Proceedings of the 2012 Winter Simulation Conference
C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, eds.

SIMULATION OF COALESCENCE WITH STRATIFIED SAMPLING

Rami El Haddad

Université Saint-Joseph
BP 11-514 Riad El Solh
Beyrouth 1107 2050, LIBAN

Rana Fakhereddine

Université de Savoie
LAMA, UMR 5127, Campus universitaire
73376 Le Bourget-du-Lac, FRANCE

Christian Lécot

Université de Savoie
LAMA, UMR 5127, Campus universitaire
73376 Le Bourget-du-Lac, FRANCE

Arthur Soucemarianadin

Université de Grenoble
LEGI, UMR 5519, Domaine universitaire
38400 Saint-Martin-d'Hères, FRANCE

Moussa Tembely

Concordia University
Department of Mechanical and Industrial Engineering
1455 boulevard De Maisonneuve Ouest, Montréal (Québec) H3G 1M8, CANADA

ABSTRACT

We analyze a stratified strategy for numerical integration and for simulation of coalescence. We use random points which are more evenly distributed in the unit cube than usual pseudo-random numbers. They are constructed so that only one point of the set lies in specific sub-intervals of the cube. This property leads to an improved convergence rate for the variance, when they are used for integrating indicator functions. A bound for the variance is proved and assessed through a numerical experiment. We also devise a Monte Carlo algorithm for the simulation of the coagulation equation. We start with an initial population of particles whose sizes are sampled from some initial distribution, and these sizes evolve according to the coalescence dynamics; the random numbers used are the stratified points described above. The results of some numerical experiments show a smaller variance, when compared to a Monte Carlo simulation using plain random samples.

1 INTRODUCTION

Coagulation models have applications in domains of technology and engineering where sprays are used. Sprays are flows produced by the ejection of liquid drops from so called atomizers and are used in many industrial fields such as biotechnology, pharmacology, electronic printing or fuel cell manufacturing. The efficiency of atomization depends on the spray characteristics, among which drop size distribution is one of the most important. The classical method for modeling drop size distribution is generally empirical as one just fits a curve through the data collected for a wide range of atomizers and operating conditions. This approach is time-independent, but collision and coalescence in the spray may modify the drop size distribution in a transient way. The coagulation equation can be employed to predict the evolution of drop size distribution.

Models of coalescence were introduced by von Smoluchowski (1916) in his work on coagulation processes in colloids. He proposed the following infinite system of differential equations for the evolution

978-1-4673-4781-5/12/$31.00 ©2012 IEEE

of the number $N_0 c(i,t)$ of clusters of mass $i = 1,2,3\ldots$

$$\frac{\partial c}{\partial t}(i,t) = \frac{1}{2}\sum_{1\le j<i} K(i-j,j)c(i-j,t)c(j,t) - \sum_{j\ge 1} K(i,j)c(i,t)c(j,t). \tag{1}$$

Here N_0 is the total number of clusters at time $t = 0$, so that $\sum_{i\ge 1} c(i,0) = 1$, and $K(i,j)$ is the coagulation kernel. The mass i is discrete and belongs to $\mathbb{N}^* := \{1,2,3,\ldots\}$; so a cluster of mass i consists of i elements of unit mass. Numerical solution of the Smoluchowski's coagulation equation is a difficult task for deterministic methods, and stochastic algorithms have been proposed by Babovsky (1999) and Eibeck and Wagner (2000). The Monte Carlo (MC) schemes take a system of test particles which interact and form larger particles according to the dynamics described in (1). Time is discretized and at each time step, random numbers are used to find out which particles interact and to determine the size of the new particles. Despite the versatility of MC methods, a drawback is their slow convergence. An approach to acceleration is to change the choice of random numbers used. Quasi-Monte Carlo (QMC) methods use quasi-random numbers instead of pseudo-random numbers. The efficiency of a QMC method depends on the quality of the quasi-random points that are used. These points should form a low-discrepancy point set. We recall from Niederreiter (1992) some basic notations and concepts. If $s \ge 1$ is a fixed dimension, then $I^s := [0,1)^s$ is the s-dimensional half-open unit cube and λ_s denotes the s-dimensional Lebesgue measure. For a point set U consisting of $u_1,\ldots,u_N \in I^s$ and for a Lebesgue-measurable subset A of I^s we define th *local discrepancy* by $D_N(A,U) := N^{-1}\sum_{k=1}^N 1_A(u_k) - \lambda_s(A)$, where 1_A is the indicator function of A. The *discrepancy* of the point set U is defined by $D_N(U) := \sup_A |D_N(A,U)|$, the supremum being taken over all subintervals of I^s. The idea of (t,m,s)-nets is to consider point sets U for which $D_N(A,U) = 0$ for a large family of intervals A. Such point sets should have a small discrepancy. For an integer $b \ge 2$, an interval of the form $I_\ell := \prod_{i=1}^s [(\ell_i - 1)b^{-d_i}, \ell_i b^{-d_i})$, with integers $d_i \ge 0$ and $1 \le \ell_i \le b^{d_i}$ for $1 \le i \le s$ is called an *elementary interval in base b*. If $0 \le t \le m$ are integers, a (t,m,s)*-net in base b* is a point set U consisting of b^m points in I^s such that $D_N(I_\ell,U) = 0$ for every elementary interval I_ℓ in base b with measure $\lambda_s(I_\ell) = b^{t-m}$, i.e., $\sum_{i=1}^s d_i = m-t$. The effectiveness of QMC methods has limitations. First, they are valid for integration problems, they may not be directly applicable to simulations, due to the correlations between the points of a quasi-random sequence. This problem can be overcome by writing the desired result as an integral. This leads to a second limitation: the improved accuracy of QMC methods may be lost for problems in which the integrand is not smooth. It is necessary to take special measures to make optimal use of the greater uniformity associated with quasi-random sequences. This was achieved by Lécot (1991) and El-Haddad, Lécot, and L'Ecuyer (2008) through the additional step of reordering the particles at each time step. A third drawback of the QMC approach is that it is often difficult to estimate the error. There is no confidence interval and the deterministic error bounds are usually very pessimistic. A possibility is to use randomized QMC algorithms where a random parameter is introduced in the construction of the quasi-random points: this approach was initiated by Cranley and Patterson (1976) for numerical integration; in the same context, see the review article of L'Ecuyer and Lemieux (2000). Randomized QMC algorithms for Markov chains have been introduced and studied by L'Ecuyer, Lécot, and Tuffin (2006), L'Ecuyer, Lécot, and Tuffin (2008) and L'Ecuyer, Lécot, and L'Archevêque-Gaudet (2009).

