Institut de Mathématiques de Marseille
Institut des Sciences Financières et d'Assurance
ED 184: Ecole doctorale en Mathématiques et Informatique de Marseille

Approximations polynomiales de densités de probabilité et applications en assurance

THESE DE DOCTORAT

présentée par

Pierre-Olivier Goffard

pour obtenir le grade de

Docteur de l'université d'Aix-Marseille

Discipline : Mathématiques appliquées

Soutenue le 29/06/2015 devant le jury :

Claude Lefèvre	Professeur à l'Université Libre de Bruxelles	Rapporteur
Partice Bertail	Professeur à l'Universités Paris X	Rapporteur
Dominique Henriet	Professeur à l'école Centrale Marseille	Examinateur
Xavier Guerrault	Actuaire chez Axa France	Responsable entreprise
Denys Pommeret	Professeur à l'université d'Aix-Marseille	Directeur de thèse
Stéphane Loisel	Professeur à l'université de Lyon 1	Co-directeur de thèse

Table des matières

Liste des figures

Liste des tableaux

Résumé

Cette thèse a pour objet d'étude les méthodes numériques d'approximation de la densité de probabilité associée à des variables aléatoires admettant des distributions composées. Ces variables aléatoires sont couramment utilisées en actuariat pour modéliser le risque supporté par un portefeuille de contrats. En théorie de la ruine, la probabilité de ruine ultime dans le modèle de Poisson composé est égale à la fonction de survie d'une distribution géométrique composée. La méthode numérique proposée consiste en une projection orthogonale de la densité sur une base de polynômes orthogonaux. Ces polynômes sont orthogonaux par rapport à une mesure de probabilité de référence appartenant aux Familles Exponentielles Naturelles Quadratiques. La méthode d'approximation polynomiale est comparée à d'autres méthodes d'approximation de la densité basées sur les moments et la transformée de Laplace de la distribution. L'extension de la méthode en dimension supérieure à 1 est présentée, ainsi que l'obtention d'un estimateur de la densité à partir de la formule d'approximation.

Cette thèse comprend aussi la description d'une méthode d'agrégation adaptée aux portefeuilles de contrats d'assurance vie de type épargne individuelle. La procédure d'agrégation conduit à la construction de *model points* pour permettre l'évaluation des provisions *best estimate* dans des temps raisonnables et conformément à la directive européenne Solvabilité II.

Mots-clé : distributions composées, théorie de la ruine, Familles Exponentielles Naturelles Quadratiques, polynômes orthogonaux, méthodes numériques d'approximation, Solvabilité II, provision *best estimate, model points.*

Abstract

This PhD thesis studies numerical methods to approximate the probability density function of random variables governed by compound distributions. These random variables are useful in actuarial science to model the risk of a portfolio of contracts. In ruin theory, the probability of ultimate ruin within the compound Poisson ruin model is the survival function of a geometric compound distribution. The proposed method consists in a projection of the probability density function onto an orthogonal polynomial system. These polynomials are orthogonal with respect to a probability measure that belongs to Natural Exponential Families with Quadratic Variance Function. The polynomiam approximation is compared to other numerical methods that recover the probability density function from the knowledge of the moments or the Laplace transform of the distribution. The polynomial method is then extended in a multidimensional setting, along with the probability density estimator derived from the approximation formula.

An aggregation procedure adapted to life insurance portfolios is also described. The method aims at building a portfolio of model points in order to compute the best estimate liabilities in a timely manner and in a way that is compliant with the European directive Solvency II.

Keywords : compound distributions, ruin theory, Natural Exponential Families with Quadratic Variance Function, orthogonal polynomials, numerical methods, Solvency II, best estimate liabilities, model points.

Introduction

Ce travail a été réalisé dans le cadre d'une thèse convention CIFRE, issue d'un partenariat entre la compagnie d'assurance AXA et Aix-Marseille Université. Il comprend une partie opérationelle concrétisée par un projet de recherche et développement au sein d'AXA France et une partie académique visant à l'élaboration de méthodes numériques pour la gestion des risques des compagnies d'assurance.

