An introduction to stochastic particle integration methods: with applications to risk and insurance

P. Del Moral, G. W. Peters, C. Vergé

Abstract This article presents a guided introduction to a general class of interacting particle methods and explains throughout how such methods may be adapted to solve general classes of inference problems encountered in actuarial science and risk management. Along the way, the resulting specialized Monte Carlo solutions are discussed in the context of how they complemented alternative approaches adopted in risk management, including closed form bounds and asymptotic results for functionals of tails of risk processes.

The development of the article starts from the premise that whilst interacting particle methods are increasingly used to sample from complex and high-dimensional distributions, they have yet to be generally adopted in inferential problems in risk and insurance. Therefore, we introduce a range of methods which can all be interpreted in the general framework of interacting particle methods, which goes well beyond the standard particle filtering framework and Sequential Monte Carlo frameworks.

For the applications we consider in risk and insurance we focus on particular classes of interacting particle genetic type algorithms. These stochastic particle integration techniques can be interpreted as a universal acceptance-rejection sequential particle sampler equipped with adaptive and interacting recycling mechanisms. We detail how one may reinterpret these stochastic particle integration techniques under a Feynman-Kac particle integration framework. In the process, we illustrate how such frameworks act as natural mathematical extensions of the traditional change of probability measures, common in designing importance samplers for risk management applications.

P. Del Moral
INRIA Research Director, INRIA Bordeaux-Sud Ouest, & Bordeaux Mathematical Institute.
Professeur chargé de cours, Ecole Polytechnique (CMAP) Palaiseau.
e-mail: Pierre.Del_Moral@inria.fr

G.W. Peters
Department of Statistical Science, University College London;
Associate Fellow, Oxford-Man Institute, University of Oxford;
Associate Fellow, Commonwealth Scientific and Industrial Research Organization, CSIRO-CMIS,
e-mail: gareth.peters@ucl.ac.uk

Ch. Vergé
ONERA - The French Aerospace Lab, F-91761 PALAISEAU
e-mail: christelle.verge@onera.fr
1 Introduction to Stochastic Particle Integration

The intention of this paper is to introduce a class of stochastic particle based integration techniques to a broad community, with a focus on risk and insurance practitioners. We will demonstrate that a range of problems in risk and insurance can directly benefit from the development of such methods. A key motivation for this endeavour is the fact that stochastic particle integration models have been extensively used in engineering, statistics and physics under sometimes different names, such as: particle filters, bootstrap or genetic filters, population Monte Carlo methods, sequential Monte Carlo models, genetic search models, branching and multi-level splitting particle rare event simulations, condensation models, go-with-the winner, spawning models, walkers, population reconfigurations, pruning-enrichment strategies, quantum and diffusion Monte Carlo, rejuvenation models, and many others. However, they have not yet been routinely applied to develop solutions in important financial domains such as those we discuss in this tutorial type overview.

We begin with an introduction to the fundamental background for interacting particle systems by highlighting key papers in their development in a range of different scientific disciplines, before introducing aspects of these stochastic methods to risk and insurance. It is the intention of this article to explain the key papers and ideas in a general Feynman-Kac interacting particle framework which is much more encompassing than the special subset of the well known particle filter based algorithms. We proceed through a selection of key features of the development of interacting particle systems, focusing on a sub-class of such methods of relevance to the application domain explored in this manuscript, risk and insurance.

The origins of stochastic particle simulation certainly starts with the seminal paper of N. Metropolis and S. Ulam [54]. As explained by these two physicists in the introduction of their pioneering article, the Monte Carlo method is, "essentially, a statistical approach to the study of differential equations, or more generally, of integro-differential equations that occur in various branches of the natural sciences". The links between genetic type particle Monte Carlo models and quadratic type parabolic integro-differential equations has been developed in the beginning of 2000's in the series of articles on continuous time models [24, 26].

The earlier works on heuristic type genetic particle schemes seem to have started in Los Alamos National Labs with works of M.N. Rosenbluth and A.W. Rosenbluth [70], and earlier works by T.E. Harris and H. Kahn [45]. We also quote the work on artificial life of Nils Aall Barricelli [5, 6]. In all of these works, the genetic Monte Carlo scheme is always presented as a natural heuristic resampling type algorithm to generate random population models, to sample molecular conformations, or to estimate high energy particle distributions, without a single convergence estimate to ensure the performance, nor the robustness of the Monte Carlo sampler.

The mathematical foundations, and the performance analysis of all of these discrete generation particle models is a rather recent development. The first rigorous study in this field seems to be the article [11] published in 1996 on the applications of particle methods to non-linear estimation problems. This article provides the first proof of the unbiased property of particle likelihood approximation models.
Title Suppressed Due to Excessive Length

(lemma 3 page 12); and adaptive resampling criteria w.r.t. the weight dispersions (see remark 1 on page p.4). We also quote the first articles presenting heuristic type particle filters [42, 47], and a series of earlier research reports [29, 30, 31, 27].

For an in-depth description of the origins of particle methods and their applications we refer to the following studies [18, 28]. These articles also contain new stochastic models and methods including look-ahead type strategies (section 4.2.2), reducing the variance using conditional explorations w.r.t. the observation sequences (example 3 p. 40), local errors transport models (see the proof of theorem 1 on page 11) and mean field models w.r.t. the occupation measures of random trees (section 3.2).

A more detailed review of particle models in discrete and continuous time can be found in [23, 19]. In the research monograph, [19], the reader will find a detailed discussion on particle models and methods including acceptance-rejection with recycling particle strategies, interacting Kalman filters a.k.a. Rao-Blackwellized particle filters (section 2.6, and section 12.6.7), look-ahead type strategies (section 12.6.6), genealogical tree models and branching strategies (section 11), and interacting Metropolis-Hasting models (chapter 5).

The practitioner will find in the research books [23, 19, 20, 22] a source of useful convergence estimates as well as a detailed list of concrete examples of particle approximations for real models, including restricted Markov chain simulations, random motions in absorbing media, spectral analysis of Schrodinger operators and Feynman-Kac semi-groups, rare event analysis, sensitivity measure approximations, financial pricing numerical methods, parameter estimation in HMM models, island particle models, interacting MCMC models, statistical machine learning, Bayesian inference, Dirichlet boundary problems, non-linear filtering problems, interacting Kalman-Bucy filters, directed polymer simulations, stochastic optimization, and interacting Metropolis type algorithms. For further discussion on the origins and the applications of these stochastic models, we refer the reader to the following texts [3, 12, 13, 21, 59, 33, 34, 48, 55, 62], and the references therein.

Despite this, particle methods are yet to be routinely or widely introduced to the areas of risk and insurance modelling. The initial examples that have been developed are detailed in [63], where a special sub-class of such methods for an important set of risk management problems was explained. It is therefore the intention of this paper to highlight aspects of this class of problems and the stochastic particle solutions that will aid further development of these approaches in risk modelling.

2 Motivation for Stochastic Particle Solutions:
Examining How Such Methods May Complement Risk Asymptotics

In the following subsections we provide motivation and context that will explain how and why risk management and actuarial sciences can benefit from the development of interacting particle system solutions. In particular, we will focus on a few key estimation problems that form an important subset of generic problems
faced by practitioners in these domains. This will involve consideration of single risk loss processes described under a Loss Distributional Approach (hereafter LDA) framework, see discussion in [56], [71], [64] and the books [49] and [72] for the background on such modelling approaches in risk. For basic discussions on how such problems relate to a large class of non-life insurance problems see examples in [62].

2.1 The Loss Distributional Approach and Risk Management: a tale of light to heavy tails

In this section we first motivate and introduce the context of LDA modelling in risk and insurance. Then we present three key challenges associated with working with such LDA models that are commonly encountered by risk and insurance practitioners, thereby explaining some important inference challenges faced by such practitioners. Next, we provide a brief specifically selected survey of closed form analytic results known in the actuarial and risk literature for sub-classes of such LDA models as the Single Loss Approximations (hereafter SLA). We first detail the closed form solution for the light tailed severity distribution case. Then we explain how such results that are applicable in the light tailed case cannot be obtained in such a form in the heavy tailed sub-exponential risk process settings. Consequently, we briefly present the results recently developed in actuarial literature for the heavy tailed case corresponding to the first order and second order asymptotic approximations, see comprehensive discussions in a general context in [1], [16], and the books [4] and [14].

The fact that SLA approximations are inherently asymptotic in nature, and may be inaccurate outside of the neighbourhood of infinity, typically means that in practice risk managers must resort to numerical procedures to estimate risk measures and capital, see discussions in [68]. It is in these cases we will explain and motivate the utility of interacting particle based solutions.

Consider the widely utilised insurance model known as a single risk LDA model. This represents the standard under the Basel II/III capital accords [7] and involves an annual loss in a risk cell (business line/event type) modelled as a compound distributed random variable,

\[ Z_t^{(j)} = \sum_{s=1}^{N_t^{(j)}} X_s^{(j)}(t), \]

for \( t = 1, 2, \ldots, T \) discrete time (in annual units) and index \( j \) identifies the risk cell. Furthermore, the annual number of losses is denoted by \( N_t^{(j)} \) which is a random variable distributed according to a frequency counting distribution \( P^{(j)}(\cdot) \), typically Poisson, Binomial or Negative Binomial. The severities (losses) in year \( t \) are repre-
sent by random variables \(X_s^{(j)}(t), s \geq 1\), distributed according to a severity distribution \(F^{(j)}(\cdot)\) and there are \(N_t^{(j)}\) of them in year \(t\).

In general, we will suppress the risk cell index \(j\) and time index \(t\) unless explicitly utilised. Therefore, we denote by \(F(x)\) a distribution with positive support for the severity model characterizing the loss distribution for each random variable \(X_s\) for \(s \in \{1, 2, \ldots, N\}\). We denote the annual loss (aggregated loss) by \(Z\) with annual loss distribution \(G = F_Z\) and the partial sum of \(n\) random losses by \(S_n = F^{n}\) where \(F^{n}\) denotes the \(n\)-fold convolution of the severity distribution for the independent losses. The densities, when they exist, for the severity distribution and annual loss distributions will be denoted by \(f(x)\) and \(f_Z(x)\) respectively.

We assume that all losses are i.i.d. with \(X_s^{(j)}(t) \sim F(x)\) where \(F(x)\) is continuous with no atoms in the support \([0, \infty)\). As a consequence, linear combinations (aggregation) of losses in a given year satisfy

\[
S_n(t) = \sum_{s=1}^{n} X_s^{(j)}(t) \sim F_{S_n}(x)
\]

and have the following analytic representation:

\[
F_{S_n}(x) = (F * F * \cdots F)(x) = \int_{[0, \infty)} F^{(n-1)}(x-y)dF(x).
\]

In [39] it is shown that if \(F(x)\) has no atoms in \([0, \infty)\) then the \(n\)-fold convolution of such severity distributions will also admit no atoms in the support \([0, \infty)\). The implications of this for interacting particle based numerical procedures is that it ensures numerical techniques are well defined for such models. In particular the ratios of densities on the support \([0, \infty)\) are well defined. This is important as it is often required for interacting particle methods. In addition we note that continuity and boundedness of a severity distribution \(F(x)\) is preserved under \(n\)-fold convolution. Hence, if \(F(x)\) admits a density \(\frac{d}{dx} F(x)\) then so does the distribution of the partial sum \(F_{S_n}\), for any \(n \in \{1, 2, \ldots\}\) and compound process (random sum) \(F_Z\).

In practice the choice of severity distribution \(F(x)\) should be considered carefully for each individual risk process. As discussed in [68] it is common to consider sub-exponential severity distributions that we denote by membership \((F(x) \in \mathcal{F})\). The sub-exponential family of distributions \(\mathcal{F}\) defines a class of heavy tailed severity models that satisfy the limits

\[
\lim_{x \to \infty} \frac{1 - F^{n}\(x\)}{1 - F(x)} = n,
\]

if and only if,

\[
\lim_{x \to \infty} \frac{1 - F^{2}(x)}{1 - F(x)} = 2.
\]

Alternatively, one may characterize the family of distributions \(F \in \mathcal{F}\) by those that satisfy asymptotically the tail ratio
Severity models $F \in \mathcal{F}$ are of interest in high consequence loss modelling since they include models with *infinite mean loss* and *infinite variance*. In addition, the class $\mathcal{F}$ includes all severity models in which the tail distribution under the log transformed r.v., $F(\log(x))$, is a slowly varying function of $x$ at infinity.

