



A new algorithm for the loss distribution function with applications to Operational Risk Management

Dominique Guegan, Bertrand Hassani

► To cite this version:

Dominique Guegan, Bertrand Hassani. A new algorithm for the loss distribution function with applications to Operational Risk Management. 2009. halshs-00384398v1

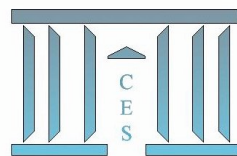
HAL Id: halshs-00384398

<https://shs.hal.science/halshs-00384398v1>

Submitted on 15 May 2009 (v1), last revised 30 Nov 2009 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



A new Algorithm for the Loss Distribution Function with applications to Operational Risk Management

Dominique GUEGAN, Bertrand HASSANI

2009.23



A new Algorithm for the Loss Distribution Function with applications to Operational Risk Management

D. Guégan*, B. Hassani†

April 24, 2009

Abstract

Operational risks inside banks and insurance companies is currently an important task. The computation of a risk measure associated to these risks lies on the knowledge of the so-called Loss Distribution Function. Traditionally this distribution function is computed via the Panjer algorithm which is an iterative algorithm. In this paper, we propose an adaptation of this last algorithm in order to improve the computation of convolutions between Panjer class distributions and continuous distributions. This new approach permits to reduce drastically the variance of the estimated VaR associated to the operational risks.

Keywords: Operational risk - Panjer algorithm - Kernel - Numerical integration - Convolution.

Introduction

The purpose of this article is - among the Basle Advanced Measurement Approach framework - to execute a Value-at-Risk evaluation on operational risks historical data. Indeed, the Basle II accords propose financial institutions to develop internal models in order to manage operational risks and compute some indicators. Our purpose is to participate to this last objective proposing

*Paris School of Economics, CES-MSE, Université Paris 1 Panthéon-Sorbonne, 106 boulevard de l'Hopital 75647 Paris Cedex 13, France, e-mail: dguegan@univ-paris1.fr

†CNCE and CES-MSE, Université Paris 1 Panthéon-Sorbonne, 106 boulevard de l'Hopital 75647 Paris Cedex 13, phone: +33 (0) 153403396 France, e-mail: bertrand.hassani@cnce.caisse-epargne.fr

a new approach to compute operational risks.

Operational risk is a very wide concept of risk arising from the failure of a business function. It is probably the most pernicious form of risks because it is indirectly responsible of numerous failures in financial institutions. Obviously it is not a new risk. Nevertheless, it is only recently, that institutions focus on tools which permit to quantify these operational risks using approaches developped in the insurance industry and from the value-at-risk techniques. Most financial fiascos can be traced to a combination of market or credit risks and failures of controls - in other words, they involve some form of operational risks. An illustration of such a risk is rogue trading. The spectacular bankrupts of the Barings bank and others involving fraud can be attributed to operational risk. Internal frauds arose at the highest level of the organization and was due to poor corporate governance. This recognition is bringing greater interest in operational risk and, this paper brings a new contribution for the computation of this kind of risks.

In order to find a robust way to compute the Loss Distribution Function associated to these operational risks, there exist several methods, some use classical methods to estimate this distribution function, others focus on extreme value theory. Here, we privilege the former approach using a new methodology. The loss distribution that permits to characterize the severities and the frequencies of the specific events has no easy analytical form because it is generally calculated as a compound distribution. Two kinds of classes of distributions are mainly used to model the events that are associated to operational risks: the discrete distribution functions such as Poisson, Binomial or Negative Binomial distributions for the frequencies of events and the continuous distribution functions - for instance Lognormal or Weibull - to characterize the severities of these events. To compute such compound distributions, convolution methods associated with Monte Carlo simulations are extensively used, [Frachot et al. (2001)] for details.

The usual convolution method is based on the Fast Fourier Transformed, or on the Heckman-Myers algorithm through the Laplace transform, [Heckman and Meyers (1983)] or finally on the Panjer's algorithm, [Panjer (1981)]. In this paper, we assume that the convolution of laws characterizing the frequencies, and the severities of the events provides a global loss distribution function that characterizes the operational risks and we will use it to propose a risk management strategy.

To compute this global loss distribution, we focus on a new faster and accurate algorithm.

Our approach mixes the use of the Panjer's algorithm with kernel density estimation, and Monte Carlo simulations in order to accurate and make the computations quicker than in the classical methods based only on the usual Panjer's Algorithm. Indeed, in the computations, problems arise when the frequencies are high: we might face several problems, specifically computer's limit, and most of time we cannot insure convergence because we cannot perform a sufficient number of iterations. Then, variance reduction becomes an important task. The computation's velocity of Panjer's algorithm is interesting, but it remains an initialization problem on which we focus, increasing the size of the information set to avoid the lack of robustness in the classical method. Thus, in this paper to avoid these drawbacks, we proceed in the following way: after building a first information set using Monte Carlo simulations, we complete it using a nonparametric kernel partial mesh. Finally, we apply the Panjer's algorithm on this new lattice.

The paper is organized as follows. In Section 1 we recall the notion of operational risks, and also the Loss Distribution Function. Section 2 presents the methodology that we use to approximate the Loss Distribution Function and specifies the numerical integration error. Section 3 presents experiments which permit to illustrate the interest of our methodology. Section 4 concludes.

1 Definitions and Recalls

An operational risk is a risk arising from execution of company's business functions. As such, it is a very wide concept including risks such as legal, physical or environmental frauds. The notion of operational risk is most commonly found in risk management programs of financial institutions that must organize it according to Basel II accords, and it is divided into credit, market and operational risk management. In many cases, credit and market risks are handled through a company's financial department, whereas operational risk management is coordinated centrally but most of the time divided and located in different operational units.

