0 = 0$. Given a fixed maturity T, say $T = 5yr$, we define $t_i = \widetilde{t}_i \wedge T$ for each $i = 0, \cdots, n$. Consequently, $t_i < T$ and $L_{t_i} = i$ if and only if there are at least i jumps of L before the maturity T on a given trajectory. The background Markov chain $\{X_i\}_{0 \leq i \leq n}$ introduced in Subsection 2.1 and used in this section is then defined by:

$$X_i = (t_i, L_{t_i}), \qquad 0 \leq i \leq n. \tag{19}$$

3.2. **Homogeneous Groups Model.** Despite its usefulness as a toy model for which explicit computations can easily be carried out, the previous model lacks in realism as it ignores the fact that L is typically not Markovian. In view of its shortcomings, we now consider a generalization for which the loss process L is given by a function of a *higher dimensional* Markov process Λ. To be more specific, we work with the model for credit portfolios used by [5, 2, 16, 10] among others. In this model, the pool of n credit names is organized in d homogeneous groups of ν obligors (so $n = \nu d$).

The group cumulative default processes L^l, $l = 1, \cdots, d$, $L^l(t)$ giving the number of defaults in the l-th group up to time t, are jointly modeled as a d-variate Markov point process $\Lambda = (L^1, \cdots, L^d)$. To alleviate the complexity of the simulation code, we exclude simultaneous jumps of the L^l's (see, e.g., Bielecki et al. [2]). For each l, the intensity of L^l is assumed to be of the form $\lambda^l(t, \Lambda_t)$, for some deterministic intensity function $\lambda^l = \lambda^l(t, \iota)$, where $\iota = (i_1, \cdots, i_d) \in \{0, 1, \cdots, \nu\}^d$. We denote by t_i the ordered sequence of the jump times of Λ, capped at T, and we define the background Markov chain $\{X_i\}_{0 \leq i \leq n}$ by:

$$X_i = (t_i, \Lambda_{t_i}), \qquad 0 \leq i \leq n. \tag{20}$$

Observe in particular that $L_{t_i} = V(\Lambda_{t_i}) = \sum_l L^l_{t_i}$, so the portfolio loss process L at time t_i is given by a simple function of X_i.

Notice that for $d = 1$, we recover the Markovian local intensity model of the previous subsection. At the other end of the spectrum, for $d = n$ (i.e. when each group has only a single element), we are in effect modeling the vector of default indicator processes of the pool. As d varies between 1 and n, we thus get a variety of models of credit risk, ranging from pure top-down models for $d = 1$, to pure bottom-up models for $d = n$ (see Bielecki et al. [1]).

Knowing the group loss-probability vectors $q(t)$, the related portfolio loss-probabilities $p = p_i(t) = \mathbb{P}\{L_t = i\}$ for $i = 0, \cdots, n$, follow in a straightforward way. However the infinitesimal generator $\mathcal{G}_t$ of the Markov process Λ now appears as a $(\nu+1)^d \otimes (\nu+1)^d$-dimensional matrix, and even if this matrix is very sparse (since the components of Λ may only jump by one and only one at a time), its dimension is prohibitive in most

cases as far as deterministic numerical methods are concerned. For instance, in the case of $d = 5$ groups of $\nu = 25$ names, one gets $(25+1)^{2\times 5} = 11881376^2$, which rules out the use of personal computers if not for specialized libraries for the manipulation of sparse matrices. The computation of the loss distribution by deterministic numerical schemes is thus precluded by the curse of dimensionality for d greater than a few units (depending on ν). So for high values of d, Monte Carlo methods appear to be the only viable alternative.

As ***benchmark model*** , we shall use henceforth the Local Intensity Model with $n = 125$ names in the pool. To be more specific we shall consider the case of homogenous obligors with individual (pre-default) instantaneous default intensities of the form $\lambda_t = a\exp(bL_t/n)$, for non-negative parameters a and b. The special case where $b = 0$, which is dealt with in Section 4, thus corresponds to a case of independent obligors, whereas the case $b > 0$, further considered in Section 5, corresponds to a situation of default contagion (as typically observed under the pricing measure, particularly so since the sub-prime crisis).

Observe that such a set-up can equivalently be encoded as any homogeneous groups model with $\lambda^l(t,\iota) = (\nu - i_l)a\exp(bL_t/n)$ for every $l = 1, \cdots d$, provided d and ν satisfy the relation $\nu d = n$. We shall use this observation later to compare the simulation results obtained in encodings of our benchmark model with various nominal dimension d (specifically $d = 1$ or $d = 5$) , and also, to compare them with exact values computed by analytic procedures relying on the one-dimensional formulation of the model. Recall that in general, there is no way one can check simulation results obtained in the homogeneous groups model for d greater than a few units, unless we consider, as done here, particular cases reducible to a lower-dimensional model.

4. **Case of Independent Obligors**

In this section we consider the case of homogenous obligors who default independently of each other, with individual (pre-default) instantaneous intensity at time t equal to $1/n$.

4.1. **Algorithms.**

4.1.1. ***IS Algorithms***. For IS algorithms, we use in the special case of independent obligors a simple twist transformation which consists in speeding up the arrivals of the defaults. This is done by scaling up the intensities by a factor α. In the case of our benchmark model with nominal dimension $d = 1$, a simple form of Girsanov's theorem for point processes gives

$$I_f = \widetilde{\mathbb{E}}\left\{ f(L_T)\exp[-L_T\log(\alpha) + (\alpha-1)\int_0^T \lambda(L_t)dt]\right\}.$$

Consequently, the IS estimate of I_f for the value $\alpha > 0$ of the free parameter is obtained by generating Monte Carlo samples ω_α using the intensity function $\lambda_\alpha = \alpha\lambda$

and then computing

$$\hat{I}_f^\alpha = \frac{1}{m}\sum_{j=1}^{m} f(L_T(\omega_\alpha^j))e^{-L_T(\omega_\alpha^j)\log(\alpha)+(\alpha-1)\int_0^T \lambda(L_t(\omega_\alpha^j))dt}.$$

An analogous formula, though slightly more involved, holds in the case of an encoding of our benchmark model in the form of a homogeneous group model with $d = 5$.

