

Université de Montréal

*Approche par la distribution des pertes pour le risque opérationnel :
Un modèle Bayésien de Mélange de Distribution avec fusion des
Données internes et externes pour la sévérité des pertes.*

Par :
David Romain DJOUMBISSIE

GENTRE DE DOCUMENTATION
1 JUN 2005
SCIENCES ECONOMIQUES U DE M

Département de Sciences Économiques, Faculté des Arts et Sciences.

*Rapport de recherche présenté à la Faculté des études supérieures en vue de l'obtention du
grade de Maîtrise en Économie, Option Économie Financière.*

Avril 2005

C'est l'occasion pour moi d'adresser mes sincères remerciements à M. René Garcia, M. Bryan Campbell, M. Jean François Angers, Le CIRANO et le Département de Sciences Économiques de L'Université de Montréal pour m'avoir permis de mener dans d'excellentes conditions cette étape de ma formation.

TABLE DES MATIÈRES

<i>Tableaux et Figures</i>	<i>i</i>
<i>Sigles et abréviations</i>	<i>ii</i>
<i>Résumé</i>	2
<i>Summary</i>	3
<i>Avant-propos</i>	4
<i>I- Risque opérationnel et Mesure par la distribution de perte</i>	9
I.1- Identification du risque opérationnel	9
I.2- Description de l'approche par la distribution des pertes	14
I.2.1- Calcul de la Distribution de perte pour le risque opérationnel.....	14
I.2.2- Mise en œuvre de la LDA dans le cas du risque opérationnel	17
<i>II- Modélisation avec mélange de distribution et Estimation</i>	21
<i>par MCMC</i>	21
II.1- Justification et Description du modèle	21
II.1.1- Justification de l'approche	22
II.1.2- Description du modèle	24
II.2- Sélection et calcul des distributions a priori et a posteriori	26
II.2.1- Spécification avec la distribution Lognormal	26
II.2.2- Spécification avec la Distribution Gamma	30
II.3- Calibrage et Estimation	33
II.3.1- Sélection du nombre de composantes par le critère de Bayes	34
II.3.2- Estimation du modèle par MCMC	35
<i>III- Présentation et analyse des résultats</i>	41
III.1- Analyse de la convergence	42
III.2- Estimation des paramètres et de la Densité des données	45
III.3- Sensibilité des résultats par rapport aux hypothèses.	49
III.4- Classification des observations et Calcul de la VAR	51
<i>Conclusion</i>	55
<i>Bibliographie</i>	56
<i>ANNEXES</i>	58

LISTE DES TABLEAUX

Tableau 2.1: Marge d'erreur en fonction du niveau de confiance et de la taille de l'échantillon (Loi Gamma avec deux paramètres).....	23
Tableau 3.1 : quelques statistiques descriptives des Données de pertes simulées.....	41
Tableau 3.2 : Valeur estimée et quantiles des moyennes μ_1, μ_2, μ_2 des trois composantes du mélange	46
Tableau 3.3 : Valeur estimée et quantiles des écarts types $\sigma_1, \sigma_2, \sigma_3$ des trois composantes du mélange	47
Tableau 3.4 : Valeur estimée et quantiles des proportions P_1, P_2, P_3 des trois composantes du mélange	48
Tableau 3.5.a : Sensibilité par rapport à la distribution a priori de la moyenne.....	49
Tableau 3.5.b : Sensibilité par rapport à la variance de la moyenne a priori $\mu \rightarrow N(a,b)$	49
Tableau 3.5.c : Sensibilité par rapport au paramètre d'échelle de la Variance a priori $\sigma^2 \rightarrow IG(c,d)$	50
Tableau 3.6 : pourcentage moyen a posteriori des observations pour chaque source de données.	51
Tableau 3.6.a : Quantile de la distribution de perte	52
Tableau 3.6.b : Quantile de la distribution de perte	52
Tableau 3.6.c : Quantile de la distribution de perte	53
Tableau 3.6.d : Quantile de la distribution de perte	53
Tableau 3.6.e : quelques statistiques descriptives des Données de pertes simulées(cas2).....	53
Tableau 3.6.f : Comparaison des quantiles de la distribution de perte (le nombre de groupe a posteriori est $K=2$).....	54

LISTE DES FIGURES

figure 1 : traces des moyennes μ_1, μ_2, μ_2 des trois composantes du mélange (T=40000, T=60000).....	42
figure 2 : traces des écarts types $\sigma_1, \sigma_2, \sigma_3$ des trois composantes du mélange (T=40000, T=60000).....	43
figure 3 : traces des proportions P_1, P_2, P_3 des trois composantes du mélange (T=40000, T=60000).....	44
figure 4 : Histogramme des données et estimation de la densité associée.....	45
figure 5 : Densités des moyennes μ_1, μ_2, μ_2 des trois composantes.....	46
figure 6 : Densités des écarts types $\sigma_1, \sigma_2, \sigma_3$ des trois composantes.....	47
figure 7 : Densités des proportions P_1, P_2, P_3 des trois composantes.....	48

SIGLES ET ABREVIATIONS

AMA : Approches des mesures avancées.

BCBS : Comité de Bâle surveillance bancaire.

BRI : Banque des Règlements Internationaux.

COSO : Committee of Sponsoring Organizations of the Treadway Commission.

COBIT : Control Objectives for Information and related Technology .

MCMC : Méthode de Monte Carlo par Chaîne de Markov.

LDA : Loss Distributional Approach.

RJMCMC : Méthode de Monte Carlo par Chaîne de Markov et Saut Réversible.

Résumé

Le nouvel accord de Bâle(2003) dans le souci de stabiliser davantage et assurer une saine croissance du système financier international fait du risque opérationnel le troisième facteur de risque auxquels les banques devront dorénavant faire face dans leurs activités au quotidien. Ce risque est considéré comme étant le « *le risque de pertes résultant de carences ou de défaillances attribuables à des procédures, personnels et systèmes internes ou à des évènements extérieurs* ». Malgré le caractère dynamique et assez vague de certains aspects inhérents à ce risque, l'identification ne cause plus de véritable problème et la principale approche de mesure et de gestion vers laquelle les institutions devraient tendre à terme est connue sous le nom « Approche de mesure Avancée ». elle est mise en œuvre soit à travers des modèles paramétriques, soit des modèles causals. A ce jour la méthode LDA de par ses nombreux avantages(robustesse, intelligibilité, coût,...) apparaît comme l'une des plus adaptées. C'est en effet une approche statistique utilisant la composée de la distribution de fréquence et de sévérité pour estimer la distribution de perte et déduire en fonction du niveau de confiance fixé la valeur à risque. Cependant son élaboration est assez contrariée par la grande difficulté à disposer d'un historique de perte consistant (trois à cinq ans). L'un des plus grands défis pour chaque structure réside dans la capacité à intégrer aux données internes, les données du reste de l'industrie financière afin d'obtenir une distribution de perte robuste et mieux prendre en compte les pertes extrêmes. Sur le plan théorique, le modèle de mélange de distribution représente un des outils potentiellement adaptés pour de telles études. Dans un cadre bayésien, nous utilisons un algorithme de MCMC pour estimer la distribution de la sévérité des pertes par un modèle de mélange de distribution. Malgré le fait que les données de pertes soient simulées, des résultats raisonnables sont obtenus dans un des cas étudiés, mais ce résultat n'est pas confirmé et la valeur à risque est considérablement surestimée lorsque l'échantillon construit est relativement homogène(en terme d'ordre de grandeur). Il nous semble par ailleurs nécessaire d'approfondir davantage les aspects suivants : l'application sur des données réelles, le type de distribution théorique, l'utilisation de l'algorithme RJMCMC plutôt que le MCMC, la pondération spécifique des différentes sources dans le mélange et l'orientation de façon à intégrer spécifiquement les préoccupations liées aux pertes extrêmes.

Mots clés : Risque opérationnel, MCMC, mélange de distribution.

Summary

The new agreement of Basel in the concern to stabilize more and to assure a healthy growth of the international financial system made by the operational risk third risk factor in which banks should from now on face in their activities in everyday life. This risk is considered as being " *the risk of losses resulting from deficiencies or from failings attributable to procedures, staffs and internal systems or in outside events* ". In spite of the dynamic and rather vague character of certain aspects inherent to this risk, identification does not cause any more real problem and the main approach of measure towards which institutions should aim eventually is known under the name " Approach of Advanced measure ". It is implemented or through statistical models (LDA,...) or causal models (Scorecard,...). This day the method LDA from its many advantages (robustness, comprehensibility, cost, ...) seems as one of the most adapted. It is indeed a statistical approach using the compound of the distribution of frequency and severity to estimate the distribution of loss and to deduct according to the fixed reliable level value at risk. However the elaboration is relatively opposed by the difficulty having a substantial historic of losses (three - five years). One of the biggest challenges lies in the capacity to be integrated into internal data, data of the rest of the financial industry to obtain robust Estimators and to better take into account the extreme losses. On the theoretical plan, the model of mixture of distribution represents one of the tools potentially adapted to such studies. In an Bayesian approach , we use MCMC algorithm to estimate the distribution of the severity of losses by a model of mixture of distribution. In spite of the data of losses are feigned, reasonable results are obtained in one of studied cases, but this result is not confirmed and value at risk is considerably overestimated when the constructed sample is relatively homogeneous (in term of order of height). It seems to us besides necessary to deepen more the following aspects: application on real data, type of theoretical distribution, use of the algorithm RJMCMC rather than MCMC, specific level-headedness from the various sources in the mixture and orientation so as to integrate specifically preoccupations bound to extreme losses.

Keywords : operational Risk, MCMC, mixture distribution

Avant-propos

A la fin de la dernière décennie , le secteur financier a connu de profondes mutations causées entre autres par : l'expansion de la technologie, l'automatisation des opérations et leurs réalisations à temps réel, la croissance du commerce électronique, la multiplicité et la complexité des produits, la globalisation financière et le développement de nouveaux marchés, la taille de plus en plus grande issue des fusions-acquisitions. Ces mutations ont entraîné de nouvelles catégories de pertes assez diversifiées, récurrentes pour certaines, très sévères pour d'autres et le plus souvent relativement difficiles à anticiper et à élucider les véritables facteurs. A titre d'exemple le groupe Risk Management du Comité de Bâle suite à une collecte auprès de 89 Banques révèle pour l'exercice 2001 plus de 47000 évènements de pertes avoisinant 8 milliards \$US.

Soucieux de stabiliser davantage et de permettre une saine croissance de ce secteur, le Comité de Bâle pour la Surveillance bancaire (BCBS) au terme de nombreux travaux, consultations et études a abouti à ce qui constitue dorénavant le troisième facteur de risque auxquels les banques sont exposées dans leurs activités quotidiennes et ce après le risque de crédit et de marché. Par la même occasion, tout un diagnostic approfondi du secteur a été fait et le dispositif d'adéquation des fonds propres substantiellement reformulé, un nouveau cadre réglementant la quasi-totalité des activités a vu le jour sous l'appellation de « le Nouvel accord de Bâle ou Bâle II [2003] ».

Le risque vu comme concept étant déjà relativement ambigu, la tâche ne sera que davantage ardue du fait de la complexité , la subjectivité, le caractère dynamique et parfois assez vague des aspects inhérents au risque opérationnel, néanmoins une avancée considérable a été faite par le BCBS. Ainsi ce risque couvre les pertes potentielles dues aux imperfections dans les procédures, les systèmes internes, actions du personnel ou résultant d'évènements extérieurs . Par ailleurs une proposition d'identification suffisamment élaborée de manière à le distinguer des autres risques a été adoptée ainsi que plusieurs recommandations au sujet de la

mesure et du management. L'Approche vers laquelle toutes les institutions devraient tendre à long terme est dénommée « Approche de Mesures Avancées » et s'articule autour de deux principales orientations.

- La première orientation tournée vers une vision du risque au niveau de l'industrie (Top-Down) est basée sur les modèles construits à partir de la volatilité des revenus et d'évaluation des actifs ;

- La deuxième, axée sur une vision du risque spécifique à chaque structure (Bottom-up) est basée sur les modèles paramétriques (LDA pour Loss Distribution Approach) aboutissant à une courbe de distribution des pertes et le calcul de la VAR, puis sur les modèles factoriels ou causals (SCD pour Scorecard Approach) mettant en évidence les principaux facteurs de risque, les interactions et leurs impacts respectifs sur le niveau et la maîtrise du risque.

Jusqu'alors très peu exploré dans les milieux académique et professionnel, l'étape ci-dessus mentionnée sera suivie par un véritable engouement des chercheurs, praticiens et institutions. Parmi les nombreuses réalisations, nous pouvons citer celles de :

Cruz [1998,2000,2002], auteur de l'un des premiers articles académiques en 1998 dans le domaine, il passe en revue l'ensemble des méthodes économétriques et probabilistes susceptibles d'être adéquates pour la modélisation, la mesure et la gestion du risque opérationnel.

Alexander [2000,2003] avec l'association d'un groupe d'auteur étudie différents aspects du risque opérationnel et insiste entre autre : sur la capacité des réseaux bayésiens à servir de cadre adéquate pour l'analyse rigoureuse de ce dernier, puis sur les principaux facteurs (type d'activité, catégories de pertes, environnement de contrôle et environnement d'affaire) à prendre en compte dans une perspective globale d'analyse et de gestion dudit risque.

A ce jour une avancée certaine a été faite au niveau théorique, mais il n'en demeure pas moins que de nombreuses zones d'ombres subsistent et plusieurs axes de recherches sont par ailleurs d'actualité. Entre autres insuffisances, nous pouvons mentionner :

- Le problème de fusion et d'utilisation simultanée des données interne et externe

provenant de multiples sources hétérogènes et indispensables pour des études plus approfondies et robustes;

- Le problème crucial de spécification et de modélisation des dépendances entre les fréquences des pertes, entre les sévérités des pertes, et entre fréquences et sévérités de pertes;
- Le prolongement au delà de la VAR dans l'analyse et le management du risque au moyen d'un modèle global type causal permettant une meilleure identification et validation des différents facteurs de risques, une mise en évidence des interactions, de leurs évolutions et de leurs différents impacts sur le risque opérationnel;
- La validation des différents modèles et éventuellement l'utilisation simultanée de ces modèles pour des meilleures prévisions à travers une agrégation bayésienne(BMA).
- L'éventuelle élaboration d'un environnement de gestion capable d'intégrer à la fois les risques opérationnel, de crédit et de marché.

Comme mentionné plus haut, une des priorités pour toute l'industrie aujourd'hui est la mise en œuvre dans un cadre relativement simplifié d'un outil opérationnel à moindre coût permettant d'identifier rigoureusement le profil de risque et donc d'assurer une meilleure gestion et une allocation optimale du capital. L'unanimité jusqu'à présent se fait autour de l'approche LDA, cependant la mise en œuvre connaît de nombreuses difficultés dont : la faible taille des bases de données pour plusieurs types de risques ; les méthodes et critères d'enregistrements de données, seules les pertes à partir d'un certains seuils sont enregistrées ; les limites de l'approche traditionnelle au regard des problèmes précédents et des résultats empiriques insatisfaisants ; l'utilisation simultanée des données externes et internes parfois issues des sources assez hétérogènes et de tailles assez variées , etc.