To further understand LDA modelling with sub-exponential severity models it will be beneficial to recall the notion of asymptotic equivalence in which a probability distribution function $F(x)$ is asymptotically equivalent to another probability distribution function $G(x)$, denoted by $F(x) \sim G(x)$ as $x \to \infty$ if it holds that, $\forall \varepsilon > 0, \exists x_0$ such that $\forall x > x_0$ the following is true

$$\left| \frac{F(x)}{G(x)} - 1 \right| < \varepsilon.$$  

(5)

Furthermore, we say that a probability distribution function is *max-sum-equivalent*, denoted by $F \sim_M G$, when the convolution of the tail distribution of two random variables is distributed according to the sum of the two tail distributions asymptotically,

$$1 - (F \ast G)(x) = (\overline{F} \ast \overline{G})(x) \sim \overline{F}(x) + \overline{G}(x), \quad x \to \infty.$$  

Then for the class of heavy tailed sub-exponential LDA models we have that a probability distribution function $F$ will belong to the sub-exponential class $\mathcal{F}$ if $F \sim_M F$, i.e. it is max-sum-equivalent with itself and that the class $\mathcal{F}$ is closed under convolutions. The implications of this for LDA models is clear when one observes that sub-exponential LDA models are compound process random sums comprised of an infinite mixture of convolved distributions,

$$G(x) = \sum_{n=0}^{\infty} p_n F^{n*}(x),$$  

(6)

for a suitable series $\{p_n\}$, (e.g. convergent sequence satisfying Kolmogorov three series theorem). Using the property of max-sum equivalence one can show the practically relevant asymptotic equivalence between the severity distribution $F$ and the annual loss distribution $G$ in which selecting $F \in \mathcal{F}$ results in $G \in \mathcal{F}$ and

$$\lim_{x \to \infty} \frac{G(x)}{F(x)} = \lambda.$$

This asymptotic equivalence relationship between the severity distribution $F$ and the annual loss distribution $G$, present for sub-exponential LDA models, greatly benefits the formulation of asymptotic approximations of tail functionals used in the estimation of bank capital.

Based on these aforementioned properties we can obtain asymptotic approximations to the annual loss distribution tails which typically fall under one of the following classifications:
• “First-Order” and “Second-Order” Single Loss Approximations: recently discussed in [10], [16], [17] and references therein.
• “Higher-Order” Single Loss Approximations: see discussions in [8] and recent summaries in [1] and references therein.
• Extreme Value Theory (EVT) Single Loss Approximations (Penultimate Approximations): the EVT based asymptotic estimators for linear normalized and power normalized extreme value domains of attraction were recently discussed in [17].
• Doubly Infinitely Divisible Tail Asymptotics given $\alpha$-stable severity models discussed in [66] and [60].

We now briefly detail the first and second order asymptotics that are known in the risk literature for light and heavy tailed severity distributions in LDA models. Then we explain how stochastic particle methods can be utilised to complement such closed form expressions in practical banking models and scenarios.

2.1.1 A Light Tale of Light Tails

Here we recall some asymptotic results known for light tailed models as these will inform the results obtained in the heavy tailed expansions. A useful result in the light tailed case was provided by [36] where they consider frequency distributions $p_n = \Pr(N = n)$ satisfying

$$p_n \sim w^n n^\gamma C(n), \quad \text{as} \quad n \to \infty,$$

for some $w \in (0, 1)$, $\gamma \in \mathbb{R}$ and a function $C(n)$ slowly varying at $\infty$. Then, if there exists $\kappa > 0$, such that the Laplace transform of the severity

$$L_X(s) = \mathcal{L}[F(x)] = \int_0^{\infty} \exp(-sx)dF(x), \quad \forall s \in \mathbb{R},$$

matches the radius of convergence of the generating function of the frequency distribution,

$$w^{-1} = L_X(-\kappa)$$

with $-L_X'(-\kappa) < \infty$, one can state the following asymptotic equivalence for the compound process tail distribution,

$$F_{Z_n}(x) \sim \frac{x^\gamma \exp(-\kappa x)C(x)}{\kappa(-wL_X'(-\kappa))^{\gamma+1}}, \quad x \to \infty.$$

This light tailed asymptotic result demonstrates that the behaviour of the compound loss distribution tail is determined by either the frequency or the severity depending on which has the heavier tail. In addition, it is clear that the Poisson distribution tail is too light for this result to be valid, since the radius of convergence of
generating function is infinite. There are therefore alternative expansions developed for compound Poisson risk processes such as the saddlepoint approximation. 

**So how do light tailed results motivate the context we are considering in sub-exponential LDA models?**

In the sub-exponential heavy tailed setting the Laplace transform does not exist and hence these results do not apply. This is unfortunate, since rates of convergence and approximation accuracy are studied for such results. There are many important examples of LDA models for which these light tailed results do not apply, these include severity distributions with power law tail decay (Pareto, Burr, log gamma, Cauchy, $\alpha$-Stable, tempered stable and $t$-distribution). In the sub-exponential model setting it is often possible to develop alternative asymptotic results, however asymptotic convergence rates are typically not available. This is one area where particle integration methods can also be informative and complementary in the study of such LDA closed form asymptotics.

### 2.1.2 A Heavier Tale of Heavy Tails

In this subsection we briefly detail the asymptotic first and second order tail results for the LDA models when sub-exponential severity distributions are considered. The sub-exponential LDA first order tail asymptotics involve obtaining a closed form approximate expression for $F_Z(x)$, see details in [9], [17]. To proceed, consider the annual loss distribution $G(z) = F_Z(z)$ under LDA formulation with the severity distribution satisfying $F \in \mathcal{F}$,

$$G(z) = F_Z(z) = \sum_{n=0}^{\infty} \Pr[Z \leq z | N = n] \Pr[N = n] = \sum_{n=0}^{\infty} p_n F^{(n)}(z).$$  

(7)

Furthermore, to ensure convergence when evaluating Equation (7) it is required that one assumes that for some $\epsilon > 0$, the following condition is satisfied

$$\sum_{n=0}^{\infty} (1 + \epsilon)^n p_n < \infty.$$

Then the right tail of the annual loss distribution $F_Z(z)$ for the annual loss random variable $Z$, is approximated according to a SLA given by,

$$F_Z(x) = \mathbb{E}[N] F(x) (1 + o(1)) \quad \text{as} \quad x \to \infty,$$

To understand this basic result of the first order tail asymptotic $F_Z(x)$ consider the following two steps:

1. Obtain an upper bound on the asymptotic ratio of $F_{S_n}(x)$ and severity $F(x)$ for all $n \in \mathcal{J}$. Typically one can apply Kesten’s Bound which states that for sub-exponential severity distributions $F$ there exists a constant $K = K(\epsilon) < \infty$ for
\( \varepsilon > 0 \) s.t. \( \forall n \geq 2 \) the following bound holds ([15])

\[
\frac{F^{*n}(x)}{F(x)} \leq K(1+\varepsilon)^n, \quad x \geq 0.
\]

2. Then utilise Kesten’s bound to motivate the application of dominated convergence theorem to interchange the order of summation and limit and recall the characterization of heavy tailed sub-exponential severity models to obtain \( F_Z(x) \sim \mathbb{E}[N]F(x) \) since,

\[
\lim_{x \to \infty} \frac{F_Z(x)}{F(x)} = \lim_{x \to \infty} \sum_{n=1}^{\infty} p_n \frac{F^{*n}(x)}{F(x)} = \sum_{n=1}^{\infty} np_n = \mathbb{E}[N].
\]

As discussed in [16], and the papers therein, the second order asymptotic results can be developed in a wide class of risk models by considering the following further assumptions.

**Assumption 1:** \( F \) is zero at the origin \((x = 0)\) and satisfies that both the tail distribution \( F \) and density \( f \) are sub-exponential.

**Assumption 2:** The frequency distribution \( N \sim F_N(n) \) is such that its probability generating function given by

\[
p_N(v) = \mathbb{E}[v^n] = \sum_{n=0}^{\infty} \mathbb{P}(N = n)v^n,
\]

is analytic at \( v = 1 \).

Examples of severity models widely used in risk and insurance settings that satisfy such assumptions include: Log-Normal, Weibull (heavy tailed), Benktander Type I and Type II, Inverse Gaussian, \( \alpha \)-Stable, Halphen Family and certain members of Generalized Hypergeometric family.

Given a distribution satisfying Assumption 1 and Assumption 2, then two second order results are obtained. One for finite mean loss models and the other for infinite mean loss models. If the loss r.v. has finite mean \( \mathbb{E}[X] < \infty \) then the following result can be derived, see [57] and [74] for details,

\[
\lim_{x \to \infty} \frac{F_Z(x) - \mathbb{E}[N]F(x)}{f(x)} = \mathbb{E}[X]\mathbb{E}[(N - 1)N]. \tag{8}
\]

Alternatively, if the loss r.v. has an infinite mean but the severity density satisfies the regular variation condition \( f \in RV_{-1/\beta-1} \) for \( 1 \leq \beta < \infty \) then,

\[
\lim_{x \to \infty} \frac{F_Z(x) - \mathbb{E}[N]F(x)}{f(x) \int_0^x F(s) ds} = c_\beta \mathbb{E}[(N - 1)N],
\]

with \( c_1 = 1 \) and \( c_\beta = (1-\beta) \frac{\Gamma^2(1-1/\beta)}{\Gamma(1-2/\beta)} \) for \( \beta \in (1, \infty) \).
2.2 Inferential Challenges for Risk and Insurance: Asymptotics and the Role for Stochastic Particle Integration

The asymptotic approximation methods just surveyed were developed in the actuarial literature to tackle the serious statistical and computational challenges posed by estimation of tail quantiles and expectations for heavy tailed LDA models. The continued interest in such asymptotic results primarily stems from the fact that such closed form expressions bypass the significant computational challenges involved in estimation of risk measures for such heavy tailed annual loss distributions under traditional integration methods, Fourier methods, recursions (Panjer) or basic Monte Carlo approaches. However, they do have associated issues, see discussions in [43].

The properties of such asymptotic single loss approximation estimates are still an active subject of study with regard to explicit approximation errors, asymptotic rates of convergence and sensitivity to parameter estimation. To understand these features for loss approximations as well as to provide an alternative estimation approach for tail functionals we propose the application of interacting particle methods.

As summarized in [68] these single loss approximations can be utilised to form estimation of risk and capital approximations by obtaining an expression for the LDA model quantile function. For example, based on second order asymptotic results in the heavy tailed LDA models, one can show that if the severity distribution $F$ satisfies Assumption 1 and Assumption 2 with a finite mean, and the hazard rate $h(x) = \frac{f(x)}{1-F(x)}$ is of regular variation $h \in RV_{-\beta}$ for $\beta \geq 0$, then as $\alpha \to 1$ one has for the inverse of the annual loss distribution the result (see [1]),

$$F^{-1}_Z(\alpha) = F^{-1} \left( 1 - \frac{1 - \alpha}{E[N]} \{1 + \tilde{c}_\beta g_1 \left( F^{-1}(\tilde{\alpha}) \right) + o \left( g_1 \left( F^{-1}(\tilde{\alpha}) \right) \right) \} \right)^{-1} \tag{9}$$

where $\tilde{\alpha} = 1 - (1 - \alpha)/E[N]$ and

$$g_1(x) = \begin{cases} \frac{f(x)}{1-F(x)}, & \text{if } E[X] < \infty, \\ \frac{\int F(x)dx f(x)}{1-F(x)}, & \text{if } E[X] = \infty; \end{cases}$$

$$\tilde{c}_\beta = \begin{cases} \frac{E[X]E[(N-1)]}{E[N]}, & \text{if } E[N] < \infty, \\ \frac{E[X]E[(N-1)]}{E[N]}, & \text{if } E[N] = \infty. \end{cases}$$

Using this result it is then possible to consider asymptotic approximations of key risk management quantities known as risk measures which are used in the allocation of capital and reserving in all financial institutions and stipulated as standards under regulatory accords in both Basel II/III and Solvency II.