More specifically, the Bâle II treaty defines operational risk as the risk of loss resulting from inadequate or failed internal processes, people and systems, or from external events. This par-

ticular way of framing risk management is of particular relevance to the banking regime where regulators are responsible for establishing safeguards to protect against systemic failure of the banking system and the economy. The Basel Committee specifies the operational risk as: "The risk of loss resulting from inadequate or failed internal processes, people and systems or from external events." However, the Basel Committee recognizes that operational risk is a term that has a variety of meanings and therefore banks are allowed to implement specific and internal definitions for operational risks. Of course, some minimum guideline should be respected. The Basel II definition of operational risk clearly excludes strategic risk and reputation risk.

Basle accords propose to use different ways to quantify operational risks. We can distinguish, the scenarios: they permit to access loss due to extremal risk or internal risk for which we have no data; the expertises: experts give their point of view on -for example- the maximum potential loss and the mean loss; finally, the use of the Loss Distribution Function (LDF) which is based on historical data. In this paper we focus on this last approach, we provide statistical and mathematical solutions to compute the LDF. In fine, we will use this LDF to compute VaR that is the cash amount that the loss will not exceed when a failure occurs given a probability level [Jorion (2006)]. The Loss Distribution Function appears as a compound of the loss severity distribution function $F_{bl,et}$, ("bl" corresponding to the business line and "et" to the event type), and of the loss frequency distribution function $P_{bl,et}$.

As a matter of fact an amount of loss is most of the time positive. Some specific insurance might reverse the fact and we could face gains, but it is very unusual. For stability reason, we will assume that we have only losses. Thus, the major part of the losses are concentrated on the right tail of the distribution function and in order to take into account this fact, we retain the Log-normal distribution $F_{bl,et}$ - to model the occurrences probabilities of severities - whose density is equal to $f_{bl,et}(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$ for $x>0$, $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}$. To fit the loss frequency distribution function, we use Poisson distribution whose density is equal to $p_{bl,et}(k; \lambda) = e^{-\lambda} \frac{\lambda^k}{k!}$, $\lambda \in \mathbb{R}$, $k \in \mathbb{N}$. Thus, the Global Loss Distribution Function $G_{bl,et}$, appears like the mixture of the two previous distribution functions, and is equal to:

$$G_{bl,et} = \sum_{k=1}^{\infty} p_{bl,et}(k, \lambda) F_{bl,et}^{\otimes k}, \quad x > 0, \quad (1.1)$$

with

$$G_{bl,et} = 0, x = 0.$$

Here \otimes denotes the convolution operator on the distribution functions and $F^{\otimes n}$ the n -fold convolution of F with itself. We denote $g_{bl,et}$ the density of $G_{bl,et}$, then:

$$g_{bl,et} = \sum_{k=1}^{\infty} p_{bl,et}(k, \lambda) f_{bl,et}^{\otimes k}, x > 0. \quad (1.2)$$

Based on a discret representation of this distribution, the method we propose permits to reduce the variance of the estimates obtained thanks to the traditional algorithms.

2 A new numerical algorithm to compute the LDF

In the following, all the discussion concerns the quicker and best way to compute $G_{bl,et}$, $\forall bl, et$. As soon as, we have no close analytical form for (1.1), our objective is to develop a quick and robust algorithm to compute the LDF. We propose, inside this algorithm, a new step which permits to build a sequence of non-parametric densities that we use to increase the velocity of the algorithm. In the following, to simplify the notations, we will denote G the global loss distribution and g its density.

2.1 Details of our procedure

1. The first step consists to use Monte Carlo methods [Fishman (1996)] to simulate i realisations of the Poisson distributions denoted (N_1, \dots, N_i) , and to compute $\sum_{i=1}^j N_i$ realisations following a logNormal distribution: (L_1, \dots, L_j) . Finally, we get a realisation of a Loss distribution $\hat{T}L_i$ as $\hat{T}L_i = \sum_{j=1}^{N_i} L_j$, which provides an empirical representation of the Global Loss Distribution \hat{G} ; its density is denoted \hat{g} .
2. This first approximation of the true LDF provides an estimate of the associated VaR measure whose bias and variance are important, as we will show in the next section thus, in order to minimize this bias, the Panjer algorithm is used to compute iteratively the convolution (1.1). The iteration is done point by point, and finally we obtain a new approximation of G that we still denote \hat{G} .

- (a) Empirically it has been observed that using this new approximation for G provides an estimated VaR whose variance is still important. In order to make the method more robust, we propose a new step permitting to increase the information set on which the iterations are done, and then we observe a reduction of the variance. We proceed in the following way:
- (b) Using the support of the density \hat{g} obtained in step one, we build a sequence of rolling subsets A_k , $k = 1, 2, \dots$ on which we adjust non-parametric densities $\tilde{g}^{\alpha, \beta}(k)$ using Epanechnikov kernel, [Wand and Jones (1995)], and we get a sequence of densities $(\tilde{g}^{\alpha, \beta}(k))$. Here, α represents the Panjer point and β the kernel point between two Panjer points. We use these successive densities to improve the first approximation of the LDF. This step is illustrated through the Figure 1.
- (c) To build this sequence of densities $(\tilde{g}^{\alpha, \beta}(\cdot))$, the kernel that we use is defined on a finite support and can be expressed as follows:

$$K(t) = \begin{cases} \frac{3}{4} \frac{(1 - \frac{1}{5}t^2)}{\sqrt{5}}, & \text{for } |t| \leq \sqrt{5} \\ 0, & \text{otherwise.} \end{cases} \quad (2.1)$$

In this method, we use a bandwidth of $15^{\frac{1}{5}} \approx 1.3510$, which has been retained after using a cross validation method based on a AMISE criteria [Wand and Jones (1995)].