Cette approche purement statistique est largement utilisée dans le domaine des assurances et de façon pratique, elle utilise les données de pertes antérieures en s'appuyant sur le travail d'identification et de standardisation des types de risques par unité d'affaire(Bâle II) pour : modéliser la distribution de fréquence de perte, la distribution de sévérité de perte, puis combiner pour déterminer la valeur à risque par ligne d'affaire et au niveau global.

Au delà de l'approche traditionnelle consistant à faire une hypothèse sur les distributions de fréquence et de sévérité, estimer les paramètres par maximum de vraisemblance et procéder à différents tests sur les distributions sélectionnées;

Carol Alexander(2003) propose une estimation bayésienne avec distribution subjective construite à partir des données externes et une objective sur les données internes.

Baud, Frachot et Roncalli (2003) suggèrent une fusion de toutes les données et une maximisation de la vraisemblance tronquée conditionnelle à autant de seuil que de source de donnée (le seuil étant également un paramètre de maximisation).

Si la distribution de la fréquence peut être appréhendée sans grande difficulté par une loi de poisson ou une loi binomiale négative, celle de la sévérité constitue certainement la principale source d'imprécision . Dans la logique d'intégration de toutes les données, nous nous focaliserons davantage sur la modélisation de cette dernière au détriment des autres aspects donc entre autres : l'agrégation et le calcul des pertes totales pour une institution , la prise en compte éventuelle des différentes formes de dépendance. En effet et bien que ce soit assez discutable nous travaillons sous les hypothèses (Bâle II) : pour une cellule donnée,

- Les montants de pertes sont indépendamment distribués ;
- La fréquence des pertes est également indépendante de la sévérité.

Les modèles avec mélange de distribution sont couramment utilisés avec succès dans plusieurs domaines pour : modéliser les données issues de différentes sources hétérogènes et de proportions différentes (Basford et Maclachlan, 88; Green et Richardson, 97), améliorer la précision et mieux prendre en compte les queues de distributions (Chen et. Fournier, 99). Il sera ainsi question pour nous d'exploiter cette méthode à travers un modèle hiérarchique intégrant toutes les données de pertes de l'industrie financière pour estimer la distribution des pertes et procéder par classification des observations pour déduire la distribution de perte propre à chaque institution.

Le travail est structuré en trois parties :

La première fait une synthèse de la définition et de l'identification du risque opérationnel en s'appuyant sur des travaux de Carol Alexander et les accords de Bâle, ensuite elle permet de

décrire l'approche LDA dans sa formulation mathématique, les différentes approches de mise en œuvre et leurs limites éventuelles dans le cas de la mesure du risque opérationnel.

Dans la deuxième partie, nous justifions le choix du modèle de mélange de distribution pour appréhender la sévérité de perte, ce modèle est ensuite décrit et la spécification complète est faite avec la famille de distribution Lognormal et Gamma. Par ailleurs nous présentons brièvement l'algorithme MCMC avant d'aborder son utilisation pour le calibrage et l'estimation du modèle.

La troisième partie est essentiellement destinée à la présentation et l'analyse des résultats. Nous abordons sur des données de pertes simulées les aspects liés à la convergence, l'estimation des paramètres a posteriori et de la densité des données, la sensibilité des résultats par rapports aux choix a priori et enfin la déduction de la distribution de perte interne par classification des observations et le calcul de la Valeur à Risque.

I- Risque opérationnel et Mesure par la distribution de perte

Bien que le risque opérationnel ne soit pas une expression et un concept nouveau dans le monde de l'industrie et plus précisément celui de l'industrie financière, c'est en 2003 au terme de plusieurs années de travaux et concertations que la banque de règlements internationaux(BRI) parvient à une identification claire et suffisamment rigoureuse pour servir de point de départ à un travail de mesure et de gestion de ce type de risque. Nous insisterons sur son identification, sa formulation mathématique et les différentes étapes de la mise en œuvre.

I.1- Identification du risque opérationnel

La nouvelle conception du ratio de solvabilité des banques par la BRI repose sur trois piliers clés : amélioration du calcul des risques et leur couverture par les fonds propres, affinage du pilier1 par un renforcement du processus de surveillance prudentielle entre banque et régulateur national, enfin amélioration de la transparence financière pour permettre une plus grande discipline du marché. Elle permet ainsi de disposer dorénavant d'une définition du risque opérationnel différente de celle du risque de crédit et de marché. Si la BRI dans son devoir d'assurer une stabilité et une saine croissance du secteur financier exige à travers le nouveau dispositif d'adéquation des fonds propres un cadre formel d'analyse, de mesure et gestion optimal du risque, il n'en demeure pas moins que le principal bénéficiaire d'un tel outil reste les banques car il permet entre autres de mieux appréhender le profil de risque, d'assurer un meilleur suivi et donc une meilleure allocation du capital entraînant par ailleurs de meilleurs résultats financiers.

Le risque opérationnel se définit ainsi (Bâle II) comme « *le risque de pertes résultant de carences ou de défaillances attribuables à des procédures, personnels et systèmes internes ou à des évènements extérieurs* ».

Antony Peccia(2003) dans une perspective d'analyse et de gestion globale identifie quatre principaux facteurs susceptibles de faciliter cette tâche :

▪ **Le type d'activité ou ligne d'affaire**

La nomenclature standard est établie par la BRI et comprend huit rubriques:

- ✓ le Financement des Entreprises ;
- ✓ les Services bancaires de détails;
- ✓ les Paiements et règlements;
- ✓ les Négociations et ventes;
- ✓ le Courtage de détail;
- ✓ les Services bancaires commerciaux;
- ✓ les fonctions d'agent;
- ✓ la Gestion d'actifs .

▪ **Les différents événements générateurs de pertes (Catégories de pertes)**

La nomenclature adoptée est également définie par la BRI, elles sont au nombre de sept :

- ✓ les Fraudes Externes ;
- ✓ les Fraudes internes ;
- ✓ les dommages aux actifs corporels ;
- ✓ les clients, produits et pratiques commerciales ;
- ✓ les pratiques en matière d'emploi et de sécurité sur le lieu du travail ;
- ✓ les dysfonctionnements de l'activité et des systèmes ;
- ✓ l'exécution, la livraison et la gestion des processus.

▪ **L'environnement de contrôle**

La description proposée est une adaptation des travaux de COSO et COBIT¹ dont les principaux sont:

- ✓ Les politiques et les directives ;
- ✓ La qualité du Personnel ;
- ✓ Les limites dans la mesure du risque ;
- ✓ Les procédures ;
- ✓ Les sources d'informations indépendantes ;

¹ COSO : Committee of Sponsoring Organizations of the Treadway Commission.
COBIT : Control Objectives for Information and related Technology .

- ✓ La communication et la surveillance.

- **L'environnement d'affaire**

Ce facteur n'est pas suffisamment formalisé et n'est par ailleurs pas connu sous une forme standard comme les précédents. Néanmoins, ce sont :

- ✓ La complexité du produit ;
- ✓ La vitesse de changement ;
- ✓ La complexité du système ;
- ✓ La concentration du système ;
- ✓ La géographie ;
- ✓ La normalisation et la complexité du système judiciaire.

Le BCBS fournit ainsi un cadre standard permettant d'identifier et de mesurer le risque opérationnel fondé essentiellement sur les deux premiers facteurs (Ligne d'affaire/Catégorie de risque) ci-dessus . Trois approches sont par ailleurs proposées pour la mesure et les exigences en fonds propres sont décroissantes avec la complexité et la sensibilité de l'approche retenue.

- **L'Approche de l'Indicateur de Base**

Exigence en fonds propres = produit net bancaire moyen sur les trois dernières années *15% .
C'est la plus simple, la moins rigoureuse du point de vue analyse et suivi du risque et bien évidemment la plus exigeante en fonds propres.

- **L'Approche Standard**

Exigence en fonds propre = $\sum_{i=1}^8 PNB_i * \beta_i$; $i=1, \dots, 8$ et désigne la ligne d'affaire i , voir tableau

ci-dessous. Elle est un peu plus approfondie que la précédente , le type d'activité est pris en compte et les pourcentages sont fixés en fonction du degré de risque propre à ce dernier. Cependant elle ne débouche pas toujours sur un véritable cadre d'analyse et de suivi du risque.

Lignes d'affaires	Pourcentage fixe (β_i)
Financement des entreprises	18%
Négociation et vente	18%
Banque de détail	12%
Banque commercial	15%
Paiement et règlement	18%
Fonctions d'agent	15%
Gestion d'actifs	12%
Courtage de détail	12%

BCSC, 2003.

▪ **L'Approche des Mesures Avancées (AMA)**

C'est la plus élaborée et la moins exigeante en fonds propres. Elle constitue par ailleurs un véritable outil de diagnostic, de mesure et de suivi du risque. Les banques doivent passer de l'approche la plus simple à la plus complexe ; elles n'obtiendront l'autorisation d'utiliser l'AMA que si un certain nombre de critères nécessaires pour une application optimale sont satisfaits. Ces critères se regroupent en deux grandes catégories :

Les critères qualitatifs :

- ✓ Une fonction de gestion de risque opérationnel indépendante ;
- ✓ L'implication des dirigeants;
- ✓ L'intégration dans la gestion des risques au quotidien ;
- ✓ La notification régulière des expositions et des pertes ;
- ✓ Un programme régulier d'analyse de scénarios ;
- ✓ La documentation sur les contrôles et procédures internes ;
- ✓ Les audits internes et/ou externes.

Les critères quantitatifs :

- ✓ La prise en compte des pertes sévères mais rares et le calibrage à partir des pertes attendues et inattendues.
- ✓ Les systèmes et bases de données cohérents avec la définition du risque opérationnel;
- ✓ Un système d'information approprié;
- ✓ Les procédures en cas de changement de taille;

- ✓ Les procédures pour l'usage des données externes ;
- ✓ Une revue périodique des méthodologies et paramètres ;
- ✓ L'historique des données sur cinq ans ;
- ✓ La reconnaissance possible des corrélations, assurances et ajustements qualitatifs.

L'AMA peut être mise en œuvre de différentes façons en utilisant une ou plusieurs des méthodologies suivantes : les modèles paramétriques, les modèles causals, les modèles basés sur la théorie des comportements, sur la volatilité du revenu, ou sur l'évaluation des actifs. Cette approche bien évidemment est générale et ne permet pas de mener une analyse spécifique à chaque institution de manière à identifier clairement un profil de risque et concevoir un plan de management adapter à cette dernière. C'est ainsi que nous retenons dans le cadre de ce travail un modèle paramétrique et plus précisément l'approche par la distribution des pertes, bien qu'elle soit assez exigeante en pratique, c'est une approche relativement simple, assez sensible au niveau de risque, intelligible, utilisée avec satisfaction dans le domaine de l'assurance depuis plusieurs années et bien évidemment « gourmande » en données de pertes .

I.2- Description de l'approche par la distribution des pertes

La LDA est une méthode purement statistique permettant d'estimer la distribution de perte et de déduire en fonction du niveau de confiance fixé la valeur à risque. Cette distribution est essentiellement construite en s'appuyant sur la composée de la distribution de la fréquence et celle de la sévérité des pertes.

I.2.1- Calcul de la Distribution de perte pour le risque opérationnel

Dans le cas particulier du risque opérationnel (BRI ; 2001,2003), les calculs sont d'abord effectués pour chaque cellule (constituée d'une ligne d'affaire et d'un type de risque donnée) . Ainsi les distributions de fréquence et de sévérité sont estimées à partir des pertes antérieures puis utilisées pour prédire la distribution de perte de la prochaine année. La valeur à risque de chaque cellule de niveau de confiance 99,9% en découle et celle de l'institution par sommation sur l'ensemble des cellules.

Soient les variables aléatoires suivantes :

$L_{i,j}$: la perte totale de la cellule (i, j), c'est à dire de la ligne d'affaire i et du type de risque j sur la période t à t+1 de fonction de répartition $F_{i,j}$;

$X_{i,j}(k)$: le montant de la k^{ème} perte de la cellule (i, j) et de fonction de répartition $H_{i,j}$;

$N_{i,j}$: le nombre de perte pour de la cellule (i, j) sur la période t à t+1 et de probabilité associée $p_{i,j}$;

$$L_{i,j} = \sum_{k=1}^{N_{i,j}} X_{i,j}(k) \text{ et}$$

$$F_{i,j}(x) = \sum_{n=1}^{\infty} p_{i,j}(n) H^n_{i,j}(x); \quad (1)$$

H^n désigne la convolution, ou encore $G(X_1 + X_2 + \dots + X_n)$.

G étant la fonction de répartition de la somme de N variables aléatoires définie de la manière suivante :

Soient $Z = X+Y$ et K la loi jointe du couple (X,Y) ;

$$G(z) = \int_x K(x, z-x) dx = \int_y K(z-y, y) dy$$

Sous l'hypothèse d'indépendance des pertes, la perte totale au niveau de chaque institution est relativement simple et s'obtient par sommation sur l'ensemble des cellules:

$$L = \sum_{i=1}^I \sum_{j=1}^J L_{i,j} \quad \text{et la fonction de répartition associée est}$$

$$F(x) = F_{i,j}^{I,J}(x) \quad \text{qui est elle-même la convolution telle que définie plus haut.}$$

La difficulté à ce niveau réside dans le calcul l'expression (1), Frachot, Georges et Roncali (2001) mentionnent qu'en générale il est impossible de calculer analytiquement. Trois approches d'approximation sont régulièrement utilisées : la méthode de Monte Carlo par simulation successive suivant les lois des deux distributions, la méthode récursive pour les convolutions d'ordre élevé (Panjer, 1981) et celle basée sur les propriétés de la fonction caractéristique (Heckman et Meyers, 1983). Nous reviendrons sur la méthode de Monte Carlo qui sera par ailleurs utilisée dans le cadre de notre étude.

Une fois l'approximation faite, la VAR de chaque cellule avec niveau de confiance α est :

$$Var(i, j, \alpha) = \begin{cases} F^{-1}_{i,j}(\alpha) \\ \text{ou} \\ \inf\{x | F_{i,j}(x) \geq \alpha\} \end{cases}$$

et la Var globale est :

$$Var(\alpha) = \begin{cases} F^{-1}(\alpha) \\ \text{ou} \\ \inf\{x | F(x) \geq \alpha\} \end{cases}$$

Le BRI propose une simple sommation des Var, $Var(\alpha) = \sum_{i=1}^I \sum_{j=1}^J Var(i, j, \alpha)$.

Frachot, Georges et Roncali (2001) démontrent par ailleurs que cette sommation correspond au cas où les risques sont parfaitement corrélés et positivement.

La LDA s'appuie essentiellement sur deux principales distributions, et exploite des algorithmes numériques pour l'approximation de la plupart des expressions ci-dessus.

