REPUBLIQUE DU CAMEROUN
Paix – Travail – Patrie
****

UNIVERSITE DE YAOUNDE I
ECOLE NORMALE SUPERIEURE
DEPARTEMENT DE Mathematiques
****

REPUBLIC OF CAMEROUN
Peace – Work – Fatherland
****

UNIVERSITY OF YAOUNDE I
HIGHER TEACHER TRAINING COLLEGE
DEPARTMENT OF Mathematics
****

Approche de calcul de la charge totale sinistres par le modele individuel

Mémoire de D.I.P.E.S II de Mathematiques

Par :

KOTSAP TCHOFFO Roussel
Licencie en Mathematiques appliquees

Sous la direction
FOTSO Simeon
Charge de cours

Année Académique
2015-2016

♠ Dédicace ♠

A

mes parents

♠ Remerciements ♠

Tout d'abord, je tiens à remercier spécialement Dr. Siméon FOTSO, Chargé de Cours à l'École Normale Supérieure de Yaoundé. En tant que Directeur de mémoire, pour son aide et sa disponibilité sans faille qui m'ont permis de surmonter les difficultés rencontrées.

Mes remerciements vont à l'endroit de mon tuteur, M. NOLACK PATRICE pour m'avoir accueilli sous son toit comme un fils et pour tous ses conseils.

Je remercie tous mes amies de Douala, pour leurs encouragements et apports multiformes durant mon séjour à Yaoundé.

Je remercie aussi mes ainés académiques de la 54^e promotion, plus précisément : TOUYEM Hilaire, pour son aide et encouragement dans l'élaboration de ce travail.

Je remercie tout mes camarades de classe (55^e promotion MATHÉMATIQUES V) pour leurs aides, conseils et assistances dans les moments difficiles que j'ai traversé durant ma formation.

J'adresse particulièrement mes plus sincères remerciements à mon Père M. KOTSAP, pour ces encouragements et conseils, sans oublier mes frères et sœurs, et à mes deux grandes familles : paternelle et maternelle, pour leurs soutiens morales et financiers.

♠ Table des matières ♠

♠ Résumé ♠

Dans ce mémoire, nous présenterons quelques principales lois de probabilité et démarches qu'un actuaire peut utiliser dans le cadre de la modélisation de la charge totale sinistre sur un portefeuille ayant plusieurs polices d'assurance. Nous utilisons le logiciel R et le logiciel Matlab pour mettre en pratique la méthode de convolutions et les méthodes d'approximations à travers quelques exemples.

Mots clés : modèle individuel, modèle collectif, charge totale sinistres, portefeuille d'assurance, police d'assurance.

♠ Abstract ♠

In this paper, we present some laws of probability and approaches that an actuary may use in the context of modeling aggregate loss of a portfolio with several insurance policies. We use the R software and Matlab software to practice the method of recursive convolution and methods of approximation through some examples.
keys words : Aggregate loss, portofolio of ansurance, ansurance police, collectif risk modèle, individual risk modèle.

♠ Abréviations et notifications ♠

1. Les abréviations sont essentiellement utilisées dans les chapitres un et deux,
   c'est l'exemple de : **v.a** qui signifie "variable aléatoire", **v.a.r** qui signifie "variable aléatoire réelle", **i.e** qui signifie "c'est à dire".
   les notations fgp et fgm, désignent : "la fonction génératrice de probabilité" et "la fonction génératrice des moments" respectivement.
2. Nous ne faisons pas de distinction entre les termes : loi de probabilité et distribution de probabilité. Ainsi nous écrivons souvent : "les lois qui modélisent ...." ou "les distributions qui modélisent...". De même la notion : "charge sinistres totale" qui s'écrit aussi "charge totale sinistres"
3. Nous utilisons la forme calligraphique pour nommer les lois de probabilité,
   par exemple : " $\mathcal{BN}$ pour la loi binomiale négative, $\mathcal{P}$ pour la loi de Poisson, $\mathcal{E}$ pour la loi Exponentielle.

♠ Table des figures ♠

♠ Liste des tableaux ♠

APPROCHE DE CALCUL DE LA CHARGE TOTALE SINISTRES PAR LE MODÈLE INDIVIDUEL

Mémoire de DI.P.E.S II de mathématiques

De

KOTSAP TCHOFFO Roussel

Matricule : **06A0055**

Licencié en Mathématiques appliquées

Sous la direction de :

Dr Siméon FOTSO

Chargé de Cours

Ecole Normale Supérieure, Université de Yaoundé I

Année Académique 2015-2016

♠ Introduction Générale ♠

Comme toute autre entreprise, les compagnies d'assurance cherchent à réaliser des profits. Puisque les réclamations constituent leur source de revenu le plus important, un assureur doit, pour ce faire, déterminer avec autant de précision que possible la distribution de la somme de ces réclamations. En théorie du risque, on l'appelle la distribution du montant total des sinistres. Une fois connue, cette distribution permet de fixer le niveau des primes en répartissant le risque entre diverses entités (assuré, assureur et réassureur) afin de minimiser pour chacun l'impact des pertes associées aux sinistres éventuels.

Une compagnie d'assurance vend des polices d'assurances (accords de volontés destinés à créer des rapports obligatoires entre des parties) qui sont en générale de deux types : Les polices d'assurance vie, ayant pour objet la vie de l'assuré et où la survenance du risque est plus certaine. Les polices d'assurance non-vie, qui regroupe plutôt les opérations d'assurance n'ayant pas pour objet la vie de l'assuré, c'est à dire les assurances des choses ou biens, les assurances de responsabilité ou de dettes, et les assurances de personnes. En termes techniques, on parle d'assurances Incendie, Accident et Risques Divers **(IARD)**. Ici la survenance du risque est probable, avec une probabilité comprise entre 0 et 1.

Cependant, nous parlerons beaucoup plus de l'assurance non-vie dans notre travail.En effet si on dispose d'un portefeuille de plusieurs polices d'assurance ou contrats dans lequel le i^{ime} sinistre, se produisant (ou étant déclaré), a un coût X_i pour la compagnie d'assurance, comment l'assureur peut-il modéliser le montant total des sinistres imputables à tous les assurés ?

Il est question de modéliser la charge totale sinistres sur un portefeuille de plusieurs polices d'assurances. Pour résoudre ce problème, on peut procéder de deux façons :
Le modèle collectif, qui considère la distribution qui modélise le nombre de sinistre, les distributions qui modélisent le montant d'un sinistre et S_N la charge sinistre total qui est s'exprime en fonction .

Le modèle individuel qui vise à représenter le montant total des sinistres à payer par la compagnie d'assurance sur une période donnée en sommant assuré par assuré, les montants des sinistres subis par chaque individu sur cette période.

Tout en insistant sur le modèle individuel, notre travail se subdivise en trois (03) chapitres : Le premier chapitre porte sur le rappel de quelques outils de probabilité, notamment les principales lois discrètes permettant de modéliser le nombre de sinistres et sur les lois continues qui modélisent le montant d'un sinistre.

Dans le deuxième chapitre, nous présentons les différentes méthodes de modélisation de la charge totale sinistre dans le modèle individuel.

la méthode de convolution, qui est une opération qui calcule la distribution de la somme de deux variables aléatoires. Pour un portefeuille d'assurance.

les méthodes de d'approximations, qui requièrent uniquement de connaître les premiers moments de la distribution de la charge totale sinistres.

La méthode récursive de De Pril quant à elle, exigent de connaitre la distribution de la fréquence et de la sévérité des réclamations. Dans la mesure où ces dernières sont exactes, cette méthode récursive produit des résultats exacts.

Enfin, au troisième chapitre, nous utiliserons les codes R pour ajuster la modélisation de la charge totale sinistre sur un portefeuille de plusieurs polices d'assurance.

Chapitre Un

LA CHARGE TOTALE SINISTRES

La charge sinistre totale est le montant total de tous les sinistres d'un portefeuille sur une période donnée (généralement un an). Elle est évaluée par les assureurs lorsqu'ils font le bilan de fonctionnement de leur structure d'assurance. Dans ce chapitre nous faisons un rappel sur les distributions qui modélisent le nombre de sinistres (*claim frequency*), les distributions qui modélisent le montant d'un sinistre (*claim severity*) et enfin nous présenterons les deux modèles qui permettent de calculer la charge sinistre totale.

1.1 Les outils de probabilité et statistique utilisés en actuariat

Dans cette section, nous passerons en revue quelques outils de probabilité et statistique qui sont fortement utilisés par les actuaires.

Fonction densité de probabilité et fonction de répartition

Définition 1.1. *Soit une variable aléatoire (v.a.) définie sur un espace de probabilité* $(\Omega, \mathcal{A}, P)$*. On appelle* ***fonction de répartition de X*** *et notée* F_X*, la fonction définie par :*

$$F_X(x) = P(X \leq x). \tag{1.1}$$

Rapppelons que : $X(\Omega) = \{X(w)/w \in \Omega\}$, est le support de X.

Remarque 1.1.1. -Si X est une v.a. discrète, alors sa fonction de répartition est donnée par :

$$\forall x \in [x_k; x_{k+1}[, F_X(x) = \sum_{x \in X(\Omega), x_i \leq x_k} P(X = x_i)$$

.

Définition 1.2. *Soit* X *une v.a.continue,* F_X *une fonction de répartition. Si* F_X *est différentiable, alors la fonction densité de probabilité notée* f_X *est définit par :*

$$f_X = \frac{dF_X}{d_x} \tag{1.2}$$

Propriété 1.1. *Soit $f(x)$, la fonction de densité de probabilité de la v.a. continue X. De manière formelle, $f(x)$ est une densité de probabilité si et seulement si elle satisfait les deux conditions suivantes :*

- $f(x) \geq 0$*, pour toutes les valeurs de* $x \in \mathbb{R}$.
- $\int_{\mathbb{R}} f(x)dx = 1.$

Remarque 1.1.2. -Si X est une v.a. continue, alors elle admet pour fonction de répartition F_X est donnée par :

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du.$$

où $f_X(.)$ est la fonction densité de probabilité.

-Si X est une v.a. mixte, alors sa fonction de répartition est donnée par :

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du + \sum_{x \in X(\Omega), x_i \leq x_k} P(X = x_i)$$

l'espérance d'une variable aléatoire

Définition 1.3. *Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité et X une variable aléatoire positive.L'espérance de la mesure de probabilité P notée : $E(X)$, est l'intégrale de X par rapport à P et est donnée par :*

$$E(X) = \int_{\Omega} X(x)dP(x) = \int XdP. \tag{1.3}$$

Remarque 1.1.3. -si X est une v.a. discrète, alors l'espérance est donnée par :

$$E(X) = \sum_{x \in X(\Omega)} xP(X = x).$$

-Si X est une v.a. continue, alors l'espérance est donnée par :

$$E(X) = \int_{\mathbb{R}} xf_X(x)dx.$$

-Si X est une v.a. mixte, alors l'espérance est donnée par :

$$E(x) = \int_{\mathbb{R}} xf_X(x)dx + \sum_{x \in X(\Omega)} xP(X = x).$$

Les moments d'une variable aléatoire

Définition 1.4. *Le moment simple d'ordre $k \geq 1$ d'une v.a.positive X, noté m_k est définie par :*

$$m_k = E(X^k). \tag{1.4}$$

Remarque 1.1.4. -si X est une v.a. discrète, alors m_k est donnée par :

$$m_k = \sum_{x \in X(\Omega)} x^k P(X = x).$$

-Si X est une v.a. continue, alors m_k est donnée par :

$$m_k = \int_{\mathbb{R}} x^k f_X(x) dx.$$

-Si X est une v.a. mixte, alors m_k est donnée par :

$$m_k = \int_{\mathbb{R}} x^k f_X(x) dx + \sum_{x \in X(\Omega)} x^k P(X = x).$$

Définition 1.5. *Pour une v.a. à valeur entières, le* ***moment factoriel d'ordre k*** *est défini par :*

$$\mu_{(k)} = E[X(X-1)...(X-k+1)]. \tag{1.5}$$

Les moments simples s'expriment en fonction des moments factoriels, ainsi nous avons :

$$\begin{aligned} m_1 &= \mu_{(1)} \quad \text{qui est la moyenne} \\ m_2 &= \mu_{(2)} + \mu_{(1)} \\ m_3 &= \mu_{(3)} + 3\mu_{(2)} + \mu_{(1)} \\ m_4 &= \mu_{(4)} + 6\mu_{(3)} + 7\mu_{(2)} + \mu_{(1)}. \end{aligned} \tag{1.6}$$

Définition 1.6. *Le* ***moment centré d'ordre k*** *est défini par :*

$$\mu_k = E[(X - E(X))^k]. \tag{1.7}$$

Nous pouvons ainsi déduire les moments centrés d'ordre k, des moments simples par les expressions suivantes :

$$\begin{aligned} \mu_2 &= m_2 - m_1^2 \quad \text{qui est la variance;} \\ \mu_3 &= m_3 - 3m_1 m_2 + m_1^2 \\ \mu_4 &= m_4 - m_1 m_4 + 6m_1^2 m_2 - 3m_1^4. \end{aligned} \tag{1.8}$$

Où les paramètres $\mu_1, \mu_2, et \mu_3$ sont utilisés pour caractériser la forme d'une distribution.

Les Fonctions Génératrices

Définition 1.7. *Soit X une variable aléatoire. On appelle* ***fonction génératrice des moments de X*** *notée : fgm, la fonction M_X définie par :*

$$\forall t \in \mathbb{R}, M_X(t) = E(e^{tX}). \tag{1.9}$$

Définition 1.8. *Soit X une variable aléatoire à valeur entière positive, on appelle* ***fonction génératrice de probabilité de*** *X notée : fgp, la fonction P_X définie par :*

$$\forall t \in \mathbb{R}, P_X(t) = E(t^X). \tag{1.10}$$

En posant $s = e^t$, nous voyons que $M_X(t) = P_X(e^t) \quad et \quad P_X(t) = M_X(lnt)$, puisque :

$$M_X(t) = E[(e^{tX})] = E[(e^t)^X] = E[(s^X)] = P_X(s) = P_X(e^t)$$

puis

$$M_X(lnt) = E[(e^{lnt))^X}] = E[(t^X)] = P_X(t)$$

.

Remarque 1.1.5.