The aim of the present paper is to construct and analyze a stratified MC strategy for approximate integration and for simulation of coalescence. The random points used retain some properties of nets, as described before. The difference with the previous works of L'Ecuyer, Lécot, and Tuffin (2006) and L'Ecuyer, Lécot, and Tuffin (2008) is twofold. Firstly we are not focusing on variance reduction; we are interested here in estimating the variance as a function of the number of random points. Secondly, we compare theoretical bounds and numerical estimates of the variance, with the same properties of the random points used. In Section 2 we analyze a MC method using stratified sampling for approximate integration. Since we have experienced that some simulation methods can be reduced to numerical integration of indicator functions of sub-domains of I^s, we consider here the case of the integral of an indicator function, and we analyze the variance of the approximate measure of a subset of I^s. In Section 3, we propose a

stratified MC method for Smoluchowski's coagulation equation, using the stratified points. Each step of the simulation is formulated as a numerical integration in I^3. In order to benefit from the great uniformity of stratified points, the particles are sorted by increasing size before performing MC quadrature. We carry out numerical experiments based on a comparison of the method with a standard MC scheme.

2 STRATIFIED MONTE CARLO INTEGRATION

Many applications require the evaluation of integrals. The dimensional effect is the phenomenon whereby deterministic quadrature rules deteriorate in performance as the dimension increases. Monte Carlo (MC) methods do not suffer so much from the dimensional effect. Stratified sampling and Latin hypercube sampling (LHS) are techniques for increasing efficiency of MC methods: see, e.g., (Haber 1966; McKay, Beckman, and Conover 1979; Cheng and Davenport 1989; Tong 2006). We are interested here in sets of $N = n^s$ random points in I^s such that

(P1) in every interval $I_\ell := \prod_{i=1}^s [(\ell_i - 1)n^{-1}, \ell_i n^{-1})$ (for $1 \le \ell_i \le n$ and $1 \le i \le s$) lies only one point,
(P2) in every interval $I^{i-1} \times [(k-1)n^{-s}, kn^{-s}) \times I^{s-i}$ (for $1 \le i \le s$ and $1 \le k \le n^s$) lies only one point.

See Figure 1 for an example in dimension $s = 2$. The proposed method is a combination of stratification in the N cubic boxes (property P1) and of LHS (property P2), in the sense that the N random points have the two properties simultaneously. These points are evenly spread over the unit cube, and retain some properties of the nets of QMC methods; it is hoped that they can lead to a variance reduction.

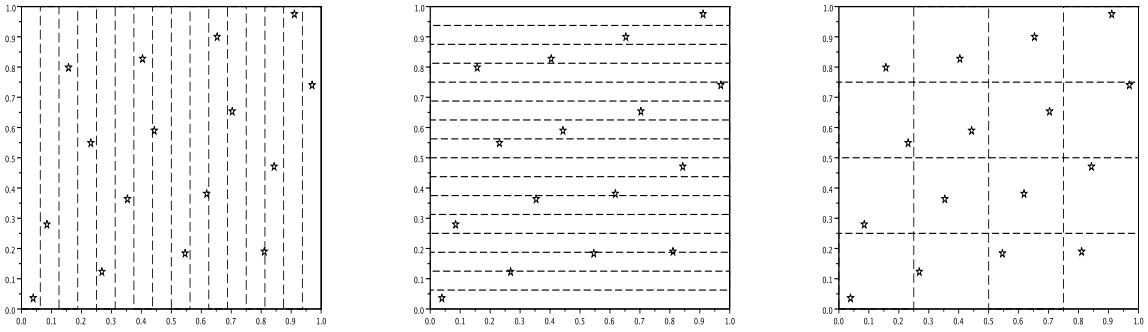

Figure 1: An example of a stratified sample of 4^2 points ($\star$) in dimension $s = 2$.

In the following, we restrict ourselves to $s = 3$ for simplicity (in addition, we encounter only three-dimensional integrals in Section 3). In this cases, $N = n^3$ and the sets may be generated as follows. Let $\sigma^1, \sigma^2, \sigma^3$ be random bijections $\{1, \dots, n\}^2 \to \{1, \dots, n^2\}$ and $u_\ell^1, u_\ell^2, u_\ell^3$ be uniform random variables over I, for $\ell = (\ell_1, \ell_2, \ell_3)$ with $1 \le \ell_i \le n$; we assume that all these variables are independent. Then we set

$$V_\ell := \left(\frac{\ell_1 - 1}{n} + \frac{\sigma^1(\ell_2, \ell_3) - 1}{n^3} + \frac{u_\ell^1}{n^3}, \frac{\ell_2 - 1}{n} + \frac{\sigma^2(\ell_1, \ell_3) - 1}{n^3} + \frac{u_\ell^2}{n^3}, \frac{\ell_3 - 1}{n} + \frac{\sigma^3(\ell_1, \ell_2) - 1}{n^3} + \frac{u_\ell^3}{n^3} \right), \quad (2)$$

for $\ell = (\ell_1, \ell_2, \ell_3)$ with $1 \le \ell_i \le n$.