For example, one may now utilise this approximation to the annual loss quantile function to obtain estimates of common risk measures and regulatory capital, see [2] and [52]. Examples of such risk measures include the Value-at-Risk (VaR) which is defined for a level $\alpha \in (0,1)$ and corresponds to the quantile of the annual loss distribution,
\[
\text{VaR}_Z(\alpha) = F_Z^{-}(\alpha) = \inf \{ z \in \mathbb{R} : F_Z(z) \geq \alpha \}
\]

\[
\approx F_Z^{-} \left( 1 - \frac{1 - \alpha}{E[N]} [1 + o(1)] \right) \sim F_Z^{-} \left( 1 - \frac{1 - \alpha}{E[N]} \right),
\]

where \( F^{-}(\cdot) \) is the generalized inverse, see [37]. A second alternative, which includes the Expected Shortfall as a special case, is the \textit{Spectral Risk Measure (SRM)}, which for a weight function \( \phi : [0, 1] \rightarrow \mathbb{R} \) is given by

\[
\text{SRM}_Z(\phi) = \int_0^1 \phi(s) \text{VaR}_Z(s) ds
\]

\[
\approx \mathcal{X}(\alpha, \phi_1) F_Z^{-} \left( 1 - \frac{1 - \alpha}{E[N]} \right) \sim \mathcal{X}(\alpha, \phi_1) \text{VaR}_Z(\alpha),
\]

with \( \forall t \in (1, \infty) \) a function \( \phi_1(1 - 1/t) \leq K t^{1/(\beta + 1 - \epsilon)} \) for some \( K > 0 \) and \( \epsilon > 0 \) where

\[
\mathcal{X}(\alpha, \phi_1) = \int_1^{\infty} s^{1/\beta - 2} \phi_1(1 - 1/s) ds.
\]

### 2.2.1 The Role for Stochastic Particle Methods

Though the asymptotic results presented are elegant for some LDA models and efficient to evaluate in closed form, they do warrant careful consideration in their application, see discussions in [36]. In practice it may often be the case that one requires calculation of VaR, ES and Spectral Risk Measures at levels which do not satisfy such asymptotic properties, rendering such approximations inaccurate. In addition, though not yet a regulatory requirement, it is always good practice to consider the uncertainty associated with the estimation of the tail functionals and quantiles. This can be achieved via statistical confidence intervals, however these are non-trivial to obtain under such asymptotic expansion results. Thirdly, as discussed in [1] and [4], the asymptotic rates of convergence of such approximations are still only known in a little-oh Landau sense and therefore do not inform or guide the applicability of such results. Finally, there is a significant interest in diversification benefits that may be gained through the modelling of tail dependence features in the multi-variate risk process setting. Extending these asymptotic results to multiple risk processes coupled with a copula structure makes the derivation of asymptotic approximations highly challenging, see [44].

It is for these reasons that we argue stochastic particle based numerical solutions for the estimation of risk measures and tail functionals in LDA structures can be of direct utility to complement such asymptotic results. \textit{However, as all practitioners will know, the naive implementation of standard Monte Carlo and stochastic integration approaches to such problems will produce often poor results even for a considerable computational budget, see discussions in [50].} We therefore require specific interacting particle methods to provide accurate and computationally efficient solutions.
3 Selected Topics in Stochastic Integration Methods

In this section we will introduce practitioners to a variety of stochastic integration methods, presenting them formally from a mathematical perspective and making clear the properties of such methods. Note, in this section the notation adopted is utilised to reflect that which is utilised in the statistics and probability literature where much of the formal study of these methods has taken place.

3.1 Standard Monte Carlo Techniques for Risk and Insurance

Here we consider a conceptually simple problem involving a \( d \)-dimensional random variable denoted by \( X \) and some measurable subset denoted by \( A \subset \mathbb{R}^d \). Now suppose we want to compute the quantity \( \mathbb{P}(X \in A) := \mathbb{P}_X(A) \). For example in an LDA model in risk one may naturally consider defining \( A \) according to the interval for the annual loss \( Z \) given by \( A = [F^{-1}(\alpha), \infty) \) for some quantile level \( \alpha \in [0, 1] \) which is typically very close to one. Then we wish to evaluate the probability that the annual loss for a risk process falls in this interval. If \( \alpha \) is close to one then such a probability calculation poses a challenging computational task involving rare-event simulation.

The simplest and least computationally efficient approach to such a computation would involve a basic Monte Carlo simulation. To understand this approach we further assume that it is straightforward to generate a sequence \( (X_i)_{1 \leq i \leq N} \) of independent copies of the random variable \( X \). In this situation, the traditional Monte Carlo approximation of the distribution \( \mathbb{P}_X \) is given by the empirical measures

\[
\mathbb{P}_X^N = \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{X_i} \xrightarrow{N \uparrow \infty} \mathbb{P}_X.
\]

Now we define a generic, bounded, measurable test function \( \varphi \) on \( \mathbb{R}^d \) that will be used through the remaining sections. Then we can say, more precisely, that the convergence can be understood as the weak convergence of empirical measures, in the sense that the sequence of random variables

\[
\mathbb{P}_X^N(\varphi) := \int \varphi(x) \mathbb{P}_X^N(dx) = \frac{1}{N} \sum_{1 \leq i \leq N} \varphi(X_i)
\]

converges almost surely, to the limiting integrals

\[
\mathbb{P}_X(\varphi) = \int \varphi(x) \mathbb{P}_X(dx) = \mathbb{E}(\varphi(X)).
\]

Using indicator functions of cells in \( \mathbb{R}^d \), the shape of the measure \( \mathbb{P}_X \) can be obtained by plotting the histograms of the samples \( X_i \) in each dimension. By the strong law of large numbers, the above convergence is also met for integrable functions w.r.t. the measure \( \mathbb{P}_X \).
For indicator functions \( \varphi = 1_A \), sometimes we make a slight abuse of notation and we set \( \mathbb{P}^N_X(A) \) and \( \mathbb{P}_X(A) \) instead of \( \mathbb{P}_X^N(1_A) \) and \( \mathbb{P}_X(1_A) \). From the above discussion, we already have that

\[
\mathbb{P}^N_X(A) := \frac{1}{N} \sum_{1 \leq i \leq N} 1_A(X^i) \rightarrow_{N \uparrow \infty} \mathbb{P}_X(A) = \mathbb{E}(1_A(X)).
\]

The following properties are readily checked

\[
\mathbb{E}(\mathbb{P}^N_X(A)) = \mathbb{P}_X(A) \quad \text{and} \quad \text{Var}(\mathbb{P}^N_X(A)) = \frac{1}{N} \mathbb{P}_X(A) \left(1 - \mathbb{P}_X(A)\right).
\]

In addition, an \( N \)-approximation of the conditional distribution of \( X \) w.r.t. the event \( \{X \in A\} \) is given by

\[
\frac{1}{\mathbb{P}^N_X(A)} 1_A(x) \mathbb{P}^N_X(dx) \rightarrow_{N \uparrow \infty} \frac{1}{\mathbb{P}_X(A)} 1_A(x) \mathbb{P}_X(dx) = \mathbb{P}(X \in dx | X \in A). \quad (12)
\]

The l.h.s. terms in the above display are well defined as soon as \( \mathbb{P}^N_X(A) > 0 \). For rare event probabilities \( \mathbb{P}_X(A) \), say of order \( 10^{-6} \), the practical implementation of this Monte Carlo algorithm meets the difficulty that we need too many samples to estimate \( \mathbb{P}_X(A) \) using the proportion of success of such an event occurring only once in millions of attempts. It is therefore in general not recommended to consider such basic Monte Carlo techniques when studying or estimating the asymptotic risk measures discussed in this paper.

We illustrate this basic Monte Carlo on a standard model in risk and insurance based on the Poisson-Log Normal LDA model of a single risk process. This example, though simple, is both widely utilised in practice and also illustrative of the complementary role of the asymptotic approximations and the role Monte Carlo plays, since this specific model admits a closed form expression for the survival quantile of the annual loss under the first order asymptotic.

**Example 1 (Single Risk LDA Poisson-Log-Normal Family).** Consider the heavy tailed severity model, selected to model the sequence of i.i.d. losses in each year \( t \), denoted \( \{X_t(t)\}_{t=1:N_t} \), and chosen to be a Log-Normal distribution \( X_t \sim LN(\mu, \sigma) \) where the two parameters in this model correspond to parametrizing the shape of the distribution for the severity \( \sigma \) and the log-scale of the distribution \( \mu \). The survival and quantile functions of the severity are given by

\[
f(x; \mu, \sigma) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad x > 0; \quad \mu \in \mathbb{R}; \quad \sigma > 0,
\]

\[
\mathcal{F}(x; \mu, \sigma) = 1 - F(x) = \int_x^\infty \frac{1}{\sqrt{2\pi\sigma u}} \exp\left(-\frac{1}{2\sigma^2} (\log(u) - \mu)^2\right) du = \frac{1}{2} + \frac{1}{2} \text{erf}\left[\frac{\ln x - \mu}{\sqrt{2\sigma^2}}\right], \quad x > 0; \quad \mu \in \mathbb{R}; \quad \sigma > 0,
\]

\[
Q(p) = \exp\left(\mu + \sigma \Phi^{-1}(p)\right), \quad 0 < p < 1.
\]
Therefore the closed form SLA for the VaR risk measure at level $\alpha$ would be presented in this case under a first order approximation for the annual loss $Z = \sum_{n=1}^{N} X_i$ according to Equation (13)

$$\text{VaR}_\alpha [Z] = \exp \left[ \mu - \sigma \Phi^{-1} \left( \frac{1 - \alpha}{\lambda} \right) \right].$$

We illustrate the basic Monte Carlo solution for the VaR for a range of quantile levels of the annual loss distribution, we display these along with the measured confidence intervals in the point estimators after a long run of 5,000,000 samples of annual years so that the Monte Carlo accuracy was sufficient. We compare these to the first order SLA asymptotic result on the quantile levels $\alpha \in \{0.70, 0.75, 0.80, 0.85, 0.9, 0.95, 0.99, 0.995, 0.9995\}$, where the 99.5% and 99.95% quantile levels do in fact correspond to regulatory standards of reporting in Basel II/III.

This example provides a clear motivation for consideration of particle methods, especially in cases where the SLA results are not accurate. One can see that even in this relatively simple example, depending on the values of the parameters in the LDA risk model, the asymptotic VaR approximation may or may not be accurate at quantile levels of interest to risk management. Therefore, even small amounts of parameter uncertainty in the LDA model estimation may manifest in significantly different accuracies in the SLA capital estimates. Further examples for higher order asymptotics and other models are provided in [68].

Since the rate of convergence of SLA’s is still an active topic of research for such approximations. This simple study illustrates the fact that in practice the only way to ensure accuracy of such methods, for a given set of estimated or specified parameters, is to complement these approximations with a numerical solution such as a Monte Carlo or more efficient interacting particle solution.

### 3.2 Importance Sampling Techniques for Risk and Insurance

One could argue that the second most widely utilised class of stochastic integration methods considered in risk and insurance settings would be Importance Sampling, see for example [40] and in insurance settings [63]. Here we consider the understanding of these classes of methods in the context of risk and insurance estimation of tail functions. This involves considering undertaking sampling using another random variable for which the occurrence probability of the desired event $\mathbb{P}(Y \in A) := \mathbb{P}_Y(A)$ is closer to 1. This well known importance sampling strategy often gives efficient results for judicious choices of twisted measures $\mathbb{P}_Y$. Nevertheless, in some practical situations, it is impossible to find a judicious $\mathbb{P}_Y$ that achieves a given efficiency. Furthermore, this importance sampling technique is intrusive, in the sense that it requires the user to change the reference statistical or physical model into a twisted sampling rule.
Fig. 1 Annual Loss VaR Capital Estimate versus quantile level for Poisson-Log Normal LDA Risk Process. Top Plot: Severity distribution $\mu = 1$, $\sigma = 0.5$, $\lambda = 3$. Bottom Plot: Severity distribution $\mu = 1$, $\sigma = 5$, $\lambda = 3$.

To be more precise, sampling $N$ independent copies $(Y_i)_{1 \leq i \leq N}$ with the same dominating probability measure $\mathbb{P}_Y \gg \mathbb{P}_X$, the traditional Monte Carlo approximation is now given by

$$
\mathbb{P}_Y^N \left( 1_A \frac{d\mathbb{P}_X}{d\mathbb{P}_Y} \right) := \frac{1}{N} \sum_{1 \leq i \leq N} 1_A(X^i) \frac{d\mathbb{P}_X}{d\mathbb{P}_Y}(Y^i) \overset{N \uparrow \infty}{\longrightarrow} \mathbb{P}_Y \left( 1_A \frac{d\mathbb{P}_X}{d\mathbb{P}_Y} \right) = \mathbb{P}_X(A).
$$

The following properties are readily checked

$$
\mathbb{E} \left( \mathbb{P}_Y^N \left( 1_A \frac{d\mathbb{P}_X}{d\mathbb{P}_Y} \right) \right) = \mathbb{P}_X(A),
$$
and
\[ \text{Var} \left( \mathbb{P}_Y \left( 1_A \frac{d\mathbb{P}_X}{d\mathbb{P}_Y} \right) \right) = \frac{1}{N} \left( \mathbb{P}_X \left( 1_A \frac{d\mathbb{P}_X}{d\mathbb{P}_Y} \right) - \mathbb{P}_X(A)^2 \right). \]

It is easily checked that
\[ \mathbb{P}_Y(dx) = \frac{1}{\mathbb{P}_X(A)} 1_A(x) \mathbb{P}_X(dx) \Rightarrow \text{Var} \left( \mathbb{P}_Y^{N} \left( 1_A \frac{d\mathbb{P}_X}{d\mathbb{P}_Y} \right) \right) = 0. \]

In other words, the optimal twisted measure \( \mathbb{P}_Y \) is given by the unknown conditional distribution of \( X \) w.r.t. the event \( \{ X \in A \} \). In practice, we try to find a judicious choice of twisted measure that is easy to sample, with a probability distribution that resembles as closely as possible the desired conditional distribution.