In particular, for $f = \delta_\ell$ (Dirac mass at the loss level ℓ in $\{0, \cdots, n\}$), we thus get IS estimates $p_\ell^m(T, \alpha)$ of the probability of a portfolio loss equal to ℓ at time T.

4.1.2. ***IPS Algorithms***. We follow the approach of Subsection 2.4 applied to the the background Markov chain defined by (20) (which in the case $d = 1$ reduces to (19)), and with weight functions w_i therein chosen of the form

$$w_i(y_i) = w_i(x_{i-1}, x_i) = e^{\alpha(\ell_i - \ell_{i-1})}, \qquad y_i = (x_0, x_1, \cdots, x_i) \tag{21}$$

where ℓ_i represents the loss in the portfolio at step i in the algorithm. So

$$w_i(Y_i) = w_i(X_{i-1}, X_i) = e^{\alpha(L_{t_i} - L_{t_{i-1}})} = \begin{cases} 1 & \text{if} \quad t_i = T \\ e^\alpha & \text{if} \quad t_i < T \end{cases}$$

(recall that L_{t_i} is a function of the state X_i of the Markov chain in our model). As in the IS method, $\alpha > 0$ is a free parameter whose choice we shall try to take advantage of.

To be more specific, we have the following IPS algorithm for the computation of the probability $\mathbb{P}(L_T = \ell)$, for $\ell = \{0, \cdots, n\}$. Notice that because of the special form of the weight functions w_i, we only need to keep track of the last two components of ζ_i^j, which we call *father* and *son* for the sake of definiteness.

Initialization: For every $j = 1, \cdots, m$, set $\xi_{0,1}^j = 0$ and simulate a pair $(t_1^j, \xi_{1,1}^j)$ starting from $\xi_{0,1}^j$ at time 0 as defined above, using the dynamics of Λ (alias L, in case $d = 1$) for this simulation step. **Loop:** Assuming the m (time, father and son)-particles $(t_{i-1}^j, \xi_{i-2,i-1}^j, \xi_{i-1,i-1}^j)$ already simulated:

Selection: Sample independently m (father and son)-particles

$$(\hat{t}_{i-1}^j, \hat{\xi}_{i-2,i-1}^j, \hat{\xi}_{i-1,i-1}^j)$$

with possible repetitions according to the Gibbs measure defined by the weights

$$w^\alpha(\xi_{i-2,i-1}^j, \xi_{i-1,i-1}^j)\delta_{t_{i-1}^j, \xi_{i-2,i-1}^j, \xi_{i-1,i-1}^j}$$

normalized to one, recall formula (15);

Mutation: For every $j = 1, \cdots, m$, set $\xi_{i-1,i}^j = \hat{\xi}_{i-1,i-1}^j$ and simulate a pair $(t_i^j, \xi_{i,i}^j)$ starting from $(\hat{t}_{i-1}^j, \hat{\xi}_{i-1,i-1}^j)$ using the dynamics of Λ (alias L, in case $d = 1$) for this simulation step.

Termination: Exit from the loop when $i = n$, and compute the following estimate of the loss probability $\mathbb{P}(L_T = \ell)$, for every $\ell \in \{0, \cdots, n\}$:

$$\widetilde{p}_\ell^m(T, \alpha) = \mathbb{E}_n^m \left\{\delta_\ell(L(\xi_{n,n}))\exp(-\alpha L(\xi_{n-1,n}))\right\} \prod_{i=1}^{n-1} \mathbb{E}_i^m w^\alpha(\xi_{i-1,i}, \xi_{i,i}) \tag{22}$$

where for every fixed i the notation $\mathbb{E}_i^m$ refers to the empirical distribution defined by the ζ_i^j's as j ranges from 1 to m.

Remark. Regarding the simulation of Λ in these algorithms, it is crucial to exploit the fact that, according to our assumptions, the components of Λ do not jump simultaneously. So the simulation of Λ may be done "component by component", in time $\mathrm{O}(d)$ (see Bielecki et al. [2]).

4.1.3. ***Choice of*** α. In each case (IS and IPS), we introduce a twisted measure (used explicitely in the case of IS and implicitly in the case of IPS) parameterized by a constant α. The range of values of ℓ for which the estimate $p_\ell^m(T,\alpha)$, resp. $\widetilde{p}_\ell^m(T,\alpha)$, is significantly different from 0, is expected to vary with α. Consequently we run a certain number μ of Monte Carlo (IS or IPS) loops corresponding to different values of α. Then, for every loss level ℓ and for every method (IS or IPS), we retain the estimator $p_\ell^m(T,\alpha)$ or $\widetilde{p}_\ell^m(T,\alpha)$ of the related probability with the highest *significance*, in the sense of the value $\alpha(\ell)$ having given rise to the greatest number of trajectories at level ℓ of the loss. To be more specific, our final IS estimate of $\mathbb{P}(L_T=\ell)$ is

$$p_\ell^m(T) = p_\ell^m(T,\alpha(\ell)) \tag{23}$$

where

$$\alpha(\ell) = \arg\max_\alpha \#\{j;\, 0\le j\le m,\, L_T(\omega_\alpha^j,\alpha)=\ell\}, \tag{24}$$

and likewise for our final IPS estimate $\widetilde{p}_\ell^m(T)$ of $\mathbb{P}\{L_T=\ell\}$.

For each encoding of our benchmark model (with nominal dimension $d=1$ or $d=5$), we implemented a straight explicit importance sampling method and an interacting particle system method. We also used numerical matrix exponentiation to compute the exact values of the loss probabilities.

In each simulation experiment, we ran $\mu=11$ Monte Carlo loops (one standard MC loops and $\mu=10$ IS or IPS MC loops), yielding eleven different estimates $p_\ell^m(T,\alpha)$ or $\widetilde{p}_\ell^m(T,\alpha)$ of the probabilities $\mathbb{P}\{L_T=\ell\}$, for every $\ell\in\{0,\cdots,n\}$. Then, for every ℓ, we retained the estimator $p_\ell^m(T,\alpha(\ell))$ or $\widetilde{p}_\ell^m(T,\alpha(\ell))$ of the related probability with the highest significance as explained above.

The values of the parameters we used are given in Tables 1 and 2.

T	n	$\lambda(t,i)$	μ	α (Step)/ IS	α (Step)/ IPS
5y	125	$1-\frac{i}{n}$	11	1	0.4

TABLE 1. Parameter Values for our benchmark model with nominal dimension $d=1$ (local intensity model).