If $g : I^3 \to \mathbb{R}$ is a square-integrable function, we want to evaluate

$$\mathscr{I} := \int_{I^3} g(x) d\lambda_3(x).$$

For the usual MC approximation, let $\{U_1, \dots, U_N\}$ be N independent random variables uniformly distributed over I^3. Then $X := N^{-1} \sum_{k=1}^N g \circ U_k$ is an unbiased estimator of $\mathscr{I}$ (crude MC estimator). Now let

$Y := N^{-1}\sum_\ell g\circ V_\ell$ be the stratified estimator of $\mathscr{I}$. We have

$$E[g\circ V_\ell] = \frac{1}{n^6}\sum_m \int_{I^3} g\left(\frac{\ell_1-1}{n}+\frac{m_1-1}{n^3}+\frac{u_1}{n^3}, \frac{\ell_2-1}{n}+\frac{m_2-1}{n^3}+\frac{u_2}{n^3}, \frac{\ell_3-1}{n}+\frac{m_3-1}{n^3}+\frac{u_3}{n^3}\right) du,$$

where the sum extends over all triples $m=(m_1,m_2,m_3)$ of integers with $1\le m_i\le n^2$. Hence

$$E[g\circ V_\ell] = n^3\int_{I_\ell} g(u)du, \tag{3}$$

consequently $E[Y]=\mathscr{I}$ and Y is another unbiased estimator of $\mathscr{I}$. We specialize now to the case when g is an indicator function: let A be a measurable subset of I^3 and $g:=1_A$. Then

$$\mathrm{Var}(X) = \frac{1}{N}\lambda_3(A)\big(1-\lambda_3(A)\big) \le \frac{1}{4N}.$$

We want to estimate $\mathrm{Var}(Y)$. For $\ell=(\ell_1,\ell_2,\ell_3)$ with $1\le\ell_i\le n$ and $m=(m_1,m_2,m_3)$ with $1\le m_i\le n^2$, let $I_{\ell,m} := \prod_{i=1}^3[(\ell_i-1)n^{-1}+(m_i-1)n^{-3}, (\ell_i-1)n^{-1}+m_in^{-3})$; then $I_\ell = \bigcup_m I_{\ell,m}$.

Proposition 1 Let $A\subset I^3$ be such that

$$A=\{u\in I^3 : u_1<f_1(u_2,u_3)\}=\{u\in I^3 : u_2<f_2(u_1,u_3)\}=\{u\in I^3 : u_3<f_3(u_1,u_2)\},$$

where all f_i are Lipschitz continuous functions $\bar{I}^2\to\bar{I}$, with a Lipschitz constant c (for the maximum norm). Let $\{V_\ell : 1\le\ell_1\le n, 1\le\ell_2\le n, 1\le\ell_3\le n\}$ be defined by (2). If $Y:=N^{-1}\sum_\ell 1_A\circ V_\ell$, then

$$\mathrm{Var}(Y)\le\frac{c+2+6(c+2)^2}{N^{4/3}}.$$

Proof. We can write

$$\mathrm{Var}(Y) = \frac{1}{N^2}\sum_\ell \mathrm{Var}(1_A\circ V_\ell) + \frac{1}{N^2}\sum_{\ell\neq\ell'}\mathrm{Cov}(1_A\circ V_\ell, 1_A\circ V_{\ell'}).$$

Using (3), we obtain:

$$\frac{1}{N^2}\sum_\ell \mathrm{Var}(1_A\circ V_\ell) = \sum_\ell W_0(\ell),$$

where $W_0(\ell)=\lambda_3(A\cap I_\ell)n^{-3}-(\lambda_3(A\cap I_\ell))^2$ and so

$$\sum_\ell |W_0(\ell)| \le \frac{1}{n^6}\#\{\ell : I_\ell\not\subset A \text{ and } I_\ell\cap A\neq\emptyset\}.$$

Here, $\#E$ denotes the number of elements of a set E. We have the following analogues of (3). If $g:I^3\to\mathbb{R}$ is a square-integrable function, then for $\ell\neq\ell'$:

- if $\ell_1 \neq \ell_1', \ell_2 = \ell_2', \ell_3 = \ell_3'$,

$$\begin{aligned} E[g\circ V_\ell\ g\circ V_{\ell'}] = \frac{1}{n^6(n^2-1)^2} & \sum_{\substack{m_1=m_1'\\ m_2\neq m_2', m_3\neq m_3'}} \\ & \int_{I^3} g\left(\frac{\ell_1-1}{n}+\frac{m_1-1}{n^3}+\frac{u_1}{n^3}, \frac{\ell_2-1}{n}+\frac{m_2-1}{n^3}+\frac{u_2}{n^3}, \frac{\ell_3-1}{n}+\frac{m_3-1}{n^3}+\frac{u_3}{n^3}\right) du \\ & \cdot \int_{I^3} g\left(\frac{\ell_1'-1}{n}+\frac{m_1'-1}{n^3}+\frac{u_1'}{n^3}, \frac{\ell_2'-1}{n}+\frac{m_2'-1}{n^3}+\frac{u_2'}{n^3}, \frac{\ell_3'-1}{n}+\frac{m_3'-1}{n^3}+\frac{u_3'}{n^3}\right) du' \\ = \frac{n^{12}}{(n^2-1)^2} & \sum_{\substack{m_1=m_1'\\ m_2\neq m_2', m_3\neq m_3'}} \int_{I_{\ell,m}} g(u)du \int_{I_{\ell',m'}} g(u')du', \end{aligned}$$

  and similar expressions for $\ell_1 = \ell_1', \ell_2 \neq \ell_2', \ell_3 = \ell_3'$ or $\ell_1 = \ell_1', \ell_2 = \ell_2', \ell_3 \neq \ell_3'$,
- if $\forall i \neq j\ (\ell_i, \ell_j) \neq (\ell_i', \ell_j')$,

$$\begin{aligned} E[g\circ V_\ell\ g\circ V_{\ell'}] = \frac{1}{n^6(n^2-1)^3} & \sum_{m_1\neq m_1', m_2\neq m_2', m_3\neq m_3'} \\ & \int_{I^3} g\left(\frac{\ell_1-1}{n}+\frac{m_1-1}{n^3}+\frac{u_1}{n^3}, \frac{\ell_2-1}{n}+\frac{m_2-1}{n^3}+\frac{u_2}{n^3}, \frac{\ell_3-1}{n}+\frac{m_3-1}{n^3}+\frac{u_3}{n^3}\right) du \\ & \cdot \int_{I^3} g\left(\frac{\ell_1'-1}{n}+\frac{m_1'-1}{n^3}+\frac{u_1'}{n^3}, \frac{\ell_2'-1}{n}+\frac{m_2'-1}{n^3}+\frac{u_2'}{n^3}, \frac{\ell_3'-1}{n}+\frac{m_3'-1}{n^3}+\frac{u_3'}{n^3}\right) du' \\ = \frac{n^{12}}{(n^2-1)^3} & \sum_{m_1\neq m_1', m_2\neq m_2', m_3\neq m_3'} \int_{I_{\ell,m}} g(u)du \int_{I_{\ell',m'}} g(u')du', \end{aligned}$$