1. Si l'intégrale ne converge pas la fgm n'existe pas ; ce qui explique pourquoi certaines v.a. n'ont pas de fgm.
2. Si la fgm de X existe dans un intervalle ouvert au voisinage de 0, alors les moments de X existent et peuvent être obtenus par dérivations successives de la fgm de X en t. Le résultat est alors obtenu en prenant $t = 0$. En effet, lorsqu'on dérive à l'ordre k la fonction génératrice on obtient :

$$\begin{aligned} M_X(t) &= \frac{d^k M_X(t)}{dt^k} \\ &= \frac{d^k}{dt^k} E(e^{tX}) \\ &= E\left[\frac{d^k}{dt^k} e^{tX}\right] car, \frac{d^k}{dt^k} \quad \text{est une constante pour l'intégrale} \quad E(e^{tX}) \end{aligned}$$

d'où :

$$M_X^{(k)}(t) = E(X^k e^{tX}). \tag{1.11}$$

En prenant $t = 0$, on obtient :

$$M_X^{(k)}(0) = E(X^k). \tag{1.12}$$

3. Si $X_1, X_2, X_3, ...X_n$ sont des v.a. indépendantes de fgm respectives $M_1(.), M_2(.), M_3(.), ..., M_n(.)$, et $X = X_1 + X_2 + ... + X_n$, alors la fgm de X est :

$$\begin{aligned} M_X(t) &= E(e^{tX}) \\ &= E(e^{tX_1+tX_2+...+tX_n}) \\ &= E\left(\prod_{i=1}^{n} e^{tX_i}\right) \\ &= \prod_{i=1}^{n} E(e^{tX_i}) \\ &= \prod_{i=1}^{n} M_i(t) \end{aligned}$$

   Donc

$$M_X(t) = \prod_{i=1}^{n} M_i(t). \tag{1.13}$$

4. Si $X_1, X_2, X_3, ...X_n$ sont des v.a.indépendantes et identiquement distribuées (**i.i.d**) de fgm M(t). c'est à dire : $M_i(t) = M(t), i = 1, 2, 3, ..., n$, alors,

$$M_{X_1+X_2+...+X_n}(t) = [M(t)]^n. \tag{1.14}$$

   Il s'agit ici de présenter les propriétés de la fgm :

5. Si la fgm d'une v.a.X existe pour tout t dans un intervalle ouvert voisinage de 0, alors tous les moments de X existent.

6. Si les fgm de deux v.a.X_1 et X_2 sont identiques pour tout t dans un voisinage ouvert de 0, alors les distributions de X_1 et X_2 sont identiques. De même, si deux distributions sont identiques, elles doivent avoir le même fgm.

Soit la fgp d'une v.a. X connue, il est possible de déterminer sa loi de probabilité en utilisant l'expression de P_X :

$$\forall t \in \mathbb{R} \quad P_X(t) = \sum_{x=0}^{\infty} t^x f_X(x). \tag{1.15}$$

Ainsi la k^{ieme} dérivée de $P_X(t)$ est :

$$P_X^{(k)}(t) = \frac{d^k}{dt^k}\left(\sum_{x=0}^{\infty} t^x f_X(x)\right) = \sum_{x=0}^{\infty} x(x-1)...(x-k+1)t^{x-k} f_X(x). \tag{1.16}$$

Si on calcule $P_X^{(k)}(t)$ en $t = 0$, tous les termes de la sommation s'annulent sauf pour $x = k$, qui est $k!f_X(k)$.D'où

$$P_X^{(k)}(0) = k(k-1)...2.1 f_X(k), \tag{1.17}$$

or en écrivant $f_X(k) = P(X = k) = p_k$, on obtient :

$$P_X^{(k)}(0) = k!p_k. \tag{1.18}$$

Donc étant donnée la fgp d'une v.a. discrète X, on peut obtenir sa loi de probabilité par la formule :

$$p_k = \frac{P_X^{(k)}(0)}{k!}, \quad k = 1, 2, ... \tag{1.19}$$

connaissant la fgp d'une v.a. discrète X, sa probabilité peut se calculer à partir de l'équation (1.19) i.e $p_k = \frac{P_X^{(k)}(0)}{k!}$. Ainsi, la fgp et la fgm sont des fonctions très utiles pour caractériser une distribution.

Tout ces outils Mathématiques cités ci-dessus sont importants pour un actuaire, c'est grâce à ces outils qu'il modélise le montant d'un sinistre et le nombre de sinistre.

1.2 Présentation de la charge sinistre totale en Assurance

Sa d Considérons un portefeuille de plusieurs polices d'assurance non-vie, la charge sinistre totale est la somme des montants de tout les sinistres enregistrés dans ce porte feuille pendant une période donnée.étermination nécessite de connaitre nombre de sinistre du portefeuille et le montant d'un sinistre. Pour cela nous allons dans cette partie : définir la fonction de queue, présenter les distributions qui permettent de modéliser le nombre de sinistre, présenter les distributions qui permettent de modéliser le montant d'un sinistre. Enfin présenter les deux modèles d'approches du calcul de la charge totale sinistres. Ces deux modèles sont : le modèle collectif et le modèle individuel qui, fait l'objet de notre travail.

1.2.1 La fonction de Queue

En plus de la fonction de répartition, son complémentaire est appelé **fonction queue** en actuariat non-vie (en biostatistique et en assurance vie, cette même fonction est appelée fonction de survie lorsque X représente la durée de vie d'un individu). Notée $\overline{F}_X$, celle-ci est définie comme suit.

Définition 1.9. *Étant donnée une variable aléatoire X, la fonction de queue associée vaut :*

$$\overline{F}_X(x) = 1 - F_X(x) = Pr[X > x], x \in \mathbb{R}; \tag{1.20}$$

$\overline{F}_X(x)$*, représente donc la probabilité que* X *prenne une valeur supérieure à* x*.*

On peut voir $\overline{F}_X(x)$ comme la prime à payer pour recevoir une somme si X excède x. Clairement, $\overline{F}_X$ est décroissante puisque l'évènement $X > x$ est plus probable que $X > x'$ lorsque $x < x'$, i.e. $\{X > x'\} \subseteq \{X > x\}$. Lorsque le support de la fonction de répartition F du montant de sinistre est $\mathbb{R}^+$, on mesure généralement le risque associé à telle ou telle fonction de répartition par l'épaisseur des queues de distribution (i.e. par la masse de probabilité répartie sur les régions ((c, $+\infty$), pour de grandes valeurs de c). Ainsi, on parlera de queue épaisse (ou lourde) quand $\overline{F}_X$ tend lentement vers 0 lorsque $x \longrightarrow +\infty$ et de queue légère quand $\overline{F}_X$ tend rapidement vers 0 lorsque $x \longrightarrow +\infty$.

Cette fonction de queue impose un ordre de préférence sur le choix de l'utilisation des distributions par l'actuaire.

Le tableau suivant (Charles Suquet, 12 Avril 2007) présente certaines lois ou distributions avec leurs poids, et leur fonction de queue.

$Poids$	Loi	$Queue\overline{F}(x)$
Très léger	$Weib(\tau, \alpha), \tau > 1$	$exp\left(-\left(\frac{x}{\alpha}\right)\right)$
Léger	$Gam(\nu, \beta)$	$\int_x^{+\infty} \frac{\beta^\nu}{\Gamma(\nu)} t^{\nu-1} e^{-\beta} dt$
Intermédiaire	$Lognormale$	$P(e^Y > x), Y \sim \mathcal{N}(\mu, \sigma)$
	$Weib(\tau, \alpha), \tau < 1$	
Lourd	$Par(\theta, \alpha)\alpha > 1$	$\left(\frac{\theta}{\theta + x}\right)^\alpha$
Très lourd	$Par(\theta, \alpha)\alpha \leq 1$	

TABLE 1.1 – Poids et fonction de queue de quelques lois

1.2.2 Modélisation du nombre de sinistres

La modélisation du nombre ou de la fréquence de sinistre se fait par des distributions discrètes et comme nous parlons de nombre, ces valeurs ne peuvent être négatives. Les distributions les plus utilisées sont entre autre : la distribution POISSON, la distribution BINOMIALE NEGATIVE et la distribution BINOMIALE.

Distribution de Poisson

Sous certaines hypothèses, on montre que le processus du nombre de sinistres est un processus de poisson. La loi de Poisson peut se construire à partir d'une hypothèse unique : la probabilité de survenance d'un sinistre dans le futur proche est proportionnelle à la durée envisagée et ne dépend pas des observations passées. Elle présente en outre l'avantage de ne nécessiter qu'un seul paramètre λ. La loi de poisson est donc d'utilisation naturelle en assurance non-vie, elle est fondamentale dans la modélisation du nombre de sinistre.

Définition 1.10. *Une variable aléatoire discrète N est dite de Poisson de paramètre $\lambda > 0$, notée $P(\lambda)$, si sa densité de probabilité s'écrit :*

$$f_N(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad pour k = 0, 1, 2, ... \tag{1.21}$$

L'espérance et la variance de N sont : $E(N) = Var(N) = \lambda$.

La fgm de N est donnée par :

$$M_N(t) = exp[\lambda(e^t - 1)] \tag{1.22}$$

et sa fgp est :

$$P_N(t) = exp[\lambda(t - 1)]. \tag{1.23}$$

La relation de récurrence entre les probabilités de la distribution de Poisson est :

$$f_N(k) = (\frac{\lambda}{k})f_N(k-1) \quad avec \quad f_N(0) = e^{-\lambda}. \tag{1.24}$$

Remarque 1.2.1. En supposant que le nombre total de sinistres de chaque assuré suit une loi de Poisson de même paramètre λ, alors le nombre total de sinistres suit une loi de Poison de paramètre Nλ (avec N représentant le nombre d'assurés.)

Précisons que : si le nombre d'assuré est important, alors le nombre total de sinistre peut être approché par une LOI NORMALE.

Nous pouvons dire que la loi de Poisson P(λ) est la loi fondamentale en assurance non-vie pour tous les avantages qu'elle présente :
-elle est à la base des distributions Poisson-mélange et de l'élaboration d'un système Bonus-Malus
-Grâce à son unique paramètre, elle est facilement compréhensible et estimable.

Notons cependant qu'elle est moins flexible en raison de son unique paramètre, elle présente une queue de distribution légère, et l'égalité $E(N) = V(N)$ peut être contraignante lors de l'utilisation de la distribution.

Distribution Binomiale

Définition 1.11. *On dira qu'une variable aléatoire N admet une distribution binomiale de paramètre n, p notée B(n,p) où $n \in \mathbb{N}$ et $p \in]0,1[$, si sa densité de probabilité est donnée par :*

$$f_N(k) = P(N = k) = C_n^k p^k (1-p)^{n-k}, \textit{pour} \quad k = 0, 1, 2, ... \tag{1.25}$$

L'espérance et la variance de N sont :

$$E(N) = np \quad \text{et} \quad Var(N) = np(1-p). \tag{1.26}$$

La fgm est :

$$M_N(t) = (pe^t + 1 - p)^n. \tag{1.27}$$

Sa fgp est :

$$P_N(t) = (pt + 1 - p)^n. \tag{1.28}$$

L'expression de la densité de cette loi est la probabilité d'obtenir k *succès* au cours de n épreuves indépendantes de Bernoulli de paramètre p. Il existe une relation récursive pour $f_N(k)$ qui facilite le calcul des probabilités.

Ainsi on a :

$$f_N(k) = \left[\frac{(n-k+1)p}{k(1-p)}\right] f_N(k-1), \quad pour \quad k = 0, 1, 2, ... \tag{1.29}$$

avec $f_N(0) = (1-p)^n$.
La loi de Binomiale apparait naturellement comme la distribution modélisant le nombre sinistre N.

Exemple 1.2.1. On suppose une période de 30 jours au cours de laquelle les observations montrent qu'on peut espérer en moyenne 10 sinistres immobiliers (cambriolages ou incendies).

Nous supposons que ce nombre reste constant sur une période donnée (ce qui un peu irréaliste dans la mesure où le nombre d'incendies peut être souvent nul durant une période ou bien il y a plus de cambriolage dans la nuit qu'en journée). Nous supposons qu'il y aura au maximum un sinistre par jour. Cette expérience est ainsi vu comme une suite de 30 épreuves indépendantes où chaque jour ayant un sinistre est pris comme *succès* et celui sans sinistre est un *échec*. On a la probabilité d'avoir un sinistre par jour qui est égale à : $\frac{1}{3}$ car nous voulons avoir un nombre de sinistre espéré égal à 10. Nous constatons qu'on peut modéliser le nombre de sinistres N par *B(30,* $\frac{1}{3}$*)*. Cependant, il peut arriver que nous soyons en situation où les sinistres soit beaucoup plus dans la nuit c'est à dire qu'on aura au plus une incendie ou un cambriolage chaque milieu de la nuit(demi-nuit). Dans ce cas, nous aurons une distribution Binomiale mais avec $n = 60$ et $p = \frac{1}{6}$, pour avoir le nombre moyen de sinistre égal à 10. Nous voyons qu'en gardant constant le nombre moyen de sinistres et en faisant varier le nombre d'épreuves, la loi binomiale apparait naturellement comme la distribution modélisant le nombre de sinistre N.

En faisant $np = \lambda$, la loi binomiale tend vers la loi de Poisson de paramètre λ lorsque n tend vers ∞. En effet, si la variable aléatoire N suit *B(n,p)* alors on a :

$$lim_{n\longrightarrow+\infty}P_k = lim_{n\longrightarrow+\infty}C_n^k(\frac{\lambda}{m})(1-\frac{\lambda}{m})^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, ...$$

qui est la loi de Poisson $P(\lambda)$. Le fait que $E(N) > Var(N)$ rend l'utilisation de cette formule difficile car le plus souvent c'est l'inégalité inverse qui est rencontrée dans les données empiriques en assurance non-vie.

Distribution Binomiale Négative

Définition 1.12. *Considérons une suite d'épreuve indépendante de Bernoulli où la probabilité de succès est p. Mais, au lieu de fixer un nombre de répétitions, on continue jusqu'à obtenir un succès. Si N est la variable aléatoire du nombre d'échecs, alors sa loi de probabilité est :*

$$P(N = k) = p(k) = (1-p)^k p, \quad k = 0, 1, 2, ... \tag{1.30}$$

On dit alors que la variable aléatoire N suit une loi géométrique de paramètre p, notée : G(p).

L'espérance et la variance de N sont respectivement $E(N) = \frac{1-p}{p}$ et $Var(N) = \frac{1-p}{p^2}$. Supposons à présent qu'on reprenne l'expérience précédente, sauf qu'au lieu de continuer jusqu'au premier succès, nous continuons jusqu'au i^{ieme}succès. Alors la probabilité du nombre d'échecs et maintenant donnée par :

$$P(N=k) = C_{i-1}^{k+i-1} p^i (1-p)^k, \quad k = 0, 1, 2.... \tag{1.31}$$

Une telle distribution est appelée Loi Binomiale Négative de paramètre i et p que l'on note *BN(i,p)*.