T	$n = d \times \nu$	$\lambda^l(t, \iota)$	μ	α (Step) / IS	α (Step)/ IPS
5y	$125 = 5 \times 25$	$\frac{\nu - i_l}{n}$	11	1	0.4

TABLE 2. Parameter Values for our benchmark model with nominal dimension $d = 5$ (case of a homogeneous five groups model reducible to a local intensity model).

4.2. **Results for IS.** Figure 1 displays the results obtained with the explicit importance sampling method in the local default intensity model (benchmark model with nominal dimension $d = 1$): multiplying the local default intensity function by a factor α ranging from $\mu = 1$ to $\mu = 11$, simulating $m = 5000$ trajectories in the models with scaled intensities and applying the related payoffs corrections.

The left panel displays the concentration of loss levels hit depending on the value of α used. The number of hits increases as the color ranges from blue to red. The right part of the figure displays the exact values of the log-probabilities (black curve) together with the estimations obtained by simulation (points in color). The color of a point refers to the value of α retained for estimating the related probability, from blue for the lowest value of α, to red for the highest value. The results are very accurate, for loss levels comprised between 0 and 50 (and on this example it is actually possible to go far beyond this level by further adjustment of the parameter α).

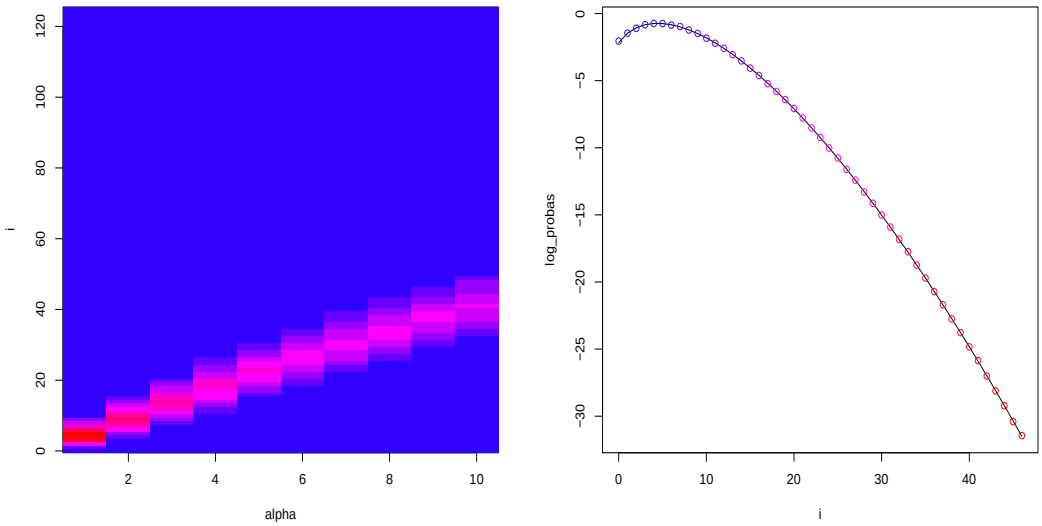

FIGURE 1. *IS losses as function of alpha (left) and log-probability esitmates (dots) superimposed on the exact probabilities (right). We used* $d = 1$ *and* $m = 5000$.

We also checked how the method handles the curse of dimensionality by assessing the robustness of the method as the model dimension increases. Figure 2 displays the results in the case of the benchmark model with $d = 5$ homogeneous classes. The loss distribution can be recovered exactly by numerical matrix exponentiation based on

the one-dimensional local intensity formulation of the model. The accuracy of the IS method is not altered in higher dimension ($d = 5$ in this experiment).

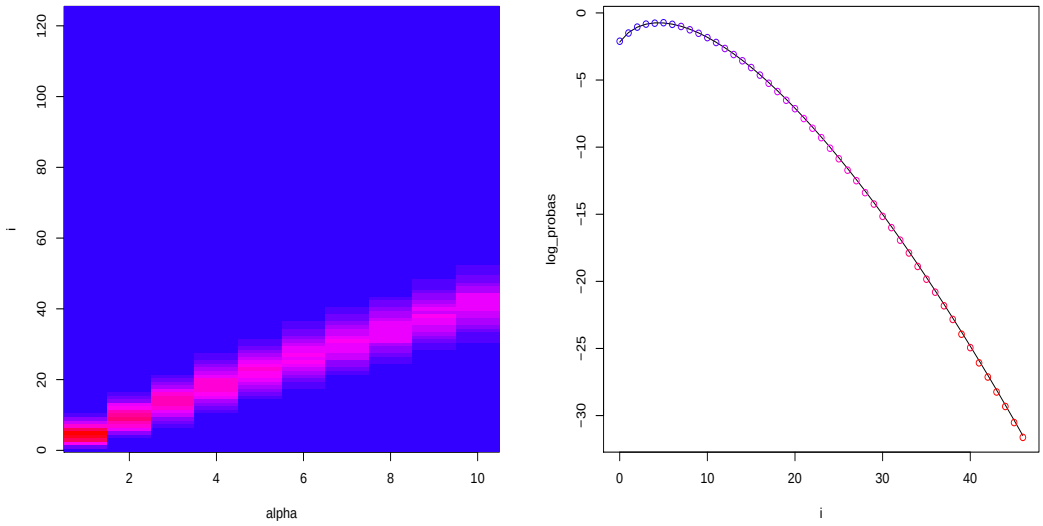

FIGURE 2. *IS losses (left) and log-probabilities (right).* $d = 5$ *and* $m = 5000$.

4.3. **Results for IPS.** Figure 3 displays the analogous results in the case of the IPS method in the local default intensity model ($d = 1$). Despite the fact that we used $m = 10^5$ simulations here (instead of $m = 5000$ as in Figures 1 and 2), for high levels of the loss the accuracy is not as good as with IS. Moreover, it seems very difficult to go higher in the loss level simply by adjusting the value of α: the related probabilities are too small and the generic IPS methodology is not able to provide reasonable estimates (see, e.g., Johansen, Del Moral and Doucet [11] for another instance of this shortcoming of the IPS).

Figure 4 is the counterpart to Figure 3 in the case of the homogeneous classes model with $d = 5$. Again the performance of the method is not significantly altered by the higher state space dimension.