Hence we obtain

$$\begin{aligned} \frac{1}{N^2}\sum_{\ell\neq\ell'} \mathrm{Cov}(1_A\circ V_\ell, 1_A\circ V_{\ell'}) = & \sum_{\substack{\ell_1\neq\ell_1'\\ \ell_2=\ell_2', \ell_3=\ell_3'}} W_1(\ell,\ell') + \sum_{\substack{\ell_2\neq\ell_2'\\ \ell_1=\ell_1', \ell_3=\ell_3'}} W_2(\ell,\ell') + \sum_{\substack{\ell_3\neq\ell_3'\\ \ell_1=\ell_1', \ell_2=\ell_2'}} W_3(\ell,\ell') \\ & + \sum_{(\ell_i,\ell_j)\neq(\ell_i',\ell_j')} W_4(\ell,\ell'), \end{aligned}$$

where the last sum is over all $i \neq j$. Here

$$W_1(\ell,\ell') = \frac{n^6}{(n^2-1)^2} \sum_{\substack{m_1=m_1'\\ m_2\neq m_2', m_3\neq m_3'}} \lambda_3(A\cap I_{\ell,m})\lambda_3(A\cap I_{\ell',m'}) - \lambda_3(A\cap I_\ell)\lambda_3(A\cap I_{\ell'}),$$

and similar expressions for $W_2(\ell,\ell')$ or $W_3(\ell,\ell')$,

$$W_4(\ell,\ell') = \frac{n^6}{(n^2-1)^3} \sum_{m_1\neq m_1', m_2\neq m_2', m_3\neq m_3'} \lambda_3(A\cap I_{\ell,m})\lambda_3(A\cap I_{\ell',m'}) - \lambda_3(A\cap I_\ell)\lambda_3(A\cap I_{\ell'}).$$

And so

$$\sum_{\substack{\ell_1\neq\ell_1'\\ \ell_2=\ell_2', \ell_3=\ell_3'}} |W_1(\ell,\ell')| \leq \frac{1}{n^6}\#\{(\ell,\ell') : \ell_1\neq\ell_1',\ \ell_2=\ell_2',\ \ell_3=\ell_3',\ I_\ell\not\subset A,\ I_\ell\cap A\neq\emptyset,\ I_{\ell'}\not\subset A,\ I_{\ell'}\cap A\neq\emptyset\},$$

Table 1: Calculation of $\lambda_s(A)$: convergence order α of the variance as a function of N.

dimension s	MC	stratified MC	theoretical bound
2	0.99	1.49	1.50
3	1.00	1.33	1.33

and similar expressions for the sums of $|W_2(\ell,\ell')|$ or $|W_3(\ell,\ell')|$,

$$\sum_{(\ell_i,\ell_j)\neq(\ell'_i,\ell'_j)} |W_4(\ell,\ell')| \leq \frac{3}{n^8}\#\{(\ell,\ell') : \forall i\neq j\ (\ell_i,\ell_j)\neq(\ell'_i,\ell'_j),\ I_\ell\not\subset A,\ I_\ell\cap A\neq\emptyset,\ I_{\ell'}\not\subset A,\ I_{\ell'}\cap A\neq\emptyset\}.$$

We have the following inferences:

- if $I_\ell\not\subset A$, there exists $(u_{\ell,2},u_{\ell,3})\in[(\ell_2-1)n^{-1},\ell_2n^{-1})\times[(\ell_3-1)n^{-1},\ell_3n^{-1})$ such that
$$nf_1(u_{\ell,2},u_{\ell,3})<\ell_1,$$
- if $I_\ell\cap A\neq\emptyset$, there exists $(v_{\ell,2},v_{\ell,3})\in[(\ell_2-1)n^{-1},\ell_2n^{-1})\times[(\ell_3-1)n^{-1},\ell_3n^{-1})$ such that
$$\ell_1<nf_1(v_{\ell,2},v_{\ell,3})+1.$$

and similar assertions with f_2 or f_3. Hence

$$\#\{\ell : I_\ell\not\subset A \text{ and } I_\ell\cap A\neq\emptyset\}\leq n^2(c+2),$$
$$\#\{(\ell,\ell') : \ell_1\neq\ell'_1,\ \ell_2=\ell'_2,\ \ell_3=\ell'_3,\ I_\ell\not\subset A,\ I_\ell\cap A\neq\emptyset,\ I_{\ell'}\not\subset A,\ I_{\ell'}\cap A\neq\emptyset\}\leq n^2(c+2)^2,$$

and same bounds when $\ell_1=\ell'_1,\ \ell_2\neq\ell'_2,\ \ell_3=\ell'_3$ or $\ell_1=\ell'_1,\ \ell_2=\ell'_2,\ \ell_3\neq\ell'_3$,

$$\#\{(\ell,\ell') : \forall i\neq j\ (\ell_i,\ell_j)\neq(\ell'_i,\ell'_j),\ I_\ell\not\subset A,\ I_\ell\cap A\neq\emptyset,\ I_{\ell'}\not\subset A,\ I_{\ell'}\cap A\neq\emptyset\}\leq (n^2(c+2))^2.$$

So the result follows. □

The hypothesis on the boundary of A is easily expressed, but quite restrictive. The conclusion remains certainly valid under not so stringent conditions (see the example below). Restrictive constraints were also imposed by L'Ecuyer, Lécot, and Tuffin (2008) for a proof in dimension $s=2$, in the context of the simulation of Markov chains. In the simple case considered here, our variance bound represents a gain of the factor $N^{-1/3}$ as opposed to simple MC. This may be generalized in dimension s, where the variance bound is of order $\mathcal{O}(N^{-1-1/s})$. The same order for the bound has been established by L'Ecuyer, Lécot, and Tuffin (2008). The gain is of diminishing importance as s becomes large and limits the use of the present stratification to problems of moderate dimension. This is precisely the case in some MC particle simulations, such as the scheme proposed in Section 3. We use a simple example to illustrate the previous analysis. We consider the subset of the unit ball $A:=\{u\in I^s : \|u\|_2<1\}$, where we denote by $\|u\|_2$ the Euclidean norm of u. In order to estimate the variance of the MC and stratified MC approximations, we can replicate the quadrature independently M times and compute the sample variance. We use $M=100,\ldots,1\,000$ and we only see small differences between the estimates. The results (for $M=1\,000$) are displayed in Figure 2. It is clear that the better accuracy due to the proposed method goes beyond an improved convergence order, at least in this very simple experiment. Assuming $\mathrm{Var}=\mathcal{O}(N^{-\alpha})$, linear regression can be used to evaluate α and the outputs are listed in Table 1. The values obtained are very close to the orders of the bounds given in Proposition 1 (despite the fact that the hypothesis is not satisfied) and its generalization to arbitrary dimensions.