Si N suit cette loi, alors :
l'espérance et la variance de N sont respectivement,

$$E(N) = \frac{i(1-p)}{p} \quad et \quad Var(N) = \frac{i(1-p)}{p^2}. \tag{1.32}$$

La fgm de N est :

$$M_N(t) = \left[\frac{p}{1-(1-p)e^t}\right]^i. \tag{1.33}$$

et sa fgp est de la forme :

$$P_N(t) = \left[\frac{p}{1-(1-p)^t}\right]^i. \tag{1.34}$$

De même on montre que la formule récursive de N est :

$$P(N=k) = \left[\frac{(k+i-1)(1-p)}{k}\right] P(N=k-1). \tag{1.35}$$

avec la valeur initiale $P(N=0) = p^i$.

Remarque 1.2.2.

1. La distribution géométrique est un cas particulier de la distribution Binomiale Négative (pour i=1).
2. L'expresion $P(N=k) = p(k) = C_{i-1}^{k+i-1} p^i (1-p)^k$, $k = 0, 1, 2...$ est la probabilité d'avoir k-échec avant le i^{ieme}-succès dans une suite d'épreuve indépendante de BERNOULLI de paramètre p. Ainsi la loi $\mathcal{BN}(i,p)$ est juste la somme de i distributions géométriques de paramètre commun p.

La distribution binomiale négative est aussi largement utilisée pour modéliser le nombre de sinistres en assurance non vie.

Exemple 1.2.2. Considérons une population d'assurés dans laquelle les individus (ou divers groupes d'individus) produisent des sinistres suivant une loi de poisson de paramètre λ; on suppose que λ varie d'un assuré à l'autre (ou d'un groupe d'assurés à l'autre). Prenons le cas de l'assurance automobile où les conducteurs assurés sont classés en deux catégories : les bons et les mauvais. Admettons que le nombre de sinistres des bons conducteurs suit une loi de Poisson $(P(\lambda_1))$ pendant que celui des mauvais conducteurs suit une loi de Poisson $(P(\lambda_2))$. Admettons de plus que 60% des conducteurs sont bons et le reste mauvais, mais que l'assureur n'a aucun moyen de faire la différence entre le deux classes. La modélisation du nombre de sinistre N dans ce cas se fera dans par un mélange de deux distributions de Poisson. Si on prend la distribution gamma pour modéliser le comportement des conducteurs, il en ressort que le nombre de sinistre par individu ou par groupes d'individus (qui est un mélange de distributions de Poisson dont les paramètres suivent une loi gamma), est distribué selon une loi binomiale négative. Donc la loi binomiale négative est un cas particulier de la famille des lois Poisson-mélanges (dont nous nous attarderons dessus dans le chapitre suivant).

Dans la distribution Binomiale Négative $\mathcal{BN}(i,p)$, $V(N) > E(N)$, ce qui la rend facilement utilisable dans la pratique. Elle est donc plus flexible que la loi de Poisson grâce à ses deux paramètres. Enfin, elle permet de modéliser naturellement la distribution de la fréquence de sinistre d'un contrat d'une classe hétérogène de risque.

Distribution de type $(a, b, 0)$

Les distributions binomiales, Poisson, et binomiale négative appartiennent à une famille de distributions discrètes positives appelée *Classe(a,b,0)* dans la littérature actuarielle.

Définition 1.13. *Une variable aléatoire N est dite de classe (a,b,0) si sa loi satisfait la relation de récurrence suivante :*

$$f_N(k) = (a + \frac{b}{k}) f_N(k-1), \quad k = 0, 1, 2, ... \tag{1.36}$$

où a et b sont constants, avec $f_N(0)$ données.

Ainsi considérons la loi binomiale où la relation récursive est donnée par :

$$f_X(x) = \left[-\frac{p}{1-p} + \frac{p(n+1)}{(1-p)x} \right] f_X(x-1) \tag{1.37}$$

En posant $a = -\dfrac{p}{1-p} \quad et \quad b = \dfrac{p(n+1)}{(1-p)}$, nous voyons aisément que la fonction de probabilité de la loi Binomiale satisfait l'équation (1.37). De la même manière, nous vérifions que les fonctions de probabilités des lois Binomiales Négatives et de Poisson vérifient aussi l'équation (1.37).

Nous résumons cela dans le tableau (Yiu-Kuen Tse, 2009) suivant :

$Distributions$	a	b	$f_X(0)$
$Binomiale$: $\mathcal{BN}(n,p)$	$-\dfrac{p}{1-p}$	$\dfrac{p(n+1)}{(1-p)}$	$(1-p)^n$
$Gometrique$: $\mathcal{GM}(p)$	$1-p$	0	p
$BinamialeN\acute{e}gative$: $\mathcal{BN}(r,p)$	$1-p$	$(r-1)(1-p)$	p^r
$Poisson$: $\mathcal{PN}(\lambda)$	0	λ	$e^{-\lambda}$

TABLE 1.2 – Distributions appartenant à la classe $(a, b, 0)$

On définit la classe $(a, b, 1)$ avec la même relation de récurrence en commençant plutôt à $k = 2$, la valeur initiale $f_N(1)$ étant connue.

1.2.3 Modélisation du montant d'un sinistre

Nous présenterons dans cette partie les principales lois utilisées pour modéliser le montant d'un sinistre.

Loi exponentielle

Définition 1.14. *Une variable aléatoire X suit une distribution exponentielle de paramètre $\lambda > 0$, notée $\mathcal{E}(\lambda)$, si sa densité de probabilité est :*

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & pour \qquad x \geq 0. \\ O, & \qquad sinon \end{cases} \tag{1.38}$$

L'espérance et la variance de X sont :

$$E(X) = \frac{1}{\lambda} \quad et \quad Var(X) = \frac{1}{\lambda^2}. \tag{1.39}$$

La fgm de X est :

$$M_X(t) = \frac{\lambda}{\lambda - t}. \tag{1.40}$$

Cette loi s'utilise très souvent dans les modèles actuariels. Sa simplicité l'impose dès qu'il s'agit de mener à terme des calculs actuariels complexes (théorie du risque, par exemple). Cependant On peut indiquer que sa densité est décroissante, elle a une queue très légère, ce qui n'est pas acceptable dans quelques applications pratiques.

Certaines distributions modélisent mal le nombre de sinistres d'un portefeuille en raison de l'hétérogénéité des assurés. Ainsi pour tenir compte de cette hétérogénéité du portefeuille, on utilise certaines lois ou Distributions appelées **lois (distributions) mélanges** qui sont crées à partir des autres lois (distributions) déjà connues.

Distribution Normale

Définition 1.15. *Une v.a. X suit une distribution Normale de paramètre $\mu \quad et \quad \sigma^2$ avec ($\mu \in \mathbb{R} \quad et \quad \sigma > 0$), notée $\mathcal{N}(\mu, \sigma^2)$, si sa densité de probabilité est de la forme :*

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \tag{1.41}$$

La fonction de répartition de X est de la forme :

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad avec \quad \phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\dfrac{-t^2}{2}} dt. \tag{1.42}$$

(Φ étant la fonction de répartition de la loi normale centrée réduite).
L'espérance et la variance sont données par les formules :

$$E(X) = \mu \quad et \quad V(X) = \sigma^2. \tag{1.43}$$

Cette formule sera utilisée pour la modélisation de la charge sinistre, mais pas pour la modélisation du montant individuel de sinistre.

Distribution Gamma

Elle a pour paramètre α (pour la forme), β (pour l'échelle) tous strictement positifs. Elle se note : $\Gamma(\alpha, \beta)$. Sa densité de probabilité est donnée par la formule suivante :

$$f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{(\alpha-1)} e^{-\dfrac{x}{\beta}}, \quad pour \quad x \geq 0. \tag{1.44}$$

La fonction Γ est appelée fonction Gamma, définie par :

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy. \tag{1.45}$$

Remarque 1.2.3. l'intégrale dans cette formule converge ce qui veut dire la fonction
Γ existe pour $\alpha > 0$.
Pour $\alpha > 1$, $\Gamma(\alpha)$ satisfait la récursion suivante :

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1). \tag{1.46}$$

Cette relation s'obtient après une intégration par partie de $\Gamma(\alpha)$. De plus, si α est un entier naturel, on obtient par induction sur la formule précédente, en faisant le produit et en simplifiant membre par membre, $\Gamma(\alpha) = 1.2.3...(\alpha - 1) = (\alpha - 1)!$. Ainsi en posant $y = \beta x \quad (avec \quad \beta > 0)$, la relation (1.56) donne :

$$\Gamma(\alpha) = \int_0^\infty (\beta x)^{\alpha-1} e^{-\beta x} \beta dx = \int_0^\infty \beta x^{\alpha-1} e^{-\beta x} dx. \tag{1.47}$$

d'où

$$\int_0^\infty x^{\alpha-1} e^{-\beta x} dx = \frac{\Gamma(\alpha)}{\beta^\alpha}. \tag{1.48}$$

L'espérance et la variance de la v.a. sont données par les formules suivantes :

$$E(X) = \alpha\beta \quad et \quad Var(X) = \alpha\beta^2. \tag{1.49}$$

La fgm de X est :

$$M_X(t) = \frac{1}{(\alpha - \beta t)^\alpha}, \quad pour \quad t < \frac{1}{\beta} \quad et \quad \alpha > 1. \tag{1.50}$$

Distribution Log-normale

Définition 1.16. *La variable aléatoire X, montant d'un sinistre, suit une loi Lognormale, notée :$\mathcal{LN}(\mu, \sigma^2)$, si et seulement si la variable aléatoire réelle lnX suit la loi Normale $\mathcal{N}(\mu, \sigma^2)$. Sa densité de probabilité est :*

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma x} exp[-\frac{1}{2\sigma^2}(lnx - \mu)^2]. \tag{1.51}$$

et sa fonction de répartition est :

$$F_x = \Phi(\frac{lnx - \mu}{\sigma}). \tag{1.52}$$

l'espérance et la variance sont :

$$E(X) = exp(\mu + \frac{\sigma^2}{2}) \quad et \quad V(X) = [exp(2\mu + \sigma^2)][exp(\sigma^2 - 1)]. \tag{1.53}$$

On utilise cette distribution pour la modélisation des sinistres graves sous la forme d'un mélange de deux lois. Elle est la plus utilisée en réassurance et en assurance non-vie.

Distribution de Weibull

Cette distribution a pour paramètre α (pour la forme) et λ(pour l'échelle) tous strictement positifs et est notée $\mathcal{W}(\alpha, \lambda)$. Une v.a.$X$ qui suit une distribution de Weibull admet pour densité de probabilité :

$$f_X(x) = (\frac{\alpha}{\lambda})(\frac{x}{\lambda})^{\alpha-1} exp\left[-(\frac{x}{\lambda})^\alpha\right], x \geq 0. \tag{1.54}$$

La fonction de répartition de X est :

$$F_X = 1 - exp\left[-(\frac{x}{\lambda})^{\alpha}\right], \quad x \geq 0. \tag{1.55}$$

L'espérance et la variance sont données par la formule :

$$E(X) = \lambda\Gamma(\frac{1}{\alpha} + 1) \quad et \quad V(X) = \lambda^2[\Gamma(\frac{2}{\alpha} + 1)] - [E(X)]^2. \tag{1.56}$$

Notons que la distribution de Weibull avec λ=1 est appelée **distribution standard de Weibull** ayant pour densité de probabilité :

$$f_X(x) = \alpha x^{\alpha-1} exp(-x^{\alpha}). \tag{1.57}$$

Distribution de Pareto

Définition 1.17. *Une v.a. X suit une distribution de Pareto de paramètre a et α, tous positifs, si sa densité de probabilité est :*

$$f_X(x) = \frac{a\alpha^a}{(x+\alpha)^{a+1}}, \quad avec \quad x \geq 0 \tag{1.58}$$

Elle est notée : $\mathcal{P}(a, \alpha)$

La fonction de répartition de X est :

$$F_X = 1 - \left(\frac{\alpha}{x+1}\right)^a, \quad x \geq 0. \tag{1.59}$$

La fonction de survie est :

$$S_X(x) = 1 - F_X(x) = \left(\frac{\alpha}{x+1}\right)^a. \tag{1.60}$$

L'espérance et la variance de X sont :

$$E(X) = \frac{\alpha a}{\alpha - 1} \quad si \quad \alpha > 1 \quad et \quad Var(X) = \frac{\alpha a^2}{(\alpha-1)^2(\alpha-2)} \quad si \quad \alpha > 2. \tag{1.61}$$

Cette distribution est la loi privilégiée en réassurance en raison de sa queue de distribution lourde. Cependant l'existence des moments n'est pas certaine.

Nous avons ainsi présenté : les distributions qui modélisent le nombre de sinistre (loi de Poisson, loi Binomiale Négative, loi Binomiale) et les distributions qui modélisent le montant d'un sinistre (Log normale, Weibull, Normal, gamma, exponentielle, Pareto) qui, assemblées constituent la charge Sinistres totale d'un portefeuille d'assurance. Il existe deux principales méthodes d'approche du calcul de la charge sinistres totale : le modèle collectif et le modèle individuel.

1.2.4 Le modèle collectif

Le modèle collectif consiste à approcher le modèle individuel non plus en regardant si chaque police fait défaut ou pas, mais en comptabilisant un nombre aléatoires de montants de sinistres indépendantes et identiquement distribués (i.i.d). On définit ainsi la charge sinistre totale sur une période donnée (un an en général) dans le modèle collectif par la variable aléatoire positive :

$$S_N = \sum_{i=1}^{N} X_i. \tag{1.62}$$

Où N est une variable aléatoire à valeurs dans $\mathbb{N}$ représentant le nombre de sinistres sur une période donnée (un an en général), et pour $i \geq 1$, X_i est une variable aléatoire à valeurs dans $\mathbb{R}^+$ représentant le coût du i-ème sinistre, avec la convention selon laquelle la somme est nulle si $N = 0$. Les $(X_i)_{i\geq 1}$ sont supposés indépendantes et identiquement distribués, et indépendantes de N (indépendance fréquences-coûts). L'indépendance fréquences-coûts est valable si le portefeuille est homogène. Si le portefeuille ne l'est pas, alors on peut tenir compte de cette hétérogénéité par des lois mélanges et composées.

1.2.5 Le modèle individuel

Le modèle individuel revient à raisonner police par police. Notons par n le nombre de polices dans le portefeuille, la charge sinistres totale est donnée par :

$$S_n = \sum_{i=1}^{n} S_i. \tag{1.63}$$

où S_i désigne la charge totale des sinistres ayant atteint la police i sur la période considérée. Le montant de sinistre S_i engendré par la police i peut s'exprimer sous la forme

$$S_i = I_i Y_i \tag{1.64}$$

où I_i vaut 1 si la police i a été touchée par au moins un sinistre, 0 sinon; et où Y_i représente la somme des montants de sinistres qui ont touché la police i si au moins un sinistre est survenu. Les variables aléatoires $I_1, I_2, ..., I_n, Y_1, Y_2, ..., Y_n$ sont supposées mutuellement indépendantes. Ce modèle fera l'objet du chapitre suivant.