Figures 5 and 6 show the results obtained using $m = 20000$ and $m = 5000$ Monte Carlo samples instead of $m = 10^5$ above. IPS runs with too small a number of samples do not ensure accuracy over the desired range of loss levels.

5. General Case with Default Contagion

We now assess the impact of default contagion. In presence of strong contagion between obligors, as observed on the credit markets since the start of the sub-prime crisis, the portfolio loss distribution has a very different structure than in the independent case. To address this issue we now consider an homogenous portfolio of credit risk with individual (pre-default) instantaneous intensity at time t equal to $\lambda_t = a \exp(bL_t/n)$ for a possibly *positive* parameter b. Note that b was always taken equal to zero in the previous section.

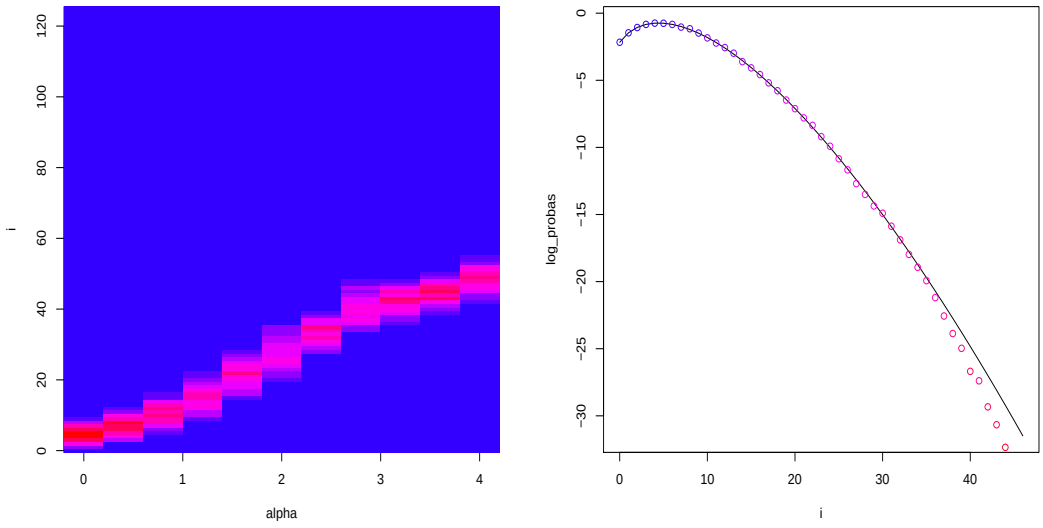

FIGURE 3. *IPS losses (left) and log-probabilities (right). We used* $d = 1$ *and* $m = 10^5$.

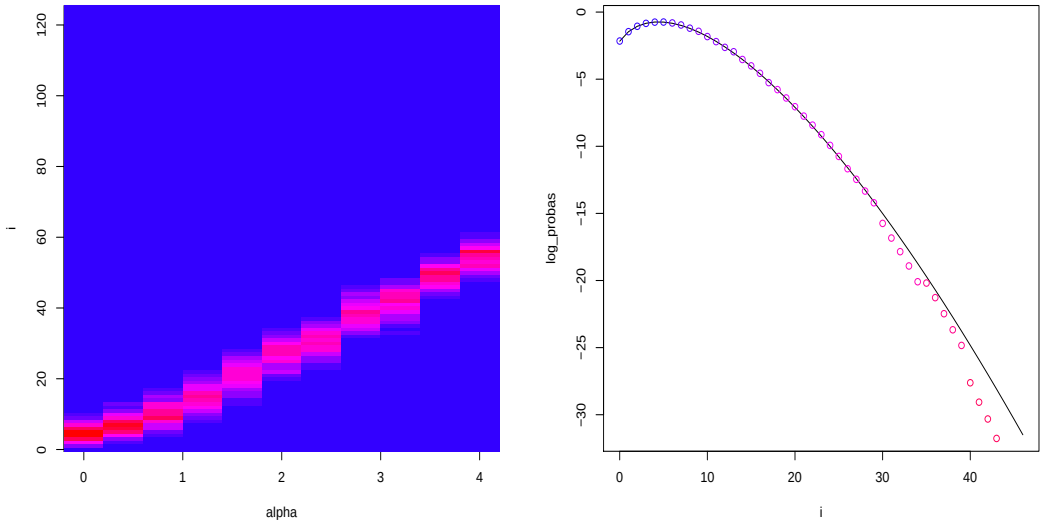

FIGURE 4. *IPS losses and log-probabilities. We used* $d = 5$ *and* $m = 10^5$.

We already discussed the fact that dimensionality matters little regarding the performance of the importance sampling methods (IS as well as IPS) at hand. In fact, we believe it is the main advantage of simulation approaches over deterministic methods. We now focus on the effect of contagion. We shall thus only work with the one-dimensional encoding of the model (local intensity model of Subsection 3.1), so throughout this section $d = 1$ and $\nu = n = 125$.

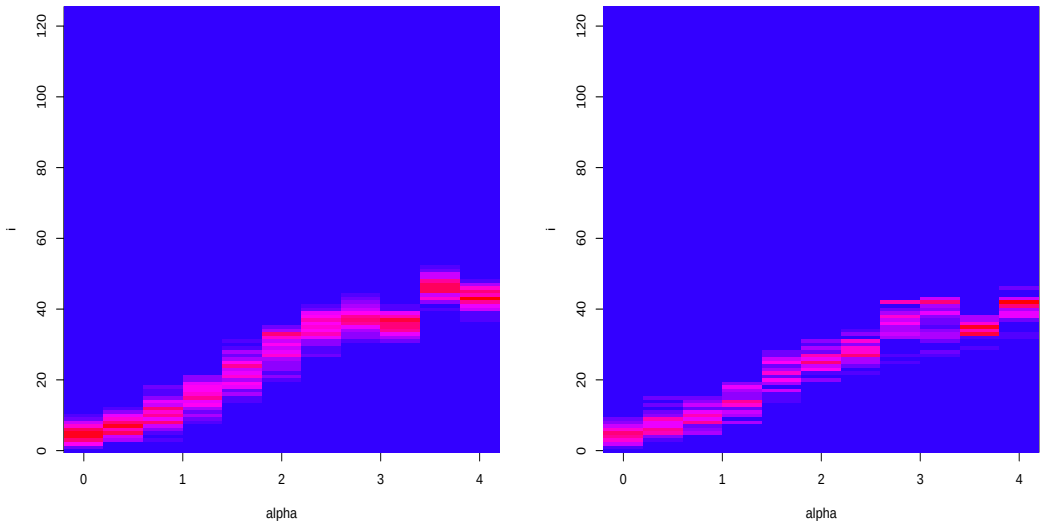

FIGURE 5. *IPS Losses,* $d = 5$. *Left panel:* $m = 20000$; *Right panel:* $m = 5000$.