3. The Panjer algorithm is iterated on this new information set through the following recursive formula:

$$g(x) = p_1 f(x) + \int_0^x (a + \frac{by}{x}) f(y) g(x-y) dy, \quad \text{where } x > 0. \quad (2.2)$$

In expression (2.2), p_1 represents the Poisson distribution applied on the first Panjer point and f is the logNormal density. The expression (2.2) can be rewritten as:

$$g(x) = \lambda e^{-\lambda} f(x) + \frac{\lambda}{x} \int_0^x y f(y) g(x-y) dy \quad x > 0, \quad (2.3)$$

where $g(x) := \Phi(x) + \frac{\lambda}{x} \Psi(x, y)$, $\Phi(x) = \lambda e^{-\lambda} f(x)$ and, $\Psi(x, y) = \int_0^x y f(y) g(x-y) dy$.

4. We approximate the integrate in (2.3) using the trapezium rule [Rahman and Schmeisser (1990)]. Being in any point x_n , we get :

$$\tilde{g}(x_n) = \lambda e^{-\lambda} f(x_n) + \frac{\lambda}{x_n} \sum_{\alpha=1}^{n-1} \sum_{\beta=0}^M \frac{x_{\alpha, \beta} f(x_{\alpha, \beta}) \tilde{g}(x_n - x_{\alpha, \beta}) + x_{\alpha, \beta+1} f(x_{\alpha, \beta+1}) \tilde{g}(x_n - x_{\alpha, \beta+1})}{2} (x_n - x_{\alpha, \beta}) \quad (2.4)$$

where n is the number of Panjer points, $n \in [1, N]$. $M = 2^\zeta$, $\zeta \in \mathbb{N}$, is the number of points between two Panjer points. The first point in the relationship (2.4) is denoted $x_{1,0}$ and is called a Panjer point. During the integration process, we use successively the Panjer points x_0, x_1, x_2, \dots and the kernel points $\tilde{x}_{1\beta}, \tilde{x}_{2\beta}, \tilde{x}_{3\beta}, \dots$ obtained previously in the step (2b). The integration procedure works in the following way.

- (a) Initialization (Step 0): from $x_{1,0}$, we build $\tilde{x}_{11}, \dots, \tilde{x}_{1M}$ and we use the full sequence $x_0, \tilde{x}_{11}, \dots, \tilde{x}_{1M}$ to compute $x_{2,0}$.
 - (b) Step 1 : We define the set $I_1 = (g(x_0), g(x_1), \tilde{g}_{21}, \dots, \tilde{g}_{2M})$ on which we compute $g(x_2)$ using the relationship (2.4).
 - (c) Steps 2 to n : At each step, we define a sequence of new information sets I_k , $k = 2, \dots, n$ on which we build the sequence $(g(x_0), g(x_1), \dots, g(x_n))$.
5. The density associated to $g(x_n)$ provides the Global Loss Density.

2.2 Numerical integration error

Thanks to the previous algorithm, we have built an approximation of the LDF using several steps. We specify now the errors we made at each step. In the Panjer recursive formula (2.3), $\Phi(x)$ is a constant, therefore the error is generated by the numerical integration of Ψ . Consequently, in the following, we only present the terms from which the error is springing. Some details are given in the Appendix.

1. In the first step, when we use the Monte Carlo method, we consider a discret version of the lognormal distribution, thus:

$$g_i = \frac{\lambda}{i} \sum_{j=1}^i j f_j g_{i-j}. \quad (2.5)$$

Then, we evaluate the difference between (2.3) and (2.5), denoted $E1$:

$$E1 = \left| \lambda e^{-\lambda} f(x) + \frac{\lambda}{x} \int_0^x y f(y) g(x-y) dy - \frac{\lambda}{i} \sum_{j=1}^i j f_j g_{i-j} \right|. \quad (2.6)$$

This error is bound by a constant $C > 0$ and cannot be use in our objective to approach the LDF for risk management.