★ ★ Chapitre Deux ★ ★

DISTRIBUTION DE LA CHARGE TOTALE SINISTRES

Nous avons présenté au chapitre précédent les distributions qui modélisent le nombre de sinistre, les distributions qui modélisent le montant d'un sinistre, et les deux modèles qui permettent de déterminer la distribution de la charge totale sinistres. Dans ce chapitre nous allons nous focaliser sur la modélisation de la charge totale sinistres par le modèle individuel. Nous présenterons tout d'abord les outils de probabilité (distributions mélanges), les hypothèses du modèle individuel, ensuite la formule de l'espérance et de la variance de la somme S_n, et enfin les différentes méthodes de calcul de la charge totale sinistres à savoir : la méthode de convolution, les méthodes d'approximations et la méthode récursive de De PRIL.

2.1 Mélanges de Distributions

Définition 2.1. *Soit Θ une v.a de fonction de répartition F_Θ et $A \subset \mathbb{R}$ tel que $P(\Theta \in A) = 1$, et $(F(.|\Theta))_{\Theta\in A}$ une collection de fonctions de répartitions pour $A \subset \mathbb{R}$. On dit que X suit un mélange de lois (avec Θ comme loi de mélange) si pour tout $x \in \mathbb{R}$, pour tout $\theta \in A$,*

$$P(X \leq x|\Theta = \theta) = F(x|\theta). \tag{2.1}$$

Dans ce cas, pour tout $x \in \mathbb{R}$,

$$F_X(x) = P(X \leq x) = \int_{\Theta\in A} F(x|\theta)dF_\Theta(\theta). \tag{2.2}$$

En d'autres termes, soit Θ une variable aléatoire de densité de probabilité $u(.)$ telle que $u(\theta) \in \mathbb{R}^+$
pour $\theta \in \mathbb{R}$ et $\int_{-\infty}^{+\infty} u(\theta)d\theta = 1$. Soit $(f(.|\theta))_{\theta\in\mathbb{R}}$ une collection de densités de probabilités.

Définition 2.2. *On dit que X suit un mélange de lois (avec Θ comme loi de mélange) si pour tout $x \in \mathbb{R}$, pour tout $\theta \in A \subset \mathbb{R}$,*

$$f_X(x) = \int_{A\subset\mathbb{R}} f(x|\theta)u(\theta)d\theta. \tag{2.3}$$

Exemple 2.1.1. : Soit une v.a. X telle que $X|\Theta = \theta$ suit une loi $\mathcal{E}(\theta)$ et Θ (v.a de la loi de mélange) suit $\Gamma(\alpha, \beta)$. On a alors :

$$\begin{aligned}
f_X(x) &= \int_0^\infty \left[\frac{1}{\Gamma(\alpha)\lambda^\alpha}\theta^{\alpha-1}e^{-\dfrac{\theta}{\beta}}\right] d\theta \\
&= \int_0^\infty \frac{\theta^\alpha exp\left[-\alpha\left(x+\dfrac{1}{\beta}\right)\right]}{\Gamma(\alpha)\lambda^\alpha} d\theta \\
&= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\lambda^\alpha}\left[\frac{\beta}{\beta x+1}\right]^{\alpha+1} \\
&= \frac{\alpha}{\lambda^\alpha}\left[\frac{\beta}{\beta x+1}\right]^{\alpha+1}.
\end{aligned}$$

Si on pose $\lambda = \dfrac{1}{\beta}$, la dernière expression ci-dessus devient :

$$\frac{\alpha}{\lambda^\alpha}\left[\frac{\beta}{\beta x+1}\right]^{\alpha+1} = \frac{\alpha\lambda^\alpha}{(x+\lambda)^{\alpha+1}} \tag{2.4}$$

qui est la densité de probabilité d'une loi de Pareto $\mathcal{P}(\alpha, \lambda)$.

Remarque 2.1.1. Lorsque Θ est une variable aléatoire continue, on parle de **mélange continu**. La moyenne et la variance de X correspondent au cas particulier du résultat général suivant, dont la démonstration est basée sur les notions d'espérances et de variances conditionnelles, puis sur la formule de décomposition de la variance.

Proposition 2.1. *Soit X une variable aléatoire suivant un mélange de lois (avec Θ comme la loi de mélange), Alors son espérance est donnée par :*

$$E[X] = E[E[X|\Theta]] \tag{2.5}$$

et sa variance est donnée par :

$$Var[X] = E[Var[X|\Theta]] + Var[E[X|\Theta]]. \tag{2.6}$$

Preuve 2.1.1. Nous admettons (2.5) et établissons (2.6).

On a :

$$\begin{array}{rrcl} & Var[X|\Theta] & = & E[X^2|\Theta] - (E[X|\Theta])^2 \\ \text{ce qui implique} & E[Var[X|\Theta]] & = & E[E[X^2|\Theta]] - E[(E[X|\Theta])^2] \\ & & = & E[X^2] - E[(E[X|\Theta])^2](a) \quad \text{d'après (2.5)} \end{array}$$

de plus

$$\begin{array}{rrcl} & Var[E[X|\theta]] & = & E[(E[X|\Theta])^2] - (E[E[X|\Theta]])^2 \\ & & = & E[(E[X|\Theta])^2] - (E[X|\Theta])^2 \quad \text{d'après (2.5)} \\ \text{d'où} & Var[E[X|\theta]] & = & E[(E[X|\Theta])^2] - (E[X])^2(b) \end{array}$$

En additionnant (a) et (b), on obtient :

$$E[X^2] - (E[X])^2 = Var[X].$$

2.1.1 Quelques exemples de lois Poisson-mélange

- Si on a un portefeuille de polices d'assurance, chaque élément dans le portefeuille peut avoir sa propre propension au risque.
- Si tous les éléments du collectif suivent la même loi (loi de poisson), et si on choisit au hasard un élément, sa fréquence sinistre suivra la loi de Poisson, et le collectif aussi. On dira alors qu'on a un collectif **Homogène**. Cependant, peut être un élément du collectif a une propension au risque différente. C'est à dire une valeur λ différente, qu'on ne connait pas. Mais on peut connaitre la distribution de λ dans le collectif, c'est à dire la fonction de densité $f(\lambda)$. Dans ce cas, si on choisit au hasard un élément du collectif, la fréquence de sinistres suivra une distribution Poisson-mélange.

Définition 2.3. ***Distribution poisson-mélange***

Soit un couple (N,Λ) de v.a., avec$N|\Lambda$ qui suit une loi de poisson, et $\Lambda > 0$, de densité de probabilité f_Λ et de fonction de répartition F. La densité marginale de N a pour expression :

$$f_N(k) = P(N = k) = \int_0^\infty P(N = k|\Lambda = \lambda) f_\Lambda(\lambda) d\lambda = \int_0^\infty e^{-\lambda} \frac{(\lambda)^k}{k!} f_\Lambda(\lambda) d\lambda, \quad k = 0, 1, 2, ... \tag{2.7}$$

La distribution ainsi obtenue est une distribution ***Poisson-mélange***

Soit la v.a. N nombre de sinistres pour un contrat d'assurance (automobile par exemple), supposons que N est distribuée selon une loi de Poison $\mathcal{P}(\lambda)$. Comme l'assureur ignore les habitudes de conduite des assurés, on introduit une incertitude quant au paramètre de la loi de Poisson. Ainsi, λ apparait comme une réalisation d'une v.a. Θ qui influence de ce fait la v.a N. On suppose que la loi conditionnelle de N est donnée par : $N|\Theta = \theta$ qui suit $P(\lambda\theta)$ avec $\lambda > 0$. On a alors :

$$E[N|\Theta] = \lambda\theta, \quad Var[N|\Theta] = \theta\lambda \quad et \quad P_{N|\Theta}(t) = E[t^N|\Theta] = exp[\theta\lambda(t-1)] \tag{2.8}$$

Dans ces conditions l'espérance et la variance de N sont obtenus en conditionnant sur Θ, ainsi :

$$E[N] = E_\Theta[E[N|\Theta]] = E[\Theta\lambda]. \tag{2.9}$$

et

$$Var[N] = E_\Theta[Var(N|\Theta)] + Var_\Theta(E[N|\Theta]) = E[\Theta\lambda] + Var_\Theta[\Theta\lambda] \tag{2.10}$$

On identifie la loi de N par l'intermédiaire de sa fgp qui est donnée par :

$$P_N(t) = E[t^N] = E_\Theta[E[t^N|\Theta]] = E[e^{\Theta\lambda(t-1)}] = M_\Theta(\lambda(t-1)) \tag{2.11}$$

La v.a.Θ peut être continue ou discrète :

– -Si Θ est continue de fonction de densité f_Θ, alors la densité de probabilité de N est donnée par :

$$f_N(k) = P(N = k) = \int_0^\infty e^{-\lambda\theta}\frac{(\lambda\theta)^k}{k!}f_\Lambda(\lambda\theta)d\lambda, \quad k = 0, 1, 2, ... \tag{2.12}$$

– -Si Θ est discrète, la fonction de masse de probabilité de N est donnée par :

$$P(N = k) = \sum_\theta e^{-\lambda\theta}\frac{(\lambda\theta)^k}{k!}P(\Theta = \theta), \quad k = 0, 1, 2, ... \tag{2.13}$$

Donc, le choix de la loi de Θ peut avoir un impact important sur le comportement de N.
Nous présenterons dans la suite quelque distributions issues de la famille des Poisson-mélanges.

Loi mélange Poisson-gamma

la distribution binomiale négative provient du mélange des distributions Poisson et Gamma. Soit $N|\Theta = \theta$ une v.a.suivant une loi de Poisson $\mathcal{P}(\theta)$ où Θ est distribuée selon une loi $\Gamma(\alpha; \beta)$. Alors, la loi de N est donnée par :

$$\begin{aligned} f_N(k) &= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}e^{-\theta}\frac{\theta^k}{k!}d\theta \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)k!}\int_0^\infty \theta^{\alpha+k-1}e^{-\theta(\beta+1)}d\theta \\ f_N(k) &= \frac{\beta^\alpha}{\Gamma(\alpha)k!}\frac{\Gamma(\alpha+k)}{(\beta+1)^{\alpha+k}}; \end{aligned} \tag{2.14}$$

or en appliquant l'équation (1.46), on a :

$$\begin{aligned} \Gamma(\alpha+k) &= (\alpha+k-1)\Gamma(\alpha+k-1) \\ &= (\alpha+k-1)(\alpha+k-2)\Gamma(\alpha+k-2) \\ \Gamma(\alpha+k) &= (\alpha+k-1)(\alpha+k-2)...\alpha\Gamma(\alpha) \end{aligned} \tag{2.15}$$

et en remplaçant $\Gamma(\alpha+k)$ par sa valeur dans la formule (2.14) de $f_N(k)$, on aboutit à :

$$\begin{aligned} f_N(k) &= \frac{\alpha(\alpha+1)...(\alpha+k-1)}{k!}\beta^{\alpha}\left(\frac{1}{\beta+1}\right)^{\alpha+k} \\ &= \frac{\alpha(\alpha+1)...(\alpha+k-1)}{k!}\left(\frac{\beta}{\beta+1}\right)^{\alpha}\left(\frac{1}{\beta+1}\right)^{k} \end{aligned} \tag{2.16}$$

En comparant avec l'équation (1.31), nous voyons que N suit une loi binomiale négative $\mathcal{BN}\left(\alpha; \dfrac{\beta}{\beta+1}\right)$. Donc, la distribution binomiale négative et une distribution mélange Poisson-gamma.

Loi mélange Poisson-inverse gaussienne

Soit $N|\Lambda = \lambda$ une v.a. qui suit une loi de Poisson $\mathcal{P}(\lambda)$ où Λ est distribuée selon une loi inverse-gaussienne de paramètre μ et λ.

Définition 2.4. *Une v.a X est dite de loi inverse-gaussienne de paramètre $\mu > 0$ et $\lambda > 0$, si sa densité de probabilité est donnée par :*

$$f_X(x) = \begin{cases} \left(\dfrac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} exp\left[\dfrac{-\lambda(x-\mu)^2}{2\mu^2 x}\right] & si \quad x > 0 \\ 0 & sinon. \end{cases} \tag{2.17}$$

Elle est notée : $\mathcal{IG}(\mu;\lambda)$; où λ est appelé paramètre de forme tandis que μ est l'espérance.

L'espérance et la variance de X sont respectivement

$$E(X) = \mu, \quad Var(X) = \frac{\mu^3}{\lambda}. \tag{2.18}$$

La fonction de répartition de cette loi est :

$$F_X(x) = \Phi\left[\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}-1\right)\right] + exp\left(\frac{2\lambda}{\mu}\right)\Phi\left[-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}+1\right)\right]. \tag{2.19}$$

où $\Phi(.)$ est la fonction de répartition de la loi normale.

Revenons au mélange de la loi de Poisson sur la loi inverse-gaussienne, on obtient une nouvelle distribution appelée **loi de poisson inverse-gaussienne** que l'on peut noter $\mathcal{P}-\mathcal{IG}(\mu,\lambda)$. Ainsi pour une v.a N qui suit cette loi, il en découle que :

$$E[N]=\mu, \quad Var[N]=\mu(1+\lambda), \tag{2.20}$$

où $E[N]$ et $Var[N]$ sont : la valeur de l'espérance et de la variance ; et enfin sa fonction génératrice des probabilités est :

$$P_X(t)=exp\left[\frac{\mu}{\lambda}\left(1-\sqrt{1+2\lambda(1-t)}\right)\right]. \tag{2.21}$$

2.2 Les hypothèses du modèle individuel

- **Présentation du modèle**
  Le modèle individuel vise à représenter le montant total des sinistres à payer par la compagnie d'assurance sur une période (un an en général) en sommant assuré par assuré les montants des sinistres subis par chaque individu sur cette période.
  Soit un portefeuille ayant n contrats ou assurés et soit S_n la v.a.représentant la charge sinistres de ce portefeuille au cours de la période considérée. Pour $k=1,2,3,...,n$, les X_k désignent les v.a montant cumulé des sinistres du k^{ieme} assuré pendant la période considérée. Dans ce modèle, la charge totale sinistres S_n est donnée par le formule :

$$S_n=\sum_{k=1}^{n}X_k. \tag{2.22}$$

- Comme hypothèse, les variables aléatoires$X_k \quad k=1,2,...n$ sont **indépendantes et identiquement distribuées** (i.i.d) de même loi que X montant d'un sinistre.