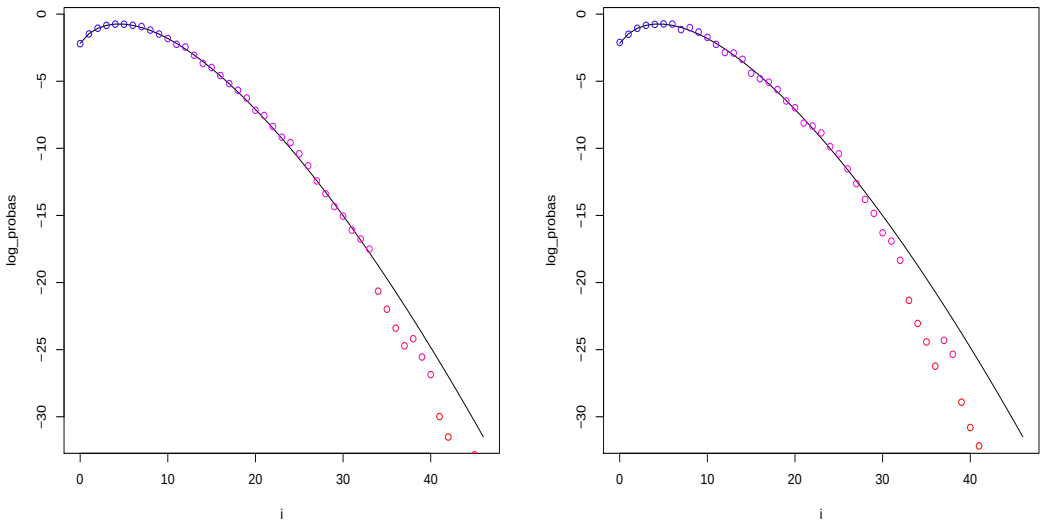

FIGURE 6. *IPS Values,* $d = 5$. *Left panel:* $m = 20000$; *Right panel:* $m = 5000$.

5.1. **Armageddon.** Figure 7 gives plots of the loss distribution for $T = 5yr$ in two different scenarios. For the left panel we used the values $a = 0.01$ and $b = 0$. This corresponds to a case of independent obligors as already considered in the previous section. For the right panel, we used the values $a = 0.01$ and $b = 13$ which correspond to a case of extreme contagion. These distributions were computed by matrix exponentiation of the one-dimensional model generator ($126 \otimes 126$ matrix). Note the different scales on the ordinate axes of the two plots.

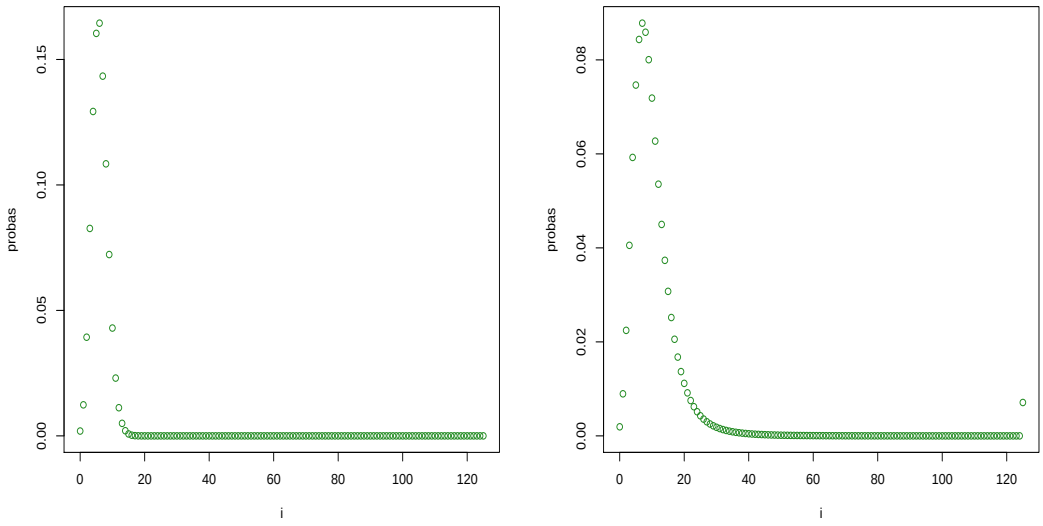

FIGURE 7. *5y loss distributions. Left panel: Independent obligors ($a = 0.01, b = 0$); Right panel: Extreme contagion ($a = 0.01, b = 13$).*

In the case of independent obligors (left panel), the structure of the loss distribution is basically that of a Poisson distribution (truncated at the level n). The right-tail of the distribution goes exponentially fast to zero, which makes high levels of the loss extremely rare. The probability of the Armageddon event (everyone defaulted in the portfolio) by the time of maturity $T = 5yr$, is equal to $1.04e-164$. This is why importance sampling (IS or IPS) methods are a *must* in this case.

In the case of extreme contagion (right panel), we observe the so-called *Armageddon effect*: the default of all the obligors within a finite time horizon becomes an event with significant probability, $7.106e-03$, of the order of one percent in the present situation. This is very important for pricing CDO tranches. Moreover there are no extremely rare levels of the loss any more. The less likely loss level is the level $i = 115$, with a loss probability of $1.108e-06$. For such a model, importance sampling methods are not strictly needed, since a standard Monte Carlo method with 10^6 samples will basically do a good job at estimating the $5yr$ loss distribution with a reasonable accuracy over the whole range of the loss levels. In this case, the usefulness of importance sampling methods is variance reduction, providing estimates for the $5yr$ loss distribution with $m = 10^4$ samples instead of $m = 10^6$.