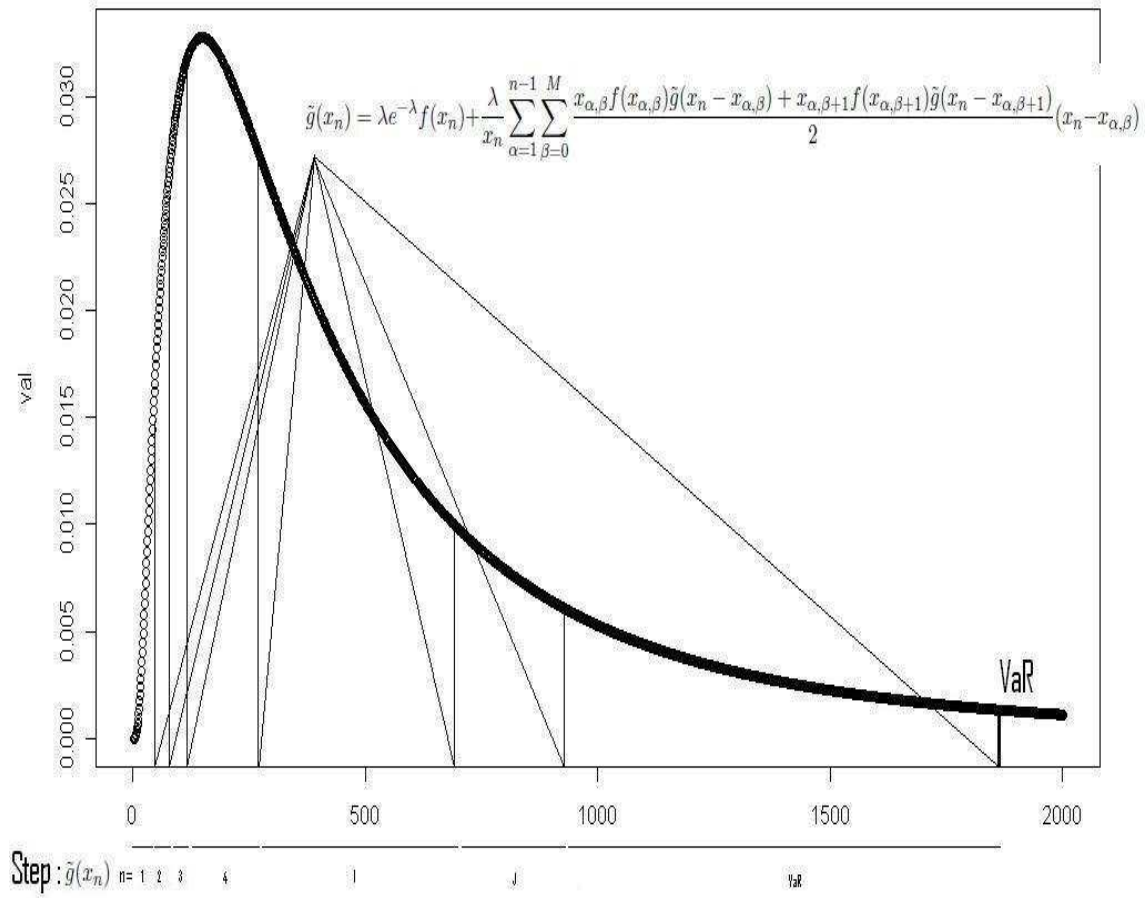


Figure 1: Details of the steps 2-3-4 to approximate the LDF

2. In the second step, when we use directly the Panjer algorithm, we consider the trapezoidal rule and we get the error E2:

$$E2 = \left| \int_0^x y f(y) g(x-y) dy - \sum_{\alpha=1}^{n-1} \frac{x_{\alpha} f(x_{\alpha}) \tilde{g}(x_n - x_{\alpha}) + x_{\alpha+1} f(x_{\alpha+1}) \tilde{g}(x_n - x_{\alpha+1})}{2} (x_n - x_{\alpha}) \right| \quad (2.7)$$

This error can be bounded as follows:

$$E2 \leq \left(\frac{1}{12n} \right) \cdot \max_{y \in [0, n]} |\theta''(x, y)| \quad (2.8)$$

where,

$$\theta(x, y) = y f(y) g(x - y), \quad (2.9)$$

and the second derivative $\theta''(\cdot)$ is computed with respect to y . Using Leibnitz formula (see the Appendix), we can show that $\theta''(\cdot)$ is finite. Thus, the term E2 tends to zero as n (the sample size) tends to ∞ . We observe that the use of the trapezoidal rule permits to reduce drastically the error $E1$. Nevertheless this procedure does not permit to get consistent results for the VaR measure because of the lack of information at the beginning of the procedure, which is endogenous to the Panjer algorithm.

3. Asymptotically ($n \rightarrow \infty$), the error $E2$ tends to zero. Nevertheless, in practice, we work with a finite sample, thus it is important to get a small error for finite samples. In order to solve this problem, we increase the initial information set on which we apply the Panjer algorithm, and the application of the trapezoidal rule provides the error E3: :

$$E3 = \left| \Psi(x, y) - \sum_{\alpha=1}^{n-1} \sum_{\beta=0}^M \frac{x_{\alpha, \beta} f(x_{\alpha, \beta}) \tilde{g}(x_n - x_{\alpha, \beta}) + x_{\alpha, \beta+1} f(x_{\alpha, \beta+1}) \tilde{g}(x_n - x_{\alpha, \beta+1})}{2} (x_n - x_{\alpha, \beta}) \right| \quad (2.10)$$

and :

$$|E3| \leq \left(\frac{n}{12M^2} \right) \cdot \max_{y \in [0, n]} |\theta''(x, y)|, \quad (2.11)$$

with $M \gg n$. Then $E2 \gg E3$. We can remark that the rate of convergence in (2.11) is in M^{-1} with $M \gg n$. This last error goes to zero very quickly and depends on the lattice that we use inside the algorithm. Thanks to the new lattice we have introduced in the algorithm, we robustify the computation of the LDF improving its approximation.

3 Results

In order to prevent the institution from a loss which could imply it to go bankrupt, it has been proposed to proceed in a VaR computation in order to decide the economical capital allowed for this kind of risks. Several methods can be used to achieve this purpose. The new method we described previously permits to reach our goal in a faster and more accurate way. We use the previous method to compute the VaR associated to this LDF. The results we present are based on simulation experiments, but the proposed parameters have been found in real cases. First, we recall the definition of VaR for operational risks.

In financial mathematics and financial risk management, Value at Risk is a widely popular measure for the risks of losses. For a given LDF, a given probability and a given time horizon, the VaR is defined as a threshold value such that the probability, that the loss due to operational risk over the given time horizon, exceeds this value for a given probability level. Formally, we can define it as :

Definition 3.1. *Given a confidence level $\alpha \in [0, 1]$, the VaR of the portfolio is given by the smallest number l such that the probability that the loss L exceeds l is not larger than $(1 - \alpha)$.*

$$VaR_{(1-\alpha)\%} = \inf(l \in \mathcal{R} : P(L > l) \leq (1 - \alpha)) \quad (3.1)$$

The regulator imposes a security threshold of 99.9%, therefore $\alpha = 0.1\%$. We will use it here.

In this work, we have chosen the logNormal distribution of parameters (μ, σ) and the Poisson distribution of parameter (λ) to compute the LDF distribution function. In our example, we assume several values for the parameters.