Another approach is to use the occupation measure of a judiciously chosen Markov Chain Monte Carlo (hereafter MCMC) sampler with prescribed target measure
\[ \eta(dx) := \mathbb{P}(X \in dx \mid X \in A). \]

Of course, the first candidate is to take a sequence of independent copies of random variables with common distribution \( \eta \). Several exact sampling techniques can be used, including the inversion of the repartition function, change of variables principles, the coupling from the past, and acceptance-rejection techniques. A random sample \( X_i \) with distribution \( \mathbb{P}_X \) is accepted whenever it enters in the desired subset \( A \). In this interpretation, we need to sample \( N \) independent copies of \( X \) to obtain \( \mathcal{N} := N \times \mathbb{P}_X(A) \) independent samples with common law \( \eta \). However, for probabilities \( \mathbb{P}_X(A) \) of order \( 10^{-6} \), this method requires millions of samples, so we consider more computationally efficient solutions.

### 3.3 Markov chain Monte Carlo for Risk and Insurance

MCMC samplers have been used in insurance applications in non-life reserving models for example in Chain Ladder models [65], [69], [38] and Paid Incurred Claims models [53] and [61], in Operational Risk models in [64], [67] and in credit risk modelling for example in [51]. Hence, we now present the fundamental mathematical description of the underlying Monte Carlo algorithm that is developed for all the risk and insurance applications discussed in these references.

MCMC algorithms are based on sampling a Markov chain with invariant measure \( \eta \). In this context, the limiting measure \( \eta \) is often called the target measure. It is not difficult to construct these random processes. For instance, let us assume that the law of \( X \) is reversible w.r.t. some Markov transition \( K(x, dy) \). In this case, starting from the set \( A \), we sample a sequence of random states using the Markov proposal \( K \), rejecting sequentially all the states falling outside the set \( A \). The algorithm is well defined as soon as \( K(x, A) = K(1_A)(x) > 0 \), and the resulting Markov chain \( X_n \) coincides with the Metropolis-Hasting algorithm with probability transition given by the following formulae.
\[ M(x, dy) := K(x, dy) \cdot 1_A(y) + \left(1 - \int K(x, dz) \cdot 1_A(z)\right) \delta_x(dy). \]

It is not difficult to check that \( \eta \) is an invariant measure of the chain with transition \( M \), that is we have that
\[
(\eta M)(dy) := \int \eta(dx) M(x, dy) = \eta(dy).
\]

Note, the exact acceptance-rejection method discussed above corresponds to the special case
\[
K(x, dy) = \mathbb{P}(X \in dy) = \delta_x(dy).
\]

In more general situations, the proposal transition \( K(x, dy) \) amounts of moving randomly around the starting point \( x \). The individual (sometimes also called the walker) makes a number of tentative steps until it succeeds to enter into the desired set \( A \). In general, the random state at that (random) hitting time of \( A \) is not distributed according to \( \eta \). Roughly speaking, when the proposal transition \( K \) is based on local moves, the individual tends to hit the set \( A \) near the boundary of \( A \). To be more precise, starting from an initial state \( X_0 = x \in \mathbb{R}^d - A \), the hitting time
\[
T := \inf \{ n \geq 0 : X_n \in A \},
\]
is a geometric random variable with distribution
\[
\mathbb{P}(T = n \mid X_0 = x) = (1 - K(x, A))^n - K(x, A),
\]
and we have
\[
\mathbb{E}(\varphi(X_T) \mid X_0 = x) = K_A(\varphi)(x) := K(\varphi \cdot 1_A)(x)/K(1_A)(x).
\]

When the chain enters in \( A \), it remains for all times confined to the set \( A \). In addition, under some weak regularity conditions on the Markov transition \( K \), the target measure \( \eta \) is approximated by the occupation measures of the states; that is, we have the following asymptotic convergence result
\[
\frac{1}{n+1} \sum_{0 \leq p \leq n} \delta_{X_p} \rightarrow_{n \rightarrow \infty} \eta \quad \text{and} \quad \mathbb{P}(X_n \in dy \mid X_0 = x) := M^n(x, dy) \rightarrow_{n \rightarrow \infty} \eta(dy). \tag{14}
\]

In the above display, \( M^n(x, dy) \) stands for the \( n \) compositions of the integral operator \( M \) defined by the induction formulae
\[
M^n(x, dy) = \int M^{n-1}(x, dz)M(z, dy) = \int M(x, dz)M^{n-1}(z, dy)
\]
with the convention \( M^0(x, dy) = \delta_x(dy) \), for \( n = 0 \). It is of course out of the scope of this article to prove the ergodic theorem stated in the l.h.s. of (14).
3.4 Sequential Monte Carlo for Risk and Insurance

Application of Sequential Monte Carlo (hereafter SMC) methods in risk and insurance modelling is still relatively underdeveloped, hence the motivation for this article. In the context of risk modelling see the example in [62] and the references therein for more discussion. We start this section with a motivating class of algorithms targeting rare-event simulation via the restriction of a target measure to a contracting, increasingly rare set, such as a tail event.

SMC methods are acceptance-rejection techniques equipped with a recycling mechanism that allows a gradual sampling of a population of individuals w.r.t. a sequence of probabilities with increasing complexity. We illustrate this methodology in the situation discussed above. Let us choose a decreasing sequence of subsets \((A_p)_{0 \leq p \leq n}\) joining \(A_0 = \mathbb{R}^d\) to the desired lower subset \(A_n = A\):

\[
A_0 = \mathbb{R}^d \supset A_1 \supset A_2 \supset \ldots \supset A_{n-1} \supset A_n = A.
\]

Now, let’s try to sample sequentially random copies of the random variable \(X\) w.r.t. the conditioning events \(\{X \in A_p\}\), with \(p \leq n\). To get one step further, we let \(\eta_p\) be the sequence of measures \(\eta_p(dy) := P(X \in dx \mid X \in A_p)\) with \(p \leq n\).

By construction, \((\eta_p)_{0 \leq p \leq n}\) is a decreasing sequence of measures w.r.t. the absolutely continuous partial order relation \(\mu \ll \nu\) between probability measures \(^1\); that is, we have that

\[
\eta_n \ll \eta_{n-1} \ll \ldots \ll \eta_2 \ll \eta_1 \ll \eta_0 = \text{Law}(X).
\]

**Example 2 (Single Risk LDA Doubly-Infinitely Divisible Poisson-\(\alpha\)-Stable Family).**

Consider a single risk LDA model, then such a sequence of measures may correspond to construction of a sequence for the annual loss distribution. As an example, consider the sequence given by

\[
\eta_n(dz) := F_Z(dz \mid Z \in A_n),
\]

where \(A_n = [\text{VaR}_Z(\alpha_n), \infty)\) is one set, corresponding to the \(n\)-th element in the strictly increasing sequence \((\alpha_p)_{0 \geq p \geq n}\) as \(\alpha_p \uparrow 1\), which results in a contracting sequence of subsets \(A_0 = [0, \infty) \supset A_1 = [\text{VaR}_Z(\alpha_1), \infty) \supset \ldots \supset A_n = [\text{VaR}_Z(\alpha_n), \infty)\).

Given samples from this measure it is then simple to see that one could estimate quantities such as \(F_Z(\text{VaR}_Z(\alpha_n))\) which would be the normalizing constant of this probability distribution when restricted to the set \(A_n\). As an explicit example we consider the \(\alpha\)-Stable severity model in a Poisson LDA framework with strictly positive support. Consider the \(\alpha\)-Stable severity model with parameters \(\alpha \in [0, 2]\), \(\beta \in [-1, 1]\), \(\gamma > 0\) and \(\delta \in \mathbb{R}\) for the i.i.d. \(\alpha\)-Stable distributed random loses with common \(\alpha\). Then w.l.o.g. the density function of an \(\alpha\)-Stable severity distribution (standardized such that \(\gamma = 1\) and \(\delta = 0\)) can be evaluated point-wise according to

\(^1\) we recall that \(\mu \ll \nu\) as soon as \(\nu(A) = 0 \Rightarrow \mu(A) = 0\), for all measurable subset \(A \subset \mathbb{R}^d\)
the series expansions [75] [Equation 2.4.6, p. 89]

\[
f_X(x; \alpha, \beta, 1, 0; S(0)) = \begin{cases} 
\frac{1}{\pi} \sum_{n=1}^{\infty} (-1)^{n-1} \frac{F(n, \beta+1)}{T(n+7)} \sin(n\pi\beta) x^{n-1}, & \text{if } \alpha > 1, \beta \in [-1, 1], x \in \mathbb{R}, \\
\frac{1}{\pi} \sum_{n=1}^{\infty} (-1)^{n-1} n! b_n x^{n-1}, & \text{if } \alpha = 1, \beta \in (0, 1], x \in \mathbb{R}, \\
\frac{1}{\pi} \sum_{n=1}^{\infty} (-1)^{n-1} n! x^{-n-1}, & \text{if } \alpha < 1, \beta \in [-1, 1], x \in \mathbb{R}^+,
\end{cases}
\]

where the coefficients \( b_n \) are given by

\[
b_n = \frac{1}{\Gamma(n+1)} \int_0^\infty \exp(-\beta u \ln u) u^{n-1} \sin \left[ (1 + \beta) u \frac{\pi}{2} \right] du.
\]

The resulting LDA model annual loss distribution \( F_Z \) is given by

\[
Z = \sum_{i=1}^N X_i \sim F_{ZN} = \sum_{n=1}^\infty \exp(-\lambda) \frac{\lambda^n}{n!} S_\alpha \left( x; \beta_n, \gamma_n, \delta_n; S(0) \right),
\]

where the parameters of each mixture component are analytically known as expressions of the base severity model according to

\[
\gamma^2 = \sum_{i=1}^N |\gamma_i|^2, \quad \bar{\beta} = \frac{\sum_{i=1}^N \beta_i |\gamma_i|^2}{\sum_{i=1}^N |\gamma_i|^2}, \quad \bar{\delta} = \begin{cases} 
\frac{\sum_{i=1}^N \delta_i + \tan \frac{\pi}{2} \left( \beta \gamma - \sum_{i=1}^N \beta_i |\gamma_i| \right)}{\sum_{i=1}^N |\gamma_i|^2} & \text{if } \alpha \neq 1 \\
\frac{\sum_{i=1}^N \delta_i + \bar{\beta} \log \gamma - \sum_{i=1}^N \beta_i \gamma_i \log |\gamma_i|}{\sum_{i=1}^N |\gamma_i|^2} & \text{if } \alpha = 1.
\end{cases}
\]

Hence, one observes that as a result of closure under convolution of the \( \alpha \)-stable severity model, the resulting distribution for the annual loss can be presented exactly as a mixture representation, see discussions in [60]. Now, consider the Levy subfamily of models in which we consider \( Z \sim S(0.5, 1, \gamma; \delta; S(0)) \) with positive real support \( x \in \delta, \infty \]. The density and distribution functions are analytic and given respectively, for \( \delta < x < \infty \), by

\[
f_X(x) = \sqrt{\frac{\gamma}{2\pi}} \frac{1}{(x-\delta)^{3/2}} \exp \left( -\frac{\gamma}{2(x-\delta)} \right), \quad F_X(x) = \text{erfc} \left( \sqrt{\frac{\gamma}{2(x-\delta)}} \right),
\]

where \( \text{erfc}(x) = 1 - \text{erf}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \). Under this severity model, the distribution of the annual loss process \( Z \), is represented by a compound process model with LDA structure in which the frequency is \( N \sim Po(\lambda) \) and the severity is \( X_i \sim S(0.5, 1, \gamma; \delta; S(0)) \). The exact density of the annual loss process can then be expressed analytically as a mixture density comprised of \( \alpha \)-stable components with Poisson mixing weights for \( N_i^{(j)}>0 \) given by,

\[
f_Z(z) = \sum_{n=1}^\infty \exp(-\lambda) \frac{\lambda^n}{n!} \left[ \sqrt{\frac{\gamma_n}{2\pi}} \frac{1}{(z-\delta_n)^{3/2}} \exp \left( -\frac{\gamma_n}{2(z-\delta_n)} \right) \right] \mathbb{I} \left[ \bar{\delta}_n < z < \infty \right]
\]

with \( \bar{\beta}_n = 1 \) and
We also notice that the chain \( \text{the random states } X \text{ the terminal state Law} \)

As before, the chain \( \pi \text{ the invariant measure} \)

We assume that we have a dedicated MCMC style probability transition \( M \)

### 3.4.1 A Markov Chain Monte Carlo Model

We assume that we have a dedicated MCMC style probability transition \( M_p \) with invariant measure \( \eta_p = \eta_p M_p \), for any \( p \leq n \). We start drawing a sequence of random states \( (X_p)_{0 \leq p \leq n_1} \) with transitions \( M_1 \), and initial condition \( \eta_0 \). For a sufficiently large time horizon \( n_1 \), both the occupation measure \( \frac{1}{n_1} \sum_{1 \leq p \leq n_1} \delta_{\gamma_p} \) and the law of the terminal state \( \text{Law}(X_{n_1}) = \eta_0 M_1^{n_1} := \pi_1 \) will approximate the target measure \( \eta_1 \).