Les principales caractéristiques de la charge totale sinistres S_n sont présentés dans la partie suivante.

2.2.1 Espérance et variance de S_n

l'espérance et la variance de la v.a.S_n de la formule (2.22) sont données par les formules suivantes :

$$E(S)=nE(X) \quad et \quad Var(S)=nVar(X). \tag{2.23}$$

Ces deux formules nous montent que pour calculer l'espérance et la variance de S, nous avons besoin de l'espérance et la variance de la v.a X. Soit θ la probabilité d'avoir un sinistre et $(1-\theta)$

la probabilité de ne pas avoir de sinistre. Soit Y le montant d'un sinistre qui est une variable aléatoire continue d'espérance μ_Y et de variance σ_Y^2. Nous avons $X = Y$ avec la probabilité θ et $X = 0$ avec la probabilité $(1-\theta)$. Nous pouvons donc écrire :

$$X = IY \tag{2.24}$$

où I est la v.a. (suivant une loi de Bernoulli) indépendamment distribuée de Y, qui est donnée par la formule :

$$I = \begin{cases} 0 & \text{avec la probabilité} \quad 1-\theta \\ 1 & \text{avec la probabilité} \quad \theta \end{cases} \tag{2.25}$$

L'espérance et la variance de I sont données par les formules :

$$E(I) = \theta \quad et \quad Var(I) = \theta(1-\theta). \tag{2.26}$$

Ainsi l'espérance de X est :

$$E(X) = E(I)E(Y) = \theta\mu_Y \tag{2.27}$$

et sa variance est :

$$Var(X) = \mu_Y^2\theta(1-\theta) + \theta\sigma_Y^2. \tag{2.28}$$

Preuve 2.2.1.

$$\begin{aligned}
Var(X) &= Var(IY) \\
&= E[(IY)^2] - [E(IY)]^2 \\
&= E(I^2)E(Y^2) - [E(I)]^2[E(Y)]^2 \\
&= E(I^2)E(Y^2) - [E(Y)]^2[E(I)]^2 + [E(Y)]^2E(I^2) - [E(Y)]^2E(I^2) \\
&= [E(Y)]^2\left(E(I^2) - [E(I)^2]\right) + E(I^2)\left(E(Y^2) - [E(Y)]^2\right) \\
&= [E(Y)]^2Var(I) + E(I^2)Var(Y) \\
&= \mu_Y^2\theta(1-\theta) + \theta\sigma_Y^2, \quad \text{d'après(2.27).}
\end{aligned}$$

En plongeant les équations (2.24) et (2.25) dans l'équation (2.20), on obtient ainsi les expressions de l'espérance et de la variance de S qui sont :

$$E(S) = n\theta\mu_Y \quad et \quad Var(S) = n[\mu_Y^2\theta(1-\theta) + \theta\sigma_Y^2]. \tag{2.29}$$

Exemple 2.2.1. Yiu-Kuen Tse (2009 : 89)

Considérons un portefeuille d'assurance comportant 500 polices indépendantes dont le montant de sinistres suit une distribution exponentielle de paramètre $\lambda = 0.5$. Considérons une chance

de 0,2 pour que survienne un sinistre. Calculons la variance et l'espérance de leur charge totale sinistres.
L'espérance et la variance pour le distribution exponentielle sont :

$$\mu_Y = \frac{1}{\lambda} = \frac{1}{0.5} = 2. \qquad \sigma_Y^2 = \frac{1}{\lambda^2} = \frac{1}{(0.5)^2} = 4.$$

D'où la variance et l'espérance et la variance du montant du sinistre d'une police est :

$$E(X) = (0.2)(2) = 0.4, \quad et \quad Var(X) = (2)^2(0.2)(1-0.2) + (0.2)(4) = 1.44$$

Donc l'espérance et la variance de la charge sinistre total sont :

$$E(S) = (500)(0.4) = 200 \quad et \quad Var(S) = (500)(1.44) = 720.$$

2.3 Modélisation du montant de la charge sinistres totale

Ayant des variables aléatoires indépendantes, il est souvent intéressant pour l'actuaire de considérer leur somme. L'une des techniques générales pour calculer la somme de plusieurs variables aléatoires indépendantes est la technique de convolution.

2.3.1 Méthode de convolution

Présentation de la formule

La technique de convolution permet dans certaines situations de donner une expression de la loi de probabilité de la somme de variables aléatoires indépendantes. Dans le cas échéant, elle permet de donner un algorithme qui permet de calculer les probabilités de la somme de variables aléatoires en utilisant les lois de probabilités de chaque variable. La méthode de convolution est utilisée dans le cas des v.a.discrètes, dans celui des v.a.continues et enfin dans le cas des v.a.suivant une loi de distribution mélange.

- **cas des v.a.discrètes**

Définition 2.5. *X et Y sont deux variables aléatoires discrètes à valeurs entières de lois* *de probabilités f et g et $Z = X + Y$. Pour tout entier z, on a :*

$$\begin{aligned} P(Z=z) &= \sum_{n=-\infty}^{+\infty} P(X=n \quad et \quad Y=z-n) \\ &= \sum_{n=-\infty}^{+\infty} P(X=n/Y=z-n)P(Y=z-n). \end{aligned} \tag{2.30}$$

Si de plus les v.a. X et Y sont indépendantes, on a :

$$\begin{aligned} P(Z=z) &= \sum_{n=-\infty}^{+\infty} P(X=n)(Y=z-n) \\ &= \sum_{n=-\infty}^{+\infty} f(n)g(z-n). \end{aligned} \tag{2.31}$$

la formule (2.31) s'appelle le produit de convolution de f et g et s'écrit :

$$(f*g)(z) = \sum_{-\infty}^{+\infty} f(n)g(z-n). \tag{2.32}$$

*Il est évident de constater que le produit de convolution est **commutatif** c'est à dire, $f*g = g*f$ car la somme est commutatif. De plus il est **associatif** c'est à dire, $(f*g)*h = f*(g*h)$, car la somme est associative.*

Exemple 2.3.1. Si X suit une loi de poisson de paramètre λ_1, Y suit une loi de poisson de paramètre λ_2, avec X et Y indépendantes, alors $X+Y$ suit une loi de poisson de paramètre $\lambda_1+\lambda_2$. i.e

$$P(X+Y=k) = \frac{(\lambda_1+\lambda_2)^k e^{-(\lambda_1+\lambda_2)}}{k!}$$

Ce résultat se généralise à la somme de n lois de poisson indépendantes.

La formule (2.32) se généralise par la somme de n variables aléatoires indépendantes $X_1, X_2 ... X_n$.

Si $f_1, f_2 ... f_n$ sont les lois de probabilités de $X_1, X_2 ... X_n$, alors la loi de probabilité de $S_n = \sum_{i=1}^{n} X_i$ est :

$$f_1 * f_2 * ... * f_n$$

ou

$$f^{*n}$$

dans le cas où $X_1, X_2, ..., X_n$ suivent la même loi.

Définition 2.6. *Soient $F(.)$ et $G(.)$ des fonctions de répartitions des v.a. X et Y, alors le produit de convolution de ces fonctions de répartitions (notée $F*G$) sera :*

$$(F*G)(z) = \sum_{-\infty}^{+\infty} F(n)G(z-n). \tag{2.33}$$

Exemple 2.3.2. (convolution de distributions discrète) Rob Kass,(2009 : 27)

Considérons $x = 0, 1, 2, 3, 4$ de densité de probabilité respectives $f_1(x), f_2(x), f_3(x)$ à l'aide de la formule (2.32), complétons le tableau suivant :

x	$f_1(x)$	$f_2(x)$	$f_3(x)$	$(f_1 * f_2)(x)$	$(f_1 * f_2 * f_3)(x)$
0	$1/4$	$1/2$	$1/4$	$1/8$	$1/32$
1	$1/2$	0	0	$2/8$	$2/32$
2	$1/4$	$1/2$	$1/2$	$2/8$	$4/32$
3	0	0	0	$2/8$	$6/32$
4	0	0	$1/4$	$1/8$	$6/32$

TABLE 2.1 – calcul de convolution des densités de probabilité

– **cas des v.a.continues**

Définition 2.7. *Étant donnés deux fonctions de répartitions F_X et F_Y des v.a. indépendantes respectives X et Y, la fonction de répartition de la somme $X+Y$ est appelé produit de convolution de F_X et F_Y notée : $F_X * F_Y$.*

$F_X * F_Y$ est défini par :

$$\begin{aligned}(F_X * F_Y)(z) &= Pr[X + Y \leq z] \\ &= \int_{-\infty}^{\infty} Pr[X + Y \leq z|X = x]dF(x) \\ &= \int_{-\infty}^{\infty} Pr[Y \leq z - x|X = x]dF(x) \\ &= \int_{-\infty}^{\infty} Pr[Y \leq z - x]dF(x) \\ &= \int_{-\infty}^{\infty} G(z - x)dF(x) \\ &= F * G(z)\end{aligned}$$

Donc

$$(F_X * F_Y)(z) = \int_{-\infty}^{\infty} G(z - x)dF(x) \tag{2.34}$$

Définition 2.8. *Soient F_X et F_Y deux fonctions de répartition ayant pour densité de probabilité f_X et f_Y respectivement. La densité de probabilité de $X + Y$ notée $f_X * f_Y$ est donnée par la formule :*

$$(f_X * f_Y)(z) = \frac{d}{dz}(F_X * F_Y)(z) = \int_{-\infty}^{\infty} f_Y(z - x)f_X(x)dx. \tag{2.35}$$

Remarque 2.3.1. pour X et Y positifs nous aurons plutôt :

$$(F_X * F_Y)(z) = \int_0^z F_Y(z-x)dF_X(x) \quad et \quad (f_X * f_Y)(z) = \int_0^z f_Y(z-x)f_X(x)dx. \tag{2.36}$$

En générale, la densité de probabilité de
$S_n = \sum_{i=1}^{n} X_i$ (où $X_1, X_2, ..., X_n$ sont n v.a. de densité de probabilités $f_1, f_2...f_n$), est notée :

$$f_1 * f_2 * ... * f_n$$

ou

$$f^{*n}$$

si les $X_1 * X_2 * ... * X_n$ suivent la même loi . Elle est déterminée de manière récursive par la formule :

$$f^{*n}(x) = \int_{-\infty}^{+\infty} f^{*(n-1)}(x-y)f_n(y)dy = \int_{-\infty}^{+\infty} f_n(x-y)f^{*(n-1)}(y)dy. \tag{2.37}$$

où $f^{*(n-1)}(x)$ est la densité de probabilité de $S_n = \sum_{i=1}^{n-1} X_i$.
Remarque 2.3.2. Pour $X_1, X_2, ..., X_n$ positifs nous aurons :

$$f^{*n}(x) = \int_0^x f^{*(n-1)}(x-y)f_n(y)dy = \int_0^x f_n(x-y)f^{*(n-1)}(y)dy \tag{2.38}$$

Exemple 2.3.3. Soient X et Y des v.a. indépendantes suivant des lois exponentielles de paramètres respectifs a et b. Déterminons la loi de $S = X + Y$.
Les v.a.X et Y sont de densités :

$$f(x) = ae^{-ax}1_{[0,+\infty[}(x), \quad g(x) = be^{-by}1_{[0,+\infty[}(y)$$

Comme X et Y sont indépendantes, la densité de $X + Y$ est :

-Si $a \neq b$

$$\begin{aligned}
(f * g)(x) &= \int_{-\infty}^{+\infty} g(y)f(x-y)dy \\
&= \int_{-\infty}^{+\infty} be^{-by}1_{[0,+\infty[}(y)ae^{-a(x-y)}1_{[0,+\infty[}(x-y) \\
&= abe^{-ax}1_{[0,+\infty[}(x)\int_0^x e^{(a-b)y}dy \\
&= \frac{abe^{-ax}}{a-b}\left(e^{(a-b)x}-1\right)1_{[0,+\infty[}(x) \\
&= \frac{ab}{a-b}\left(e^{-bx}-e^{-ax}\right)1_{[0,+\infty[}(x)
\end{aligned}$$

Notons que dans ce calcul, on a utilisé l'égalité entre produit d'indicatrices i.e. :

$$1_{[0,+\infty[}(x-y)1_{[0,+\infty[}(y) = 1_{[0,x[}(y)1_{[0,+\infty[}(x)$$

- Si $a = b$, la densité est :

$$\begin{aligned}
(f * g)(x) &= \int_{-\infty}^{+\infty} g(y)f(x-y)dy \\
&= a^2 \int_{-\infty}^{+\infty} e^{-ay}1_{\{y\geq 0\}}e^{-a(x-y)}1_{\{x-y\geq 0\}}dy \\
&= a^2 1_{\{x\geq 0\}} \int_0^x e^{-ay}e^{-a(x-y)}dy \\
&= a^2 1_{\{x\geq 0\}}e^{-ax} \int_0^x dy \\
&= a^2 x 1_{\{x\geq 0\}}e^{-ax}
\end{aligned}$$

– **cas des v.a.suivant une distribution mélange**

Ce cas est très utilisé dans le cadre du modèle individuel. Soit X_i la variable aléatoire suivant un distribution mélange, soit la probabilité de sa fonction densité :

$$f_{X_i}(x) = \begin{cases} 1-\theta_i, & si \quad x = 0 \\ \theta_i f_{Y_i}(x), & si \quad x > 0 \end{cases} \tag{2.39}$$

où $f_{Y_i}(.)$ est la densité de probabilité d'une variable aléatoire continue positive. D'après la formule (2.36), nous pouvons dire que pour n v.a.$X_1, X_2, ..., X_n$ la fonction de répartition de $\sum_{i=1}^{n} X_i$ est donnée par la formule récursive :

$$F^{*n}(x) = \int_0^x F^{*(n-1)}(x-y)dF_{X_n}(y) = \int_0^x F_{X_n}(x-y)dF^{*(n-1)}(y). \tag{2.40}$$

où $F^{*(n-1)}(x)$ est la fonction de répartition de $\sum_{i=1}^{n-1} X_i$. En prenant la première égalité de la formule (2.42) et en y introduisant la probabilité de la fonction densité de la formule (2.39), on obtient :

$$F^{*n}(x) = \int_0^x F^{*(n-1)}(x-y)f_{X_n}(y)dy + (1-\theta)F^{*(n-1)}(x). \tag{2.41}$$

En particulier, si $X_1, X_2, ..., X_n$ sont i.i.d avec $\theta_i = \theta$, et $f_{Y_i}(x) = f_Y(x)$,pour tout $i = 1, 2, ..n$, alors,

$$F^{*n}(x) = \theta \int_0^x F^{*(n-1)}(x-y)f_Y(y)dy + (1-\theta)F^{*(n-1)}(x). \tag{2.42}$$

Cette expression est donc celle de la fonction de répartition des variables aléatoires qui suivent une distribution mélange.