- Figure 2 : $(\mu = 0, \sigma = 1)$, and $\lambda = 25$
- Figure 3 : $(\mu = 0.1, \sigma = 1.2)$, and $\lambda = 500$
- Figure 4 : $(\mu = 5, \sigma = 2)$, and $\lambda = 25$

We need a benchmark in order to compare the values got for the VaR by the different methods. The benchmark will be the VaR obtained using Monte Carlo simulations with a number of

iterations $\Omega = 10^6$. We work with 1000, 10000, 100000 and 1000000 Poisson realisations, and we use the previous algorithm to get the LDF. We change the generator's seed 100 times to create Box Plots. They provide the dispersion and the standard deviation of the results. Four methods are implemented to compute the VaR:

1. Monte Carlo simulations (MC). We simulate the LDF using (1.1).
2. Kernel method (K). We simulate the LDF using (1.1) smoothed with an Epanechnikov kernel.
3. The new algorithm described in Section 4.1 (H).
4. The Panjer's method (P). We use the expression (2.3) to compute the LDF.

In Figure 2, we provide results obtained for the first set of parameters. In bold line the value of the VaR obtained by each method for different sample sizes, and a "box" which illustrates the dispersion around this VaR. In Figure 1 (d), we observe that the methods, (a), (b), (c) are competitive : the values of the VaR are very close. The value obtained with Panjer method is higher. Now we can also remark that with our algorithm we reduce drastically the dispersion around the VaR. This last result is fundamental with respect of management risk strategy. Indeed, it permits to reduce the costs compared to Panjer algorithm and the calculation time comparing with traditional Monte Carlo methods.

In Figure 3, the results provided with the set of parameters ($\mu = 0.1, \sigma = 1.2$), and $\lambda = 500$ shows that the results are always more interesting than with the Panjer algorithm (bias and variance are reduced), and the variance of the Hybrid algorithm is still lower than using Monte Carlo method, even if the difference does not appear so large.

In Figure 4, the results provided with the set of parameters ($\mu = 5, \sigma = 2$), and $\lambda = 25$ show that the Panjer algorithm, and our new algorithm are both competitive comparing with the results obtained with the Monte Carlo Simulations. It is important to note that our algorithm is always as good as the best of the three others.

Whatever the set of values used in the several previous experiments, the new algorithm permits to reduce, in all cases, the bias and the variance of the estimated VaR Measure.

4 Conclusion

The approach proposed in this paper is innovative in the sense that it has been developed to deal with traditional drawbacks, such as variance (Monte Carlo), speed of computation and results precision. The new method, proposed in section two, is a recursive reduction variance method, which is associated to a new form of interpolation based on kernel smoothing. The trade-off we had from the use of historical algorithms is solved by our new hybrid one. We have a large variance reduction, therefore the number of iterations is reduced. Consequently, the simulation time decrease is nearly of $\approx 30\%$ comparing with the others: this point is fundamental for practitioners. At the same time the accuracy of the VaR results compared to a VaR computed from a Monte Carlo simulation of a million iterations is very good.

Furthermore, as we have shown, the numerical integration error we would have faced with the Panjer algorithm is reduced by the kernel interpolation. This error tends to zero as the number of interpolation points grows, and we have no initialisation bias we would face with a traditional Riemann sum. Nevertheless we have made some assumptions concerning the choice of the distributions that we can bypass and in a companion paper, we will introduce extremal distributions to fit in a better way the right tail. Furthermore, to increase the computation speed, it might be interesting to assess the LDF initial support and consequently bypass the Monte Carlo simulation step.

5 Appendix

5.1 The Panjer algorithm: a brief description

Panjer (1981) provides a recursive formulation to compute high order convolutions. The main result is the following. Assuming that there exists two constants a and b such that

$$p_n = p_{n-1}(a + \frac{b}{n}), \text{ where } n = 1, 2, 3, \dots, \quad (5.1)$$

[!]