The stratified sampling strategy, with only one point in the $N=n^s$ cubic boxes (property P1) without the LHS property (P2), has been analyzed by El-Haddad, Fakhereddine, and Lécot (2013). The same order $\mathcal{O}(N^{-1-1/s})$ for the variance bound is obtained. But this approach fails to improve crude MC results when it is applied to particle simulation, such as the algorithm proposed in the next section.

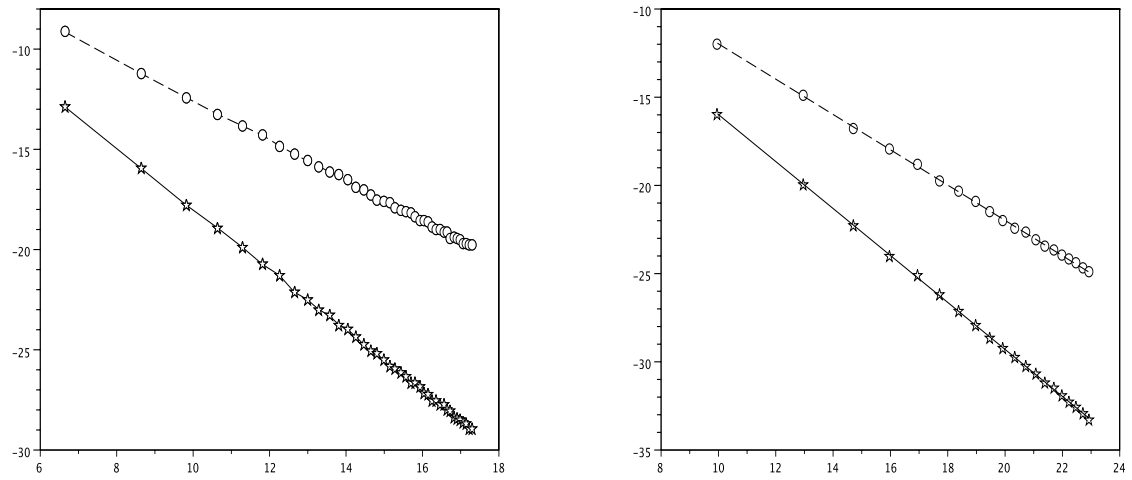

Figure 2: Sample variance of $M = 1\,000$ independent copies of the calculation of $\lambda_s(A)$ as a function of N; $s = 2$ and $10^2 \leq N \leq 400^2$ (left), $s = 3$ and $10^3 \leq N \leq 200^3$ (right). Log-log plots of MC ($\circ$) and stratified MC ($\star$) outputs.

3 SIMULATION OF COAGULATION

We go back to Equation (1). We assume that the coagulation kernel $K(i,j)$ is nonnegative and symmetric: $K(i,j) \geq 0$ and $K(i,j) = K(j,i)$. Multiplying (1) by i and summing over all i, indicates that mass is conserved

$$\frac{d}{dt} \sum_{i \geq 1} ic(i,t) = 0,$$

provided an interchange of summation order on the right is valid. Let $C_1 := \sum_{i \geq 1} ic(i,0)$; rather than approximating the density of clusters $c(i,t)$, one can approximate the *mass density* $f(i,t) := ic(i,t)/C_1$, which satisfies the following equation for $i = 1,2,3\ldots$

$$\frac{\partial f}{\partial t}(i,t) = \sum_{1 \leq j < i} \widetilde{K}(i-j,j) f(i-j,t) f(j,t) - \sum_{j \geq 1} \widetilde{K}(i,j) f(i,t) f(j,t), \tag{4}$$

where $\widetilde{K}(i,j) := C_1 K(i,j) j^{-1}$. Equation (4) has been used by Babovsky (1999) for constructing a stochastic algorithm for Smoluchowski's coagulation equation. If E is a subset of $\mathbb{N}^* := \{1,2,3,\ldots\}$, we denote by s_E the sequence

$$s_E(i) = \begin{cases} 1 & \text{if } i \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Equation (4) can be given the following form

$$\frac{d}{dt} \sum_{i \geq 1} f(i,t) s_E(i) = \sum_{i,j \geq 1} \widetilde{K}(i,j) f(i,t) f(j,t) \left(s_E(i+j) - s_E(i) \right), \tag{5}$$

for any $E \subset \mathbb{N}^*$. We denote by f_0 the initial mass density. We have $\sum_{i\geq 1} f_0(i) = 1$. We choose an integer $n \geq 1$ and we put $N := n^3$. We write δ_j for the unit mass at j

$$\delta_j(i) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

A point set J^0 of N *particles* $j_1^0, \ldots, j_N^0$ is sampled from the probability distribution f_0. Then

$$f^0 := \frac{1}{N} \sum_{k=1}^{N} \delta_{j_k^0} \approx f_0.$$

The particles with the same j_k^0 do not represent the number of clusters of mass j_k^0 in the system, but rather represent the mass of clusters of mass j_k^0 in the system. In the simple case of a monodisperse initial condition

$$f_0(1) = 1,\ f_0(2) = f_0(3) = \cdots = 0,$$

we set

$$j_1^0 = \cdots = j_N^0 = 1.$$

We assume that the kernel $\widetilde{K}(i,j)$ is bounded and we put

$$\widetilde{K}^\star := \sup_{i,j\geq 1} \widetilde{K}(i,j).$$

We choose a time step Δt such that $\Delta t \widetilde{K}^\star < 1$. Computations are still possible for unbounded kernels: see the second numerical experiment below. For an integer p, we set $t_p := p\Delta t$ and $f_p(i) := f(i,t_p)$. If we assume that we have a point set J^p of N particles $j_1^p, \ldots, j_N^p$ such that

$$f^p := \frac{1}{N} \sum_{k=1}^{N} \delta_{j_k^p} \approx f_p, \tag{6}$$

we compute f^{p+1} as follows.