Cependant bien que l'écriture de la formule de convolution s'écrit facilement, le calcul de cette valeur est très complexe. Pour implémenter la méthode de convolution en pratique, on **discrétise** tout d'abord la distribution continue, puis on applique la formule de convolution des v.a. discrètes (formule(2.32)généralisée).

Les limites de la méthode

Le calcul par la méthode de convolution de la charge sinistre totale est très fastidieuse pour n très grand. .

2.3.2 Les méthodes d'approximations

Basées sur des hypothèses simplificatrices qui ne conviennent pas nécessairement pour représenter de manière satisfaisante les diverses caractéristiques de la distribution du montant total des réclamations, les approximations ont l'avantage de fournir un résultat intermédiaire rapide éclairant sur le choix d'une méthode alternative plus précise. Nous nous attarderons beaucoup plus dans cette partie la méthode d'ajustement des moments. Elle consiste à choisir une loi de probabilité qui possède des caractéristiques semblables à celles de la distribution de la charge totale sinistres.On ajuste les paramètres de la loi de probabilité choisie de sorte que les moments calculés soient équivalents pour les deux distributions. Louis Thibaut, (1997 :9)

Approximation normale

Pour approximer la distribution de S_n qui est la somme de n variables aléatoires i.i.d, nous utilisons la méthode d'approximation Normale. Nous savons que la méthode d'approximation de la fonction de répartition est basée sur le **Théorème Central Limite**. Parmi les lois des grands nombres que nous connaissons, le Théorème Central Limite est le plus important en statistique.

Théorème 2.1. *Rob Kass,(2009 :30)*
Si $X_1, X_2, ..., X_n$ sont n v.a.indépendants et identiquement distribuées avec pour espérance μ et pour variance $\sigma^2 < \infty$, alors :

$$\lim_{n \longrightarrow \infty} Pr\left[\sum_{i=1}^{n} X_i \leq n\mu + x\sigma\sqrt{n}\right] = \Phi(x). \tag{2.43}$$

Où ϕ est la fonction de répartition de la loi Normale.

Ainsi, si la somme a une variance finie, nous pouvons approximer la fonction de répartition de $S_n = X_1 + ... + X_n$ par :

$$F_{S_n}(s) \approx \Phi\left(s; \sum_{i=1}^{n} E[X_i], \sum_{i=1}^{n} Var[X_i]\right). \tag{2.44}$$

En effet,

$$\begin{aligned} Pr(S_n \le s) &= Pr\left(\frac{S_n - E(S_n)}{\sqrt{Var(S_n)}} \le \frac{s - E(S_n)}{\sqrt{Var(S_n)}}\right) \\ &\simeq Pr\left(Z \le \frac{s - E(S_n)}{\sqrt{Var(S_n)}}\right) \\ &= \Phi\left(\frac{s - E(S_n)}{\sqrt{Var(S_n)}}\right). \end{aligned} \tag{2.45}$$

Donc la formule (2.44) s'écrit encore de la façon suivante :

$$Pr(S_n \le s) \approx \Phi\left(\frac{s - E(S_n)}{\sqrt{Var(S_n)}}\right). \tag{2.46}$$

Exemple 2.3.4. Yiu-kuen Tse, (2009 :96)

Revenons à l'exemple 2.2.1 et calculons la fonction de répartition de S_n pour $s = 180$ et $s = 230$ en utilisant l'approximation normale.

Dans l'exemple précédent nous avons déterminer l'espérance et la variance qui sont respectivement $E(X) = 200$ et $Var(S) = 720$. Ainsi $\sqrt{Var(S)} = \sqrt{720} = 26,8328$. D'après l'approximation normale :

$$Pr(S_n \le 180) \simeq Pr\left(Z \le \frac{180,5 - 200}{26,8328}\right) = \Phi(-0,7267) = 0,2337$$

et

$$Pr(S_n \le 230) \simeq Pr\left(Z \le \frac{230,5 - 200}{26,8328}\right) = \Phi(1,1367) = 0,8722$$

Cependant les résultats obtenus par la méthode de convolution sont 0.2772 et 0.8968. La différence entre ces deux résultats est due à la discrétisation de la distribution du montant d'un sinistre (loi exponentielle) et l'utilisation de l'approximation normale supposant les données du portefeuille suffisamment nombreuses.

Limite de la formule d'approximation normale :

Même pour un portefeuille très grand, l'approximation normale peut donner des résultats insatisfaisants. On peut remarquer que la distribution du montant total des sinistres est généralement asymétrique. Comme la distribution d'une loi normale est symétrique, elle ne peut pas convenir adéquatement pour estimer la variable aléatoire S_n.

Approximation Log-normale

La loi log-normale permet d'estimer la distribution du montant total des réclamations avec plus de précision. Cette dernière possède un avantage par rapport à la loi normale puisqu'elle est asymétrique. Le niveau d'asymétrie d'une distribution se mesure à l'aide du coefficient d'asymétrie.

Définition 2.9. *Le coefficient d'asymétrie (ou Skewness en anglais) d'une variable aléatoire X, notée $\gamma[X]$, est défini par :*

$$\gamma[X] = \frac{E[(X - E[X])^3]}{\left(\sqrt{Var[X]}\right)^3}. \tag{2.47}$$

ou encore

$$\gamma[X] = \frac{m_3}{\sigma^3}. \tag{2.48}$$

(avec m_k = le moment centré d'ordre 3 et σ = l'écart type de la v.a X.)

Pour la loi log-normale, ce coefficient est toujours positif, ce qui le rapproche de celui des distributions du montant total des réclamations que l'on retrouve en pratique. Les premiers moments de la loi log-normale sont :

$$\begin{cases} E[X] &= e^{\mu+\frac{1}{2}\sigma^2} \\ Var[X] &= e^{2\mu+2\sigma^2} - e^{2\left(\mu+\frac{1}{2}\sigma^2\right)} \\ \gamma[X] &= \dfrac{e^{3\sigma^2} - e^{\sigma^2+2}}{e^{\frac{3}{2}\sigma^2} - 1} \end{cases} \tag{2.49}$$

Où les termes de gauche sont calculés en fonction du modèle de risque considéré, tandis que ceux de droite correspondent aux formules d'espérance et de variance pour une variable aléatoire log-normale. Pour évaluer la fonction de répartition approximative, on ajuste les paramètres μ et σ^2 selon le système d'équations suivant :

$$\begin{cases} E[X] &= e^{\mu+\frac{1}{2}\sigma^2} \\ Var[X] &= e^{2\mu+2\sigma^2} - e^{2\left(\mu+\frac{1}{2}\sigma^2\right)} \end{cases} \tag{2.50}$$

Une fois les coefficients trouvés, soit :

$$\begin{aligned} \mu &= ln(E^2[S]) - \frac{1}{2}ln(E^2[S] + Var[S]) \\ \sigma^2 &= ln(E^2[S] + Var[S]) - ln(E^2[S]), \end{aligned} \tag{2.51}$$

il ne reste plus qu'à remplacer ceux-ci dans la fonction de répartition de la loi log-normale pour obtenir l'approximation suivante :

$$F_S(x) = \Phi\left(\frac{lnx - [ln(E^2[S]) - \frac{1}{2}ln(E^2[S] + Var[S])]}{\sqrt{ln(E^2[S] + Var[S]) - ln(E^2[S])}}\right). \tag{2.52}$$

limite de l'approximation Lognormale :
Malgré le fait que le coefficient d'asymétrie de cette approximation soit de même signe que celui des distributions généralement observées, il n'en demeure pas moins que ce dernier peut être très différent. Louis Thibaut, (1997 : 12)

Approximation gamma

La loi gamma est une autre loi de probabilité souvent utilisée pour estimer la distribution du montant total des réclamations. On rappelle que la fonction de densité de probabilité de la loi gamma se paramètres α et λ, notée $g(x;\alpha,\lambda)$ et donnée par :

$$g(x;\alpha,\lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$$

où $x \in \mathbb{R}, \alpha > 0, \lambda > 0$. Nous avons :

$$\begin{aligned} E[X] &= \frac{\alpha}{\lambda} \\ Var[X] &= \frac{\alpha}{\lambda^2} \end{aligned}$$

Pour ajuster les paramètres adéquatement, il suffit d'isoler α et λ dans le système d'équations suivant :

$$\left\{ \begin{aligned} E[S_n] &= \frac{\alpha}{\lambda} \\ Var[S_n] &= \frac{\alpha}{\lambda^2} \end{aligned} \right.$$

où $E[S_n]$ et $Var[S_n]$ sont la moyenne et la variance des valeurs données. Ainsi on estime la loi de probabilité de la v.a S_n par une loi gamma de paramètre $\alpha = \dfrac{E^2[S_n]}{Var[S_n]}$ et $\lambda = \dfrac{E[S_n]}{Var[Sn]}$.

Limite de l'approximation Gamma

L'approximation gamma n'offre aucune flexibilité pour la coefficient d'asymétrie puisque ce dernier est fixé indirectement en résolvant le système précédent. Il est toutefois possible d'ajuster ce paramètre en utilisant une généralisation de la loi gamma appelée **loi gamma translatée** définie comme suit :

Définition 2.10. *Si une variable aléatoire X suit une loi gamma translatée de paramètre $\alpha, \quad \lambda \quad et \quad x_0$, alors :*

$$f_X(x) = g(x - x_0; \alpha, \lambda), \tag{2.53}$$

où $x \in \mathbb{R}, \alpha > 0, \lambda > 0, \quad x_0 \in \mathbb{R}$.

Les premiers moments de la loi gamma translatée sont :

$$E[X] = \frac{\alpha}{\lambda} + x_0, \quad Var[X] = \frac{\alpha}{\lambda^2}; \quad E[(X - E[X])^3] = \frac{2\alpha}{\lambda^3}. \tag{2.54}$$

A l'aide de ces moments on forme le système d'équations suivant :

$$\left\{ \begin{array}{rcl} E[S_n] & = & \frac{\alpha}{\lambda} + x_0 \\ Var[S_n] & = & \frac{\alpha}{\lambda^2} \\ E[(S_n - E[S_n])^3] & = & \frac{2\alpha}{\lambda^3} \end{array} \right. \tag{2.55}$$

Une fois résolu, on obtient les paramètres de la loi gamma translatée à utiliser, soient :

$$\alpha = \frac{4Var^3[S_n]}{E^2[(S_n - E[S_n])^3]}, \quad \lambda = \frac{2Var[S_n]}{E^2[(S_n - E[S_n])^3]}, \quad x_0 = E[S_n] - \frac{2Var^2[S_n]}{E[(S_n - E[S_n])^3]}. \tag{2.56}$$

Ainsi, la fonction de répartition s'estime par :

$$F_{S_n}(x) \simeq G(x - x_0; \alpha; \lambda), \tag{2.57}$$

où $G(x; \alpha; \lambda)$ est la fonction de répartition d'une loi gamma de paramètre α et λ. Cette dernière approximation a comme principal intérêt d'offrir la possibilité d'ajuster les coefficients d'asymétrie. Louis Thibaut,(1997 : 14)

Remarque 2.3.3. Lorsque $\alpha \longrightarrow \infty, \quad \lambda \longrightarrow \infty, \quad x_0 \longrightarrow -\infty$, alors $G(x - x_0; \alpha; \lambda)$ converge vers une loi normale de paramètres $\mu = x_0 + \frac{\alpha}{\lambda} \quad et \quad \sigma^2 = \frac{\alpha}{\lambda^2}$. En ce sens, l'approximation gamma translatée est plus générale que l'approximation normale.

l'approximation Normal-puissance ou *(Normal-Power)*

Comme l'approximation lognormale et l'approximation gamma, ses paramètres se déterminent aussi par la méthode d'ajustement des moments. Si $E[S_n] = \mu$, $Var[S_n] = \sigma^2$ et $\gamma_{S_n} = \gamma$, alors pour $s \geq 1$,

$$Pr\left[\frac{S - \mu}{\sigma} \leq s + \frac{\gamma}{6}(s^2 - 1)\right] \approx \Phi(s) \tag{2.58}$$

cette formule est équivalente à :pour $x \geq 1$,

$$Pr\left[\frac{S - \mu}{\sigma} \leq x\right] \approx \Phi\left(\sqrt{\frac{9}{\gamma^2} + \frac{6}{\gamma}\frac{x - \mu}{\sigma} + 1} - \frac{3}{\gamma}\right) \tag{2.59}$$

La formule (2.58) est utilisée pour approximer la fonction densité de probabilité de S_n tandis que la formule (2.59) est utilisée pour approximer les quantiles. Rob Kass, (2009 : 33)

Nous reviendrons sur l'application de cette méthode d'approximation dans le chapitre suivant.

2.3.3 La formule récursive de DE PRIL

En 1986, Nelson De Pril a présenté une formule récursive permettant d'évaluer la distribution du montant total des réclamations pour le modèle individuel du risque. En plus d'offrir une plus grande rapidité de calcul que la formule de convolution standard, cet algorithme procure une évaluation exacte de la distribution du montant total des sinistres.

Présentation de la formule

Considérons un portefeuille de polices d'assurance indépendantes produisant au plus un sinistre durant une période donnée. la méthode de De Pril consiste à ranger le porte feuille sous forme de tableau à deux entrées $I \times J$ (disposition utilisée pour la première fois par Petter S. Kornya en 1983) Louis Thibaut, (1997 :52).

		(distribution du montant d'un sinistre)					
		$f_1(x)$	$f_2(x)$	$\cdots$	$f_i(x)$	$\cdots$	$f_I(x)$
	θ_1				$\vdots$		
	θ_2				$\vdots$		
(probabilié de sinstres)	θ_3				$\vdots$		
	$\vdots$				$\vdots$		
	θ_j	$\cdots$	$\cdots$	$\cdots$	n_{ij}		
	$\vdots$						
	θ_J						

TABLE 2.2 – Portefeuille d'assurance utilisé par De Pril

où :

-les f_i : représentent les distributions des montants de sinistre des polices de la ligne $i = 1, 2, ...I$; pour $x = 1, 2, ...n$;

-θ_j lignes : représentent les probabilités pour lesquelles la police de la ligne j produise un sinistre.

- n_{ij} : le nombre indépendant de polices de la distribution du montant i et de probabilité j.

-$1 - \theta_j$: probabilité pour que la police de la ligne j ne produise pas de sinistres ; $j = 1, 2, ...J$.