5.2. **Algorithms.** An interesting point is that the specific IS (based on the idea of multiplying the intensities by a 'large' factor to favor defaults) or IPS (based on the choice of weight function favoring the defaults) algorithms that we used in the independent case of section 4.1 become completely inappropriate in a strong contagion context. Indeed in the context of a model with extreme contagion, favoring the defaults simply leads to concentrate all the trajectories on the Argameddon scenario. So the probabilities of all the events but Argameddon will be essentially estimated by zero in any IS or IPS scheme based on the idea of favoring the defaults. As for Argameddon, its probability is very well estimated by standard Monte Carlo with a

small number of trajectories (since this probability is of the order of one percent), therefore putting more trajectories on this event is also useless in this regard.

We do not report the related numerical results but they fully confirm the intuition. In order to implement IS and IPS algorithms covering the whole range of the loss levels in case $b > 0$ (for high b's in particular), we thus need to devise new (explicit or implicit) importance sampling strategies.

5.2.1. ***Algorithm for IS***. In order to cover the whole range of loss levels, we now resort for IS to Markovian changes of probability measures (see, e.g., [2] for general formulas) such that the aggregated portfolio jump intensity under the twisted measure becomes a constant of the form $\alpha\frac{n}{T}$, where α is fixed in $(0, 1)$. The average proportion of defaults at the loss horizon T under the twisted measure is thus of the order of α. Varying α, it is possible to cover the whole range of loss levels.

5.2.2. ***Algorithm for IPS***. As for IPS the idea to cover the whole range of loss levels is to use weight functions of the form (to be compared with (21))

$$w_i(y_i) = \exp(-\mathbb{1}_{t_i<T} \arctan(\ell_i - \alpha)) \tag{25}$$

where ℓ_i represents the loss in the portfolio at step i in the algorithm and α is a free parameter. The IPS algorithm is thus analogous to that of section 4.1.2, except for the use of the weight function w defined by (25) and the fact that we do not need a father-and-son algorithm anymore. We thus favor the trajectories with loss level at T of the order of α.

5.3. **Results for IS.** We tested the IS algorithm described Subsection 5.2.1 in two cases. Figure 8 gives the results for independent obligors ($a = 0.01, b = 0$), while Figure 9 gives the results in a case of strong contagion ($a = 0.01, b = 14$).

The fact that it is possible to cover the whole range of loss levels by varying α in the IS algorithm of section 5.2.2 is clearly visible on the left panels of Figures 8 and 9, which were produced using the five values $\alpha = 1/6, \cdots, 5/6$.

In the case of independent obligors, this IS method accordingly succeeds very well in estimating the $T = 5yr$ loss distribution on the whole range of loss levels (right panel of Figure 8).

But, oddly enough, this is not true any more in the case of extreme contagion. Indeed the right panel of Figure 9 shows that the IS method is completely inefficient at estimating the probabilities of loss levels with probabilities less than 10^{-3}, though $m = 10^4$ simulations were used in this experiment. The reason of this negative result is that the weights involved in the change of measures become extreme, creating huge fluctuations and a large variance, rendering the method essentially useless in practice.

5.4. **Results for IPS.** We now report the results obtained with the IPS algorithm described in Subsection 5.2.2, using the values $\alpha = 0, \cdots 125$.

In the independent case (Figure 11) the performances of the IPS algorithm are good for computing not too small probabilities (until the loss level $i = 36$ with exact probability $2.12e - 18$, on this specific example). For higher levels of the loss, the

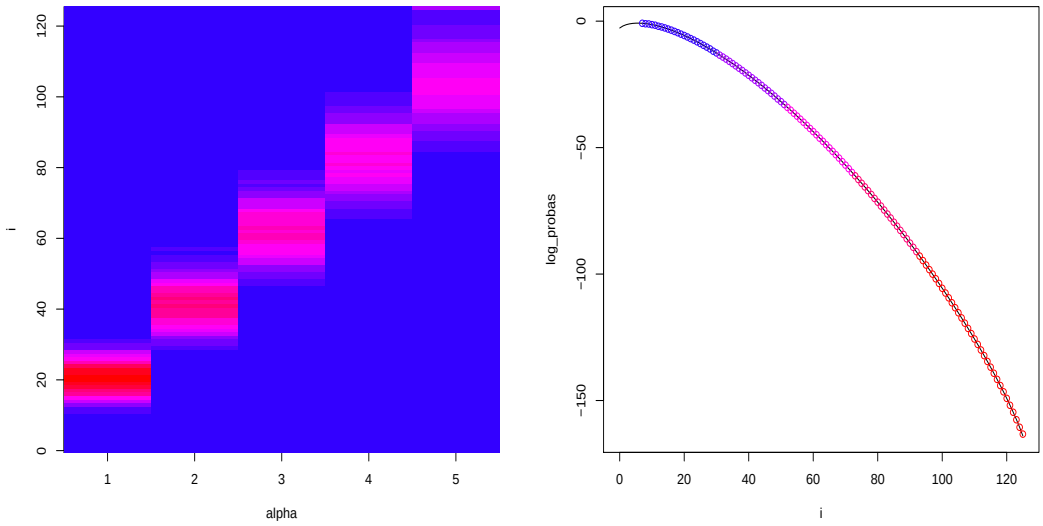

FIGURE 8. *IS losses and log-probabilities for independent obligors* ($a = 0.01, b = 0, m = 10000$).