Generating a stratified sample: We compute a new $\{V_\ell : 1 \leq \ell_1 \leq n, 1 \leq \ell_2 \leq n, 1 \leq \ell_3 \leq n\}$, given by (2), independent of all samples used previously.

Relabeling the particles: We order particles by size

$$j_1^p \leq j_2^p \leq \ldots \leq j_N^p.$$

This type of sorting was used by Lécot (1991), and El-Haddad, Lécot, and L'Ecuyer (2008) for a QMC approach and by L'Ecuyer, Lécot, and Tuffin (2006) and L'Ecuyer, Lécot, and Tuffin (2008) in a randomized setting. It guarantees convergence: since the algorithm can be described by a series of numerical integration, the sorting helps in minimizing the amplitude of the jumps of the function to be integrated. Sorting strategies were discussed by L'Ecuyer, Lécot, and L'Archevêque-Gaudet (2009).

Coagulation: Equation (5) is discretized using the Euler scheme, and we define g^{p+1} by

$$\frac{1}{\Delta t}\left(\sum_{i\geq 1} g^{p+1}(i) s_E(i) - \sum_{i\geq 1} f^p(i) s_E(i)\right) = \sum_{i,j\geq 1} \widetilde{K}(i,j) f^p(i) f^p(j) \left(s_E(i+j) - s_E(i)\right),$$

for any $E \subset \mathbb{N}^*$, and so

$$\sum_{i\geq 1} g^{p+1}(i) s_E(i) = \frac{1}{N} \sum_{k_1=1}^{N} \left(1 - \frac{\Delta t}{N} \sum_{k_2=1}^{N} \widetilde{K}(j_{k_1}^p, j_{k_2}^p)\right) s_E(j_{k_1}^p) + \frac{\Delta t}{N^2} \sum_{k_1=1}^{N} \sum_{k_2=1}^{N} \widetilde{K}(j_{k_1}^p, j_{k_2}^p) s_E(j_{k_1}^p + j_{k_2}^p). \tag{7}$$

The probability g^{p+1} approximates f_{p+1}, but is not uniform, like f^p. We recover this kind of approximation if we use a MC estimate. We first express the right-hand side of (7) as the three-dimensional integral of an indicator function $\mathscr{I}_E^{p+1}$.

Integration: Let $1_{k_1,k_2}$ be the indicator function of $[(k_1-1)N^{-1},k_1N^{-1})\times[(k_2-1)N^{-1},k_2N^{-1})$, and $I^p_{k_1,k_2}$ denote the indicator function of $\left[0,\Delta t\widetilde{K}(j^p_{k_1},j^p_{k_2})\right)$. For any $E\subset\mathbb{N}^*$, define

$$\mathscr{I}_E^{p+1}(u):=\sum_{k_1=1}^{N}\sum_{k_2=1}^{N}1_{k_1,k_2}(u_1,u_2)\left(\left(1-I^p_{k_1,k_2}(u_3)\right)s_E(j^p_{k_1})+I^p_{k_1,k_2}(u_3)s_E(j^p_{k_1}+j^p_{k_2})\right),$$

for $u=(u_1,u_2,u_3)\in I^3$. Then

$$\sum_{i\geq1}g^{p+1}(i)s_E(i)=\int_{I^3}\mathscr{I}_E^{p+1}(u)du.$$

We obtain f^{p+1} by

$$\forall E\subset\mathbb{N}^*\quad\sum_{i\geq1}f^{p+1}(i)s_E(i)=\frac{1}{N}\sum_{\ell}\mathscr{I}_E^{p+1}(V_\ell).$$

It is possible to summarize the calculation on a time step as follows. If $u\in I$, let $k(u):=\lfloor Nu\rfloor$. Then, for $\ell=(\ell_1,\ell_2,\ell_3)$ with $1\leq\ell_i\leq n$, we have:

$$j^{p+1}_{k(V_{\ell,1})}=\begin{cases}j^p_{k(V_{\ell,1})}+j^p_{k(V_{\ell,2})} & \text{if } V_{\ell,3}<\Delta t\widetilde{K}(j^p_{k(V_{\ell,1})},j^p_{k(V_{\ell,2})}),\\ j^p_{k(V_{\ell,1})} & \text{otherwise.}\end{cases}$$

The numbers $V_{\ell,1}$ and $V_{\ell,2}$ select particles; particle $k(V_{\ell,1})$ has for coagulation partner particle $k(V_{\ell,2})$, and the coagulation probability is $P_c:=\Delta t\widetilde{K}(j^p_{k(V_{\ell,1})},j^p_{k(V_{\ell,2})})$. Then $V_{\ell,3}$ is used to select an event:

- if $0\leq V_{\ell,3}<P_c$, particles $k(V_{\ell,1})$ and $k(V_{\ell,2})$ coalesce,
- if $P_c\leq V_{\ell,3}<1$, no coalescence occurs.

Because of the LHS property (P2), each particle is considered once for possible coalescence, during each time step.

The corresponding MC scheme is as follows: there is no reordering of particles and, for $1\leq k\leq N$,

$$j^{p+1}_k=\begin{cases}j^p_k+j^p_{L_k} & \text{if } U_k<\Delta t\widetilde{K}(j^p_k,j^p_{L_k}),\\ j^p_k & \text{otherwise.}\end{cases}$$

Here $L_1,\dots,L_N$ are independent random samples drawn from the uniform distribution on $\{1,\dots,N\}$, while $U_1,\dots,U_N$ are independent random samples drawn from the uniform distribution on I.

We want to compare the previous stratified scheme with a usual MC algorithm. It has long been recognized that three particular kernels $K(i,j)$ are mathematically tractable (Aldous 1999): for monodisperse initial configuration $f_0=\delta_1$, explicit solutions of Smoluchowski's coagulation equation are available. In the following we restrict our consideration to the kernels $K(i,j)=1$ and $K(i,j)=i+j$ (note that the latter does not satisfy the hypothesis of bounded $\widetilde{K}$). Let us define the moments of the solution:

$$C_0(t):=\sum_{i\geq1}c(i,t)=\sum_{i\geq1}\frac{f(i,t)}{i}\quad\text{and}\quad C_2(t):=\sum_{i\geq1}i^2c(i,t)=\sum_{i\geq1}if(i,t).$$

At time t_p, $C_0(t_p)$ and $C_2(t_p)$ are approximated according to (6) as follows:

$$C_0(t_p)\approx\frac{1}{N}\sum_{k=1}^{N}\frac{1}{j^p_k}\quad\text{and}\quad C_2(t_p)\approx\frac{1}{N}\sum_{k=1}^{N}j^p_k$$

and we compare MC and stratified MC strategies.