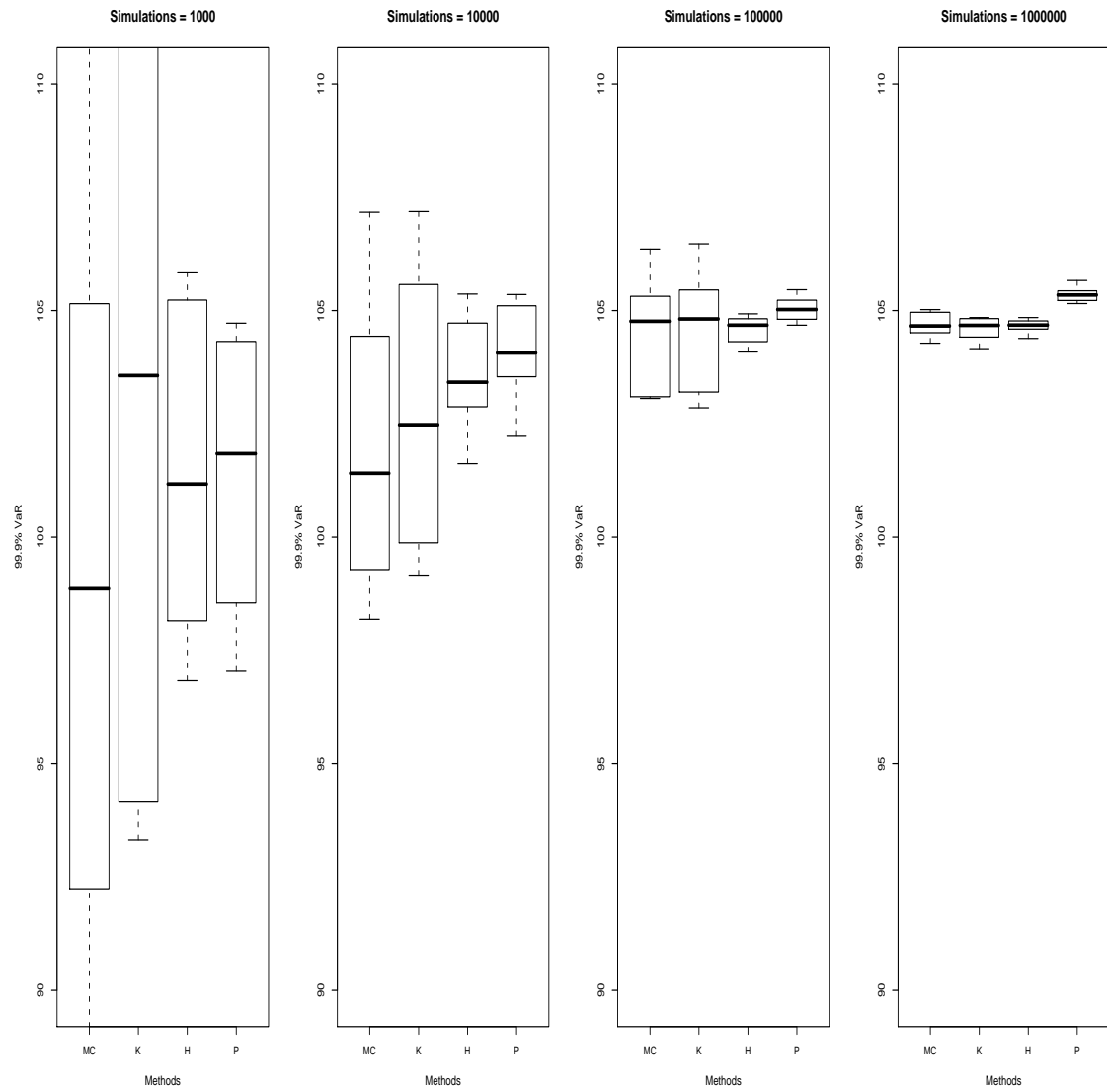


Figure 2: Box plots obtained from the four methods: $(\mu = 0, \sigma = 1)$, and $\lambda = 25$

[!]

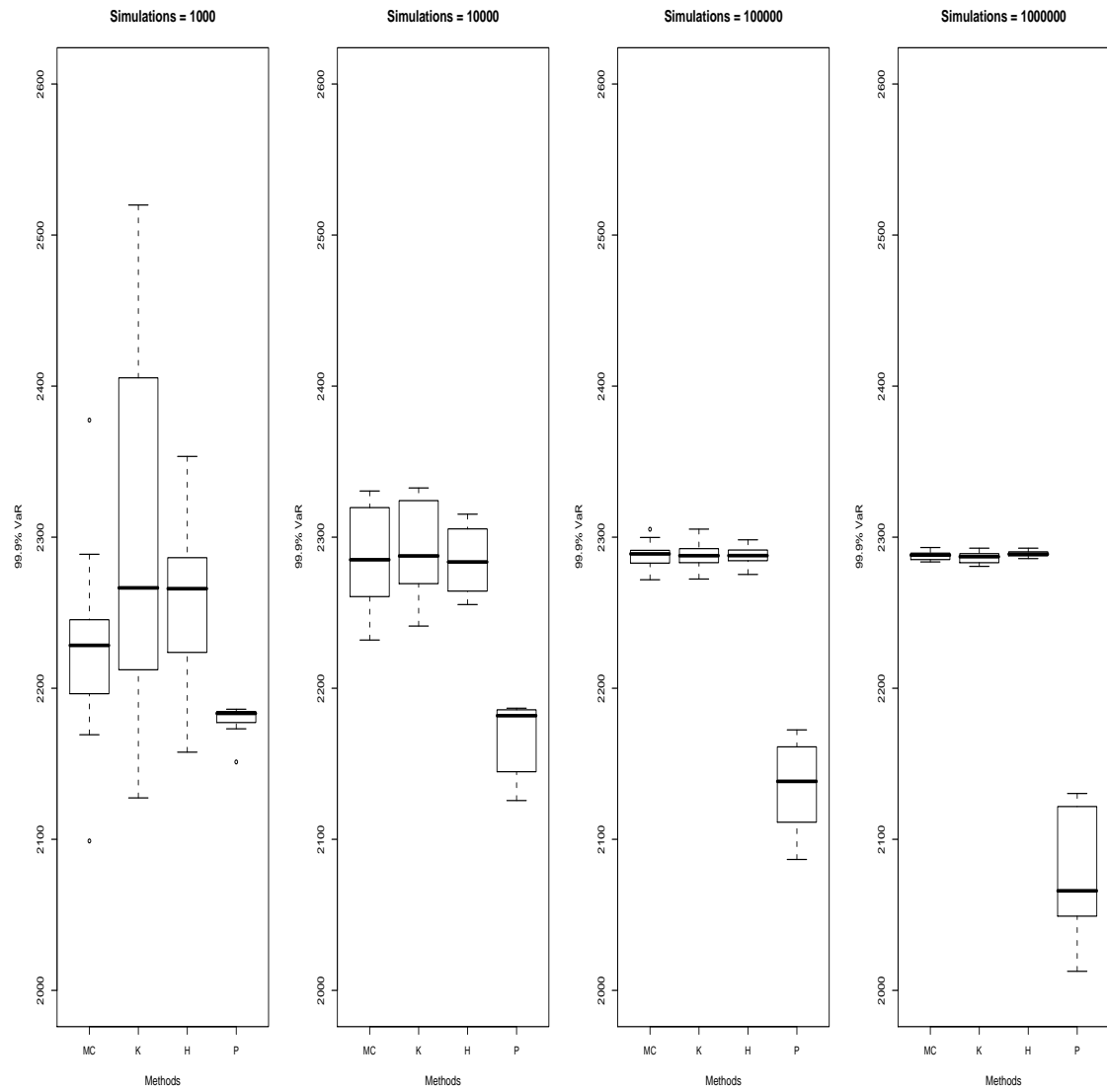


Figure 3: Box plots obtained from the four methods: $(\mu = 0.1, \sigma = 1.2)$, and $\lambda = 500$

[!]