C'est par ailleurs une approche fondée sur un certains nombres d'hypothèses sous-jacentes.

Carol Alexander(2003) cite quatre principaux :

- Les pertes antérieures constituent un bon proxy pour les pertes futurs;
- L'historique des pertes est suffisamment longue (trois à cinq ans pour le BRI);
- Toutes les cellules contiennent assez d'événements de grande ampleur pour assurer une modélisation robuste de la distribution de la sévérité;
- Les données externes sont nécessaires pour améliorer la robustesse des estimations.

Sous ces hypothèses, l'approche LDA présente bien évidemment les inconvénients d'exigence d'effort considérable pour les opérations de collecte de données et d'acquisition des données externes ; puis d'insuffisance dans l'anticipation des nouvelles formes de pertes car basée essentiellement sur les pertes antérieures. Néanmoins elle a de nombreux avantages. Entre autres, les fondements mathématiques rigoureux et similaires à ceux utilisés pour la modélisation du risque de crédit et de marché, la capacité d'intégrer les pertes les plus récentes, les ajustements qualitatifs et les modèles causals , sans oublié la possibilité d'effectuer des analyses de sensibilité et de déterminer les coûts/bénéfices relatifs à différents cas envisageables.

I.2.2- Mise en œuvre de la LDA dans le cas du risque opérationnel

La mise en œuvre proprement dite est précédée d'une étape cruciale de collecte et de validation des données. Sous les hypothèses que les données externes sont nécessaires pour une estimation robuste des distributions, une des difficultés à ce stade est de trouver une méthodologie adaptée et robuste permettant la fusion des données provenant des institutions ayant parfois des tailles et des procédures de gestion assez variées. Nous pouvons citer : l'utilisation des facteurs d'échelles qui est un coefficient calculé en fonction des caractéristiques propres à chaque institution; l'utilisation des données externes comme échantillon subjective à priori et des données internes comme échantillon objective; la fusion en considérant que les données sont enregistrées avec autant de seuils que d'institutions et donc l'inférence sur la base d'un échantillon tronquée; et enfin celle que nous utiliserons dans ce travail, une fusion de toutes les données dans une logique de mélange de distribution et une estimation de la distribution de perte de chaque institution dans une seconde étape par classification. Ainsi une fois l'échantillon valide la suite peut être subdivisée en quatre grandes étapes :

- **Modélisation de la distribution de la sévérité des pertes**

C'est l'étape la plus importante et la plus sensible, la qualité et la robustesse de tout le processus de mesure du risque et donc de l'allocation optimal du capital dépend en grande partie de cette étape. Elle est par ailleurs la plus difficile et est abordée de différentes façons.

La méthode traditionnelle et la plus simple se limite aux données internes, une distribution théorique (lognormal, Gamma, weibul) est sélectionnée, l'estimation des paramètres se fait par la technique du maximum de vraisemblance et les tests statistiques (Khi-deux, Kolmogorov-Smirnov) permettent de valider la distribution. Elle est assez simple et de nombreux logiciels permettent la mise en œuvre rapide, cependant les résultats empiriques ne sont pas satisfaisants(Gourde et Dahen, 2003), et de nombreuses cellules possèdent très peu de données et il est difficile de capter les pertes extrêmes avec une telle approche.

Une méthode alternative (Frachot, Georges et Roncali) est établie sur la base que seules les pertes au dessus d'un certain seuil sont enregistrés , les données de toutes les sources sont fusionnées tout en considérant qu'elles sont issues de la même famille de distribution mais tronquées à de seuils différents. la technique de maximum de vraisemblance est par la suite utilisée en ajoutant les différents seuils comme des paramètres de maximisation. L'une des principales difficultés est le choix du nombre de seuil, ceci d'autant plus que les bases de données externes contiennent des pertes issues de plusieurs sources. De plus les pertes externes peuvent être très élevées comparativement aux pertes internes et biaiser considérablement les résultats.

Une troisième alternative(Alexander) est fondée sur l'approche bayésienne, la distribution a priori des paramètres est obtenue grâce aux données externes, les données internes servent d'échantillon objective pour le calcul de la vraisemblance, permettant ainsi d'estimer les distributions de paramètres a posteriori . Sous la fonction de perte quadratique, les moyennes empiriques des paramètres sur les échantillons a posteriori obtenus à partir d'un certain rang sont des estimateurs admissibles. La principale difficulté de cette approche se pose lorsqu'un véritable conflit apparaît entre les données internes et externes disponibles, entraînant ainsi une différence considérable dans l'ordre de grandeurs des paramètres a priori et de la vraisemblance.

Nous utiliserons dans le cadre de ce travail une approche bayésienne avec mélange de distribution régulièrement utilisée en statistique avec de nombreuses applications dans différents domaines et environnements semblables à celui du risque opérationnel. Les données de toutes les sources sont fusionnées et la distribution est un mélange de plusieurs distributions de pondérations différentes, issues de la même famille de distribution mais de paramètres différents. Les paramètres sont estimés par les MCMC . Sous la fonction de perte quadratique, nous obtenons également des estimateurs admissibles. L'identification de la distribution de perte de chaque institution se fait par classification (en maximisant la probabilité d'appartenance à un groupe) de chaque observation dans les différents groupes estimés. Nous reviendrons sur la justification plus loin.

- **Modélisation de la distribution de fréquence des pertes**

L'approche traditionnelle décrite plus haut est suffisante pour estimer de façon robuste la distribution de probabilité de la fréquence des pertes, les résultats empiriques sont satisfaisants et les distributions théoriques sont le plus souvent la loi de poisson ou la Binomial Négative.

Une autre alternative est proposée par les mêmes auteurs que précédemment dans la logique d'échantillon tronquée, cette fois-ci, les résultats de la distribution de sévérité sont utilisés car la vraisemblance est conditionnelle à un montant de perte supérieur au seuil estimé.

- **Agrégation des distributions(Sévérité et Fréquence) par simulation Monte Carlo**

Elle permet d'obtenir une approximation de la distribution des pertes totales annuelles pour chaque cellule. Bien évidemment nous ignorons la prise en compte des corrélations éventuelles . Au préalable il faut tenir compte de la valeur maximale de perte à prendre en considération et voir son impact par analyse de la sensibilité. L'une des approches peut consister à : générer aléatoirement la fréquence de perte journalière ou annuelle , puis générer aléatoirement le montant de chaque perte à partir des distributions obtenues précédemment. Le montant total des pertes quotidiennes et annuelles découlent en sommant ces dernières valeurs.

- **Calcul de la VAR, Analyse de la sensibilité et validation des résultats**

La Var est obtenue pour chaque cellule sur un horizon d'un an et le niveau de confiance est 99.9%. dans les étapes précédentes, de nombreux paramètres sont initialisés et parfois des distributions sont sélectionnées sans méthode systématique. Ceci laisse une grande marge de manœuvre et il est crucial à cette étape de faire varier ces différents paramètres en analysant l'impact sur les estimateurs a posteriori et sur les résultats finaux. Ces analyses associées aux connaissances des experts dans le domaine, aux réalités de l'institution, de toute l'industrie et même aux évidences intuitives doivent permettre de confirmer et de valider les modèles.

Conclusion

La LDA est une approche essentiellement statistique possédant un fondement mathématique assez rigoureux. Deux des principales contraintes sont certainement l'utilisation d'un échantillon de données suffisamment long et la difficulté d'anticiper d'éventuelles pertes extrêmes. C'est dans cette optique que nous envisageons une approche permettant de combiner les données internes et externes dans un modèle prenant en compte l'hétérogénéité des différentes sources. Malgré ses relatives faiblesses parce que construite sous l'hypothèse que les pertes antérieures permettent une meilleure approximation des pertes futures et l'impossibilité d'identifier et d'agir directement sur les facteurs de risque, elle demeure une approche rigoureuse, flexible, adaptée pour la mesure du risque opérationnel dans les institutions bancaires et peut préalablement servir d'outil rigoureux pour la mesure avant d'être associé plus tard dans un cadre plus élargi de modélisation et de gestion globale intégrant les modèles causals.

II- Modélisation avec mélange de distribution et Estimation par MCMC

Les avancées au cours des années 90 en informatique ont permis un important progrès dans de nombreux domaines donc celui de l'inférence statistique. Il est de plus en plus possible d'appréhender des phénomènes par de modèles nécessitant davantage de calculs et de manipulation assez complexes. Une des principales illustrations est le développement d'un outil tel les MCMC et son importance pour l'inférence sur les modèles de mélange, les modèles à données manquantes et les modèles bayésiens pour ne citer que ces quelques uns qui par ailleurs font l'objet de notre étude. Cette partie nous permettra de mettre en évidence l'utilisation de cette méthode pour calibrer un modèle de mélange de distribution appartenant à la même famille de loi et permettant de construire la distribution de sévérité des pertes .

II.1- Justification et Description du modèle

Avant d'aborder le développement de cette partie, nous présentons ci-dessus quelques définitions nécessaires pour la suite du travail. Pour des lectures plus approfondies, voir (Robert 92, 96, 99).

Définition 1 : Loi a posteriori

Le théorème de Bayes permet d'établir et de définir la loi a posteriori comme la loi du paramètre étant donné les observations.

$$\pi(\mu|Y) = \frac{\pi(\mu) * f(Y|\mu)}{m(Y)}$$

$m(Y)$: loi marginale des observations.

$f(Y|\mu)$: vraisemblance des observations.

$\pi(\mu)$: loi a priori des observations.

Définition 2 : Fonction de perte

Une fonction de perte est une fonction quelconque mesurant la perte ou l'erreur lorsqu'on utilise la décision d pour faire l'inférence sur μ .

$$\begin{aligned} L : \theta \times D &\rightarrow [0 \ \infty] \\ (\mu, d) &\rightarrow L(\mu, d) \end{aligned}$$

Définition 3 : perte espérée a posteriori

La perte espérée a posteriori (ou risque a posteriori) de l'action a notée $\delta(\pi, a|Y)$ est donnée par $\delta(\pi, a|Y) = E[L(\mu, a)|Y]$.

Définition 4 : Estimateur inadmissible / admissible

Un estimateur δ_0 est inadmissible s'il existe un autre estimateur δ_1 qui R-domine δ_0 , c'est à dire $R(\mu, \delta_0) \leq R(\mu, \delta_1)$, $\forall \mu$ avec inégalité stricte pour au moins une valeur de μ . S'il n'existe pas de tel δ_1 alors δ_0 est admissible.

Définition 5 : Algorithme MCMC

Un algorithme Monte Carlo par les Chaînes de Markov représente toute méthode ou procédure permettant de générer une chaîne de Markov ($X^{(t)}$) ergodique ayant pour loi stationnaire la distribution d'intérêt.

II.1.1- Justification de l'approche

Soucieux de concilier robustesse du modèle et contraintes d'ordre pratique énumérées au premier chapitre, l'approche par un mélange de distribution possède de nombreux avantages :

- C' est un cadre naturel (Green et Richardson,97) de modélisation pour un échantillon de données hétérogènes constitué des observations appartenant à une ou plusieurs populations dans des proportions variables. De nombreux cas dans différents domaines sont illustrés par Titterington, Smith and Markov (1985), McLachan et Basford(1988).

- Au regard du tableau ci-dessous résumant la marge d'erreur pour une taille de l'échantillon et un niveau de confiance donnés , une estimation bayésienne en permettant d'intégrer toutes les données de l'industrie financière, puis de prendre en compte l'information à priori contribue significativement à solutionner le problème crucial d'insuffisance des observations nécessaires pour des estimations davantage robustes.

Tableau 2.1 : Marge d'erreur en fonction du niveau de confiance et de la taille de l'échantillon (Loi Gamma avec deux paramètres).

Puissance T	α		β	
	95%	90%	95%	90%
50	44%	36%	50%	40%
2000	5.3%	4.6%	6.2%	5.3%
6000	3.2%	2.5%	3.6%	2.9%

Loi Gamma avec deux paramètres (en exploitant la loi asymptotique de l'estimateur du MV)

- Il est insensible au problème de censure ou troncature des observations régulièrement pratiquées dans l'enregistrement des pertes par les institutions financières, bref, robert(1996) et plus précisément Berger et Wolpert (1988) démontrent que la censure a généralement un effet nul sur l'inférence bayésienne.

- Le modèle de mélange permet de disposer d'un modèle global ayant une densité continue et prenant à la fois en compte les événements fréquents et ceux assez rares.

- Dans la même idée que la précédente (Chen et. Fournier, 99) démontrent dans une étude de cas réel que le mélange de distribution peut sensiblement améliorer la précision et permettre une meilleure prise en compte des queues de distribution.

- Une fois l'estimation faite et les différents groupes clairement identifiés, il est possible de procéder à une classification des observations et ainsi déterminer la distribution de probabilité de la sévérité des pertes propre à chaque institution.

- De plus la méthode bayésienne est assez flexible et en permettant une exploitation simultanée des modèles statistiques(LDA) et causals(Scorecard) contribue à une meilleure identification du profil de risque et fournit assez de visibilité nécessaire pour une gestion optimale.

II.1.2- Description du modèle

Soient $(X_1, X_2, X_3, \dots, X_n)$, N observations d'une variable aléatoire X . Nous supposons que les observations sont indépendantes, identiquement distribuées et issues de P populations (P inconnu). Le modèle postulé est :

$$\pi(X_i|W, \theta) = \sum_{j=1}^P w_j f(X_i|w_j, \theta_j) .$$

$i = 1, \dots, n$ désigne la taille de l'échantillon ; $j = 1, \dots, P$, ce nombre de groupe au terme de l'estimation n'est pas forcément égal au nombre de source des observations.

$W = (w_1, w_2, \dots, w_P)$ représente le vecteur des poids effectifs de chaque sous population dans l'échantillon.

$$\sum_{j=1}^P w_j = 1; \quad w_j \in [0,1].$$

$f(X_i|w_j, \theta_j)$ est la fonction de densité de probabilité avec le vecteur de paramètres θ_j de la sous population j .

Les principales difficultés pour ce type de modèle résident dans le problème d'identification et d'interprétation découlant de l'invariance de la vraisemblance suite à une permutation des labels de groupes et surtout dans la manipulation et le calcul des densités jointes et marginales. Bref le calcul de la constante de normalisation et de l'espérance a posteriori des paramètres de l'expression ci-dessous est impossible.

$$\pi(W, \theta|X) \propto \prod_{i=1}^n \left\{ \sum_{j=1}^P w_j f(X_i|w_j, \theta_j) \right\} \pi(W, \theta) .$$

Une des approches est étudiée par les auteurs suivants : Verdinelli et Wasserman (1991), Lavine et West(1992), Escobar et West(1995), Diebot et Robert(1990g,1994). Elle consiste à construire un second échantillon constitué d'une variable latente L , désignant le groupe d'appartenance de chaque observation .