Nous notons par S la charge totale de sinistres du portefeuille d'assurance, pour une période donnée et $f_S(s)$ la probabilité pour laquelle S soit égale à s avec $s = 0, ..., n$, où $n = \sum_{i=1}^{I} \sum_{j=1}^{J} in_{ij}$ est le montant maximal de la charge totale sinistres du portefeuille, alors $f_S(s)$ est calculée en utilisant le théorème suivant :

Théorème 2.2. *La probabilité de S, notée $f_S(s)$, satisfait les équations, connues sur le nom de formule récursive de De Pril(1985-1986)*

$$f_S(s) = \frac{1}{s}\sum_{i=1}^{min\{s,i\}}\sum_{k=1}^{\left[\frac{s}{i}\right]} f_S(s-ik)h(i,k), \quad pour \quad s = 1,2,...,n \tag{2.60}$$

où $[x]$ représente la partie entière de x et

$$h(i,k) = \begin{cases} i(-1)^{k-1}\sum_{j=1}^{J} n_{ij}\left(\frac{\theta_j}{1-\theta_j}\right)^k & pour \quad i = 1,2,...,n \\ 0 & sionon, \end{cases} \tag{2.61}$$

avec la valeur initiale de récurrence qui est :

$$f_S(0) = \prod_{i=1}^{I}\prod_{j=1}^{J}(1-\theta_j)^{n_{ij}}. \tag{2.62}$$

Les équations (2.60), (2.61), (2.62) traduisent la formule améliorée de De Pril en 1986 qui tient compte le cas où la v.a.X_i inclut des valeurs négatives. En effet De Pril présenta en 1985 deux formules (théorèmes) permettant de calculer $f_S(s)$.

-la première est présentée sous forme de deux étapes de récurrence :

$$f_S(0) = \prod_{j=1}^{J}(f_j)^{n_j} \tag{2.63}$$

et

$$sf_S(s) = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\sum_{x=1}^{s} w_{ij}(x)f_S(s-x) \quad avec \quad s = 1,2,...,m. \tag{2.64}$$

où la fonction auxiliaire w_{ij} est donnée par :

$$w_{ij}(1) = \frac{\theta_j}{1-\theta_j}f_i(1) \quad et \quad w_{ij} = -\frac{\theta_j}{1-\theta_j}\sum_{k=1}^{m} f_i(k)w_{ij}(x-k) \quad x = m+1,... \tag{2.65}$$

- la deuxième formule contenant le plus grand ordre de convolution (présentée aussi en deux étapes de récurrence) :

$$f_S(0) = \prod_{j=1}^{J}(f_j)^{n_j}. \tag{2.66}$$

et

$$sf_S(s) = \sum_{i=1}^{I}\sum_{k=1}^{s} A(i,k)\sum_{k=1}^{mins,km} xf^{*k}(x)f_S(s-x) \quad s = 1,2,...m \tag{2.67}$$

où les coefficients $A(i,k)$ sont donnée par :

$$A(i,k) = \frac{(-1)^{k+1}}{k}\sum_{j=1}^{J} n_{ij}\left(\frac{\theta_j}{1-\theta_j}\right)^k. \tag{2.68}$$

et $f^{*k}(x)$ est la convolution d'ordre k de $f_i(x)$.

les limites de la formule de De Pril

Il est clair que la compilation de cette formule nécessite énormément de temps avec la machine lorsque I est assez grand et lorsque f_i prend plusieurs valeurs. Toutefois, comme θ_j est souvent petit, $h(i,k)$ tend rapidement vers zéro lorsque k croit. Pour cela la formule (2.60) est approximée en tronquant sur la variable k. Nelson De Pril est parvenu à réduire le temps de calcul de la formule(2.60) sans trop altérer la précision des résultats. ainsi une approximation de la formule (2.60) :

$$f_S^k(s) = \frac{1}{s} \sum_{i=1}^{min\{s,I\}} \sum_{k=1}^{min\left\{K,\left[\frac{s}{i}\right]\right\}} f_S^K(s-ik)h(i,k) \tag{2.69}$$

avec

$$f_S^K(0) = f_S(0)$$

.L'erreur de cette approximation est donnée par la formule :

$$\sum_{s=0}^{m} |f_S(s) - f_S^K(s)| \leq e^{\varepsilon(r)} - 1. \tag{2.70}$$

où $\varepsilon(r)$ est donnée par la formule :

$$\varepsilon(r) = \frac{1}{r+1} \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij} \frac{1-\theta_j}{1-2\theta_j} \left(\frac{\theta_j}{1-\theta_j}\right)^{r+1} \tag{2.71}$$

Louis Thibaut, (1997 : 75) .

⋆ ⋆ Chapitre Trois ⋆ ⋆

APPLICATIONS

Dans ce chapitre, notre travail est essentiellement basé sur la présentation de quelques exemples d'application. La première partie porte sur la comparaison des lois qui modélisent le nombre de sinistre et la comparaison des lois qui modélisent le montant d'un sinistre. Puis deux exemples montrant l'utilisation des méthodes d'approximations en deuxième partie. Enfin la troisième partie, porte sur la modélisation de la charge totale sinistres S_n par la méthode convolution. Notons que les logiciels utilisés sont le $\mathcal{R}$ et MATLAB.

3.1 Modélisation du nombre de sinistre et du montant d'un sinistre

3.1.1 Modélisation du nombre de sinistre

Comme nous l'avons vu au chapitre 1, la loi de Poisson, la loi Binomiale, la loi Binomiale Négative sont des lois discrètes qui permettent de modéliser le nombre de sinistre. Nous avons montrer comment elles sont utilisées par des exemples. Ici, nous allons faire une étude comparative de ces différentes lois.

Exemple 3.1.1. Soit N le nombre de sinistre d'un portefeuille d'assurance suivant l'une ces trois lois distributions . La figure suivante présente la variation de ces trois lois pour une même espérance $E(N)$.

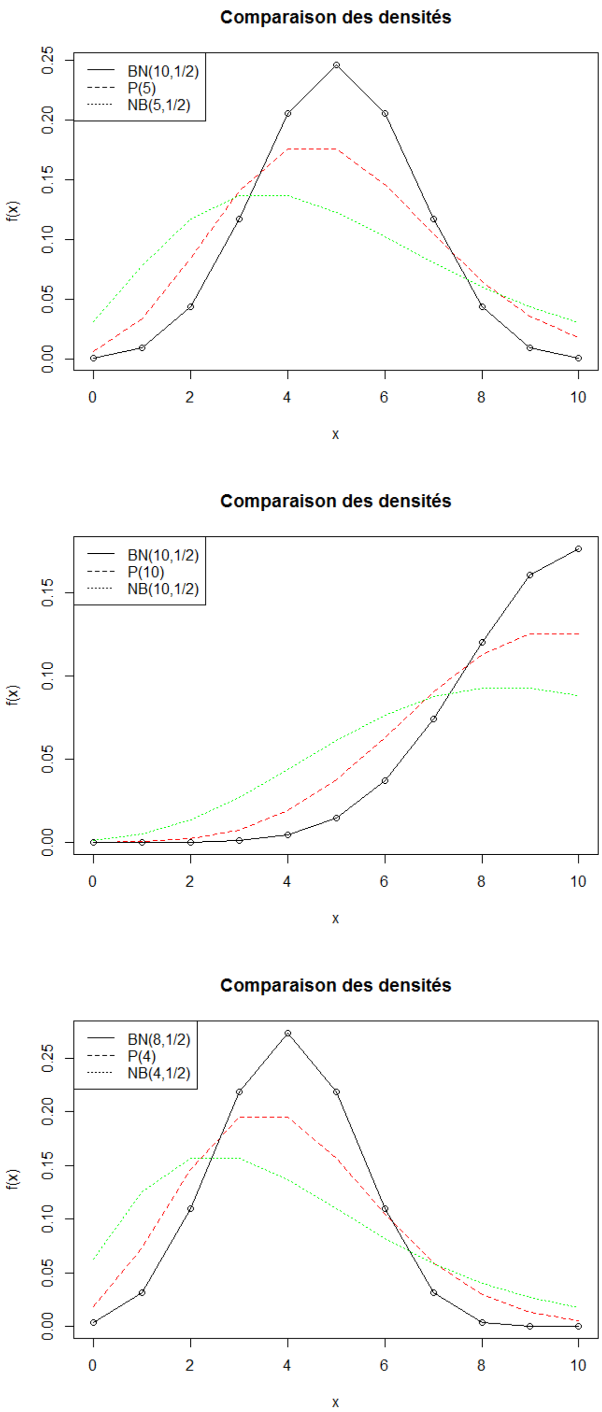

FIGURE 3.1 – Fonction de masse de probabilité des lois discrètes usuelles

Comparaison des distributions

Comme le montre la figure 3.1, ces lois usuelles ont des densités de probabilité différentes pour une même espérance.

– Pour la valeur de l'espérance $E(N) = 5$, la distribution Binomiale a la plus grande valeur de la densité et est centrée. La distribution de Poisson a la deuxième grande valeur. La distribution Binomiale négative quant à elle possède la plus petite valeur de densité.
– Pour $E(N) = 10$, la la distribution Binomiale possède toujours la plus grande valeur de la densité et reste centrée. La distribution de Poisson a la deuxième grande valeur. La distribution Binomiale négative quant à elle possède la plus petite valeur de densité et reste asymétrique.
– Pour $E(N) = 4$, les trois courbes présentent toujours les mêmes différences.

3.1.2 modélisation du montant d'un sinistre

Nous faisons ici, une étude comparative des lois qui permettent de modéliser le montant d'un sinistre. Il s'agit des lois continues à savoir : la loi exponentielle, la loi gamma, la loi de Pareto, la loi lognormale.

Soit un portefeuille d'assurance, soit X une v.a.r qui représente le montant d'un sinistre. X est modéliser par l'une des lois citées ci-dessus. La figue (3.2) représente la fonction densité de probabilité de ces différentes lois.

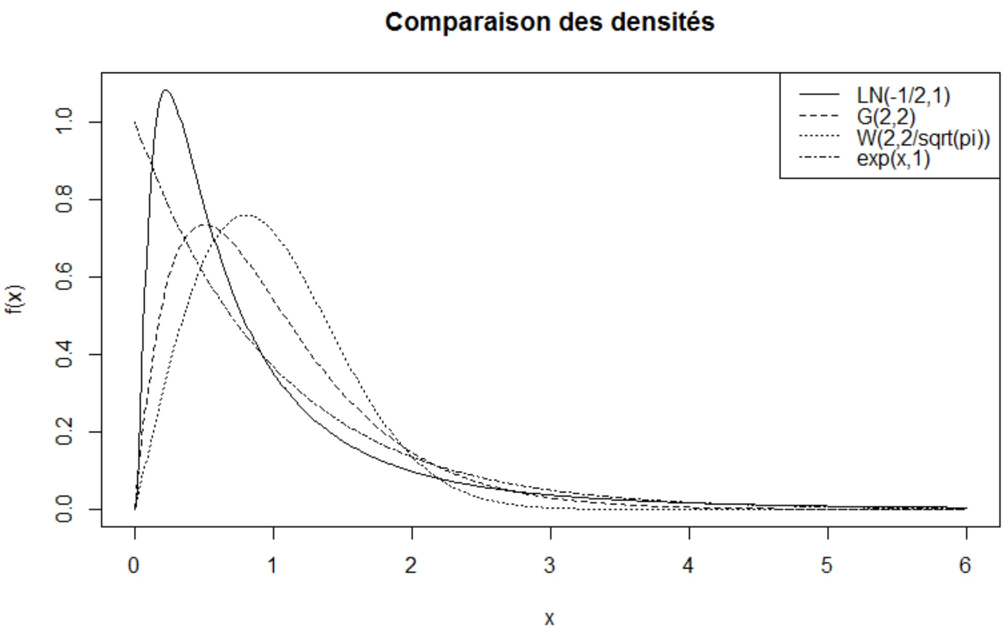

FIGURE 3.2 – Densités de lois usuelles pour des variables positives.

Cette comparaison de lois montre pour une distribution de montant de sinistre X donnée :

une seule lois peut mieux l'ajuster.

3.2 Modélisation de la charge sinistre totale

3.2.1 Applications des méthodes d'approximations

Dans cette partie nous montrons l'utilisation des différentes méthodes d'approximations pour avoir des informations approximatives sur le montant total des réclamations.

Exemple 3.2.1. Rob Kass, (2009 : 32-34).

Soit un échantillon n= 1000,jeunes individus véhiculés qui prennent chacun une assurance automobile pour une durée d'un an. La probabilité pour que survienne un sinistre est de p= 0.001 par personne cette année et le payement est de 1. Nous voulons calculer la probabilité pour laquelle le payement total des sinistres (S) soit au moins 4, sachant qu'elle suit une loi Binomiale de paramètres (1000,0.001). Déterminons cette probabilité pour un payement de 3.5 au lieu de 4 (la correction continu). Notons que la probabilité de la loi Binomiale est de : 0.01893.Cependant comme n est très grand et p très petit, on approximera cette probabilité par une loi de poisson (Poisson(np)) c'est à dire Poisson(1).

Le but de cet exemple est de trouver les valeurs qui approchent le mieux cette probabilité par les différentes méthodes d'approximations.

Approximation normale :

Nous avons : $\mu = E[S] = (\sigma)^2 = Var[S] = 1$.

Ainsi, $Pr[S \leq 3.5] = Pr\left[\frac{s-\mu}{\sigma} \leq \frac{3.5-\mu}{\sigma}\right] \approx 1 - \Phi\left(\frac{3.5-\mu}{\sigma}\right) \approx 1 - \Phi(2.5) = 0.0062$.

Cette valeurs est très inférieure à 0.01893. Ceci est dut au fait que le montant des sinistres suit une loi asymétrique ainsi que celle de S,dont le coefficient est : $\gamma_S = 1$.

Approximation Gamma translaté

Supposons que $S \sim$ Poisson(1), ce qui implique, $\mu = \sigma = \gamma = 1$ et d'après les formules (2.58), nous avons : $x_0 = \mu - \frac{2\sigma}{\gamma}, \quad \alpha = \frac{4}{\gamma^2}, \quad et \quad \beta = \frac{2}{\gamma\sigma}$. Ce qui nous donne $\mu = \sigma = \gamma = 1$ et $x_0 = -1$.

$Pr[S \geq 3.5] \approx 1 - G(3.5 - (-1); 4, 2) = 0.0212$.

Cette valeur de probabilité est plus proche de celle que nous voulons approximer.

Approximation Normale Power

Si $S \sim$ Poisson(1), alors l'approximation nous donne : $Pr[S \leq 3.5] \approx 1 - \Phi(2) = 0.0228$. Cette valeur est aussi proche de celle de la probabilité de la loi de Poisson(1).

Donc l'approximation Gamma translaté nous donne une meilleure approximation de la probabilité pour laquelle le payement total soit au moins 4.