We also notice that the chain \( (X_p)_{p_1 \leq p \leq n_1} \) is confined to the set \( A_1 \) as soon as one of the random states \( X_{p_1} \in A_1 \) hits the set \( A_1 \) for some \( p_1 \leq n_1 \).

In the second step, starting from \( X_{n_1} \) we run a sequence of random states \( (X_{n_1+p})_{0 \leq p \leq n_2} \) with transitions \( M_2 \) (and initial condition \( \pi_1 \)). For a sufficiently large time horizon \( n_2 \), both the occupation measure \( \frac{1}{n_2} \sum_{1 \leq p \leq n_2} \delta_{X_{n_1+p}} \) and the law of the terminal state \( \text{Law}(X_{n_1+n_2}) = \pi_1 M_2^{n_2} \) will now approximate the target measure \( \eta_2 \).

As before, the chain \( (X_{n_1+p})_{p_2 \leq p \leq n_2} \) is confined to the set \( A_2 \) as soon as one of the
random states \( X_{n_1 + p_2} \in A_2 \) hits the set \( A_2 \) for some \( p_2 \leq n_2 \).

\[
\eta_0 \xrightarrow{M_1^{n_1}} \eta_0 M_1^{n_1} := \pi_1 \xrightarrow{M_2^{n_2}} \pi_1 M_2^{n_2} = \pi_2 \xrightarrow{M_3^{n_3}} \pi_2 M_3^{n_3} = \pi_3 \ldots (22)
\]

### 3.4.2 An Acceptance-Rejection Markov Model

Our next objective is to better understand the evolution of the flow of measures \( \eta_p \), from the origin \( p = 0 \) up to the final time \( p = n \). Firstly, it is readily checked that

\[
P(X \in dx \mid X \in A_{p+1}) = \frac{1}{\mathbb{P}(X \in A_{p+1} \mid X \in A_p)} \mathbb{P}(X \in dx \mid X \in A_p)
\]

and

\[
P(X \in A_{p+1} \mid X \in A_p) = \int 1_{A_{p+1}}(x) \mathbb{P}(X \in dx \mid X \in A_p).
\]

Therefore, if we specifically set \( G_p(x) = 1_{A_{p+1}}(x) \), then we have that

\[
\eta_{p+1} = \Psi_{G_p}(\eta_p)
\]

with the Boltzmann-Gibbs \( \Psi_{G_p} \) transformation defined by:

\[
\eta_p(dx) \rightarrow \Psi_{G_p}(\eta_p)(dx) := \frac{1}{\eta_p(G_p)} G_p(x) \eta_p(dx).
\]

The next formula provides an interpretation of the Boltzmann-Gibbs transformation in terms of a non-linear Markov transport equation

\[
\Psi_{G_p}(\eta_p)(dy) = (\eta_p S_{p,\eta_p})(dy) := \int \eta_p(dx) S_{p,\eta_p}(x, dy)
\]

with the Markov transition \( S_{p,\eta_p} \) defined below

\[
S_{p,\eta_p}(x, dy) = G_p(x) \delta_x(dy) + (1 - G_p(x)) \Psi_{G_p}(\eta_p)(dy).
\]

In summary, we have shown that \( (\eta_p)_{0 \leq p \leq n} \) satisfies the following evolution equation

\[
\eta_0 \xrightarrow{S_{0,\eta_0}} \eta_1 \xrightarrow{S_{1,\eta_1}} \eta_2 \xrightarrow{S_{2,\eta_2}} \eta_3 \xrightarrow{S_{3,\eta_3}} \eta_4 \ldots
\]

In other words, \( \eta_p = \text{Law}(X_p^*) \) can be interpreted as the law of the random state of a Markov chain \( X_p^* \) with transitions \( S_{p,\eta_p} \); that is, we have that

\[
X_0^* \xrightarrow{S_{0,\eta_0}} X_1^* \xrightarrow{S_{1,\eta_1}} X_2^* \xrightarrow{S_{2,\eta_2}} X_3^* \xrightarrow{S_{3,\eta_3}} \ldots
\]

The Markov chain \( X_p^* \) can be interpreted as the optimal sequential acceptance-rejection scheme along the non-increasing sequence of subsets \( A_p \), in the sense that
where $X_{p+1}^*$ stands for an independent random sample with distribution $\eta_{p+1} = \Psi_{G_p}(\eta_p)$. When the sample $X_p^*$ is not in the desired subset $A_{p+1}$, it jumps instantly to a new state $X_{p+1}^{*\prime}$ randomly chosen with the desired distribution $\eta_{p+1} = \Psi_{G_p}(\eta_p)$.

Next we provide a brief discussion on the optimality property of this Markov chain model. We recall that

$$\|\eta_{p+1} - \eta_p\|_{tv} = \sup \left\{ |\eta_{p+1} - \eta_p| (f) : \text{osc}(f) \leq 1 \right\}$$

$$= \inf \left\{ \mathbb{P}(X_p \neq X_{p+1}) : (X_p, X_{p+1}) \text{ s.t. Law}(X_p) = \eta_p \text{ and Law}(X_{p+1}) = \eta_{p+1} \right\}$$

In the above display $\text{osc}(\varphi) = \sup_{x,y}(|\varphi(x) - \varphi(y)|)$ stands for the oscillation of a given function $\varphi$ on $\mathbb{R}^d$.

**Proposition 3.1** The chain $X_p^*$ with Markov transitions $S_p, \eta_p$ realizes the optimal coupling between the sequence of distributions $\eta_p$, in the sense that

$$\|\eta_{p+1} - \eta_p\|_{tv} = \|\eta_p S_p, \eta_p - \eta_p\|_{tv} = \mathbb{P}(X_p^* \neq X_{p+1}^*).$$

A proof of this assertion is provided in the appendix.

### 3.4.3 Feynman-Kac Distributions

As the reader may have noticed, the MCMC model and the acceptance-rejection Markov chain models discussed may have very poor stability properties, in the sense that the distributions of the random states may strongly depend on the initial distribution $\eta_0$. For instance, we notice that $\eta_p$ coincides with the restriction of $\eta_0$ to the subset $A_p$; more formally, we have that

$$\eta_p(dx) = \Psi_{G_{p-1}}(\eta_0) = \frac{1}{\eta_0(A_p)} 1_{A_p}(x) \eta_0(dx).$$

The sequential Monte Carlo methodology is based on combining the MCMC methodology presented in (22) with the sequential acceptance-rejection technique discussed above. To describe with some precision this method, we let $M_p$ be an MCMC transition with invariant measure $\eta_p = \eta_p M_p$. In this case, we have the evolution equation

$$\eta_{p+1} = \eta_p M_{p+1} = \Psi_{G_p}(\eta_p) M_{p+1} := \Phi_{p+1}(\eta_p).$$

Notice that $\Phi_{p+1}$ maps the set of probability measures $\eta$ s.t. $\eta(G_p) > 0$ into the set of probability measures, and it is the composition of an updating transformation $\Psi_{G_p}$ and a Markov transport equation w.r.t. $M_{p+1}$; that is, we have that
\[ \eta_p \xrightarrow{\Psi_p} \tilde{\eta}_p := \Psi_{G_p}(\eta_p) \xrightarrow{M_{p+1}} \tilde{\eta}_pM_{p+1} = \Phi_{p+1}(\eta_p). \]

The solution of this equation is given by the Feynman-Kac measures defined for any measurable function \( \varphi \) on \( \mathbb{R}^d \) by the following formulae

\[ \eta_p(\varphi) = \gamma_p(\varphi)/\gamma_p(1) \text{ with } \gamma_p(\varphi) = \mathbb{E} \left( \varphi(X_p) \prod_{0 \leq q < p} G_q(X_q) \right). \] (24)

To prove this claim, we use the Markov property to check that

\[ \gamma_{p+1}(\varphi) = \mathbb{E} \left( M_{p+1}(\varphi)(X_p) G_p(X_p) \prod_{0 \leq q < p} G_q(X_q) \right) = \gamma_p(G_pM_{p+1}(\varphi)). \]

This clearly implies that

\[ \eta_{p+1}(\varphi) = \frac{\gamma_p(G_pM_{p+1}(\varphi))}{\gamma_p(1)} \cdot \frac{\gamma_p(1)}{\gamma_p(G_p)} = \frac{\eta_p(G_pM_{p+1}(\varphi))}{\eta_p(G_p)} = \Psi_{G_p}(\eta_p)M_{p+1}(\varphi). \]

We already mention that the unnormalized measures \( \gamma_p \) can be expressed in terms of the flow of measures \( (\eta_p)_{0 \leq p \leq n} \) with the following multiplicative formulae

\[ \gamma_p(\varphi) = \eta_p(\varphi) \times \prod_{0 \leq q < p} \eta_q(G_q). \] (25)

This result is a direct consequence of the following observation

\[ \gamma_p(1) = \mathbb{E} \left( G_{p-1}(X_{p-1}) \prod_{0 \leq q < p-1} G_q(X_q) \right) = \gamma_{p-1}(G_{p-1} = \eta_{p-1}(G_{p-1}) \gamma_{p-1}(1). \]

It is readily checked that the measures \( \eta_n \) are the \( n \)-th time marginals of the Feynman-Kac measures on the path space defined by the following formulae

\[ dQ_n := \frac{1}{Z_n} \left\{ \prod_{0 \leq p < n} G_p(X_p) \right\} dP_n \] (26)

with some normalizing constants \( Z_n = \gamma_n(1) \) and the reference measures

\[ P_n = \text{Law}(X_{0:n}) \text{ with } X_{0:n} := (X_0, \ldots, X_n). \]

This class of path space measures goes beyond the MCMC model discussed above. These measures represent the distribution of the trajectories of a reference Markov process, weighted by a collection of potential functions. These functional models are natural mathematical extensions of the traditional change of probability measures, commonly used in importance sampling.
From a purely probabilistic viewpoint, these measures can be interpreted as the conditional distribution of a given Markov chain w.r.t. to a sequence of events. For instance, if we take \( G_n = 1_{A_n} \) indicator potential functions of some measurable subsets \( A_n \in E_n \), then it can be readily checked that

\[
Q_n = \text{Law}(X_0 \mid \forall 0 \leq p < n \ X_p \in A_p) \quad \text{and} \quad Z_n = \mathbb{P}(\forall 0 \leq p < n \ X_p \in A_p).
\]

For a thorough discussion on the application domains of these Feynman-Kac models, we refer the reader to the books \[12, 19, 22, 34\].

**Example 3 (Multiple LDA Risk Conditional Tail Expectations).** Consider the class of problems in risk management involving the evaluation of a coherent capital allocation. We consider the \( X \in E \) to be a random vector \( X = [Z_1(\cdot), Z_1(\cdot), \ldots, Z_d(\cdot)] \) for \( d \) LDA structured risk processes, with the space on which this random vector is defined given by \( E = [0, \infty)^d \). In this case we can consider the multi-variate distribution for the \( d \) risk processes, for which we can consider dependence if required, according to

\[
Q_n := \text{Law}(X_{0:n} \mid \forall 0 \leq p < n \ X_p \in A_p) = F(Z_0^{(1)}, Z_0^{(d)}, Z_1^{(1)}, \ldots, Z_n^{(d)} \mid \forall 0 \leq p < n \ X_p \in A_p).
\]

If one considers the event \( X_p \in A_p \) as corresponding to the sequence of multi-variate loss draws that produce the rare-event that the total loss \( Z_T = \sum_{i=1}^d Z_i^{(i)} \) gives \( Z_T \in (\text{VaR}_{Z_T}(\alpha) - \varepsilon_n, \text{VaR}_{Z_T}(\alpha) + \varepsilon_n) \) for some \( \varepsilon_n \downarrow 0 \), then one has a mechanism for calculating conditional tail expectations, relevant to assessing multi-variate risk measures, tail dependence and capital allocation problems.