$l = (l_1, l_2, l_3, \dots, l_n)$ et $l_i = j$ signifie l'observation i appartient au groupe j avec une distribution de probabilité à déterminer.

$$X_i|l_i = j \rightarrow f(X_i|\theta_j) \text{ et l'expression précédente devient}$$

$$\pi(W, \theta | X, l) \propto \prod_{i=1}^n \{ \pi(l_i = j) f(X_i | \theta_j) \} .$$

Cette transformation facilite l'exploitation du modèle et permet une meilleure utilisation des MCMC (ou de l'algorithme EM) sur le modèle hiérarchique ci-dessous qui en découle.

$$\pi(\eta, P, W, \theta, l, X) = \pi(\eta) \cdot \pi(P | \eta) \cdot \pi(W | P, \eta) \cdot \pi(\theta | P, \eta, W) \cdot \pi(l | P, \eta, W, \theta) \cdot \pi(X | P, \eta, W, \theta, l) .$$

une décomposition est nécessaire au niveau du paramètre θ , ceci pour des soucis de robustesse et de flexibilité. Zellner(1976), Angers(1987), Anger et MacGibbon(1990).

$$\pi(\theta) = \int \pi_1(\theta | \lambda) \cdot \pi_2(\lambda) \cdot d\lambda ;$$

$\lambda \rightarrow \pi_2(\lambda)$ une loi non informative.

La suite de la spécification est fortement influencée par l'utilisation de l'algorithme MCMC pour l'estimation et nécessite que les aspects liés au choix de la distribution des pertes, de la distribution a priori et a posteriori des paramètres soient abordés.

II.2- Sélection et calcul des distributions a priori et a posteriori

De prime abord toutes les distributions continues et positives peuvent être utilisées pour la modélisation de la sévérité des pertes. Cependant les travaux de nombreux auteurs dont : Carol Alexander (2001,2003); Cruz (2002); Dimon et Hela (2003, application au cas particulier de la Banque Nationale du Canada); de Frachot, Georges et Roncalli (2001, du groupe de recherche opérationnelle, Crédit Lyonnais, France) permettent de faire une première sélection. Trois distributions sont retenues pour l'étude de la sévérité, la loi Gamma, la loi Lognormale et la loi Weibul. L'accent n'étant pas mis sur la distribution de la fréquence, nous nous limiterons à la loi de Poisson car elle donne des résultats empiriques satisfaisants.

II.2.1- Spécification avec la distribution Lognormale

Le modèle avec la distribution lognormale représente un cas classique où l'application de la MCMC donne des résultats assez robuste. La description qui découle du modèle (II.1.2) est la suivante :

$$X_i | l_i = j \rightarrow LN(\mu_j, \sigma_j) .$$

Pour $Y = \log(X)$ nous obtenons $Y_i | l_i = j \rightarrow N(\mu_j, \sigma_j)$;

$$\pi(Y|P, W, \lambda, \sigma) = \prod_{i=1}^n \left\{ \sum_{j=1}^P W_j N(Y_i | \mu_j, \sigma_j) \right\} \text{ et la vraisemblance s'écrit :}$$

$$L(P, \eta, l, \mu, \sigma) = \prod_{j=1}^P \left[\prod_{i=1}^{n_j} N(Y_i | \mu_j, \sigma_j) \right]$$

$$= \prod_{j=1}^P \left[\frac{1}{\sqrt{2\pi}^{n_j} \sigma_j^{n_j}} * \exp\left(\frac{-1}{2\sigma_j^2} * \sum_{i:l_i=j}^{n_j} (Y_i - \mu_j)^2 \right) \right]$$

1- Distributions a priori des paramètres

$$\mu_j \rightarrow Std\left(a, \frac{n_0 a_1}{2b_1}, a_1\right) \quad \text{ou} \quad \mu_j \rightarrow N(a, b)$$

$$\sigma_j^2 \rightarrow IG(a_1, b_1) \quad ; \quad j = 1, 2, \dots, P$$

$$W \rightarrow \text{Dirichlet}(c, c, \dots, c) .$$

Le choix des hyper paramètres (a, a₁, b₁, c.....) se fait en fonction des observations et de l'étude de sensibilité.

2- Distributions a posteriori des paramètres

▪ Moyenne et Variance

$$\begin{aligned}\pi(\mu_j, \sigma_j^2 | Y, l, P) &= \frac{\pi(Y | \mu_j, \sigma_j^2, l, P) * \pi(\mu_j, \sigma_j^2 | l, P)}{\pi(Y | l, P)} \\ &\propto \pi(Y | \mu_j, \sigma_j^2, l, P) * \pi(\mu_j, \sigma_j^2 | l, P) \\ &= \pi(Y | \mu_j, \sigma_j^2, l, P) * \pi(\mu_j | \sigma_j^2, l, P) * \pi(\sigma_j^2 | l, P)\end{aligned}$$

Cas ou la moyenne a une loi a priori de student

$$\begin{aligned}\pi(\mu_j, \sigma_j^2 | Y, l, P) &= \prod_{j=1}^P \left[\frac{1}{\sqrt{n_j} \sqrt{2\pi} * \sigma_j^{n_j}} * \exp\left(\frac{-1}{2\sigma_j^2} * \sum_{i:l_i=j}^{n_j} (X_i - \mu_j)^2\right) \right] * \\ &\quad \frac{\Gamma\left(\frac{a+1}{2}\right) / \Gamma\left(\frac{a}{2}\right)}{\sqrt{\pi a_1}} * \left(1 + \frac{\left(\mu_j - \frac{n_0 a_1}{2b_1}\right)^2}{a a_1}\right)^{-0.5*(a+1)} * \frac{b_1^{a_1} * \exp(-b_1 / \sigma_j^2)}{\Gamma(a_1) * \sigma_j^{2(a_1+1)}}\end{aligned}$$

Après regroupement des termes identiques et quelques transformations, nous obtenons :

$$\mu_j | Y \rightarrow Std\left(\frac{n_0 a + n_j \bar{Y}_j}{n_0 + n_j}, (n_0 + n_j) 0.5(a_1 + n_j) b_{1a}, a_1 + n_j\right)$$

$$b_{1a} = b_1 + 0.5 n_j S_j^2 + 0.5 (n_0 + n_j)^{-1} n_0 n_j (a - \bar{Y}_j)^2$$

avec $n_j \bar{Y}_j = \sum_{i:l_i=j}^{n_j} Y_i$

$$S_j^2 = \sum_{i:l_i=j}^{n_j} (Y_i - \bar{Y}_j)^2$$

$$\sigma_j^2 | Y \rightarrow IG(0.5(a_1 + n_j), b_{1a}) \quad j = 1, 2, \dots, P$$

Cas ou la moyenne a une loi a priori Normale

$$\begin{aligned} \pi(\mu_j, \sigma_j^2 | Y, l, P) = \prod_{j=1}^P \left[\frac{1}{\sqrt{2\pi} * \sigma_j^{n_j}} * \exp\left(\frac{-1}{2\sigma_j^2} * \sum_{i:l_i=j}^{n_j} (X_i - \mu_j)^2\right) \right] * \\ \frac{1}{\sqrt{2\pi} * b} * \exp\left(\frac{-1}{2b^2} * \sum_{i:l_i=j}^{n_j} (\mu_j - a)^2\right) * \frac{b_1^{a_1} * \exp(-b_1 / \sigma_j^2)}{\Gamma(a_1) * \sigma_j^{2(a_1+1)}} \end{aligned}$$

Après regroupement des termes identiques et quelques transformations, nous obtenons :

$$\mu_j | Y \rightarrow N\left(\frac{\sigma_j^{-2} n_j \bar{Y}_j + ab^{-1}}{\sigma_j^{-2} n_j + b^{-1}}, \frac{1}{\sigma_j^{-2} n_j + b^{-1}}\right)$$

avec
$$n_j \bar{Y}_j = \sum_{i:l_i=j}^{n_j} Y_i$$

$$\sigma_j^2 | Y \rightarrow IG\left(a_1 + 0.5n_j, b_1 + 0.5 \sum_{i:l_i=j} (y_i - \mu_j)^2\right) \quad j = 1, 2, \dots, P$$

▪ **Pondération et Identification des groupes**

$$\pi(W | \mu_j, \sigma_j^2, P, Y) = \frac{\pi(Y | \mu_j, \sigma_j^2, P, W) * \pi(W | \mu_j, \sigma_j^2, P)}{\pi(Y | \mu_j, \sigma_j^2, P)}$$

$$\propto \pi(Y | \mu_j, \sigma_j^2, P, W) * \pi(W | \mu_j, \sigma_j^2, P)$$

$$= \prod_{i=1}^n \left[\sum_{j=1}^P W_j \frac{1}{\sqrt{2\pi} * \sigma_j} * \exp\left(\frac{-1}{2\sigma_j^2} (Y_i - \mu_j)^2\right) \right] * \frac{\Gamma(Pc)}{\Gamma(c)^P} * W_1^c * \dots * W_j^c * \dots * W_P^c$$

et nous obtenons $W|Y \rightarrow \text{Dirichlet}(n_1 + c, n_2 + c, \dots, n_p + c)$.

Par ailleurs l'observation Y_i appartient au groupe j avec la probabilité,

$$\pi_j(Y_i) = \frac{W_j f(Y_i | \mu_j, \sigma_j^2)}{\sum_{j=1}^P W_j f(Y_i | \mu_j, \sigma_j^2)} = \frac{W_j \frac{1}{\sqrt{2\pi} * \sigma_j} * \exp\left(\frac{-1}{2\sigma_j^2} * (Y_i - \mu_j)^2\right)}{\sum_{j=1}^P W_j \frac{1}{\sqrt{2\pi} * \sigma_j} * \exp\left(\frac{-1}{2\sigma_j^2} * (Y_i - \mu_j)^2\right)}$$

3- Estimation des paramètres

Nous utilisons la fonction de perte quadratique. En effet en théorie de la décision bayésienne, pour une fonction de coût $L(\theta, \delta)$ et une loi a priori des paramètres π l'estimateur bayésien $\delta^\pi(x)$ est solution de l'équation $\text{Min } E^\pi(L(\theta, \delta)|x)$.

les fonctions de pertes ou de coûts régulièrement utilisées sont les fonctions quadratique ($L(\theta, \delta) = (\theta - \delta)^2$) et absolu ($L(\theta, \delta) = |\theta - \delta|$). De nombreuses autres existent (pour une liste exhaustive et détaillée, voir Berger, 1985a). Par ailleurs dépendamment de la complexité de cette fonction de perte, l'estimateur peut être déterminé de façon explicite ou par des méthodes numériques. Dans le cas de la fonction quadratique, l'estimateur associé est la moyenne a posteriori qui est aussi établie comme étant un estimateur admissible. Ceci malgré (Berger, puis Robert, 92) le désavantage de ce type de fonction de coût qui de par sa convexité stricte pénalise fortement les grandes erreurs. Ainsi nous avons les estimateurs suivants :

$$\delta^\pi(\mu_j) = \frac{1}{T} \sum_{i=1}^T \mu_j^i ;$$

$$\delta^\pi(\sigma_j^2) = \frac{1}{T} \sum_{i=1}^T \sigma_j^{2(i)} . \quad J=1, 2, \dots, P.$$

II.2.2- Spécification avec la Distribution Gamma

Le modèle avec la loi Gamma n'est pas aussi utilisé que le précédent, l'une des principales raisons est la difficulté à identifier et générer la loi a posteriori du paramètre de forme. De nombreuses transformations sont utilisées dépendamment des observations.

Soit $X_i | l_i = j \rightarrow G(\alpha_j, \beta_j)$.

$$f(X_i) = \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)} * X_i^{\alpha_j-1} * \exp(-\beta_j X_i)$$

$$\pi(X|P, W, \lambda, \sigma) = \prod_{i=1}^n \left\{ \sum_{j=1}^P W_j G(X_i | \alpha_j, \beta_j) \right\} \text{ et la vraisemblance s'écrit :}$$

$$\begin{aligned} L(P, \eta, l, \mu, \sigma) &= \prod_{j=1}^P \left[\prod_{i=1}^{n_j} G(X_i | \alpha_j, \beta_j) \right] \\ &= \prod_{j=1}^P \left[\frac{\beta_j^{n_j \alpha_j}}{\Gamma(\alpha_j)^{n_j}} * \prod_{i: l_i=j}^{n_j} X_i^{\alpha_j-1} * \exp\left(-\sum_{i: l_i=j} \beta_j X_i\right) \right] \end{aligned}$$

1- Distributions a priori des paramètres

Nous utiliserons la loi Gamma pour le paramètre d'échelle et les lois exponentielle et Gamma pour le paramètre de forme.

$$\alpha_j \rightarrow \text{Exp}(a) \text{ ou } \alpha_j \rightarrow G(a_1, b_1) . \quad j = 1, 2, \dots, P$$

$$\beta_j \rightarrow G(b, c) .$$

$$W \rightarrow \text{Dirichlet}(d, d, \dots, d) .$$

Le choix des hyper paramètres (a, a₁, b₁, c,.....) se fait comme précédemment en fonction des observation et de l'étude de sensibilité.

2- Distributions a posteriori des paramètres

▪ Paramètres de Forme et d'Échelle

$$\pi(\alpha_j, \beta_j | X, l, P) = \frac{\pi(X | \alpha_j, \beta_j, l, P) * \pi(\alpha_j, \beta_j | l, P)}{\pi(X | l, P)}$$

$$\begin{aligned} &\propto \pi(X|\alpha_j, \beta_j, l, P) * \pi(\alpha_j, \beta_j | l, P) \\ &= \pi(X|\alpha_j, \beta_j, l, P) * \pi(\alpha_j | \beta_j, l, P) * \pi(\beta_j | l, P) \\ &= \prod_{j=1}^P \left[\frac{\beta_j^{n_j \alpha_j}}{\Gamma(\alpha_j)^{n_j}} * \prod_{i, l_i=j}^{n_j} X_i^{\alpha_j-1} * \exp\left(-\sum_{i, l_i=j}^{n_j} \beta_j X_i\right) \right] * \alpha_j \exp(-a\alpha_j) * \frac{c^b \beta_j^{b-1} \exp(-c\beta_j)}{\Gamma(b)} \end{aligned}$$

Après regroupement des termes identiques et quelques transformations, nous obtenons :

$$\beta_j | X, \alpha_j \rightarrow G\left(n_j \alpha_j + b, c + \sum_{i, l_i=j}^{n_j} X_i\right)$$

Cas ou le paramètre de forme a une loi a priori exponentiel

$$\pi(\alpha_j | X) \rightarrow \frac{\Gamma(b + n_j \alpha_j) \prod_{i, l_i=j}^{n_j} X_i^{\alpha_j-1} \exp(-\alpha_j a)}{\Gamma(\alpha_j)^{n_j} (c + \sum_{i, l_i=j}^{n_j} X_i)^{n_j \alpha_j + b}} \quad j = 1, 2, \dots, P$$

Cas ou le paramètre de forme a une loi a priori Gamma

$$\pi(\alpha_j | X) \rightarrow \frac{\Gamma(b + n_j \alpha_j) \prod_{i, l_i=j}^{n_j} X_i^{\alpha_j-1} \alpha_j^{a_1-1} b_1^{a_1} \exp(-\alpha_j b_1)}{\Gamma(a_1)^{n_j} (c + \sum_{i, l_i=j}^{n_j} X_i)^{n_j \alpha_j + b}}$$

▪ **Pondération et Identification des groupes**

$$\begin{aligned} \pi(W|\alpha_j, \beta_j, P, Y) &= \frac{\pi(X|\alpha_j, \beta_j, P, W) * \pi(W|\alpha_j, \beta_j, P)}{\pi(X|\alpha_j, \beta_j, P)} \\ &\propto \pi(X|\alpha_j, \beta_j, P, W) * \pi(W|\alpha_j, \beta_j, P) \\ &= \prod_{i=1}^n \left[\sum_{j=1}^P W_j \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)} * X_i^{\alpha_j-1} * \exp(-X_i) \right] * \frac{\Gamma(Pd)}{\Gamma(d)^P} * W_1^d * \dots * W_j^d * \dots * W_P^d \end{aligned}$$

et nous obtenons $W|X \rightarrow \text{Dirichlet}(n_1 + d, n_2 + d, \dots, n_p + d)$.