Le logiciel R est un autre moyen plus rapide de trouver ces différentes valeurs de probabilité. Pour les valeurs de $\mu, \sigma \quad et \quad \gamma$ données, nous avons effectué les calculs et trouvé les même valeurs. (Le script est donné en annexe)

Exemple 3.2.2. Rob Kass, (2009 : 32-34). Le montant total des réclamations (S) d'un échantillon de portefeuille d'assurance est estimé à 10000, l'écart-type est de 1000 et le coefficient d'asymétrie est de 1. déterminons la probabilité pour laquelle, 13000 soit un capital insuffisant pour couvrir S.

Approximation Normal Power

Nous utiliserons la formule (2.61) pour : $\mu = 10000, \quad \sigma = 1000 \quad et \quad \gamma = 1$.

$$\begin{aligned} Pr[S > 13000] &= Pr\left[\frac{S-\mu}{\sigma} > 3\right] \\ &\approx 1 - \Phi\left(\sqrt{9 + 6 \times 3 + 1} - 3\right) \\ &= 1 - \Phi(2.29) \\ &= 0.011. \end{aligned}$$

Approximation Gamma translaté

D'après les égalités (2.58) et les valeurs de μ, γ, σ données, on a $\alpha = 4$, $\beta = 0.002$, et $x_0 = 8000$, d'où :

$$Pr[S > 13000] \approx 1 - \Phi(13000 - 800; 4, 0.0002) = 0.010.$$

cette valeur de probabilité est plus petite que celle donnée par l'approximation Normal Power.

Approximation Normale

$\mu = 10000, \sigma = 1000$

$$Pr[S > 13000] \approx 1 - \Phi\left(\frac{13000 - 10000}{1000}\right) = 1 - \Phi(3) = 0.0013.$$

La plus petite probabilité dans cet exemple est celle donnée par l'approximation Normale. Nous résumons les résultats de ces deux exemples par le tableau suivant :

$Approximations$	$Normale$	$Gammatranslat$	$Normal - Power$
$Pr[S \geq 3.5]$	0.0062	0.0212	0.0228
$Pr[S > 13000]$	0.0013	0.010	0.011

TABLE 3.1 – Comparaison des probabilités approximées

3.2.2 Application de la méthode convolution

Yiu-kuen Tse (2009,91-93)

Dans cette partie, nous reprenons l'exemple 2.2.1, concernant un portefeuille ayant 500 contrats d'assurance, où le montant des distributions suit une lois exponentielle de paramètre $\lambda = 0.5$.

Déterminons par la méthode de convolution la distribution de la charge sinistre total S_n. Comme nous l'avons dit au chapitre précédent, le calcul par la méthode de convolution des lois continus est très difficile ; nous approximons la loi exponentielle en la discrétisant en des points 0,1,2,...10. pour l'appliquer nous avons le choix entre deux méthodes :

Soit Nous utilisons la commande, "*discrtise*" du package ACTUAR du logiciel R.

Soit nous écrivons un code R qui donne des valeurs approximatives de la distribution de la distribution exponentielle pour des valeurs de x données.(voir figure 3.6 en annexe).

Nous obtenons le tableau suivant :

x	0	1	2	3	4	5	6	7	8	9	10
$f_X(x)$	0.88442	0.0613	0.0372	0.0225	0.0137	0.0083	0.0050	0.0031	0.0019	0.0011	0.0017

TABLE 3.2 – Valeurs de discrétisation de la loi exponentielle

D'après le tableau, l'espérance $E[X]$= 0.3933 et la variance $Var[X] = 1.3972$ au lieu de 0.4 et 1.4 dans l'exemple 2.2.1 précédent.

Nous pouvons à présent appliquer la méthode de convolution pour les lois discrètes sur notre vecteur. A l'aide du code MATLAB (Voir figure 3.7 en annexes). Les valeurs du tableau suivant sont celles de la fonction de répartition F_X. En effet une fois après avoir effectuer une convolution à l'ordre $n = 500$, de la fonction densité du tableau précédent, on détermine les valeurs de la fonction de répartition associée aux points de discrétisation allant de 110 à 300 au pas de 10. Les résultats sont dans le tableau suivant :

x	110	120	130	140	150	160	170	180	190	200
$F_X(x)$	0.0001	0.00008	0.0035	0.0121	0.0345	0.0810	0.1613	0.2772	0.4194	0.05697
x	210	220	230	240	250	260	270	280	290	300
$F_X(x)$	0.7074	0.8181	0.8968	0.9465	0.9746	0.9890	0.9956	0.9984	0.9994	0.9998

TABLE 3.3 – Valeurs de discrétisation de la fonction de répartition

Ce tableau présente la distribution de la charge sinistre totale S_n par la méthode de convolution en des points x donnés.

♠ Implication Pédagogique ♠

Ce mémoire a été rédigé dans le cadre de l'obtention du diplôme de professeur de l'enseignement secondaire. Ainsi nous ne pouvons faire ce travail sans toute fois mentionner son apport dans le sens de la pédagogique. L'intérêt pédagogique de ce mémoire est vu sous plusieurs angles :

- Il permet de maitriser les outils informatiques tels que l'ordinateur, qui est un outil essentiel pour un enseignant du 21-ième siècle avec l'avenue des T.I.C.E (technique de l'information et de la communication de l'enseignement) dans l'enseignement.
- La maitrise des logiciels informatiques, est un très grand avantage pour un enseignant qui sait les utiliser. Effet ce mémoire est rédigé en LATEX, qui est un logiciel qui permet de produire de bonnes épreuves de Mathématiques. De plus dans ce travail il a été question pour nous d'écrire certains codes informatiques avec les logiciels R et MATLAB, pour appliquer les formules et de tracer des courbes. Ce qui permettra l'enseignant d'être efficace sur le terrain lorsqu'il s'agira d'utiliser les T.I.C.E .
- Elle permet de mieux connaitre le champs d'application de certaines matières (enseignées au secondaires) dans le monde l'emploi. Prenons l'exemple de ce mémoire qui, présente la détermination de la charge sinistre totale en assurance. Pour cela, nous avons tout d'abord présenter les outils Mathématiques qu'utilise un actuaire. Ces outils mathématiques sont entre autres : la probabilité et la statistique qui, sont des matières qui figurent dans le programme d'enseignement au secondaire. La maitrise de ce thème permettra au futur enseignant de mieux préparer un cours sur la statistique ou la probabilité suivant le système d'approche par compétence (**A.P.C**).

♠ conclusion ♠

Dans le cadre de ce mémoire, nous avons été amenés à rappeler les outils de probabilités qui permettent de mieux appréhender le domaine des mathématiques de l'assurance non-vie. Cela nous a conduit à faire un bref rappel sur les notions des fonctions de répartitions, densités de probabilités, fonctions de queue et des fonctions génératrices. Pour permettre à l'assureur de bien maîtriser la théorie du risque liée à son activité, l'actuaire devra connaitre quelques lois de probabilités utiles à la modélisation des sinistres. C'est pourquoi il a aussi été question dans ce travail de présenter des lois discrètes (Poisson,binomiale et binomiale négative) qui permettent de modéliser les nombres de sinistres N d'une part et d'autre part des lois continues (Normale, Exponentielle, Pareto, Weibull, Gamma, Lognormale), qui permettent la modélisation du montant d'un sinistre X. Il est à noter que, la modélisation de la charge totale des sinistres S_n, aussi souvent appelée charge sinistres totale exige une bonne maîtrise des techniques des lois mélanges et des lois composées.Concernant le modèle individuel en particulier, nous avons présenté les distributions mélanges avec quelques exemples du point de vue mathématique. La distribution de la charge totale des sinistres S_n s'obtient au moyen des méthodes d'approximations (Approximation Normales, Approximations Log-normale, Approximation Gamma translaté, Approximation Normale-Puissance ou *Normale Power*), de la formule récursive de De Pril et enfin de la formule de convolution. En application, nous avons utilisé les codes R pour approximer la charge totale sinistres S_n d'un portefeuille de plusieurs polices d'assurances. du code MATLAB pour modéliser les valeurs de la charge sinistre totale par la méthode de convolution. L'assureur en analysant les données d'un portefeuille d'assurance telles que la fréquence de sinistres, le coût d'un sinistre et la charge totale sinistres grâce aux outils de probabilité, pourra être à mesure de prendre des dispositions utiles pour prévoir les provisions pour risques de sa compagnie d'assurance.

Nous pensons néanmoins qu'une application des résultats de ce travail aux données réelles auprès d'une compagnie d'assurance de la place soulèverait, sans doute, des réflexions plus intéressantes ; c'est pourquoi nous proposons que des futurs travaux soient orientés dans ce sens si possible.

♠ Bibliographie ♠

[1] **Arthur Charpentier, Christophe Dutang**, (Décembre 2012) L'Actuariat avec R-Version Numérique 1-34. France : http//creativecommons.org/licenses/by-sa/3.0/fr/, 215 pages.

[2] **Yadolah Dodge** (2007) Statistique Dictionnaire encyclopédique. Paris, Springer-Verlag, 633 pages.

[3] **C.Fiszka** (2013) Cours d'introduction aux Probabilités, notes de cours POLYTECH Paris-UMPC, Université Paris VII.

[4] **Rob Kass, Marc Goovaerts, Jan Dhaene, Michel Denuit** (2009) Modern Actuarial Risk Theory Using R Second Edition, ISBN : 978-642-03407-7 London Springer Heidelberg Dordrecht London, 381 pages.

[5] **Pierre Lafaye de Micheaux, Rémy Drouillet, Benoît Liquet** (2011) Le logiciel R Maitriser le langage Effectuer des analyses statistiques. Paris, Springer, 527 pages.

[6] **Stéphane Loisel** (2011) Cours de gestion des risques d'assurances et de théorie de la ruine, notes de cours université paris II.

[7] **Nchoutsu Ndam Félix Aimé** (2014) Modélisation de la charge totale sinistre sur un portefeuille de polices d'assurances. Mémoire de DIPES II de mathématiques. Université de Yaoundé I, Ecole Normale Supérieure, Cameroun.

[8] **NGUYEN Thi To-Vong** (Juillet 2008) Flottes automobiles : Un nouveau modèle de tarification. Impact de la conservation sur la distribution du ratio sinistres à primes. Mémoire de master, CNIAM.

[9] **Louis Thibaut** (Septembre 1997) Méthodes d'évaluation de la distribution du montant total des réclamations. mémoire Maître des sciences, Université Laval, Faculté de sciences et de génie, Québec.

[10] **Jean-Yves Tourneret** (2016) Thème 1 : Analyse et Synthèse de l'information, Université de Toulouse, ENSEEIHT-IRIT-TéSA, Notes de cours.

[11] **Yiu-Kuen Tse (2009)** Nonlife Actuarial Models theory, Methods and Evaluation. Cambridge, Cambridge University Press, 542 pages.

♠ Annexes ♠

```r
### code courbes des  densités discrètes ###
x<-seq(0,10, by=1)
x
y<-dbinom(x,10,1/2)
y
y1<-dpois(x, 5, log=FALSE)
y1
y2<-dnbinom(x,5, 1/2)
y2
leg.txt <- c("BN(10,1/2)","P(5)","NB(5,1/2)")
plot(x, y, xlab="x", ylab="f(x)", main="Comparaison des densités", ylim=range(y, y1,y2
lines(x,y1, col="red", lty=2)
lines(x,y, col="BLACK", lty=1)
lines(x,y2, col="green", lty=3)
legend("topleft",leg=leg.txt, col="black",lty=1)
```

FIGURE 3.3 – code R courbe des densités de probabilité discrètes

```r
#code desités continues##
x <- seq(0,6,.01)
x
y <- dlnorm(x, -1/2, 1)
y
y2 <- dgamma(x, 2, 2)
y2
y3 <- dweibull(x, 2, 2/sqrt(pi))
y3
y4<- dexp(x,1)
y4
leg.txt <- c("LN(-1/2,1)","G(2,2)","W(2,2/sqrt(pi))","exp(x,1)")
plot(x, y, xlab="x", ylab="f(x)", main="Comparaison des densités",ylim=range(y, y2, y3,y4), col="black", typ
lines(x,y2, lty=2)
lines(x,y3, lty=3)
lines(x,y4, lty=4)
legend("topright",leg=leg.txt, col="black",lty=1:4)
```

FIGURE 3.4 – code R courbe densités des probabilités continues

```r
#code aproximation des probabilités##
x<-3.5
mu<-1
sig<-1
gam<-1
z<-(x-mu)/sig
z
c<-pbiom(x, 1000, 0.001)
c
e<-ppois(x,1)
e
f<-pnorm(z)
f
d<-pnorm(sqrt(9/gam^2+6*z/gam+1))
d
s<-pgamma(x-(mu-2*sig/gam), 4/gam^2, 2/gam/sig)
s
```

FIGURE 3.5 – code R approximation des probabilités

```r
##discretisation de la loi exponentielle###
x<-seq(1,9, by=1)
x
lam<-0.5
lam
w0<-.8+0.2*(1- exp(-0.5*lam)
w(0)
w<-0.2*(exp(-lam*(x-0.5)) - exp(-lam*(x+0.5)))
w10<-0.2*exp(-0.5*lam)
w(10)
z<-dexp(x, lam)
c<-hist(w)    ###tracage de l'histogramme##
leg.txt <- c("pexp(x,lam)")
plot(z, xlab="z",ylab="dexp",mean="loi exponentielle",col="black")
lines(x,z,col="red",lty=2) ##tracage de la courbe###
legend("topright"leg=leg.txt,col="black",lty=1)
```

FIGURE 3.6 – code R discrétisation de la loi exponentielle

```
%%code_pour_la_convolution_dordre_N
function [c]=rous(f,x, N)
M= length(x);
P= length(f);
    c=zeros(M,1);
    nb=0; k=1;
    if N==2
    while(nb~=N-1)
        i=0;k=1;
        while(i~=M)
        a=x(k);
        for n=0:a
            c(k)=c(k)+f(a-n+1)*f(n+1);
        end
        k=k+1; i=i+1;
        end
```

```
    nb=nb+1;
    end
    else
        i=0;k=1;
        while(i~=M)
        a=x(k);
        for n=0:a
            c(k)=c(k)+f(a-n+1)*f(n+1);
        end
        k=k+1; i=i+1;
        end
        nb=1;
         while(nb~=N-1)
            g=c
            c=zeros(M,1);
            i=0;k=1;
        while(i~=M)
        a=x(k);
        for n=0:a
            c(k)=c(k)+f(a-n+1)*g(n+1);
        end
        k=k+1; i=i+1;
```

```
        nb=nb+1;
        end
        end
    end
```

FIGURE 3.7 – code MATHLAB pour la formule de convolution