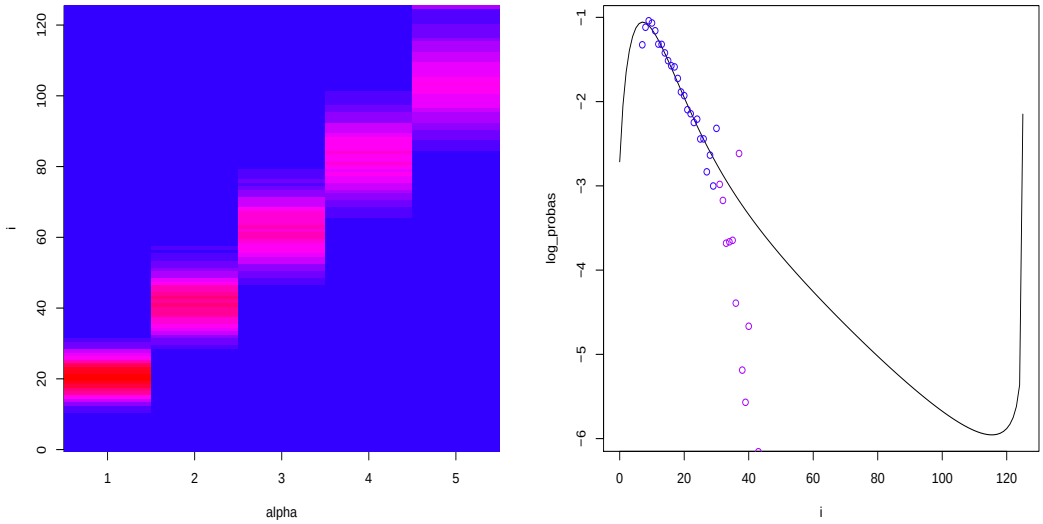

FIGURE 9. *IS losses and log-probabilities in the case of extreme contagion* ($a = 0.01, b = 13, m = 10000$).

related probabilities are too small and the generic IPS methodology is not sufficient to provide reasonable estimates. This is a known limitation of IPS estimates, already illustrated and commented upon in Section 4). However, the IPS method *captures* events of probability 10^{-5} to 10^{-6} (right panel of Figure 10) in the strong contagion case, even with only $m = 10^4$ Monte Carlo samples.

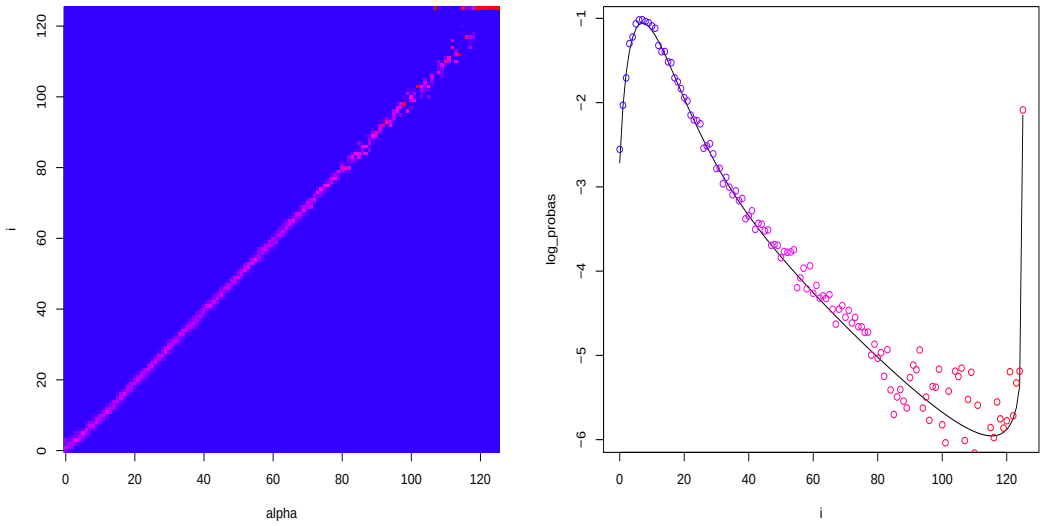

FIGURE 10. *IPS losses and log-probabilities in the case of extreme contagion* ($a = 0.01, b = 13, m = 10000$).

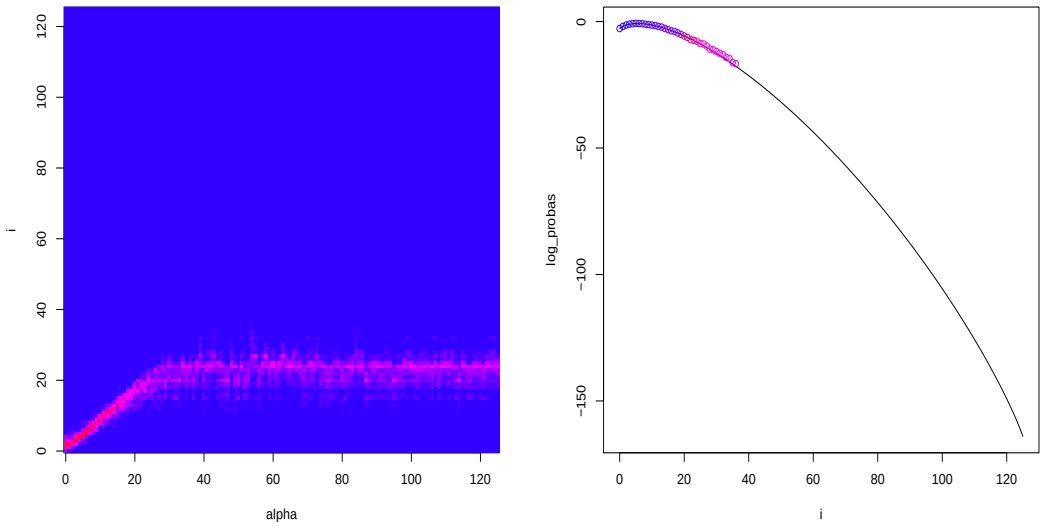

FIGURE 11. *IPS losses and log-probabilities in the case of independent obligors* ($a = 0.01, b = 0, m = 10000$).

6. **Summary and Conclusions**

We showed how to implement explicit forms of importance sampling together with interacting particle systems algorithms for the computation of credit portfolio loss probabilities in Markovian intensity models of credit risk. The need for simulation and variance reduction methods for such models is dictated by the fact that, even in relatively simple cases, the computational complexity of deterministic methods is prohibitive because of the curse of dimensionality.

We illustrated the fact that explicit IS methods *can do wonders* when models for the loss distribution are simple enough for a Girsanov like transformation to be identified, a random generator for the distorted probability structure is available and the corresponding densities can be easily computed along the samples.

However, these conditions are not always satisfied in practice and we showed that rare event probability estimation based on interacting particle systems offer a very useful alternative when no obvious or effective Girsanov change of measure is available, or when the Monte Carlo simulations are based on a computer implementation in the form of a black box which cannot be opened and modified for the purpose of importance sampling.