### 3.5 Non-linear McKean Markov Chains

The central idea behind Feynman-Kac particle samplers is to observe that any evolution equation of probability measures

\[
\eta_n = \Phi_n(\eta_{n-1})
\]

on some measurable state spaces \( E_n \) can be interpreted as the law

\[
\eta_n = \text{Law}(\overline{X}_n)
\]

of a Markov chain \( \overline{X}_n \) with initial distribution \( \eta_0 \) and Markov transitions

\[
\mathbb{P}(\overline{X}_n \in d\overline{x}_n \mid \overline{X}_{n-1} = \overline{x}_{n-1}) = K_{n,\eta_{n-1}}(x_{n-1},d\overline{x}_n).
\]

The Markov transitions \( K_{n,\eta_{n-1}} \) are chosen so that
\[ \forall n \geq 1 \quad \eta_{n-1} K_n, \eta_{n-1} = \Phi_n(\eta_{n-1}). \]

The Markov chain \( X_n \) incorporate free evolution moves according to \( M_n \), with sequential updates of the measures \( \eta_n \), so that the law of the random states \( X_n \) coincide with the desired distributions \( \eta_n \), at every time step. This chain can be interpreted as a perfect sequential sampler of the sequence of measures \( \eta_n \).

The choice of the transitions \( K_{n+1}, \eta_n \) is not unique. For instance, for the Feynman-Kac models on \( E_n = \mathbb{R}^d \) discussed above, if we take

\[ K_{n+1}, \eta_n (x, dy) := [S_n, \eta_n, M_{n+1}] (x, dy) \quad \text{or} \quad K_{n+1}, \eta_n (x, dy) := \Phi_{n+1} (\eta_n) (dy) \]

we readily check that

\[ \eta_n K_{n+1}, \eta_n = \Phi_{n+1} (\eta_n) = \Psi_G (\eta_n) M_{n+1} = \eta_n S_{n, \eta_n, M_{n+1}}. \]

We also mention that the law of the random trajectories \( (X_0, \ldots, X_n) \) are given by the so-called McKean measures

\[ \mathbb{P}_n (dx_{0:n}) = \eta_0 (dx_0) K_{1, \eta_0} (x_0, dx_1) \cdots K_{n, \eta_n} (x_{n-1}, dx_n), \]

where \( d x_{0:n} = d(x_0, \ldots, x_n) \) stands for an infinitesimal neighbourhood of the trajectory \( x_{0:n} := (x_0, \ldots, x_n) \).

We further assume that the Markov transitions \( M_n (x_{n-1}, dx_n) \) are absolutely continuous with respect to some reference measure \( \nu_n \) and we set

\[ Q_n (x_{n-1}, dx_n) := G_{n-1} (x_{n-1}) M_n (x_{n-1}, dx_n) = H_n (x_{n-1}, x_n) \nu_n (dx_n). \]

In this situation, we have the following time reversal formulae

\[ \mathbb{Q}_n (dx_{0:n}) = \eta_n (dx_n) \mathbb{M}_{n, \eta_{n-1}} (x_n, dx_{n-1}) \cdots \mathbb{M}_{1, \eta_0} (x_1, dx_0), \tag{27} \]

with the Markov transitions

\[ \mathbb{M}_{n, \eta_{n-1}} (x_n, dx_{n-1}) := \frac{\eta_{n-1} (dx_{n-1}) H_n (x_{n-1}, x_n)}{\eta_{n-1} (H_n (\ast, x_n))}. \]

We prove this backward formula using the fact that

\[ \eta_n (dx_n) = \Psi_{G_{n-1}} (\eta_{n-1}) M_n (dx_n) = \frac{\eta_{n-1} (H_n (\ast, x_n))}{\eta_{n-1} (G_{n-1})} \nu_n (dx_n), \]

from which we find that

\[ \eta_n (dx_n) \mathbb{M}_{n, \eta_{n-1}} (x_n, dx_{n-1}) = \frac{1}{\eta_{n-1} (G_{n-1})} \eta_{n-1} (dx_{n-1}) Q_n (x_{n-1}, dx_n). \]

Iterating this process, we prove (27).
3.5.1 Mean Field Particle Simulation

This section is concerned with particle approximations of the Feynman-Kac model (24) and (26). We also present a series of exponential concentration inequalities that allows one to estimate the deviation of the particle estimates around their limiting values.

In the remainder of this section \( \phi_n \) stands for some function s.t. \( \| \phi_n \| \leq 1 \), and \((c_1, c_2)\) represent two constants related to the bias and the variance of the particle approximation scheme, and \( c \) stands for some universal constant. The values of these constants may vary from line to line but they don’t depend on the time horizon. Furthermore, we assume that the Feynman-Kac model satisfies some strong stability properties. For a more detailed description of the stability properties, and the description of the quantities \((c, c_1, c_2)\) in terms of the Feynman-Kac model (24), we refer the reader to the books [19, 22].

We approximate the transitions

\[
X_n \leadsto X_{n+1} \sim K_{n+1, \eta_n}(X_n, dx_{n+1}),
\]

by running a Markov chain \( \xi_n = (\xi_1^n, \ldots, \xi_N^n) \in E^N_n \) that approximate the distribution \( \eta_n \) when \( N \uparrow \infty \)

\[
\frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_i^n} := \eta^N_n \longrightarrow_{N \uparrow \infty} \eta_n.
\]

A natural choice of particle transitions is to take at every time step a sequence of conditionally independent particles

\[
\xi_i^n \leadsto \xi_{i+1}^n \sim K_{n+1, \eta_n}(\xi_i^n, dx_{i+1}).
\]

For the Feynman-Kac models discussed above, we can chose the transitions \( K_{n+1, \eta_n} = S_n, \eta_n M_{n+1} \). In this context, the evolution of the particle algorithm is decomposed into two steps.

\[
\begin{bmatrix}
\xi_1^n \\
\vdots \\
\xi_i^n \\
\vdots \\
\xi_N^n
\end{bmatrix}
\xrightarrow{S_n, \eta_n G_n} \begin{bmatrix}
\xi_1^n \\
\vdots \\
\xi_i^n \\
\vdots \\
\xi_N^n
\end{bmatrix}
\xrightarrow{M_{n+1}} \begin{bmatrix}
\xi_1^n+1 \\
\vdots \\
\xi_i^n+1 \\
\vdots \\
\xi_N^n+1
\end{bmatrix}
\]

During the first step, every particle \( \xi_i^n \) evolves to a new particle \( \hat{\xi}_i^n \) randomly chosen with the distribution

\[
S_{\eta_n^N}(\xi_i^n, dx) := G_n(\xi_i^n) \delta_{\xi_i^n}(dx) + (1 - G_n(\xi_i^n)) \Psi_{G_n}(\eta_n^N)(dx),
\]

with the updated measures
This transition can be interpreted as an acceptance-rejection scheme with a recycling mechanism. In the second step, the selected particles $\xi^i_n$ evolve randomly according to the Markov transitions $M_{n+1}$. In other words, for any $1 \leq i \leq N$, we sample a random state $\xi^i_{n+1}$ with distribution $M_{n+1} (\xi^i_n, dx)$.

### 3.6 A Sequential Monte Carlo Formulation

Most of the SMC technology developed for Bayesian inference, is based on finding judicious sequential importance sampling representations of a given sequence of target measures, on some general state space models defined on $E_n$. More precisely, let us suppose that we are given a sequence of target measures of the following form

$$Q_n(dx_0: x_n) \propto Q_{n-1}(dx_0: x_{n-1}) \times Q_n(x_{n-1}, dx_n),$$

(28)

for some bounded positive integral operators $Q_n(x_{n-1}, dx_n)$ from $B_b(E_{n-1})$ into $B_b(E_n)$. By construction, we observe that these target measures can alternatively be defined by

$$Q_n(dx_0) := \frac{1}{Z_n} \eta_0(dx_0) Q_1(x_0, dx_1) \ldots Q_n(x_{n-1}, dx_n),$$

for some normalizing constant $Z_n$. Given a sequence of importance sampling transition $M_{n+1}$ s.t.

$$Q_n+1(x_n, \cdot) \ll M_{n+1}(x_n, \cdot),$$

for any $x_n \in E_n$, we denote by $W_n$ the sequential importance weights

$$W_n(x_n, x_{n+1}) \propto \frac{\text{Target at time (n+1)}}{\text{Target at time (n)} \times \text{Twisted transition}} \frac{Q_{n+1}(dx_{0:n+1})}{Q_n(dx_{0:n}) \times M_{n+1}(x_n, dx_{n+1})} := \frac{dQ_{n+1}(x_n, \cdot)}{dM_{n+1}(x_n, \cdot)}(x_{n+1}).$$

(29)

The corresponding change of measure has the following form

$$Q_n(dx_{0:n}) = \frac{1}{Z_n} \left\{ \prod_{0 \leq p < n} W_p(x_p, x_{p+1}) \right\} \mathbb{P}_n(dx_{0:n}).$$

(30)

We consider the Markov chain on the transition space defined by

$$X_n := (x_n, x_{n+1}) \in E_n = (E_n \times E_{n+1}).$$
In this notation, for any bounded measurable function \( \phi \) on the product state space \((E_0 \times \ldots \times E_n)\), we have the following importance sampling formulae

\[
\mathbb{E}(\phi_n(X_{0:n}) \prod_{0 \leq p < n} W_p(X_p, X_{p+1})) = \mathbb{E}(\phi_n(X_{0:n}) \prod_{0 \leq p < n} G_p(X_p))
\]

with the functions

\[
\phi_n(X_{0:n}) = \phi_n(X_{0:n}), \quad \text{and the potential functions } G_p := W_p.
\]

This implies that

\[
Q_n(\phi_n) = Q_n(\Phi_n), \quad (31)
\]

with the Feynman-Kac measure \( Q_n \) associated with the Markov chain \( X_n \) on the transition space \( E_n = (E_n \times E_{n+1}) \) and the potential functions \( G_n \). In this formulation, sequential Monte Carlo samplers coincide with the mean field particle interpretations discussed in section 3.5.1.

### 3.6.1 Some Non-Asymptotic Estimates: Finite Sample Accuracy for Particle Integration

The exponential concentration inequalities developed below are satisfied under some regularity conditions on the Feynman-Kac parameters \((G_n, M_n)\), on some general state space models defined on \( E_n \). It is clearly out of the scope to present here all the details of the proof of these inequalities. As shown in section 3.6, the importance sampling Feynman-Kac representation of a given sequence of target measures is far from unique. Roughly speaking, the twisted transitions \( M_n \) and the corresponding potential weight functions \( G_n \) have to be chosen so that the non-linear semi-group associated with evolution equation

\[
\eta_n = \Psi_{G_{n-1}}(\eta_{n-1})M_n,
\]

of the \( n \)-th time marginals \( \eta_n \) of the Feynman-Kac target measures \( Q_n \) are sufficiently stable. One way to satisfy this stability property is to choose sufficiently mixing twisted transitions, with bounded relative oscillations of the weight functions. For a more thorough discussion on these stability conditions, we refer the reader to [19, 22, 20].