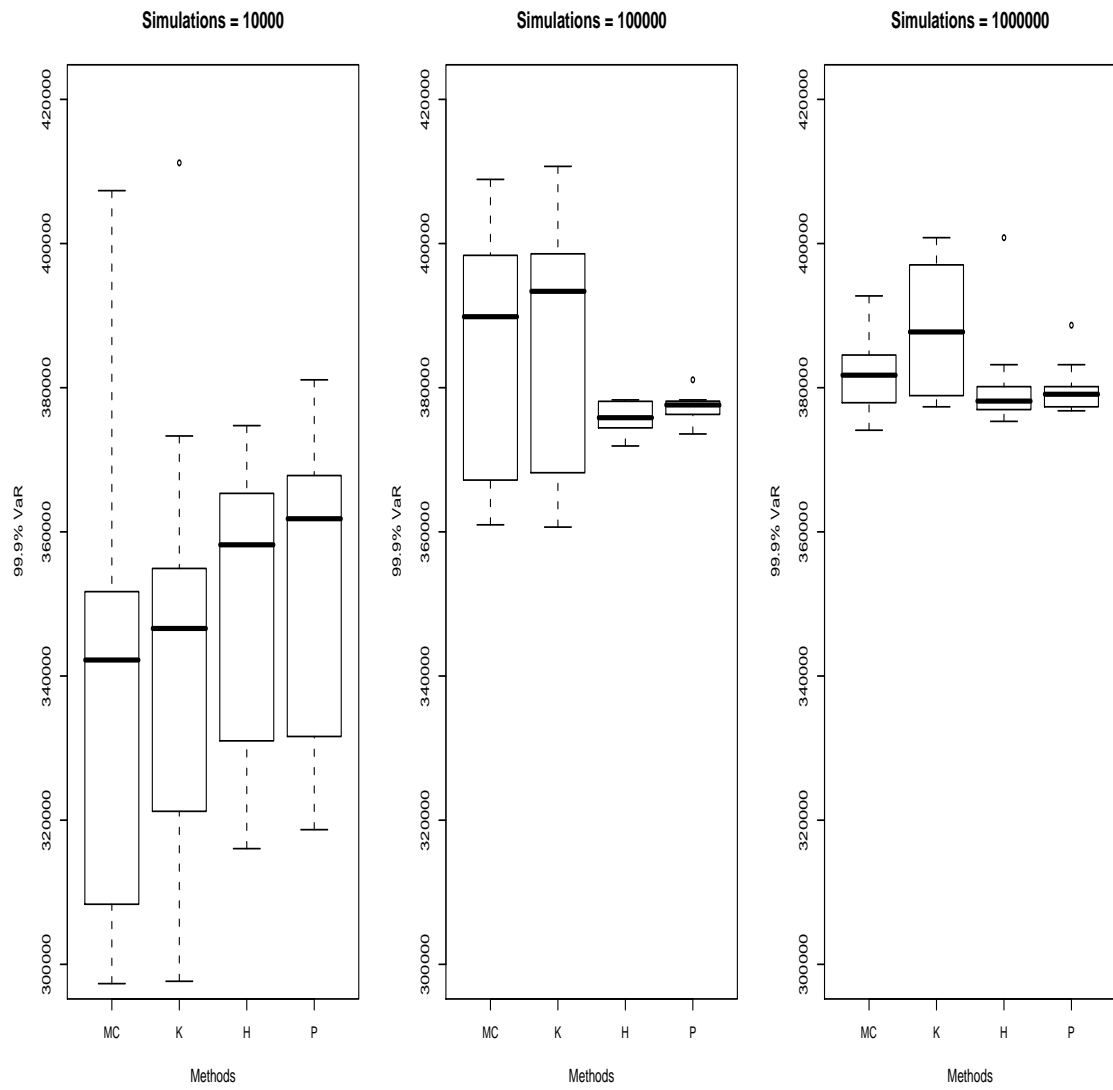


Figure 4: Box plots obtained from the four methods: $(\mu = 5, \sigma = 2)$, and $\lambda = 25$

then the following recursion holds for a regular function $g(\cdot)$:

$$g(x) = p_1 f(x) + \int_0^x \left(a + \frac{by}{x}\right) f(y) g(x-y) dy, \quad \text{where } x > 0. \quad (5.2)$$

We can apply this recursion for the computation of the LDF. We consider the family of claim number distributions satisfying the recursion (5.1). In our case the Panjer class distribution is the Poisson distribution, such as $p_n = \frac{e^{-\lambda} \lambda^n}{n!}$, where $n = 0, 1, 2, \dots$ and $p_{n-1} = \left(\frac{\lambda}{n}\right) p_n$, where $p_0 = e^{-\lambda}$, $a = 0$ $b = \lambda$. Then, using the formula (5.2), we get :

$$g(x) = \begin{cases} \sum_{n=1}^{\infty} p_n f^{*n}(x), & x > 0 \\ 0, & \text{otherwise} \end{cases} \quad (5.3)$$

and $f(x)$ is any continuous type distribution for $x > 0$. Therefore, in our particular case, we could write the distribution as:

$$g(x) = \lambda e^{-\lambda} f(x) + \frac{\lambda}{x} \int_0^x y f(y) g(x-y) dy. \quad (5.4)$$

In the case of a discret claim amount distribution, we use:

$$g(x) = \frac{\lambda}{i} \sum_{j=1}^i j f_i g_{i-j}. \quad (5.5)$$

5.2 Numerical Integration Error

Right after the Monte Carlo simulation, using \hat{G} got in step 1 of Subsection 2.1, we approximate this distribution and its density using Panjer algorithm.