Appendix A. IS and IPS for Diffusion Processes

Both IS and IPS computations rely on changes from the original probability measure to absolutely continuous *twisted* measures. Yet, despite this similarity, the two methods differ in very significant ways. For pedagogical purposes we illustrate these differences in the context of diffusion processes. Indeed, despite the fact that we work with point processes throughout the paper, the ubiquitous Girsanov theory of change of measures offers a clear platform familiar to stochastic analysts.

Let us assume for the sake of definiteness that we are trying to estimate the value of an expectation of the form $I_f = \mathbb{E}\{f\} = \mathbb{E}^{\mathbb{P}}\{f\}$ where $\mathbb{P}$ denotes the d-dimensional Wiener measure, $W = \{W_t\}_{t\in[0,T]}$ the coordinate process on Wiener's space (with finite horizon T), and f a functional of the Wiener sample path (up to time T). The typical IS and IPS procedures used to compute approximations of I can be summarized as follows:

In the case of **Importance Sampling**, we change probability measure using Girsanov's theorem. Denoting by $\widetilde{\mathbb{P}}$ the *twisted* distribution on Wiener's space given by its density with respect to Wiener's measure

$$\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} = \mathcal{E}\left(\int_0^T h'(W_t)dW_t\right) = \exp\left(\int_0^T h'(W_t)dW_t - \frac{1}{2}\int_0^T (h'(W_t))^2dt\right)$$

where h is a differentiable function with $h(0) = 0$, we can write (assuming h differentiable)

$$I_f = \widetilde{\mathbb{E}}\left\{f(W)\exp\left[-\int_0^T h'(W_t)d\widetilde{W}_t - \frac{1}{2}\int_0^T (h'(W_t))^2dt\right]\right\}.$$

where $\widetilde{\mathbb{E}}$ stands for expectation under $\widetilde{\mathbb{P}}$ and $\widetilde{W} = W - \int_0^{\cdot} h'(W_t)dt$. The strategy is then to generate Monte Carlo samples $\widetilde{\omega}^1, \cdots, \widetilde{\omega}^m$ from the twisted distribution $\widetilde{\mathbb{P}}$ and to compute the estimate

$$\begin{aligned}\hat{I}_f &= \frac{1}{m}\sum_{j=1}^m f\left(\omega^j\right)\exp\left[-\int_0^T h'(\omega_t^j)d\widetilde{\omega}_t^j - \frac{1}{2}\int_0^T (h'(\omega_t^j))^2dt\right]\\ &= \frac{1}{m}\sum_{j=1}^m f\left(\omega^j\right)\exp\left[-h(\omega_T^j) + \frac{1}{2}\int_0^T (h'(\omega_t^j)^2 - h''(\omega_t^j))dt\right]\end{aligned}$$

(26)

in which $\omega = \widetilde{\omega} + \int_0^{\cdot} h'(\omega_t)dt$, and where we postulated h twice differentiable in the second line.

In the **Interacting Particles System** approach, the change of measure is done by ways of the Feynman-Kac *twisted distributions* $\widetilde{\mathbb{P}}$ defined from a potential function V by:

$$\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} = \frac{e^{\int_0^T V(W_s)ds}}{\mathbb{E}\{e^{\int_0^T V(W_s)ds}\}}.$$

In this way, the expectation to be computed is expressed as a Feynman-Kac expectation

$$I_f = \widetilde{\mathbb{E}}\{f(W)e^{-\int_0^T V(W_s)ds}\}\mathbb{E}\{e^{\int_0^T V(W_s)ds}\}. \tag{27}$$

It is well known from the classical theory of Markov processes that multiplicative functionals of the Feynman-Kac type correspond to *killing and branching*. In the discrete time setting, these path transformations take the form of a re-sampling preceding a normal one-step transition of the Markov chain ("Selection and Mutation" in our discussion of the IPS algorithm in section 2.4). In the present context, this means that the estimate $\hat{I}_f$ related to the representation (27) can in fact be computed by generating Monte Carlo samples *from the original distribution* $\mathbb{P}$ only. As discussed at length in the paper, this represent, in many applications, an important advantage of IPS over IS.

References

[1] T.R. Bielecki, S. Crépey, and M. Jeanblanc. Up and down credit risk. *Quantitative Finance*, Forthcoming.

[2] T.R. Bielecki, S. Crépey, M. Jeanblanc, and M. Rutkowski. Valuation of basket credit derivatives in the credit migrations environment. volume Handbook in Operations Research and Management Science: Financial Engineering, pages 471–709. North Holland, 2007.

[3] R. Carmona, J.-P.Fouque, and D. Vestal. Interacting particle systems for the computation of cdo tranche spreads with rare defaults. *Finance & Stochastics*, 2009.

[4] R. Cont and A. Minca. Recovering portfolio default intensities implied by cdo quotes. Technical report, 2008.

[5] R. Frey and J. Backhaus. Pricing and hedging of portfolio credit derivatives with interacting default intensities. *International Journal of Theoretical and Applied Finance*, 11:611–634, 2008.

[6] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer-Verlag, 2004.

[7] P. Glasserman. Tail approximations for portfolio credit risk. *Journal of Derivatives*, pages 24–42, Winter 2004.

[8] P. Glasserman, W. Kang, and P. Shahabuddin. Large deviations of multifactor portfolio credit risk. *Mathematical Finance*, 17:345–379, 2007.

[9] P. Glasserman and J. Li. Importance sampling for portfolio credit risk. *Management Science*, 51:1643–1656, 2005.

[10] A. Herbertsson. Pricing portfolio credit derivatives, 2007.

[11] A. Johansen, P. Del Moral, and A. Doucet. Sequential monte carlo samplers for rare event estimation. Technical report, University of Cambridge, 2005.

[12] J.P. Laurent, A. Cousin, and J.D. Fermanian. Hedging default risks of cdos in markovian contagion models. *Quantitative Finance*, Forthcoming.

[13] P. Del Moral. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer-Verlag, 2004.

[14] P. Del Moral and J. Garnier. Genealogical particle analysis of rare events. *Annals of Applied Probability*, 15:2496–2534, 2005.
[15] W. Morokoff. Risk Analysis of CPDOs. Technical report, 2007.
[16] R. Sircar and E. Papageorgiou. Multiscale intensity models and name grouping for valuation of multi-name credit derivatives. *Applied Mathematical Finance*, 2009.

E-mail address: rcarmona@princeton.edu