Par ailleurs l'observation Y_i appartient au groupe j avec la probabilité,

$$\pi_j(X_i) = \frac{W_j f(X_i | \alpha_j, \beta_j)}{\sum_{j=1}^p W_j f(X_i | \alpha_j, \beta_j)} = \frac{W_j \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)} * X_i^{\alpha_j-1} * \exp(-\beta_j X_i)}{\sum_{j=1}^p W_j \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)} * X_i^{\alpha_j-1} * \exp(-\beta_j X_i)}$$

II.3- Calibrage et Estimation

La description du modèle s'est faite en tenant compte de la méthode de résolution qui n'est autre que la Méthode de Monte Carlo par les chaînes de Markov précédée d'une sélection du nombre de groupe par le critère de Bayes. En effet pour un tel modèle, il faut à la fois déterminer le nombre de groupe et estimer les différents paramètres. Différentes approches sont utilisées dont principalement :

- La sélection du nombre de groupe par l'un des critères suivant : BIC (critère de Sélection de Schwarz's ou Bayésien), LEC(critère empirique de Laplace), MDL(longueur minimal de Rissanen's), ... Cette étape précède l'estimation du modèle qui peut se faire soit par MCMC, soit par EM avec approximation de la vraisemblance par la méthode, de Laplace, d'itération quadratique, de fonction d'importance ou de Monte Carlo. Cependant l'algorithme EM dans sa forme classique est assez sensible aux valeurs initiales et aboutit assez souvent à un maximum local ou un point selle.

- Le RJMCMC, Méthode de Monte Carlo par Chaîne de Markov et Saut Réversible (Richardson et Green, 1997) est fondé sur l'algorithme de Hastings-Métropolis avec probabilité d'acceptation redéfini, par ailleurs elle intègre directement la variabilité de la dimension des paramètres. Dans un algorithme qualifié de dynamique, il est possible d'une étape à une autre, et en fonction de la vraisemblance, de fusionner deux groupes pour en former un seul ou alors qu'un groupe soit éclaté pour en donner naissance à plusieurs. Bien évidemment les calculs sont un peu plus complexes et la convergence est assez lente.

- Le BDMCMC, Birth-Death MCMC (Stephens, 2000) est une alternative à la précédente, intégrant également la variabilité de la dimension des paramètres, il utilise la théorie des « point process » et le processus étudié est ainsi vu comme un cas particulier de ce type de processus.

II.3.1- Sélection du nombre de composantes par le critère de Bayes

En analyse bayésienne, le facteur de Bayes obtenu à partir du théorème de Bayes est régulièrement utilisé pour effectuer les tests d'hypothèses et la sélection des modèles.

Soient un échantillon X et deux modèles M_1 et M_2 ,

$$\text{nous avons } \pi(M_i|X) = \frac{\pi(X|M_i) * \pi(M_i)}{\pi(X|M_1) * \pi(M_1) + \pi(X|M_2) * \pi(M_2)} \quad . \quad i = 1, 2$$

$$\Rightarrow \frac{\pi(M_1|X)}{\pi(M_2|X)} = \frac{\pi(X|M_1) * \pi(M_1)}{\pi(X|M_2) * \pi(M_2)}$$

et $B_{12} = \frac{\pi(X|M_1)}{\pi(X|M_2)} = \frac{\pi(M_1|X)}{\pi(M_2|X)} * \frac{\pi(M_1)}{\pi(M_2)}$ qui est en fait le rapport de vraisemblance dans

le cas classique. L'existence de plusieurs méthodes (BIC, LEC,MDL) se justifie bien évidemment par les difficultés liées aux calculs d'intégrales. En effet la densité $\pi(X|M_i)$ s'obtient par intégration sur le domaine des paramètres.

$\pi(X|M_i) = \int \pi(X|\theta_i, M_i) \pi(\theta_i|M_i) d\theta_i$. Un second aspect pouvant parfois susciter quelques confusions dans l'utilisation du facteur de Bayes est le seuil à partir duquel un modèle ou une hypothèse doit être rejeté par rapport à une autre. Néanmoins cette préoccupation n'a pas sa raison d'être dans le cas de la sélection des modèles car il est interprété comme une mesure relative du succès d'un modèle à un autre (Kass et Raftery, 1995) ou alors dépendamment des objectifs, les modèles ayant les valeurs les plus élevés sont retenus.

Pour ce qui est de la méthode d'évaluation, les principales citées plus haut sont couramment utilisées, le critère de Schwarz est une approximation basée sur le logarithme de la vraisemblance et le résultat est assez satisfaisant.

$$BIC_k = 2 \log \pi(X|M_i) \approx 2 \log \pi(X|\hat{\theta}_k, M_k) - \nu_k \log(n)$$

ν_k est le nombre de paramètres à estimer dans le modèle de mélange M_k avec k composantes et $\hat{\theta}_k$ les paramètres estimées et π la vraisemblance.

II.3.2- Estimation du modèle par MCMC

L'objectif n'étant pas d'étudier la théorie des MCMC, qui d'ailleurs est nettement plus compliquée, Robert(1996), Robert et Cassela(1999) constituent d'excellentes références. Néanmoins la théorie est basée essentiellement sur l'utilisation de chaîne de Markov ergodique de loi stationnaire. Ceci se justifie, comme mentionne ce dernier par deux principales raisons :

- La particularité que possède les algorithmes d'optimisation stochastique de produire une structure de chaîne de Markov pouvant être exploitée de façon systématique.
- La grande difficulté dans certains cas de construire rapidement un algorithme efficace de simulation suivant une loi donnée.

C'est une approche proposant un cadre de simulation rigoureux et applicable quelle que soit la contrainte ou la caractéristique de la distribution étudiée. Ainsi pour une fonction de densité f , on se donne une valeur initiale $x^{(0)}$, arbitraire et une chaîne $x^{(t)}$ est générée à partir d'un noyau de transition de loi stationnaire f , qui garantit par ailleurs la convergence en loi vers f et ce après un nombre important d'itération. Dépendamment de la complexité des fonctions de densité étudiées, la forme générale appelée algorithme de Hastings-Métropolis ou alors l'algorithme de Gibbs considéré comme cas particulier du premier est utilisé. Les différentes étapes de l'algorithme peuvent se résumer en :

- **Hastings-Métropolis**

L'algorithme ici est un procédé permettant pour une distribution f donnée de générer

$$X_1, X_2, X_3, \dots, X_N \text{ tel que } E_f(h(X)) \rightarrow \frac{1}{T} \sum_{i=1}^T h(X_i).$$

La connaissance de f étant limitée, elle se fait en exploitant une seconde loi q plus aisée à manipuler (Normale, Gamma, Bêta,.....) et le plus souvent obtenue par approximation de f à l'infini et les étapes sont :

Étape 0

Choix du nombre de simulation N et d'une valeur initiale X_0 ;

Étape 1

Générer un candidat $Y_t \rightarrow q(Y|X_t)$;

Étape 2

$$\text{Prendre } X_{t+1} \rightarrow \begin{cases} Y_t \\ X_t \end{cases} \text{ avec probabilité } \min \left\{ \frac{f(Y_t) * q(X_t|Y_t)}{f(X_t) * q(Y_t|X_t)}, 1 \right\}$$

Étape 3 répéter 1 et 2 jusqu'à convergence. (Comparaison des estimées et représentation graphique)

▪ Échantillonnage de Gibbs

Contrairement à l'algorithme de Hasting et Métropolis précédent, des améliorations sont faites en utilisant les propriétés de la fonction f .

Soit m le nombre de paramètre.

Étape 0

Choix du nombre de simulation N et d'une valeur initiale X_0 ;

Étape 1

$$X^1_{t+1} \rightarrow f_1(X^1 | X^2_t, X^3_t, \dots, X^m_t)$$

$$X^2_{t+1} \rightarrow f_2(X^2 | X^1_{t+1}, X^3_t, \dots, X^m_t)$$

Générer .

$$X^m_{t+1} \rightarrow f_m(X^m | X^1_{t+1}, X^2_{t+1}, \dots, X^{m-1}_{t+1})$$

Étape 2 répéter 1 et 2 jusqu'à convergence. (Comparaison des estimées et représentation graphique).

1- Application au modèle avec distribution Lognormale

Étape 1

Initialisation des différents paramètres .

$$\mu^0_j \rightarrow Std\left(a, \frac{n_0 a_1}{2b_1}, a_1\right), \text{ moyenne du logarithme des pertes du groupe } j$$

a est initialisé à la valeur moyenne de l'intervalle des pertes (Min + Max)/2;

n₀ = 1;

a₁ le degré de liberté peut varier autour de 1 à 4;

$$\sigma_j^{2(0)} \rightarrow IG(a_1, b_1);$$

b₁ → $G(a_2, b_2)$, nous analyserons par ailleurs l'impact en la supposant fixe ou aléatoire;

a₂ = 0.1*a₁;

b₂ = 0.1*(Max-Min)², Max et Min du logarithme du montant des pertes ;

W → Dirichlet(c, c,c),

c = 1 .

Étape 2

Détermination du groupe d'appartenance pour chaque observation .

Y_i appartient au groupe **j** avec la probabilité,

$$\pi_j(Y_i) = \frac{W_j \frac{1}{\sqrt{2\pi} * \sigma_j} * \exp\left(\frac{-1}{2\sigma_j^2} * (Y_i - \mu_j)^2\right)}{\sum_{j=1}^P W_j \frac{1}{\sqrt{2\pi} * \sigma_j} * \exp\left(\frac{-1}{2\sigma_j^2} * (Y_i - \mu_j)^2\right)}$$

Étape 3

Mise à jour de la moyenne de chaque composante par échantillonnage de Gibbs.

$$\mu^t_j | Y, \sigma^{t-1}, W^{t-1} \rightarrow Std\left(\frac{n_0 a + n_j \bar{Y}_j}{n_0 + n_j}, (n_0 + n_j) 0.5(a_1 + n_j) b_3, a_1 + n_j\right)$$

$$b_3 = b_1 + 0.5 n_j S_j^2 + 0.5 (n_0 + n_j)^{-1} n_0 n_j (a - \bar{Y}_j)^2 \text{ ou}$$

$$\mu^t_j | Y, \sigma^{t-1}, W^{t-1} \rightarrow N\left(\frac{\sigma_j^{-2} n_j \bar{Y}_j + a b^{-1}}{\sigma_j^{-2} n_j + b^{-1}}, \frac{1}{\sigma_j^{-2} n_j + b^{-1}}\right)$$

$j = 1, \dots, P$. Toutes les autres variables sont indicées par $t-1$ et représentent les valeurs courantes (temps t).

Étape 4

Mise à jour de la variance de chaque composante par échantillonnage de Gibbs.

$$\sigma_j^{2(t)} | Y, \mu^t, W^{t-1} \rightarrow IG(0.5(a_1 + n_j), b_3) \text{ ou}$$

$$\sigma_j^{2(t)} | Y, \mu^t, W^{t-1} \rightarrow IG\left(a_1 + 0.5n_j, b_1 + 0.5 \sum_{i, i=j} (y_i - \mu_j)^2\right).$$

Étape 5

Mise à jour des pondération de chaque composante par échantillonnage de Gibbs.

$$W^t | Y, \mu_j^t, \sigma_j^{2(t)} \rightarrow \text{Dirichlet}(n_1 + c, n_2 + c, \dots, n_p + c).$$

Étape 6

Mise à jour de l'hyper paramètre b_1 par échantillonnage de Gibbs.

$$b_{1j}^{(t)} | Y, \mu^t, \sigma^{2(t)}, W^t \rightarrow G\left(a_2 + 0.5Pa_1, b_2 + \sum_j \sigma_j^{-2(t)}\right) \text{ ou}$$

$$b_{1j}^{(t)} | Y, \mu^t, \sigma^{2(t)}, W^t \rightarrow G\left(a_2 + b^{-1}a_1, b_2 + \sum_j \sigma_j^{-2(t)}\right).$$

Étape 7

Calcul des estimateurs.

$$\delta^\pi(\mu_j) = \frac{1}{T} \sum_{i=T} \mu_j^i;$$

$$\delta^\pi(\sigma_j^2) = \frac{1}{T} \sum_{i=T} \sigma_j^{2(i)}. \quad J=1, 2, \dots, P.$$

Étape 8

Classification des observation.

Y_i appartient au groupe j avec la probabilité,

$$\pi_j(Y_i) = \frac{W_j \frac{1}{\sqrt{2\pi * \delta^\pi(\sigma_j^2)}} * \exp\left(\frac{-1}{2\delta^\pi(\sigma_j^2)} * (Y_i - \delta^\pi(\mu_j))^2\right)}{\sum_{j=1}^P W_j \frac{1}{\sqrt{2\pi * \delta^\pi(\sigma_j^2)}} * \exp\left(\frac{-1}{2\delta^\pi(\sigma_j^2)} * (Y_i - \delta^\pi(\mu_j))^2\right)}.$$

2- Application au modèle avec distribution Gamma

Étape 1

Initialisation des différents paramètres .

$$\alpha_j \rightarrow \text{Exp}(a) \quad \text{ou} \quad \alpha_j \rightarrow G(a_1, b_1) . \quad j = 1, 2, \dots, P$$

$$\beta_j \rightarrow G(b, c) .$$

$$W \rightarrow \text{Dirichlet}(d, d, \dots, d) .$$

a est ;

a₁

b₁ ;

b ;

c ;

d = 1 .

Étape 2

Détermination du groupe d'appartenance pour chaque observation .