La partie opérationelle de ce travail s'inscrit dans les préparatifs de l'entrée en vigueur de la directive européenne Solvabilité II. Cette directive définit un nouveau cadre prudentiel et oblige les compagnies d'assurance européennes à adopter une vue prospective de leurs engagements afin de bien définir leur niveau de provisionnement. Ayant été accueilli au sein d'une direction technique en charge de la gestion des contrats d'assurance vie du périmètre épargne individuelle, le projet de recherche et développement conduit au sein d'AXA France a eu pour but l'optimisation des temps de calcul des provisions *Best Estimate* associées aux contrats d'assurance vie de type épargne individuelle. L'assuré, lors de la souscription, effectue un premier versement pour constituer un capital initial. Ce capital est investi sur différents supports d'investissement dont la performance impacte la valeur de l'épargne au cours du temps. L'assureur s'engage à verser à l'assuré ou à ses bénéficiaires désignés la valeur de l'épargne en cas de décès ou de rachat du contrat. Le rachat du contrat est une action volontaire de l'assuré qui décide de récupérer tout ou partie de son épargne afin de l'utiliser autrement. La provision *Best Estimate* d'un contrat épargne individuelle est égale à la moyenne des *cash-flows* sortants, actualisés et pondérés par leur probabilité d'occurence au cours du temps. Le décès et le rachat sont les deux évènements qui entrainent un *cash-flow* sortant. Ces deux évènements sont respectivement associés à des probabilités de décès et de rachat fonction de l'âge de l'assuré, de son genre et aussi de l'ancienneté du contrat. Des lois de rachat et de décès sont calibrées statistiquement à l'aide d'un historique. La valeur du *cash-flow* sortant est fonction de la valeur de l'épargne au cours du temps et du rendement des actifs sur les marchés financiers. Le rendement des actifs est modélisé par l'intermédiaire de processus stochastiques. Des scénarios d'évolution de taux d'intérêts sont générés sur la base de ces processus stochastiques. La provision est évaluée pour chacun de ces scénarios, la moyenne des provisions sur les différents scénarios financiers renvoit la provision *Best Estimate*. La directive Solvabilité II préconise une approche contrat par contrat. Le temps de calcul pour un contrat est relativement faible, de l'ordre de quelques secondes, il devient problématique dans le cadre d'un portefeuille de contrats volumineux comme celui d'AXA France qui comprend 3 millions de contrats. L'autre problème réside dans la consommation de mémoire induite par le stockage des données de simulation. Les autorités de contrôle et de réglementation sont conscientes de ces problèmes et cautionnent la mise en place de techniques pour réduire les temps de calculs sous réserve de ne pas fausser significativement la valorisation des provisions. L'objectif du projet de recherche et développement mené au cours de cette thèse est la mise au point d'une procédure visant à agréger le portefeuille de contrats d'assurance vie. Le portefeuille de contrats agrégé est alors utilisé en *input* du modèle interne de projection des *cash-flows* permetant l'évaluation des provisions *Best Estimate*.