We note that the exponential concentration inequalities presented below are also valid for non necessarily stable Feynman-Kac semi-groups. Nevertheless, in this degenerate situation the constants \( c \) and \((c_1, c_2)\) depend on the time parameter. Using the concentration analysis of mean field particle models developed in [32], the following exponential estimate was proved in [22]. For any \( x \geq 0, n \geq 0 \), and any population size \( N \geq 1 \), the probability of the event

\[
[\eta_n^N - \eta_n](\varphi) \leq \frac{c_1}{N} (1 + x + \sqrt{x}) + \frac{c_2}{\sqrt{N}} \sqrt{x},
\]
For any $y$ we let

$$F_n(x) = \eta_n(1_{(-\infty,x]}) \quad \text{and} \quad F_n^N(x) = \eta_n^N(1_{(-\infty,x]}).$$

For any $y \geq 0$, $n \geq 0$, and any population size $N \geq 1$, the probability of the following event

$$\sqrt{N} \left\| F_n^N - F_n \right\| \leq c \sqrt{d (y + 1)},$$

is greater than $1 - e^{-y}$. This concentration inequality ensures that the particle repartition function $F_n^N$ converges to $F_n$, almost surely for the uniform norm. For $d = 1$, we let $F_n^\leftarrow$ be the generalized inverse on $[0,1]$ of the function $F_n$; that is, we have that

$$F_n^\leftarrow(\alpha) := \inf \{ x \in \mathbb{R} : F_n(x) \geq \alpha \}.$$

We let $F_n^\leftarrow(\alpha) = q_{n,\alpha}$ be the quantile, of order $\alpha$, and we denote by $\zeta_i^n$ the order particle statistic associated with the particle system $\xi_i^n$ at time $n$; that is, we have that

$$\zeta_1^n := \xi_\sigma^{(1)} \leq \zeta_2^n := \xi_\sigma^{(2)} \leq \ldots \leq \zeta_n^n := \xi_\sigma^{(N)},$$

for some random permutation $\sigma$. We also denote by $q_{n,\alpha}^N := \zeta_{\lfloor N\alpha \rfloor}^n$ the $\alpha$-particle quantile. By construction, we have that

$$|F_n(q_{n,\alpha}^N) - F_n(q_{n,\alpha})| \leq |F_n(q_{n,\alpha}^N) - F_n^N(q_{n,\alpha})| + |F_n^N(q_{n,\alpha}) - \alpha|$$

$$\leq \left\| F_n^N - F_n \right\| + \left(1 + \left\lfloor N\alpha \right\rfloor - \alpha \right) \leq \left\| F_n^N - F_n \right\| + 1/N.$$

This clearly implies that $q_{n,\alpha}^N$ converges almost surely to $q_{n,\alpha}$ as $N$ tends to $\infty$. In addition, for any $y \geq 0$, $n \geq 0$, and any population size $N \geq 1$, the probability of the following event

$$\sqrt{N} \left| F_n(q_{n,\alpha}^N) - \alpha \right| \leq c \sqrt{d (y + 1)} + \frac{1}{\sqrt{N}},$$

is greater than $1 - e^{-y}$.

If we interpret the mutation-selection particle algorithm as a birth and death branching process, then we can trace back in time the whole ancestral line $\gamma_n^i = (\xi_p^i)_{0 \leq p \leq n}$ of the individual $\xi_n^i$ at the $n$-th generation

$$\xi_{0,n}^i \leftarrow \xi_{1,n}^i \leftarrow \ldots \leftarrow \xi_{n-1,n}^i \leftarrow \xi_{n,n}^i = \xi_n^i.$$
The random state $\xi_{i,p}^n$ represents the ancestor of the individual $\xi_{i}^n$ at the level $p$, with $0 \leq p \leq n$, and $1 \leq i \leq N$. It is more or less well known that $\gamma_n$ coincides with the particle approximation of the Feynman-Kac model defined in (24) by replacing $X_n$ by the historical process $(X_{p})_{0 \leq p \leq n}$. This interpretation provides an alternative particle approximation scheme of the measures (26), that is we have that

$$\eta_N^n = \frac{1}{N} \sum_{1 \leq i \leq N} \delta(\xi_{i,0}^n, \xi_{i,1}^n, \ldots, \xi_{i,n}^n). \rightarrow_{N \uparrow \infty} Q_n.$$ 

More precisely, we proved in [22] the following exponential concentration estimate. For any test function $\phi_n$ on path space s.t. $\|\phi_n\| \leq 1$, for any $y \geq 0$, $n \geq 0$, and any $N \geq 1$, the probability of the event

$$[\eta_N^n - Q_n](\phi) \leq c_1 \frac{n + 1}{N} (1 + x + \sqrt{x}) + c_2 \sqrt{\frac{(n+1)}{N}} \sqrt{x},$$

is greater than $1 - e^{-x}$.

Further details on these genealogical tree models can be found in [19, 22, 25]. Mimicking formulae (25) and (27), we define an unbiased particle estimate $\gamma_N^n$ of the unnormalized measures $\gamma_n$ and a particle backward measures $Q_N^n$ by setting

$$\gamma_N^n(\phi) = \eta_N^n(\phi) \times \prod_{0 \leq q < n} \eta_q^n(G_q),$$

and

$$Q_N^n(d(x_0, \ldots, x_n)) = \eta_n^n(d(x_n)) \mathbb{M}_{n, \eta_{n-1}^N}(x_n, dx_{n-1}) \ldots \mathbb{M}_{1, \eta_0^N}(x_1, dx_0).$$

We end this section with a couple of exponential concentration estimates proved in [22]. For any $x \geq 0$, $n \geq 0$, $N \geq 1$, and any $\epsilon \in \{+1, -1\}$, the probability of the event

$$\frac{\epsilon}{n} \log \frac{\gamma_N^n(1)}{\gamma_0^n(1)} \leq c_1 \frac{n + 1}{N} (1 + x + \sqrt{x}) + c_2 \sqrt{\frac{n+1}{N}} \sqrt{x},$$

is greater than $1 - e^{-x}$. In addition, for any normalized additive functional

$$\phi_n(x_0, \ldots, x_n) = \frac{1}{n + 1} \sum_{0 \leq p \leq n} \phi_p(x_p)$$

with $||\phi_p|| \leq 1$, for $x \geq 0$, $n \geq 0$, and any population size $N \geq 1$, the probability of the event

$$Q_N^n - Q_n(\psi_n) \leq c_1 \frac{1}{N} (1 + (x + \sqrt{x})) + c_2 \sqrt{\frac{x}{N(n+1)}},$$

is greater than $1 - e^{-x}$.
4 Illustration of Interacting Particle Solutions for Risk and Insurance Capital Estimation

In this section we detail a special subset of algorithms, from within the stochastic particle integration methods, that were specifically developed to solve problems for risk and insurance in [63]. The class of recursive solutions developed is applicable to a wide range of insurance and risk settings. We provide a novel result in this illustration which extends the framework originally presented in [63] through consideration of a higher-order Panjer recursion whilst avoiding the need to perform discretisation of the severity distribution. We shall present a generic version of this approach which adopts an interacting particle solution. In addition, we illustrate how this method may be used in inference for tail quantiles of compound processes to complement the results considered for the SLA approximations.

4.1 Recursions for Loss Distributions: Panjer and Beyond

We extend the framework proposed in [63] for developing a recursive numerical solution to estimation of such risk measures through estimation of the density of the compound process. In particular, we briefly summarize an approach to transform the standard actuarial solution known as the Panjer recursion [58] to a sequence of expectations. We note that recursions for the evaluation of single risk process distributions, under discretisation, are ubiquitous in risk and insurance modelling, see discussions in [73]. We consider an advanced development that avoids the need to discretise the severity distribution via development of a stochastic particle integration based solution.

Consider the actuarial recursions for evaluating $F_Z(x)$ based around the Panjer class of frequency distribution relationships defined by

$$p_n = \left(a + \frac{b}{n}\right)p_{n-1}, \quad (32)$$

with members Poisson ($a = 0, b = l, p_0 = e^{-\lambda}$), Binomial ($a = \frac{-q}{(1-q)}, b = \frac{(m+1)q}{(1-q)}, p_0 = (1-q)m$) and Negative Binomial ($a = \frac{k}{1-q}, b = \frac{(r-1)b}{1+q}, p_0 = (1+b) - r$). In addition, we consider the higher order Panjer recursion for an extended class of frequency distributions given by the generalized Poisson distribution (GPD). The GPD model is defined via the probability mass function

$$\Pr(N = n) = p_n(\lambda, \theta) = \begin{cases} \lambda(\lambda + n\theta)^{n-1}, & \forall n = 0, 1, 2, \ldots \\ 0, & \text{if } n > m, \text{ when } \theta < 0, \end{cases}$$
with $\lambda > 0$ and $\max (-1, \lambda / m) \leq \theta < 1$ and $m \geq 4$ is the largest positive integer s.t. $\lambda + \theta m > 0$ when $\theta$ is negative, where the GPD is Poisson for $\theta = 0$; over-dispersed $\theta > 0$ and under-dispersed $\theta < 0$.

One can then derive closed form recursions for the annual loss LDA compound process distribution given by

$$f_Y(x) = p_Y f_X(x) + \int_0^x \left( a + \frac{by}{x} \right)f_X(y)f_Y(x-y)dy,$$  \hspace{1cm} (33)

or the generalized higher order Panjer recursion [41],

$$f_Y(x) = p_Y(\lambda, \theta) f_X(x) + \frac{\lambda}{\lambda + \theta} \int_0^x \left( \theta + \frac{\lambda y}{x} \right)f_X(y)f_Y(x-y)dy.$$  \hspace{1cm} (34)

To understand how these recursions are obtained, consider the convolution identity for an i.i.d. partial sum $S_{n+1} = X_1 + \ldots + X_{n+1}$ with density

$$f^{(n+1)}(x) = \int_0^x f(\tau)f^{(n)}(x-\tau)d\tau, \hspace{1cm} \forall n = 1, 2, 3, \ldots$$  \hspace{1cm} (35)

Substitute the conditional of $X_1$ when $S_{n+1} = x$,

$$f_{X_1}(\tau|X_1 + \ldots + X_{n+1} = x) = \frac{f(\tau)f^{(n)}(x-\tau)}{f^{(n+1)}(x)},$$  \hspace{1cm} (36)

into the average given $S_{n+1} = x$ to get

$$E[X_1|X_1 + \ldots + X_{n+1} = x] = \int_0^x \frac{\tau f_{X_1}(\tau)f^{(n)}_{X_1}(x-\tau)}{f^{(n+1)}_{X_1}(x)}d\tau.$$  \hspace{1cm} (37)

Then observe that with i.i.d. losses one also gets

$$E[X_1|X_1 + \ldots + X_{n+1} = x] = \frac{1}{n+1} \sum_{i=1}^{n+1} E[X_i|X_1 + \ldots + X_{n+1} = x]$$

$$= \frac{1}{n+1} E[X_1 + \ldots + X_{n+1}|X_1 + \ldots + X_{n+1} = x] = \frac{x}{n+1}.$$  \hspace{1cm} (38)

Equating these conditional mean expressions and rearranging gives

$$\frac{1}{n+1} f^{(n+1)}_{X_1}(x) = \frac{1}{x} \int_0^x \tau f_{X_1}(\tau)f^{(n)}_{X_1}(x-\tau)d\tau.$$  \hspace{1cm} (39)

Now utilise the Panjer class of frequency distributions satisfying for some $a$ and $b$,

$$Pr(N = n) = p_n = \left( a + \frac{b}{n} \right)p_{n-1}.$$  \hspace{1cm} (40)

Upon substitution and some elementary algebra one obtains the Panjer recursion.
4.2 Stochastic Particle Methods as Solutions to Panjer Recursions

Avoiding the distributional discretisation of the severity model in applying the Panjer recursion reduces the computational cost when considering heavy-tailed severity models. It also reduces the discretisation error. It was noted in [63] that the Panjer recursions could be re-expressed as linear Volterra integral equations of the second kind via the mapping

\[ x_1 = x - y, \quad g(x) = p_1 F(x), \quad f(x_1) = f_Z(x_1), \quad k(x, x_1) = \left( a + b \frac{x - x_1}{x} \right) F(x - x_1). \] (41)

where the kernel \( k : E \times E \mapsto \mathbb{R} \) and the function \( g : E \mapsto \mathbb{R} \) are known whilst the function \( f : E \mapsto \mathbb{R} \) is unknown. Furthermore, if one defines \( k^0(x, y) \equiv 1, \ k^1(x, y) \equiv k(x, y) \) and

\[ k^n(x, y) \equiv \int k(x, y) k^{n-1}(z, y) dz \]

and these kernels satisfy that

\[ \sum_{n=0}^{\infty} \int_E |k^n(x_0, x_n)| g(x_n) |dx_n| < \infty, \]

then one can identify the resolvent kernel and Neumann series through iterative expansion of the recursion to obtain for a sequence of domains \( E_{1:n} \)

\[ f(x_0) = g(x_0) + \sum_{n=0}^{\infty} \int_0^{x_0} \cdots \int_0^{x_{n-1}} g(x_n) \prod_{l=1}^{n} k(x_{l-1}, x_l) dx_{1:n}. \]

Under this formulation it was shown in [63] how to address two problems: estimation of the annual loss density over a set \( A \) and estimation of the annual loss density pointwise. These are both directly relevant to obtaining estimates of the risk measures specified for capital estimation.