X_i appartient au groupe j avec la probabilité,

$$\pi_j(X_i) = \frac{W_j \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)} * X_i^{\alpha_j-1} * \exp(-\beta_j X_i)}{\sum_{j=1}^P W_j \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)} * X_i^{\alpha_j-1} * \exp(-\beta_j X_i)} .$$

Étape 3

Mise à jour des paramètres d'échelle de chaque composante par échantillonnage de Gibbs.

$$\beta^t_j | X, \alpha^{t-1}, W^{t-1} \rightarrow G\left(n_j \alpha_j + b, c + \sum_{i, l_i=j}^{n_j} X_i\right)$$

Étape 4

Mise à jour du paramètre de forme de chaque composante par échantillonnage de Gibbs.

$$\pi(\alpha_j^t | X, \beta^t, W^{t-1}) \rightarrow \frac{\Gamma(b + n_j \alpha_j) \prod_{i, l_i=j}^{n_j} X_i^{\alpha_j-1} \alpha_j \exp(-\alpha_j a)}{\Gamma(\alpha_j)^{n_j} (c + \sum_{i, l_i=j}^{n_j} X_i)^{n_j \alpha_j + b}}$$

$j = 1, \dots, P$. Toutes les autres variables sont indicées par $t-1$ et représentent les valeurs courantes (temps t).

Étape 5

Mise à jour des pondérations de chaque composante par échantillonnage de Gibbs.

$$W^t | X, \alpha_j^t, \beta_j^t \rightarrow \text{Dirichlet}(n_1 + d, n_2 + d, \dots, n_p + d).$$

Étape 6

Calcul des estimateurs.

$$\delta^\pi(\alpha_j) = \frac{1}{T} \sum_{i=T} \alpha_j^i;$$

$$\delta^\pi(\beta_j) = \frac{1}{T} \sum_{i=T} \beta_j^i. \quad J=1, 2, \dots, P.$$

Étape 7

Classification des observations

X_i appartient au groupe j avec la probabilité,

$$\pi_j(X_i) = \frac{W_j \frac{\delta^\pi(\beta_j)^{\delta^\pi(\alpha_j)}}{\Gamma(\delta^\pi(\alpha_j))} * X_i^{\alpha_j-1} * \exp(-\delta^\pi(\beta_j)X_i)}{\sum_{j=1}^P W_j \frac{\delta^\pi(\beta_j)^{\delta^\pi(\alpha_j)}}{\Gamma(\delta^\pi(\alpha_j))} * X_i^{\delta^\pi(\alpha_j)-1} * \exp(-\delta^\pi(\beta_j)X_i)}.$$

III- Présentation et analyse des résultats

Les résultats présentés dans cette partie sont entièrement fondés sur des données simulées. Nous examinerons la convergence, les différentes densités a posteriori, la robustesse du modèle, quelques principaux quantiles pour les densités de paramètres, la sensibilité des paramètres aux hypothèses initiales et enfin nous analyserons le comportement de la valeur à risque (99.9% de la distribution de perte) en supposant connu le nombre de perte annuelle.

Faute de résultats de références pouvant faciliter éventuellement des comparaisons, le choix des variables est certes aléatoire mais elle est faite de manière raisonnable et surtout en tenant compte de l'ordre de grandeur des statistiques de pertes souvent observées.

Nous considérons ainsi trois sources de données. L'analyse se fait sous un mélange en supposant que la distribution de chaque source suit une loi lognormale. Le tableau 3.1 résume quelques statistiques importantes des pertes simulées.

Source 1 : $X_i \rightarrow \text{LN}(14, 1.5)$;

Source 2 : $X_i \rightarrow \text{LN}(8, 2)$;

Source 3 : $X_i \rightarrow \text{LN}(4, 1.75)$.

L'échantillon utilisée pour l'estimation du modèle est obtenu en fusionnant les trois échantillons issues de la simulation préalable.

Tableau 3.1 : quelques statistiques descriptives des Données de pertes simulées.

	Nombre de perte	Moyenne de Perte ¹	Perte minimale	Perte maximale
Source 1	450	3 844 000	15129	187 000 000
Source 2	700	25 826	5	3 682 700
Source 3	850	273	0	6 125

1 : nous supposons une unité de monnaie quelconque.

III.1- Analyse de la convergence

Avant la présentation des résultats, l'algorithme utilisée pour l'estimation des paramètres étant un algorithme de simulation Monte Carlo, il est important de se rassurer de la convergence des résultats ceci d'autant plus que l'on ne dispose pas de méthode formelle de vérification. Cependant la trace des estimateurs pour différentes valeurs du nombre de simulation est un outil permettant de se prononcer sur la convergence des estimateurs. Les figures 1, 2 et 3 représentent les traces respectives de la moyenne, de l'écart type et du poids de chaque composante dans le mélange. Elles sont représentées pour ($T=40000, 60000$), et la constante observée dans les différentes trajectoires confirme que le nombre de simulation est suffisant pour assurer la convergence.

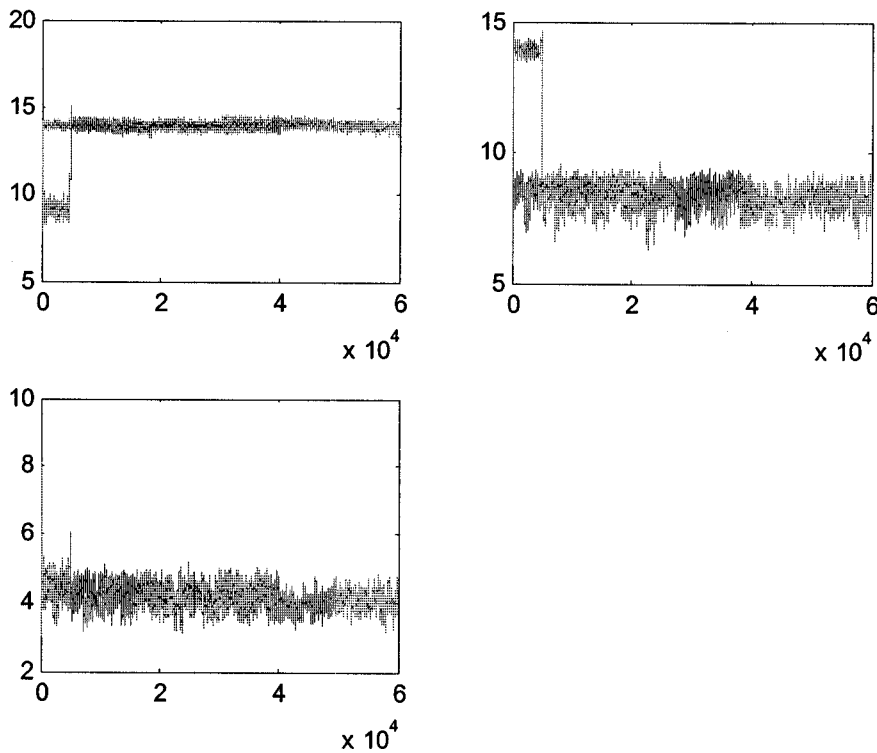


figure 1 : trace des moyennes μ_1, μ_2, μ_2 des trois composantes ($T=40000, T=60000$).

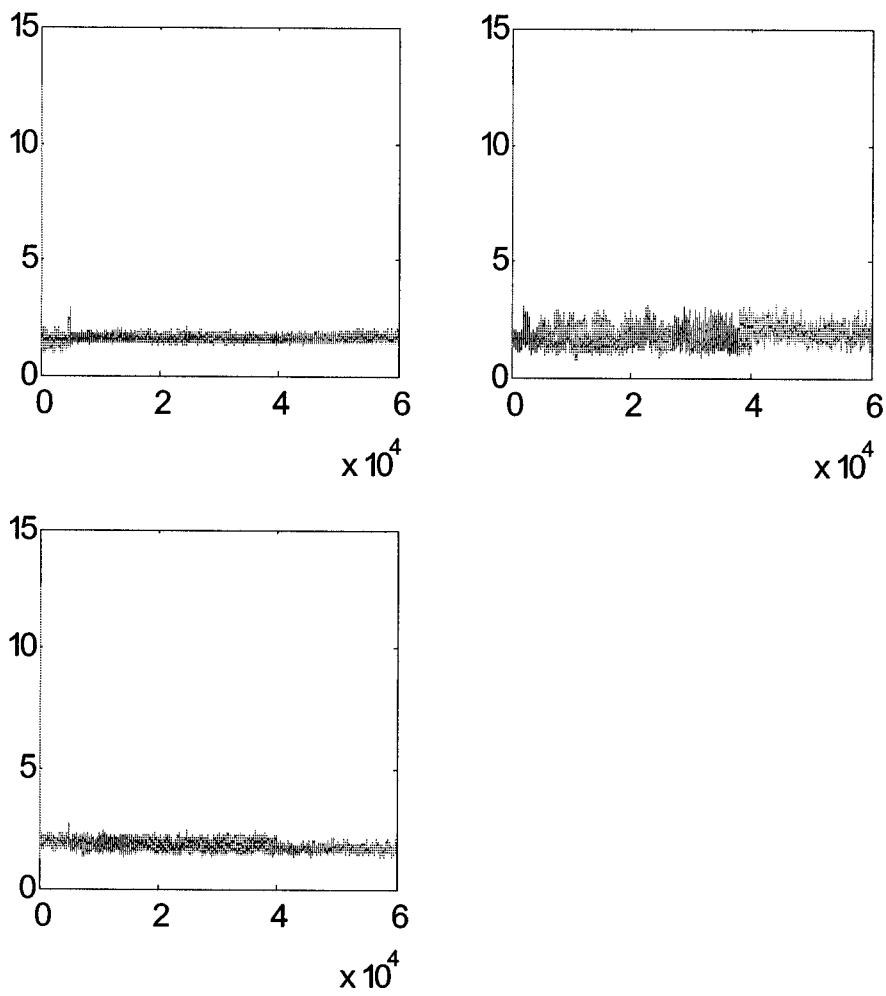


figure 2 : traces des écarts types σ_1 , σ_2 , σ_3 des trois composantes ($T=40000, 60000$).

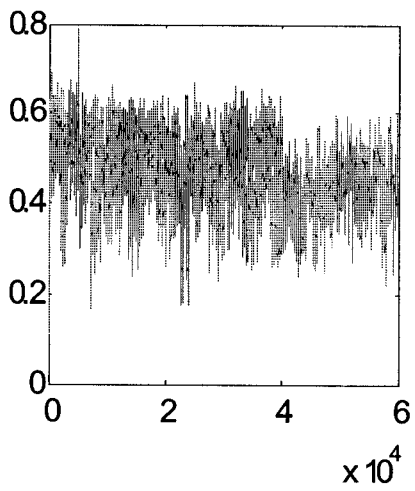
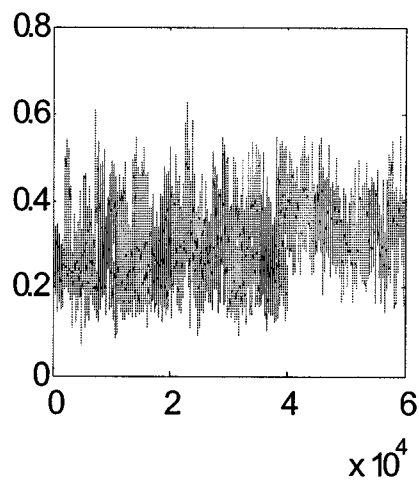
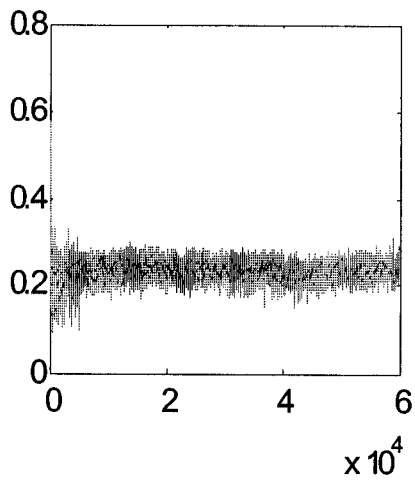


figure 3 : traces des paramètres de pondération P_1 , P_2 , P_3 des trois composantes ($T= 40000$, 60000).

III.2- Estimation des paramètres et de la Densité des données

La densité représentée ci-dessous (figure 4) est une approximation de la densité prédictive, elle est calculée avec les paramètres a posteriori. $\pi(Y_i|\hat{W},\hat{\theta}) = \sum_{j=1}^P \hat{w}_j f(Y_i|\hat{w}_j,\hat{\theta}_j)$; f étant la fonction de densité normale et $Y = \log X$. l'histogramme est celui des données de pertes simulées.

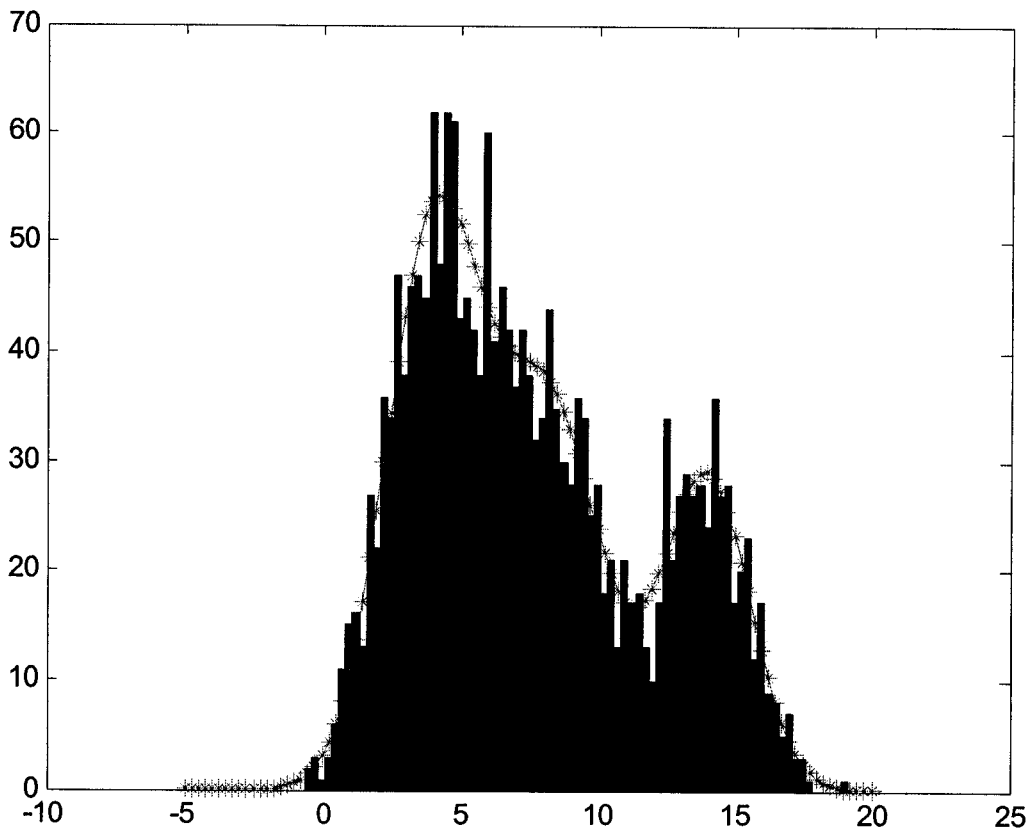


figure 4 : Histogramme des données et estimation de la densité associée.