We recall that any integrated function $\theta(\cdot)$ on a closed interval, such as $\int_0^x \theta(y) dy$, can be approximated by :

$$\sum_{j=0}^{N-1} \delta \sum_{i=1}^s \omega_i \theta(y_j + \alpha_i \delta), \quad (5.6)$$

where $\delta = \frac{x}{N}$, $\omega \in R$ and, N is the number of trapezes. Then, we rewrite the numerical integration error E_2 as :

$$\left| \int_0^x y f(y) g(x-y) dy - \sum_{j=0}^{N-1} \delta \sum_{i=1}^s \omega_i \theta(y_j + \alpha_i \delta) \right|. \quad (5.7)$$

The following theorem provides the expression of the previous approximation (5.7) as a function of the Peano Kernel, [Schatzman (1991)].

Theorem 5.1. *Considering a p -order quadrature formula and a k integer satisfying $k \leq p$. If $\theta : [y_0, y_0 + \delta] \rightarrow \mathfrak{R}$ and k continuously differentiable times, the error is,*

$$E(\theta, y_0, \delta) = h^{k+1} \int_0^1 N_k(\tau) \theta^{(k)}(y_0 + \tau\delta) d\tau, \quad (5.8)$$

where $\theta^{(k)}$ represents the k^{th} derivative of θ and $N_k(\tau)$, the Peano kernel, is given by,

$$N_k(\tau) = \frac{(1-\tau)^k}{k!} - \sum_{i=1}^s \omega_i \frac{(\alpha_i - \tau)_+^{k-1}}{(k-1)!}. \quad (5.9)$$

The next proposition recall the main Peano Kernel properties: we are particularly interested by the fourth.

Proposition 5.1. *(Peano's kernel properties) Considering a p -order quadrature formula and a number k satisfying $1 \leq k \leq p$. We have :*

1. $N'_k(\tau) = -N_{k-1}(\tau)$ for $k \geq 2$ (for $\tau \neq \alpha_i$ if $k = 2$);
2. $N_k(1) = 0$ for $k \geq 1$ if $\alpha_i \leq 1$ ($i = 1, \dots, s$);
3. $N_k(0)$ for $k \geq 2$ if $\alpha_i \geq 0$ ($i = 1, \dots, s$);
4. $\int_0^1 N_p(\tau) d\tau = \frac{1}{p!} \left(\frac{1}{p+1} - \sum_{i=1}^s \omega_i \alpha_i^p \right) = C(\text{error constant})$;
5. $N_1(\tau)$ is fragmently linear, of slope coefficient -1 with jumps of height denoted ω_i at each points α_i ($i = 1, \dots, s$).

Now, we give a bound for the error (5.7), where the function θ is defined on a given interval :

Theorem 5.2. *Let $\theta : [a, b] \rightarrow \mathfrak{R}$ k continuously differentiable and a p -order quadrature formula ($p \geq k$), then we have the following result*

$$|\tilde{E}| \leq h^k \cdot (b-a) \cdot \int_0^1 |N_k(\tau)| d\tau \cdot \max_{x \in [a, b]} |\theta^{(k)}(x)|, \quad (5.10)$$

where $h = \max_j h_j$.

From this theorem, we get the specific bound E_2 applying the previous result to (5.7):

$$|\tilde{E}| \leq \left(\frac{1}{M} \right)^2 \cdot (N) \cdot \frac{1}{12} \cdot \max_{y \in [0, N]} |\theta''(x, y)|. \quad (5.11)$$

The derivative is computed with respect to y and is obtained using the Leibniz formula which provides the k^{th} derivatives of $\theta = uv$ when u and v are C^∞ :

$$\theta^{(k)} = \sum_{n=0}^k \binom{k}{n} u^{(n)} v^{(k-n)}. \quad (5.12)$$

Then, for $u(y) = yf(y)$ and $v(y) = g(x - y) = g(x, y)$, we obtain :

$$\frac{\partial^2 \theta}{\partial y^2} = (2, 0) y f(y) \frac{\partial^2 g(x, y)}{\partial y^2} + (2, 1) \left(f(y) + y \frac{\partial f(y)}{\partial y} \right) \frac{\partial g(x, y)}{\partial y} + (2, 2) \left(2 \frac{\partial f(y)}{\partial y} + y \frac{\partial^2 f(y)}{\partial y^2} \right) g(x, y). \quad (5.13)$$

And,

$$\frac{\partial^2 \theta}{\partial y^2} < \infty \quad (5.14)$$

References

- A. Frachot, P. Georges, and T. Roncalli. Loss distribution approach for operational risk. *Working Paper, GRO, Crédit Lyonnais*, 2001.
- P. E. Heckman and G. G. Meyers. The calculation of aggregate loss distributions from claim severity and claim count distributions. *Proceedings of the Casualty Actuarial Society*, 70:22–61, 1983.
- H. H. Panjer. Recursive evaluation of a family of compound distributions. *Astin Bulletin*, 12: 22–26, 1981.
- P. Jorion. *Value at Risk: The New Benchmark for Managing Financial Risk*. 2006.
- G. S. Fishman. *Monte Carlo*. Springer, 1996.
- M. P. Wand and M. C. Jones. *Kernel Smoothing*. Chapman and Hall/CRC, 1995.
- Q. I. Rahman and G. Schmeisser. *Analytic Theory of Polynomials*. Oxford Science Publications, 1990.
- M. Schatzman. *Analyse numérique : cours et exercices*. InterEditions, 1991.
- B. W. Silverman. *Density Estimation for statistics and data analysis*. Chapman and Hall/CRC, 1986.
- P. J. Davis and P. Rabinowitz. *Methods of Numerical Integration*. Academic Press, 1984.