Table 2: Calculation of $C_0(T)$ (left) and $C_2(T)$ (right) at time $T = 1.0$: convergence order β of the variance as a function of N.

$C_0(T)$	Kernel	MC	stratified MC
	Constant	1.01	1.22
	Linear	1.06	1.15

$C_2(T)$	Kernel	MC	stratified MC
	Constant	0.99	1.28
	Linear	0.78	1.14

- $K(i,j) = 1$. The moments of the exact solution of (4) with monodisperse initial condition are $C_0(t) = 2/(2+t)$ and $C_2(t) = 2+t$.
- $K(i,j) = i+j$. The moments of the exact solution of (4) with monodisperse initial condition are $C_0(t) = \mathrm{e}^{-t}$ and $C_2(t) = \mathrm{e}^{2t}$.

In both cases, we compute $C_0(T)$ and $C_2(T)$ at time $T = 1.0$ with N particles and $P = 4\,000$ time steps. We put $N = (4m)^3$, with $1 \le m \le 8$. For estimating the variance of the MC and stratified MC approximations, we replicate the calculation independently $M = 1\,000$ times and compute the sample variance. The results are displayed in Figure 3 for the constant kernel and in Figure 4 for the linear kernel.

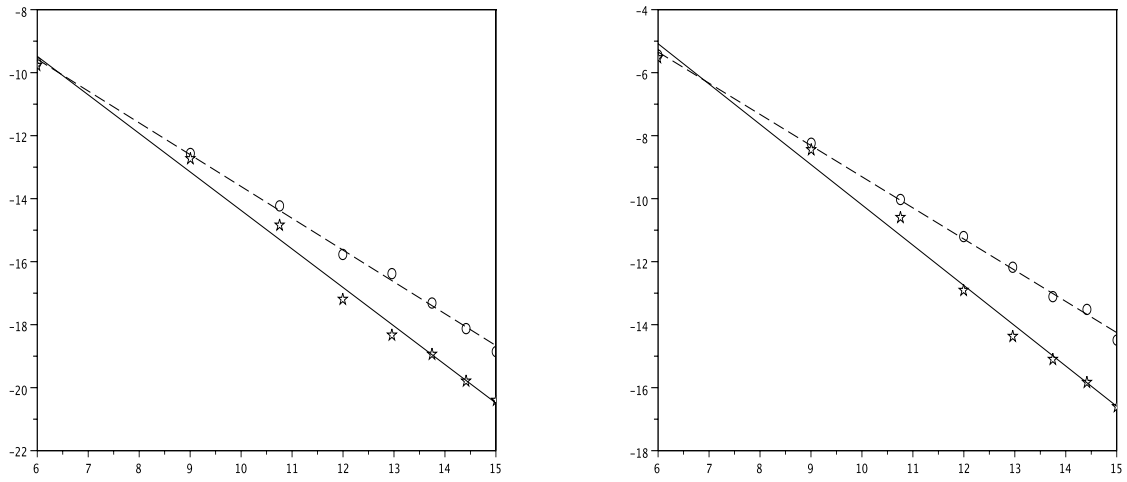

Figure 3: Constant kernel. Sample variance of $M = 1\,000$ independent copies of the calculation of the moments $C_0(T)$ (left) and $C_2(T)$ (right) as a function of N (with $4^3 \le N \le 32^3$). Log-log plots of MC ($\circ$) and stratified MC ($\star$) outputs with linear regression lines.

We see that the stratified MC method achieves a better convergence order of the variance as a function of N. Assuming $\mathrm{Var} = \mathcal{O}(N^{-\beta})$, linear regression can be used to evaluate β and the outputs are listed in Table 2.

4 CONCLUSION

We have proposed a stratified sampling technique that produces random points which are evenly distributed in the unit cube. We have proved that for approximating the measure of special subsets, the technique leads to reduced variance, when compared to usual Monte Carlo. Then we have proposed a procedure for solving Smoluchowski's coagulation equation. The approach is to use the Monte Carlo method to simulate the aggregation of clusters. A sample of test particles is chosen, time is discretized and since we approximate the mass density, the scheme works with a fixed particle number. We have considered an

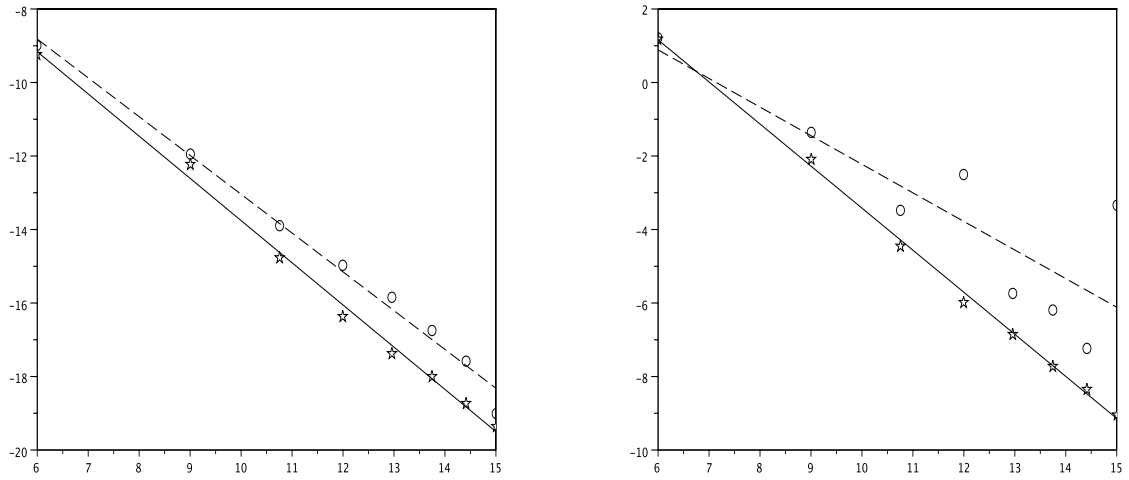

Figure 4: Linear kernel. Sample variance of $M = 1000$ independent copies of the calculation of the moments $C_0(T)$ (left) and $C_2(T)$ (right) as a function of N (with $4^3 \leq N \leq 32^3$). Log-log plots of MC ($\circ$) and stratified MC ($\star$) outputs with linear regression lines.

improvement to this method by using stratified random numbers in the implementation of the algorithm. To make optimal use of the greater uniformity of the points, the particles are reordered by size at each time step. We test our method in two cases where analytic solutions to the Smoluchowski equation are known. In both comparisons, the numerical experiments show that the variance of the simulation using stratified sampling is significantly less than the variance for a standard MC simulation.