Les figures 5, 6, 7 représentent les densités de probabilité a posteriori des paramètres estimés des différentes composantes. L'échantillon de données étant obtenu par simulation, l'estimation est relativement aisée et se traduit par des densités de paramètres uni modales. En présence de données réelles on pourrait avoir des modèles concurrents et éventuellement des

densités de paramètres avec plusieurs pics. Par ailleurs la valeur estimée et les quantiles d'ordre 1%, 5%, 95% et 99% sont respectivement résumés dans les tableaux 3.2, 3.3, 3.4 .

Nous observons entre autres que les biais sont relativement faibles pour certains estimateurs, en effet ils varient de 0.003 pour le plus bas et 0.16 pour le plus élevé.

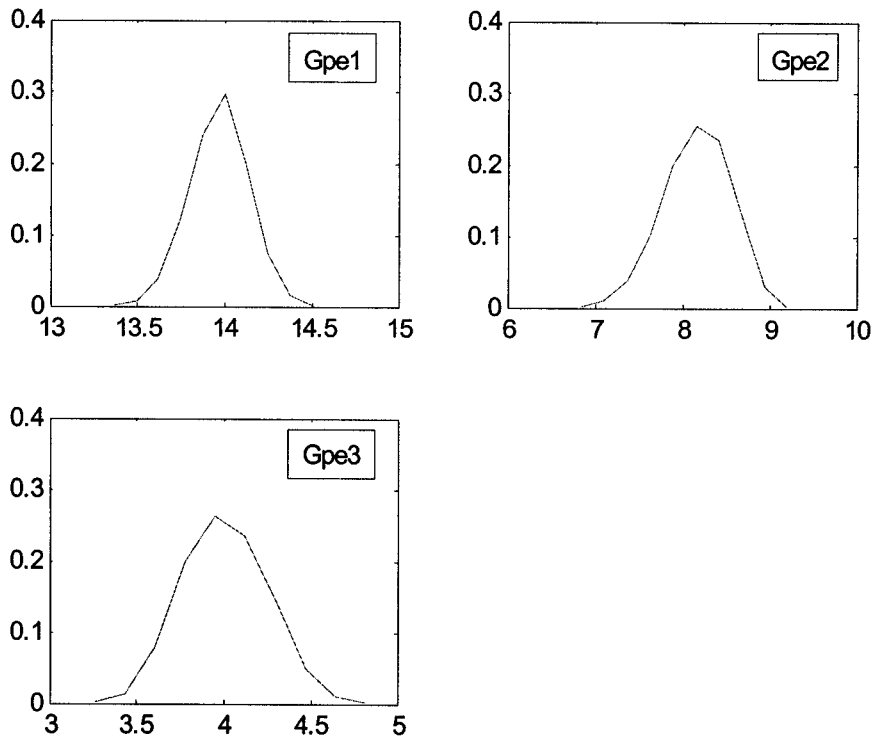


figure 5 : densités des moyennes μ_1, μ_2, μ_2 des trois composantes.

Tableau 3.2 : Valeur estimée et quantiles des moyennes μ_1, μ_2, μ_2 des trois composantes.

	Valeur théorique	Valeur estimée	1%	5%	95%	99%
Groupe1	14	13.968	13.577	13.690	14.236	14.346
Groupe2	8	8.164	7.243	7.510	8.753	8.932
Groupe3	4	4.002	3.492	3.625	4.400	4.554

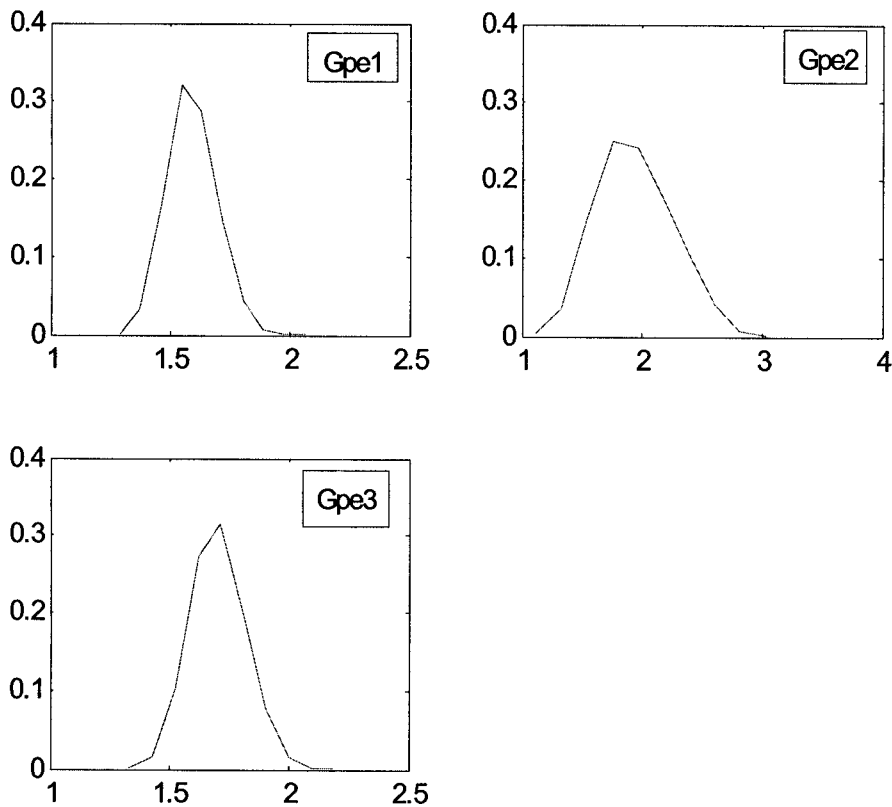


figure 6 : Densités des écarts types σ_1 , σ_2 , σ_3 des trois composantes.

Tableau 3.3 : Valeur estimée et quantiles des écarts types σ_1 , σ_2 , σ_3 des trois composantes.

	Valeur théorique	Valeur estimée	1%	5%	95%	99%
Groupe1	1.5	1.590	1.378	1.434	1.763	1.842
Groupe2	2	1.937	1.311	1.469	2.487	2.678
Groupe3	1.75	1.701	1.456	1.525	1.894	1.977

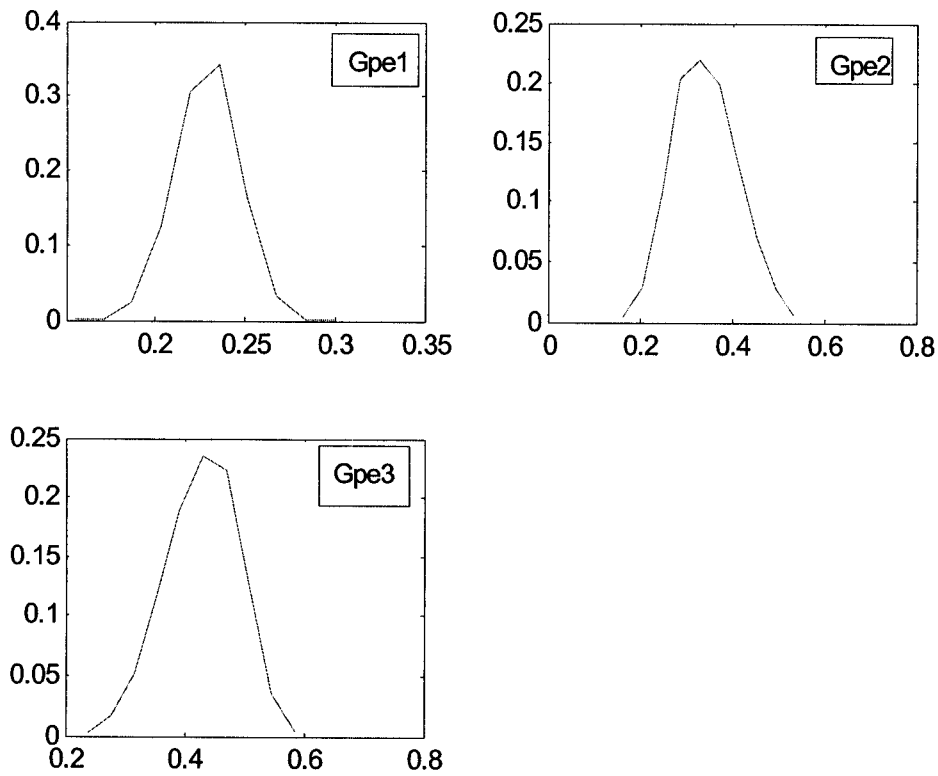


figure 7 : Densités de paramètres de pondération P_1 , P_2 , P_3 des trois composantes.

Tableau 3.4 : Valeur estimée et quantiles des proportions P_1 , P_2 , P_3 des trois composantes.

	Valeur théorique	Valeur estimée	1%	5%	95%	99%
Groupe1	0.2250	<i>0.228</i>	0.187	0.200	0.256	0.268
Groupe2	0.3500	<i>0.341</i>	0.199	0.237	0.460	0.503
Groupe3	0.4250	<i>0.429</i>	0.283	0.324	0.521	0.552

III.3- Sensibilité des résultats par rapport aux hypothèses.

Les résultats sont présentés dans le tableau 3.5.a, 3.5.b, 3.5.c. Ils résument l'analyse de la sensibilité par rapport à la distribution a priori de la moyenne (Student/Normale), à l'ordre de grandeur de la variance de la moyenne a priori et à la nature du paramètre d'échelle de la variance a priori (fixe/aléatoire). Bien évidemment, le faible biais des estimateurs (tableau 3.5.a) confirme la robustesse de la loi de Student pour la distribution a priori de la moyenne. En effet Anger(1987) démontre plus rigoureusement que sur des observations iid de loi normale, la distribution de Student pour la distribution a priori de la moyenne fournit des résultats plus robustes. Par ailleurs nous pouvons également conclure que les résultats sont sensibles à l'ordre de grandeur de la variance de la moyenne a priori et qu'il est préférable de travailler avec le paramètre d'échelle de l'écart type aléatoire plutôt que de le fixer.

Tableau 3.5.a : Sensibilité par rapport à la distribution a priori de la moyenne

	μ a priori est une Normale			μ a priori est une Student		
	μ a posteriori	σ a posteriori	P	μ a posteriori	σ a posteriori	P
Valeur estimée	13.857	1.637	0.236	13.968	1.590	0.228
	8.808	1.436	0.204	8.164	1.937	0.341
	4.518	2.037	0.558	4.002	1.701	0.429

Tableau 3.5.b : Sensibilité par rapport à la variance de la moyenne a priori $\mu \rightarrow N(a,b)$

	b=1			b=6		
	μ a posteriori	σ a posteriori	P	μ a posteriori	σ a posteriori	P
Valeur estimée	13.968	1.590	0.228	13.376	1.963	0.278
	8.164	1.937	0.341	8.881	0.984	0.112
	4.002	1.701	0.429	4.795	2.060	0.608

Tableau 3.5.c : Sensibilité par rapport au paramètre d'échelle de la Variance a priori
 $\sigma^2 \rightarrow IG(c, d)$

	d est fixe			d est aléatoire Gamma(e, f)		
	<i>μ a posteriori</i>	<i>σ a posteriori</i>	<i>P</i>	<i>μ a posteriori</i>	<i>σ a posteriori</i>	<i>P</i>
Valeur estimée	13.788	1.607	0.246	13.968	1.590	0.228
	8.061	1.611	0.299	8.164	1.937	0.341
	4.210	1.775	0.454	4.002	1.701	0.429

III.4- Classification des observations et Calcul de la VAR

La classification des observations de chaque source s'obtient en utilisant les estimateurs a posteriori des différents paramètres . En effet Y_i appartient au groupe j avec la probabilité,

$$\pi_j(Y_i) = \frac{\hat{W}_j \frac{1}{\sqrt{2\pi} * \hat{\sigma}_j} * \exp\left(\frac{-1}{2\hat{\sigma}_j^2} * (Y_i - \hat{\mu}_j)^2\right)}{\sum_{j=1}^p \hat{W}_j \frac{1}{\sqrt{2\pi} * \hat{\sigma}_j} * \exp\left(\frac{-1}{2\hat{\sigma}_j^2} * (Y_i - \hat{\mu}_j)^2\right)}$$

Le pourcentage a posteriori de chaque groupe varie légèrement, celui ci-dessous représente le pourcentage moyen pour chaque source des observations, l'impact de cette variation sur la VAR sera par ailleurs examiné .

Tableau 3.6 : pourcentage moyen a posteriori des observations pour chaque source de données.

	Valeur théorique	Valeur estimé	Nombre d'observation	Pourcentage moyen a posteriori Grp1	Pourcentage moyen a posteriori Grp2	Pourcentage moyen a posteriori Grp3
Source1	14; 1.5	13.96; 1.59	450	0.931	0.069	0
Source 2	8; 2	8.16; 1.93	700	0.035	0.773	0.192
Source3	4; 1.75	4.002; 1.70	850	0	0.172	0.827

Pour le calcul de la VAR, nous considérons le nombre de perte connu pour une année. La VAR est calculée respectivement sur la source de données initialement simulées puis sur la même source de données a posteriori après classification des observations. Par ailleurs la perte maximale est fixée.

Dans la suite, nous supposons que la fréquence des perte suit une loi de poisson de moyenne annuelle égale à 600 (ou de moyenne mensuelle 50). Nous travaillons sur les données de pertes simulées de la source2, elle contient 700 observations et suit une Lognormale (8, 2). La classification a posteriori des observations est résumée au tableau 3.6.

Par simulation Monte Carlo(N=50000) nous obtenons deux distributions de pertes :

- La première(tableau 3.6.a) en supposant que la fréquence de perte suit une loi de poisson de paramètre 600(fréquence annuelle) et que la sévérité de perte suit une Lognormale(8, 2), donnée de pertes(Source2).

- La deuxième(tableau 3.6.b, 3.6.b, 3.6.b) en maintenant la fréquence et en considérant la distribution a posteriori de la source 2.

Tableau 3.6.a : Quantile de la distribution de perte

- $f = LN(8;2)$.

	95%	98%	99%	99.5%	99.9%
Source2	15 9456305	17 100595	18 036 149	18 953 950	20 690 017

Nous faisons plusieurs calculs dans le cas du modèle de mélange. En effet nous avons mentionné plus haut que le poids des groupes varie légèrement, et que le pourcentage moyen est obtenu en répétant plusieurs fois la classification . Ceci permet en sélectionnant plus d'une combinaison d'analyser aussi l'impact du poids des différents groupes sur le valeur de la VAR. Ainsi, trois sous cas sont examinés.

- $f = 0.028*LN(13.96;1.59)+0.758*LN(8.16;1.93)+ 0.214*LN(4.002;1.70)$.

Tableau 3.6.b : Quantile de la distribution de perte

	95%	98%	99%	99.5%	99.9%
Source2 a posteriori	20 003 055	20 539 890	20 921 998	21 329 379	22 025 000

- $f = 0.035*LN(13.96;1.59)+0.773*LN(8.16;1.93) + 0.192*LN(4.002;1.70)$.