La partie académique de ce travail consiste en la mise au point d'une méthode numérique d'approximation des probabilités dans le cadre des modèles de risque en assurance non-vie. Pour un portefeuille de contrats d'assurance non vie, sur une période d'exercice donnée, le nombre de sinistres est une variable aléatoire de comptage et le montant des sinistres est une variable aléatoire positive. La charge totale du portefeuille, qui correspond au cumul des prestations à verser par l'assureur sur la période d'exercice considérée, est une quantité centrale dans la gestion des risques de la compagnie. Elle quantifie les engagements contractés par la compagnie d'assurance auprès de ses clients. Pour calibrer le montant de la prime et déterminer les marges de solvabilité, l'actuaire doit étudier la distribution du cumul des prestations et connaître son espérance, sa variance, sa fonction de survie ou encore ses quantiles. La charge totale du portefeuille admet une distribution de probabilité composée. Le problème de ces distributions de probabilité est que leur densité de probabilité n'est accessible que dans un nombre très restreint de cas. Les calculs de probabilité peuvent être effectués par le biais de méthodes de simulation de Monte-Carlo mais les temps de calcul importants associés à une bonne précision sont souvent prohibitifs et motivent la mise au point de méthodes numériques d'approximation de la densité de probabilité. Une compagnie d'assurance possède souvent plusieurs branches d'activité, chaque branche d'activité est associée à une charge totale et la gestion globale des risques implique la définition d'un vecteur aléatoire. En cas d'indépendance entre les branches d'activité, le problème se résume à considérer plusieurs problèmes en dimension 1. En revanche, s'il existe une corrélation entre les branches d'activité alors il est opportun de modéliser les charges totales conjointement via un vecteur aléatoire associé à une distribution de probabilité multivariée. Les méthodes numériques doivent alors pouvoir être étendues en dimension supérieure à 1. Cette corrélation est particulièrement pertinente dans la gestion des risques d'un réassureur qui réassure la même branche d'activité de plusieurs compagnies d'assurance. Un prolongement du modèle collectif est étudié en théorie de la ruine où un aspect dynamique est ajouté. La théorie de la ruine concerne la modélisation de la richesse globale d'une compagnie d'assurance ou la réserve financière allouée à une certaine branche d'activité ou portefeuille de contrats. La réserve financière est égale à une réserve initiale à laquelle s'ajoute les primes reçues par unité de temps moins les prestations. L'arrivée des sinistres est modélisée par un processus de comptage, le montant des prestations par des variables aléatoires positives. L'objectif est de déterminer la probabilité de ruine qui correspond a la probabilité de passage en dessous de 0 du processus en fonction de la réserve financière initiale. La complexité des variables aléatoires en jeu rend la probabilité de ruine difficile à évaluer via des formules fermées d'où la mise au point de méthodes numériques.

L'objet de ce travail est d'optimiser la précision et la rapidité des temps de calcul. La partie opérationnelle traite d'une procédure à utiliser en amont du calcul alors que la partie académique s'intéresse à la façon même de réaliser les calculs.

Le Chapitre 1 présente les distributions composées et leur utilité en actuariat, et en théorie de la ruine. La distribution composée en dimension 1 modélise le cumul du montant des prestations pour un portefeuille de contrats sur une période d'exercice donnée. Le concept de distribution composée s'étend en dimension supérieure à 1 pour permettre la modélisation jointe des risques supportés par plusieurs portefeuilles dont les charges totales sont corrélées. Le Chapitre 1 s'achève sur une introduction à la théorie de la ruine avec la présentation du modèle de ruine de Poisson composé et la définition de la probabilité de ruine ultime.