Future works include generalizations of the variance bound for numerical integration with stratified points and variance estimates when these points are used for simulation of coalescence.

REFERENCES

Aldous, D. J. 1999. "Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists". *Bernoulli* 5:3–48.

Babovsky, H. 1999. "On a Monte Carlo scheme for Smoluchowski's coagulation equation". *Monte Carlo Methods and Applications* 5:1–18.

Cheng, R. C. H., and T. Davenport. 1989. "The problem of dimensionality in stratified sampling". *Management Science* 35:1278–1296.

Cranley, R., and T. N. L. Patterson. 1976. "Randomization of number theoretic methods for multiple integration". *SIAM Journal on Numerical Analysis* 13:904–914.

Eibeck, A., and W. Wagner. 2000. "An efficient stochastic algorithm for studying coagulation dynamics and gelation phenomena". *SIAM Journal on Scientific Computing* 22:802–821.

El-Haddad, R., R. Fakhereddine, and C. Lécot. 2013. "Stratified Monte Carlo integration". In *Monte Carlo Methods and Applications*, edited by K. K. Sabelfeld and I. Dimov, 105–113. Berlin: De Gruyter.

El-Haddad, R., C. Lécot, and P. L'Ecuyer. 2008. "Quasi-Monte Carlo simulation of discrete-time Markov chains on multidimensional state spaces". In *Monte Carlo and Quasi-Monte Carlo Methods 2006*, edited by A. Keller, S. Heinrich, and H. Niederreiter, 413–429. Berlin: Springer.

Haber, S. 1966. "A modified Monte-Carlo quadrature". *Mathematics of Computation* 20:361–368.

Lécot, C. 1991. "A quasi-Monte Carlo method for the Boltzmann equation". *Mathematics of Computation* 56:621–644.

L'Ecuyer, P., C. Lécot, and A. L'Archevêque-Gaudet. 2009. "On array-RQMC for Markov chains: mapping alternatives and convergence rates". In *Monte Carlo and Quasi-Monte Carlo Methods 2008*, edited by P. L'Ecuyer and A. B. Owen, 485–500. Berlin: Springer.

L'Ecuyer, P., C. Lécot, and B. Tuffin. 2006. "Randomized quasi-Monte Carlo simulation of Markov chains with an ordered state space". In *Monte Carlo and Quasi-Monte Carlo Methods 2004*, edited by H. Niederreiter and D. Talay, 331–342. Berlin: Springer.

L'Ecuyer, P., C. Lécot, and B. Tuffin. 2008. "A randomized quasi-Monte Carlo simulation method for Markov chains". *Operations Research* 56:958–975.

L'Ecuyer, P., and C. Lemieux. 2000. "Variance reduction via lattice rules". *Management Science* 46:1214–1235.

McKay, M. D., R. J. Beckman, and W. J. Conover. 1979. "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code". *Technometrics* 21:239–245.

Niederreiter, H. 1992. *Random Number Generation and Quasi-Monte Carlo Methods*. Philadelphia, Pennsylvania: SIAM.

Tong, C. 2006. "Refinement strategies for stratified sampling methods". *Reliability Engineering & System Safety* 91:1257–1265.

von Smoluchowski, M. 1916. "Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen". *Zeitschrift für Physikalische Chemie* 92:129–168.

AUTHOR BIOGRAPHIES

RAMI EL HADDAD is an assistant professor at the Faculty of Sciences of Saint Joseph University (USJ) of Beirut where he studied mathematics. Since 2009, he is the head of the Master *Actuarial sciences and Finance*. He received his doctorate in 2008 from the Université de Savoie (France) and USJ. His research interests lie in Monte Carlo and quasi-Monte Carlo methods. His email address is rami.haddad@fs.usj.edu.lb.

RANA FAKHEREDDINE is a PhD student at the Université de Savoie, Chambéry & Annecy, France. Her research interests are the design and analysis of stratified Monte Carlo methods and their applications to differential equations used in modeling of physical processes, such as heat diffusion or coalescence phenomena. Her email address is Rana.Fakhereddine@univ-savoie.fr.

CHRISTIAN LÉCOT is a professor at the Université de Savoie, Chambéry & Annecy, France. He is mainly interested in devising and analyzing numerical methods for physics, with a specific focus on Monte Carlo and quasi-Monte Carlo techniques. Areas of application include neutron transport, plasma physics and fluid dynamics. He serves as an editor for *Mathematics and Computers in Simulation*. His email address is Christian.Lecot@univ-savoie.fr.

ARTHUR SOUCEMARIANADIN is a full professor at the University Joseph Fourier in Grenoble and he is at the Laboratory of Industrial and Geophysical Flows. His area of research is in the field of hydrodynamic instabilities with a specific focus on non-Newtonian fluids. Besides his research interests, he is presently the Director for Institutional Relations at Campus France a national agency focused on the promotion of French Higher Education & Research. His email address is arthur.soucemarianadin@ujf-grenoble.fr.

MOUSSA TEMBELY received his PhD in Fluid Mechanics from the University of Grenoble, France in 2010. He also holds a master's degree from the French aerospace engineering school, SUPAERO. His background is in fluid and solid dynamics, and his research interests include droplet dynamics, fluid instability, fluid-structure interaction, inkjet printing, spray generation, heat transfer, as well as computational fluid dynamics. He is currently a postdoctoral research fellow at Concordia University, Montréal, Canada. His email address is m_tembel@encs.concordia.ca.