Tableau 3.6.c : Quantile de la distribution de perte

	95%	98%	99%	99.5%	99.9%
Source2 a posteriori	23 714 760	24 296 057	24 704 269	2 5 142 742	25 894174

▪ $f = 0.040 * LN(13.96;1.59) + 0.725 * LN(8.16;1.93) + 0.235 * LN(4.002;1.70).$

Tableau 3.6.d : Quantile de la distribution de perte

	95%	98%	99%	99.5%	99.9%
Source2 a posteriori	27 333 684	28 0129 745	28 490 672	29 003 214	29 875 553

Au regard de ces différents calculs et ce malgré qu'ils soient effectués sur des données de pertes simulées, la VAR ou plus simplement les quantiles de la distribution sont assez sensibles au poids des différentes composantes et il est par ailleurs possible (tableau 3.6.a et 3.6c) d'obtenir une VAR assez proche de celle obtenue a partir de la distribution théorique initiale. Cependant ce résultat n'est pas confirmé (Tableau 3.6.e et 3.6.f) lorsque nous reprenons le même exercice sur trois sources de données relativement homogènes(en terme d'ordre de grandeur des pertes), le modèle de mélange surestime de façon considérable les quantiles et vu le comportement de l'algorithme utilisé il nous a semblé nécessaire d'intégrer la variabilité de la dimension des paramètres tel que c'est fait dans les algorithmes avec saut réversible (RJMCMC).

Tableau 3.6.e : quelques statistiques descriptives des Données de pertes simulées.

	Nombre de pertes	Moyenne de Pertes ¹	Perte minimale	Perte maximale
Source 1	450	134 635	105	4 176 835
Source 2	700	12 702	20	1 095 763
Source 3	850	164	0	6 002

1 : nous supposons une unité de monnaie quelconque.

Tableau 3.6.f : Comparaison des quantiles de la distribution de perte(le nombre de groupe a posteriori est $K = 2$)

	95%	98%	99%	99.5%	99.9%
Source 2	10 097 538	10 995 67	11 643 742	12 039 021	13 051 794
Source 2 a posteriori	21 936 779	23 190 344	24 121 684	24 899 891	25 452 731

Conclusion

Le succès de l'approche LDA dans le domaine de l'Assurance, la rigueur mathématique sous-jacente, l'intelligibilité et le cadre relativement simplifié de mise en œuvre font de cette approche une des plus adaptées de mesure du risque opérationnel. Cependant son fondement statistique l'expose aux exigences des données de pertes sur trois à cinq ans pour espérer obtenir des résultats robustes. C'est à ce niveau que se situe le plus grand défi car ne disposant pas de données sur de tel horizon et vu le caractère dynamique de l'environnement associé, il faut trouver des outils permettant d'intégrer toutes les données de l'industrie financière pour mener à bien cette opération.

C'est dans cette logique que pour des raisons mentionnées plus haut (II.1.1) dans la justification du modèle (prise en compte de la diversité des sources et de la taille des pertes, moins exigeante en données interne, flexibilité et possibilité de classification a posteriori), le modèle de mélange de distribution ne serait ce que théoriquement nous paraît au côté des autres approches, potentiellement adapter pour la mesure du risque par l'approche LDA. Malgré que l'estimation soit faite sur des données de pertes simulées, il est possible de tirer quelques leçons des résultats issues de la section précédente. En effet les résultats obtenus sur une des études de cas sont assez raisonnables au regard de la comparaison des différents quantiles. Cependant lorsque l'échantillon est construit de façon à obtenir des données de pertes relativement homogènes en terme d'ordre de grandeur, le modèle de mélange surestime considérablement les quantiles. Au regard du comportement de l'algorithme MCMC utilisé pour ce dernier cas, il paraît inadapté du fait de la présence des étapes nécessitant la fusion ou la division des groupes constitués .

Il nous semble par ailleurs nécessaire d'approfondir certains aspects dont entre autres :

- l'estimation du modèle de mélange sur des données réelles;
- le choix des distributions théoriques de données;
- l'utilisation d'un algorithme intégrant éventuellement la variation de la dimension des paramètres, fusion ou division des groupes constitués type RJMCMC;
- La pondération des différentes sources dans le mélange et l'orientation de manière à intégrer spécifiquement les préoccupations liées aux pertes extrêmes.

Bibliographie

Alexander C., Operational Risk: Regulation, Analysis and Management. FT Prentice Hall, 2003.

Alexander C., Bayesian Methods for Measuring Operational Risk. Discussion papers, 2000-02.
Ali Samad-Khan, Statistical/Actuarial Modeling in Operational Risk, GARP conference London, England, 2001.

Angers, J.F., development of robust Bayes estimators for a multivariate normal mean. Ph.D. Thesis, Purdue University. 1987.

Basel Committee on Banking Supervision, Consultative Document the New Basel Capital Accord. 2003.

Berger, J. Statistical Decision Theory and Bayesian Analysis (2nd edition). Springer-Verlag, NY. 1985a

Chen Y, D. Fournier, Impacts of atypical data on Bayesian inference and robust Bayesian approach in fisheries. Canadian Journal of Fisheries and Aquatic Sciences, 56, 1525-1533, 1999.

Christian Robert, L'Analyse Statistique Bayésienne, Economica. 1992.

Christian Robert, Méthodes de Monte Carlo par Chaînes de Markov, Economica, 1996.

Christian Robert, George Casella, Monte Carlo Statistical Methods, Springer, 1999.

Christian Robert, Kerrie Mengersen; Reparametrisation Issues in Mixture Modelling and their bearing on MCMC algorithms. Computational Statistics & Data Analysis, 29 1999, 325-343.

Chris Fraley, Adrian Raftery, How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis, Computer Journal, vol. 41, 1998, 578-588.

Cruz M., Modeling, Measuring and Hedging Operational Risk. Wiley finance, 2002.

Edward Dudewicz, Satyan Mishra, Modern Mathematical Statistics. Wiley serie in probability and Mathematical Statistics.

Frachot A., P. georges, T. Roncalli, Loss Distribution Approach for operational risk, GRO Crédit Lyonnais, 2001.

Frachot A., P. georges, T. Roncalli, Internal data, External data and consortium data for operational risk measurement: How to pool data properly?, GRO Credit Lyonnais, 2002.

Frachot A., P. Georges, T. Roncalli, Loss Distribution Approach in Practice, GRO Credit Lyonnais, 2003.

Gertner G., H. Zhu, Bayesian estimation in forest surveys when samples or prior information are fuzzy. Fuzzy sets and systems, 1996, 77, 277-290.

Jennifer A., A. Hoeting, D. Madigan, A. Raftery, et T. Volinsky, Bayesian Model Averaging. Statistical Science, 1999. 14 382-417.

MacLachlan, G., Basford K. Mixture Models. Dekker, NY. 1987.

Matthew Stephens, Bayesian analysis of mixture models with an unknown number of components—An alternative to reversible jump methods. Journal of statistics, 2000 v. 28, 40-74.

Norman Johnson, Samuel Kotz, Continuous univariate distribution 1 & 2. Distribution in Statistics, 1970.

Simon Gourde, Hela Dahan, Etude Exploratoire relative au capital minimum requis relié au risque opérationnel: Présentation et application de la méthode LDA. Banque National du Canada, 2003

Sylvia R., J. Green, on Bayesian analysis of mixtures with an unknown number of components. JR stat society. Serie B 59, 731-792, 1997.

ANNEXES

↳ Déterminer Zi pour chaque observation

```
function Poids = simulerZ(P1,MU1,SIGMA1,K1)
global Tail_Ech Donnees Sigma_j Mu_j P
Z1 = ones(Tail_Ech,1);
[tempon1,Sumtempon1] = P_DE_Z(P1,MU1,SIGMA1,K1);
for j1=1:K1
    for i1=1:Tail_Ech % améliorer en enlevant cette boucle.
        tempon4a=tempon1(i1,)./Sumtempon1(i1);
        tempon4b = cumsum(tempon4a);
        j4 = find(tempon4b>= rand);
        if (length(j4) >0)
            Z1(i1)=j4(1);
        end
    end
end
end
Poids = Z1;
```

```
function [PZi,SumPZi] = P_DE_Z(P3,MU3,SIGMA3,K3)
global Tail_Ech Donnees Sigma_j Mu_j P
tempon3 = zeros(Tail_Ech,K3);
for j3=1:K3
    tempon3(:,j3)= P3(j3).*normpdf(Donnees,MU3(j3),SIGMA3(j3));
end
PZi=tempon3;
SumPZi = sum(PZi,2);
```

% calcul de : taille de sous groupe, moyenne empirique*n et somme des carrées

```
function [N_j2,N_j_Xbar2]= calculerN(Z2,K2)
global Tail_Ech Donnees
for i2=1:K2
    j2 =find(Z2==i2);
    A2_0=length(j2);
    B2 =sum(Donnees(j2(1:A2_0)));
    A2= B2/A2_0;
    N_j2(i2,1)=A2_0;
    N_j_Xbar2(i2,1)=B2;
    C2 = sum((Donnees(j2(1:A2_0))-A2).^2);
    Scarre2(i2,1)=C2;
end
```

%Initialisation et fixation de l'etat pour eviter des effets nés de la dépendance

% ***** de quelques fonctions a rand et randn

```
function Inistate
state = sum(100*clock);
rand('state', state);
randn('state', state);
```



```

% Programme Principal

clear;
format long
global Tail_Ech Donnees Sigma_j Mu_j P

%          ***** Description des lois a priori          *****

% mu_j suit un student T(alpha,beta.theta/2,theta)et theta est le degré deL
% (mu_j,Sig_j^2 ) suit une normale gamma(alpha,beta,theta/2,neta) ou student
% ainsi mu_j/Sig_j^2 suit une normale N(alpha,beta.Sig_j^2); beta positif et
% Sig_j^2 suit une gamma(theta/2,neta); theta et neta positifs
% neta suit une gamma(a,b); a et b positifs
% (P1,P2,.....Pk) suit un Dirichlet (rho,rho,.....rho)

%          ***** Initialisation des hyperparamètres          *****
A = lognrnd(8,2,700,1);
B = lognrnd(14,1.5,450,1);
C = lognrnd(4,1.75,850,1);
Donnees = [log(A);log(B);log(C)];
% alpha éventuellement le milieu de l'intervalle min..max des pertes
alpha = 7.5288;
% essayer de voir pour beta, 1 éventuellement et voir sensibilité
beta = 1 ;
% voir si on peut choisir theta et neta tel que Sig_j^2 soit R^2 (min - max)perte
Sig_jcarre_I = 150.5653;
% theta/2, essayer de voir, éventuellement 2 voir aussi sensibilité;
theta= 4;
% a un multiple éventuel(0.1;0.2;...) de theta/2,theta/2 > a
a= 0.1*theta/2 ;
% b un multiple éventuel(0.1;0.2;...) de neta
b=0.1*Sig_jcarre_I ;
% rho est 1 pour la loi de dirichlet
rho = 1 ;
% initialiser la taille de l'échantillon
Tail_Ech =2000 ;

%          ***** Initialisation des moyennes a priori          *****

% revoir la taille en fonction de critère. voir un peu les papiers Rob,Rich et ...
Nbre_sim = 100000;
Sigma_j = zeros(Nbre_sim,K);
Sigma_j(1,:)= Sig_jcarre_I.*ones(1,K);
Mu_j = zeros(Nbre_sim,K);
Mu_j(1,:)= alpha + (trnd(theta)./sqrt(Sig_jcarre_I/beta));
Neta = zeros(Nbre_sim,K);
Neta(1,:)= gamrnd(a,1/b);

% générer un dirichlet(1,1,1,...1) a partir de la gamma pour P(Devroy,p594)
tempon0=zeros(Nbre_sim,K);
tempon0(1,:)= gamrnd(1,1,K,1)';
tempon0_a = sum(tempon0);

```

```
P(1,:)= tempon0(1:K)./tempon0_a;
```

```
% ***** Intérieur de la boucle *****
```

```
for I=2:Nbre_sim
```

```
    Inistate
```

```
    % Déterminer Zi pour chaque observation, bref étant a l'étape m ;
    % les valeurs de P, Mu_j, Sigma_j sont celles de l'étape précédentes
    % P(Zi=j) = pj*f(Xi/les paramètres
```

```
    tempon0_a0=P(I-1,:);
```

```
    tempon0_a1=Mu_j(I-1,:);
```

```
    tempon0_a2= sqrt(Sigma_j(I-1,:)); %modif
```

```
    Z = simulerZ(tempon0_a0,tempon0_a1,tempon0_a2,K);
```

```
    % calcul de : taille de sous groupe, moyenne empirique*n et somme des carrées
    [N_j,N_j_Xbar,Scarre]= calculerN(Z,K);
```

```
    % déterminer Mu_j(I,:) en utilisant Sigma_j précédent, N_j courant et une
    %normale(beta*alpha+N_jXbar/Beta+N_j, Beta*Neta_j*sigmacarre
```

```
    tempon0_b = (beta*alpha + N_j_Xbar')./(beta + N_j');
```

```
    tempon0_e = N_j_Xbar./N_j;
```

```
    tempon0_c = Neta(I-1,:) + 0.5*Scarre' + 0.5*(beta*N_j'.*((alpha-tempon0_e').  
^2))./(beta + N_j)';
```

```
    tempon0_c1 = (beta + N_j)'*(0.5*theta + 0.5*N_j')./tempon0_c;
```

```
    Mu_j(I,:) = tempon0_b + (trnd(theta + N_j')./sqrt(tempon0_c1)); % modif
```

```
    % déterminer Sigma_j(I,:) en utilisant neta précédent et un gamma(0.5  
*theta+0.5.N_j,
```

```
    % neta+0.5N_J*Scarre+beta*N_j(alpha-xbar)^2/(beta+N_j))
```

```
    tempon0_d = 0.5*theta + 0.5*N_j';
```

```
    Sigma_j(I,:) = gamrnd(tempon0_d,1./tempon0_c);
```

```
    Sigma_j(I,:) = 1./Sigma_j(I,:);
```

```
    % déterminer P(I,:), un dirichlet(n1+1,1,1,...nk+1)
```

```
    tempon0_g= gamrnd(N_j'+1,ones(1,K));
```

```
    tempon0_h = sum(tempon0_g);
```

```
    P(I,:)= tempon0_g(1:K)./tempon0_h;
```

```
    % déterminer neta un gamma(a+0.5*k*theta,b+somme(sigma^2))
```

```
    tempon0_i = a + 0.5*K*theta;
```

```
    tempon0_j = b + sum(1./Sigma_j(I,:));
```

```
    Neta(I,:)= gamrnd(tempon0_i,1/tempon0_j);
```

end

```
% ***** Moyenne a posteriori *****
T1 = Mu_j((20000:100000),:);
Mean_Mu_j = mean(T1,1)
T1 = Sigma_j((20000:100000),:);
Mean_Sigma_j = mean((sqrt(T1)),1)
T1 = P((20000:100000),:);
Mean_P = mean(T1,1)
```

5

5