Le Chapitre 2 présente la méthode numérique d'approximation au centre de ce travail et comprend la plupart des résultats mathématiques de cette thèse. La densité de probabilité d'une variable aléatoire est projetée sur une base de polynômes orthogonaux. L'approximation est obtenue en tronquant à un certain ordre la représentation polynomiale de la densité de probabilité. La f.d.r. et la f.d.s. sont approchées par intégration de l'approximation polynomiale de la densité de probabilité. L'application de la méthode d'approximation aux distributions composées en dimension 1 conduit à opter pour une mesure de probabilité gamma associée aux polynômes de Laguerre généralisés. Pour que l'approximation polynomiale soit valide, il est nécessaire de vérifier une condition d'intégrabilité. La Proposition 6 définit une majoration de la densité de probabilité qui permet de choisir les paramètres de la mesure de référence, voir Corollaire 2, pour assurer la validité de l'approximation polynomiale. La qualité de l'approximation dépend de la décroissance des coefficients de la représentation polynomiale. La Proposition 7 indique une convergence en $1/k$, sous certaines conditions de régularité sur la densité de probabilité. La fonction génératrice des coefficients s'exprime en fonction de la transformée de Laplace de la densité de probabilité. L'étude de la forme de cette fonction génératrice permet dans certains cas de choisir les paramètres de la mesure de référence pour accélérer drastiquement la convergence vers 0 des coefficients. L'approximation de fonctions basées sur des polynômes ou des fonctions orthogonales est un sujet déjà beaucoup étudié dans la litterature, la plus-value de ce travail réside en grande partie dans les façons de choisir les paramètres de la mesure de référence pour obtenir une meilleure approximation. La densité de probabilité multivariée d'un vecteur aléatoire peut aussi être projetée sur une base de polynômes orthogonaux. Ces polynômes sont orthogonaux par rapport à une mesure de référence construite via le produit de mesures de probabilité univariées appartenant aux Familles Exponentielles Naturelles Quadratiques (FENQ). La représentation polynomiale de la densité de probabilité multivariée est justifiée par le Théorème 7. L'approximation polynomiale découle d'une troncature des séries infinies qui définissent la représentation polynomiale. La Fonction De Répartition Multivariée (f.d.r.m.) et la Fonction De Survie Multivariée (f.d.s.m.) sont obtenues par intégration de l'approximation polynomiale. L'application de la méthode d'aproximation aux distributions composées multivariées conduisent à opter pour des mesures de références gamma et des polynômes multivariés issus du produit de polynômes de Laguerre généralisés respectivement orthogonaux par rapport aux mesures de probabilité marginales choisies. La Proposition 9 propose une majoration de la densité de probabilité multivariée. Le choix des paramètres des mesures de référence gamma est effectué conformément au Corollaire 3 de façon à satisfaire la condition d'intégrabilité qui sous-tend la validité de la représentation polynomiale. Le choix des paramètres est affiné via l'étude de la fonction génératrice des coefficients qui s'exprime en fonction de la transformée Laplace du vecteur aléatoire considéré. La représentation polynomiale en dimension 2 est reliée aux lois de probabilité de Lancaster. Le vecteur aléatoire introduit dans la Définition 12 admet une distribution particulière sur $\mathbb{R}_+^2$. La Proposition 10 établit l'appartenance de cette distribution aux lois de Lancaster de type gamma-gamma. Une illustration numérique des performances de la méthode d'approximation polynomiale sur l'approximation d'une distribution Poisson composée est proposée et une étude comparative avec les autres méthodes numériques évoquées dans ce travail est conduite. En plus de l'approximation de la densité de probabilité et de la f.d.s., une approximation polynomiale de la prime *stop-loss* usuelle est étudiée. Le Chapitre 2 s'achève sur la description de l'estimateur de la densité de probabilité basé sur la formule d'approximation polynomiale. Une discussion autour de son application concrète à l'estimation des distributions composées en fonction des données disponibles conclut le Chapitre 2.

Le Chapitre 3 présente l'application de la méthode d'approximation via la représentation polynomiale à l'évaluation de la probabilité de ruine ultime dans le modèle de ruine de Poisson composé. Une comparaison avec d'autres méthodes numériques d'approximation est effectuée. Le Chapitre 3 est un article scientifique en révision, **?**.

Le Chapitre 4 présente l'application de la méthode d'approximation via la représentation polynomiale à l'évaluation des probabilités associées à une distribution composée bivariée. L'usage de telles distributions en actuariat et plus particulièrement en réassurance est justifié. Une comparaison avec d'autres méthodes numériques d'approximation est également effectuée. Le Chapitre 4 est un article scientifique soumis, **?**.

Le Chapitre 5 décrit la procédure d'agrégation des portefeuilles de contrat d'assurance vie de type épargne. Il s'agit d'une procédure en deux étapes, comprenant une classification des contrats en portefeuille via des algorithmes de classification classiques en analyse de données, suivie de la détermination d'un contrat représentatif, appelé *model point*, pour chacune des classes. Le portefeuille de contrats agrégé est utilisé comme input du modèle interne de projection des *cash-flows*. Le Chapitre 5 est un article scientifique accepté, **?**.

Mes Publications

GOFFARD, P.-O. et X. GUERRAULT. 2015, «Is it optimal to group policyholders by age, gender, and seniority for BEL computations based on model points ?», *European Actuarial Journal*, p. 1–16.

GOFFARD, P.-O., S. LOISEL et D. POMMERET. 2015a, «Polynomial approximations for bivariate aggregate claim amount probability distributions», *Working Paper*.

GOFFARD, P.-O., S. LOISEL et D. POMMERET. 2015b, «A polynomial expansion to approximate the ultimate ruin probability in the compound Poisson ruin model», *Working Paper*.