To achieve this we convert the Neumann series above into a sequence of expectations with respect to an importance sampling distribution. This is performed by making the following associations

\[ f_0(x_0) = g(x_0), \quad \text{and} \quad f_n(x_{0:n}) = g(x_n) \prod_{l=1}^{n} k(x_{l-1}, x_l) \]

\[ \therefore f(x_0) = f_0(x_0) + \sum_{n=0}^{\infty} \int_0^{x_0} \cdots \int_0^{x_{n-1}} f_n(x_{0:n}) dx_{0:n}. \]

Now we may develop this reformulated problem as an expectation with respect to a sequence of distributions \( \{\pi(n, x_{1:n})\}_{n \geq 0} \).
\begin{equation}
f(x) = \frac{f_0(x)}{\pi(0)} \pi(0) + \sum_{n=1}^{\infty} \int_{A_{1:n}(x)} \frac{f_n(x,x_{1:n})}{\pi(n,x_{1:n})} \pi(n,x_{1:n}) \, dx_{1:n}
\end{equation}

\begin{equation}
= \mathbb{E}_{\pi(n,x_{1:n})} \left[ \frac{f_n(x,x_{1:n})}{\pi(n,x_{1:n})} \right],
\end{equation}

with the sets \( A_{1:n}(x_0) = \{(x_1,\ldots,x_n) : x_0 > x_1 > \cdots > x_n\} \) playing an analogous role to the sequence of level sets described previously.

We note that there are now two path-space based particle solutions available, those that consider estimating \( f(x) \) point-wise via an importance sampling solution on the path-space defined by

\begin{equation}
\bigcup_{n=0}^{\infty} \{n\} \times A_{1:n}(x).
\end{equation}

The other alternative involves characterizing \( f(x) \) over some interval by obtaining samples from its restriction to that interval \([x_a,x_b]\), via importance sampling on a slightly larger space

\begin{equation}
\bigcup_{n=0}^{\infty} \{n\} \times A_{1:n}([x_a,x_b]).
\end{equation}

In [63] a path space based Sequential Importance Sampling (SIS) approximation to this sequence of expectations is obtained. This involves considering a Markov chain with initial distribution \( \mu(x) > 0 \) on \( E \) and transition kernel \( M(x,y) > 0 \) if \( k(x,y) \neq 0 \) and \( M \) has absorbing state \( d \notin E \) such that \( M(x,d) = \mathbb{P}_d \) for any \( x \in E \). Under this framework the interacting particle solution to the Panjer recursion is summarized in Algorithm 1. This is directly applicable to the higher order Panjer recursions discussed above.

**Algorithm 1: Path Space Stochastic Particle Methods for Panjer Recursions**

1. Generate \( N \) independent Markov chain paths \( \{X_{0,n+1}^{(i)} \}_{i=1}^{N} \) until absorption \( X_{n+1}^{(i)} = d \).
2. Evaluate the importance weights for each particle on the path space by,

\begin{equation}
W(X_{0,n+1}^{(i)}) = \begin{cases}
\frac{1}{\mu(x_0^{(i)})} \left( \prod_{n=1}^{n^{(i)}} k(x_{n-1}^{(i)},x_n^{(i)}) \right) g(x_n^{(i)}) \frac{\mathbb{P}_d(x_n^{(i)})}{\mu(x_n^{(i)})}, & \text{if } n^{(i)} \geq 1, \\
\frac{\mathbb{P}_d(x_0^{(i)})}{\mu(x_0^{(i)})}, & \text{if } n^{(i)} = 0.
\end{cases}
\end{equation}

4.3 Stochastic Particle Solutions to Risk Measure Estimation

If we consider \( \mu(X_{0}^{(i)}) = \delta(x_0^{(i)}) \), \( \forall i \in \{1,\ldots,N\} \), the empirical measure at a point \( x_0 \) is given by
\[ \hat{f}_Z(x_0) = \frac{1}{N} \sum_{i=1}^{N} W \left( x_0, X_{1:n(i)}^{(i)} \right), \]

or over an interval by

\[ \hat{f}_Z(x_0) = \frac{1}{N} \sum_{i=1}^{N} W_1 \left( X_{0:n(i)}^{(i)} \right) \delta \left( x_0 - X_0^{(i)} \right). \]

These estimators can be used to construct unbiased Monte Carlo approximations of the expectation of \( f_Z(z) \) for any set \( A \) given by \( E \left[ \int_A \hat{f}(x_0) dx_0 \right] = \int_A f(x) dx_0. \)

Having obtained this particle based approximation, this weighted Dirac measure can then be utilised to estimate any of the required risk measures such as VaR and SRM for any desired level \( \alpha \). This can be performed in two ways, depending on whether the particle solution is obtained for the evaluation of the recursions pointwise over a fixed grid or alternatively over an interval, which could be increasing in size. In the case of an interval, or contracting set, one considers perhaps a set of interest to be \( A = [0, x_{\text{max}}] \) such that \( x_{\text{max}} >> F^{-1} \left( 1 - \frac{\alpha}{E[N]} \right) \), and then utilises this to construct an unbiased particle approximation of the distribution of the annual loss up to any level \( \alpha \in (0, 1) \). This could be obtained from growing a set \( A_1 = [0, x_1] \subset A_2 \subset \cdots \subset A = [0, x_{\text{max}}] \) recursively, as discussed in previous sections.

If the partition is made pointwise over a linear or a non-linear spacing \([0, z] = \bigcup_{m=1}^{M} [(m-1)\triangle, m\triangle] \) and the distribution evaluated pointwise, this leads to an estimation of

\[ \hat{F}_Z(z) = \sum_{m=0}^{M} \triangle f(m\triangle) \approx \frac{1}{N} \sum_{m=0}^{M} \sum_{i=1}^{N} \triangle W \left( m\triangle, X_{1:n(i)}^{(i,m)} \right). \quad (43) \]

Alternatively, if the estimation is performed over an interval \( A(x_{\text{max}}) = [0, x_{\text{max}}] \), then for any \( z < x_{\text{max}} \) one may use the construction of the resulting empirical measure to obtain,

\[ \hat{F}_Z(z) = \frac{1}{N} \sum_{i=1}^{N} W \left( X_{0:n(i)}^{(i)} \right) \mathbb{1} \left( X_{0:n(i)}^{(i)} \in [0, z] \right) \rightarrow_{N \rightarrow \infty} \int_0^z \hat{f}_Z(z) dz. \quad (44) \]

**Practical advice:** Consider a range for the support \([0, x_{\text{max}}]\) s.t. \( x_{\text{max}} >> F^{-1} \left( 1 - \frac{\alpha}{E[N]} \right) \).

From these unbiased particle approximations of the annual loss density and distribution we can reconstruct the (inverse cdf) quantile function of the annual loss LDA model. This can either be based on a random set of particle locations or on a discrete deterministic grid as follows:

**Deterministic Grid Solution:** Given partition \([0, x_{\text{max}}] = \bigcup_{m=1}^{M} [(m-1)\triangle, m\triangle] \) for some step \( \triangle \) s.t.

\[ \hat{Q}(p) = \inf \left\{ x \in \{0, \Delta, \ldots, M\Delta\} : p \leq \frac{1}{N} \sum_{m=0}^{M} \sum_{i=1}^{N} \triangle W \left( x, X_{1:n(i,m)}^{(i,m)} \right) \right\}. \quad (45) \]
Interval Solution: Construct the empirical measure over $A(\infty) = [0, \infty)$ s.t.
\[
\hat{Q}(p) = \inf \left\{ x \in \{X^{(i)}(0)\}\text{ s.t. } p \leq \frac{1}{N} \sum_{i=1}^{N} W \left( X^{(i)}(0) ; x \right) I \left( X^{(i)}(0) ; x \right) \in [0,x] \right\}
\]
(46)

$X^{(i)}(0)$ represents the order statistics for the particles.

Given the quantile function estimate we get the risk measure estimates for any $\alpha \in (0,1)$ by:

Value-at-Risk (VaR): directly obtained using the estimated quantile function!

Spectral Risk (SRM): the SRM for a weight function $\phi : [0,1] \mapsto \mathbb{R}$ is given by
\[
\text{SRM}_Z(\phi) = \frac{1}{N} \sum_{i=1}^{N} X^{(i)}(0) \phi(p_i) \triangle p_i
\]

with $p_i = \sum_{j=1}^{i} W \left( X^{(j)}(0) ; x \right)$. For additional discussions and detailed examples of this numerical approach to risk estimation, we refer the reader to the examples found in [63] and [72].

Example 4 (Poisson-Log Normal LDA Model (continued)). Consider the Poisson-Log Normal compound process detailed in Example 1. We demonstrate results for the standard Monte Carlo approach and compare results to the path-space particle solution discussed above. The Monte Carlo solution involved $N = 50$ million samples, hence it is considered effectively exact since the resulting Monte Carlo error was insignificant. In addition, a grid based solution was adopted for the particle solution with $N = 50k$ per grid point giving a total of $NT = 500k$, with a grid width = 1. The estimated quantiles (rounded to integer values) are provided in the following table for two sets of parameter settings of $\lambda = 2$, $\mu = 2$ and $\sigma = 0.5$ and $\lambda = 2$, $\mu = 2$ and $\sigma = 1$. The particle solution presents a 95% confidence interval and the single loss approximation simply reports the asymptotic approximation no explicit error can be calculated for the point estimated quantile (as discussed above).

<table>
<thead>
<tr>
<th>Quantile Level</th>
<th>Standard Monte Carlo</th>
<th>Particle Solution (Alg. 2)</th>
<th>Single Loss Approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\sigma = 0.5$</td>
<td>$\sigma = 1$</td>
<td>$\sigma = 0.5$</td>
</tr>
<tr>
<td>50%</td>
<td>14</td>
<td>16</td>
<td>15 [14,16]</td>
</tr>
<tr>
<td>80%</td>
<td>27</td>
<td>39</td>
<td>25 [26,28]</td>
</tr>
<tr>
<td>90%</td>
<td>35</td>
<td>57</td>
<td>33 [31,35]</td>
</tr>
<tr>
<td>95%</td>
<td>42</td>
<td>77</td>
<td>40 [38,43]</td>
</tr>
<tr>
<td>99%</td>
<td>57</td>
<td>129</td>
<td>55 [54,56]</td>
</tr>
<tr>
<td>99.5%</td>
<td>77</td>
<td>234</td>
<td>73 [68,79]</td>
</tr>
<tr>
<td>99.95%</td>
<td>83</td>
<td>276</td>
<td>79 [73,91]</td>
</tr>
</tbody>
</table>

Table 1 Standard Monte Carlo Solution (exact) versus Particle Solution and First Order Single Loss Approximations.
The results that we present in Table I are obtained on a linearly spaced grid. However, this can be changed to either include a non-linear spacing, placing more points around the mode and less points in the tails, or as we detailed, straight out evaluation on an interval, avoiding the discretisation of the grid. For the sake of comparison between the standard Monte Carlo and the importance sampling estimates, we histogram the standard Monte Carlo procedure samples using unit length bins. We can see two things from Table I, firstly as expected the particle based solution performs accurately under any parameter settings for a modest computational budget. When compared to the Single Loss Approximation, we see that there is two clear advantages in having a complementary particle solution, since we obtain measures of uncertainty in the quantile point estimates, trivially. Secondly, we demonstrate that the Single Loss Approximations may not be as accurate as required for even these simple models at quantiles that may be of interest to assessment and are required for reporting of capital figures under financial regulation standards.

Appendix

Optimal Coupling Updated Models

This section is mainly concerned with the proof of the coupling formula (23). By construction, we clearly have that

\[ P(X_{p+1}^* \neq X_p^*) = \eta_p(A_p - A_{p+1}) = \eta_p(1 - G_p) = 1 - \eta_p(G_p) \]

On the other hand, we have

\[
\eta_{p+1}(\varphi) - \eta_p(\varphi) = \eta_p (S_p, \eta_p) (\varphi - \varphi) = \eta_p \left( [1 - G_p] \left[ \varphi - \Psi_{G_p}(\eta_p) \right] \right)
\]

Choosing \( \varphi = 1 - G_p \), so that

\[ \Psi_{G_p}(\eta_p)(\varphi) = 1 - \Psi_{G_p}(\eta_p)(G_p) = 0 \]

and

\[ \eta_p \left( [1 - G_p] \left[ \varphi - \Psi_{G_p}(\eta_p) \right] \right) = \eta_p \left( [1 - G_p]^2 \right) = 1 - \eta_p(G_p) \]

This ends the proof of the optimal coupling formulae (23). Next, we observe that

\[ 1 - \eta_p(G_p) = 1 - \eta_0(A_{p+1})/\eta_0(A_p) \quad \text{(with} \quad \eta_0 = \text{Law}(X)) \]

from which we conclude that

\[ \eta_0(A_p) \geq \eta_0(A_{p+1}) \geq (1 - \varepsilon) \eta_0(A_p) \implies P(X_{p+1}^* = X_p^*) \geq 1 - \varepsilon \quad (47) \]
References


