

ASSURANCE NON VIE

Le modèle collectif

Support de cours 2003-2004

Frédéric PLANCHET

Octobre 2003

SOMMAIRE

1. ASSURANCE NON-VIE - LE MODELE COLLECTIF CLASSIQUE	3
1.1. HYPOTHESES ET PRESENTATION DU MODELE.....	3
1.2. LES MOMENTS DE S , LA CHARGE SINISTRE	4
1.3. DISTRIBUTIONS POISSON COMPOSEES	5
1.4. DISTRIBUTIONS CLASSIQUES	6
1.4.1. <i>Nombre de sinistres</i>	6
1.4.2. <i>Coût des sinistres</i>	9
2. DETERMINATION DE LA LOI DE S.....	15
2.1. FONCTION DE REPARTITION DE S , AVEC X A VALEURS DANS \mathfrak{R}	15
2.2. FONCTION DE REPARTITION DE S , AVEC X A VALEURS DANS N	15
2.3. LA FORMULE RECURSIVE DE PANJER.....	16
2.4. LA FFT (FAST FOURIER TRANSFORM)'.....	18
2.4.1. <i>La transformée de Fourier inverse</i>	18
2.4.2. <i>L'algorithme de la FFT</i>	18
2.5. ALGORITHME DE HECKMAN-MEYERS	19
2.6. LES METHODES DE MONTE-CARLO	20

1. ASSURANCE NON-VIE - LE MODELE COLLECTIF CLASSIQUE

Le Modèle Collectif est un modèle très utilisé en assurance non-vie. L'assurance non-vie regroupe les opérations d'assurance qui n'ont pas pour objet la vie de l'assuré. Elle est donc principalement composée des assurances de choses ou de biens, des assurances de responsabilité ou de dettes, et des assurances de personnes. La principale différence opposant l'assurance non-vie à l'assurance vie est la survenance même du sinistre, qui est le plus souvent certain en assurance vie, tandis qu'il est juste probable (avec une probabilité comprise entre 0 et 1) en assurance non-vie. De plus, en assurance non-vie, le coût du sinistre est rarement connu, ce qui est encore une particularité propre.

C'est pourquoi, un modèle de sinistralité d'un portefeuille en assurance non-vie doit permettre de simuler à la fois la fréquence de sinistres sur le portefeuille, et aussi le coût de ceux-ci. C'est l'objet même du Modèle Collectif, extrêmement usité en actuariat.

1.1. HYPOTHESES ET PRESENTATION DU MODELE

On définit une classe C de contrats, dont le cardinal est compris entre 1 et le nombre de contrats du portefeuille d'une entreprise d'assurance dans une branche donnée. Les contrats sont indistincts dans cette branche, c'est à dire que l'on considère qu'ils sont tous identiques (d'où l'importance du choix des variables tarifaires pour toute segmentation d'un portefeuille existant).

Dans le modèle collectif (ou composé), la charge sinistre S de C est fonction du nombre de sinistres N et des montants de ceux ci X_i . Formalisons cette définition :

Définition : Charge sinistre de la classe C dans le modèle collectif

- ✓ Soit S la charge sinistre de la classe C .
- ✓ Soit N la variable aléatoire à valeurs entières représentant le nombre de sinistres affectant C sur une période de temps $[0; t]$
- ✓ Soit la suite $X = (X_i)_{i \geq 0}$ des variables aléatoires réelles représentant les montants individuels de sinistre dans l'ordre de survenance, avec $X_0=0$, pour envisager simplement le cas où N réalise 0.

Alors dans le modèle collectif, on a

$$S = \sum_{i=0}^N X_i$$

Pour utiliser facilement ce modèle, deux hypothèses (très fortes) doivent être faites :

- ✓ Indépendance entre la fréquence et le coût des sinistres, c'est à dire que les variables aléatoires N et $(X_i)_{i \geq 1}$ sont supposées indépendantes.
- ✓ Indépendance et stationnarité des montants de sinistres. Les $(X_i)_{i \geq 1}$ sont indépendantes, et surtout ne varient pas avec l'effet du temps.

Remarque : Ces deux hypothèses amènent plusieurs commentaires. En effet, il est évident qu'avec l'inflation, par exemple, les coûts de sinistres sont fortement impactés sur une durée longue, ce qui contredit l'hypothèse de stationnarité. Elle peut tout de même s'appliquer avec actualisation des montants, mais encore faut-il trouver le taux d'actualisation adéquat. De même, l'hypothèse d'indépendance des coûts de sinistres est parfois très forte : il n'est pas rare qu'une jurisprudence motive le montant des indemnités des affaires futures (cf. cas de l'amiante). La première hypothèse n'est quant à elle vérifiée que si le portefeuille étudié est vraiment homogène.

On montre que (voir section 2 ci-dessous) la fonction de répartition de S est :

- Si X est à valeurs dans \mathfrak{R} :

$$F_S(x) = \sum_{n=\hat{a}}^{+\infty} P(N=n) F^{*n}_x(x)$$

Dans le cas où X est à valeurs réelles, c'est le seul résultat qui permette de connaître la distribution de S . Ce calcul analytique de la distribution de S n'est que très rarement faisable, le calcul de la convoluée de la fonction de répartition de X étant très complexe.

- Si X est à valeurs dans \mathbb{N} :

$$P(S=x) = \sum_{n=\hat{a}}^{+\infty} P(N=n) P^{*n}_x(x)$$

Ainsi, il est possible de calculer les probabilités $P(S=x)$, mais de même que précédemment, il est très difficile de résoudre ces problèmes lorsque il y a un grand nombre de sinistres. C'est pourquoi les méthodes analytiques se sont attachées à obtenir de résultats facilement utilisables sur les moments de S , principalement dans un objectif de calculs de primes.

1.2. LES MOMENTS DE S , LA CHARGE SINISTRE

Lorsque l'existence du premier moment de N et de X est assurée, on a :

$$E(S) = E(N)E(X)$$

Si les deux premiers moments de N et de X existent, alors en appliquant l'équation de décomposition de la variance on obtient :

$$V(S) = E(N)V(X) + V(N)E^2(X)$$

Ces formules, extrêmement usitées dans le passé, expriment ainsi (pour la première) l'égalité entre Prime pure $E(S)$, et Fréquence Moyenne $E(N)$ fois Coût Moyen $E(X)$. C'est le seul résultat facilement utilisable dans le modèle collectif, si on n'effectue pas de simulations informatiques. Lorsqu'il s'agit de calculer des primes d'assurance, ce résultat est suffisant, en déterminant la fréquence et le coût moyen des sinistres. Par contre, il est impossible d'utiliser les calculs analytiques présentés ici pour calculer la distribution de S , ce qui empêche aussi les calculs de probabilités de ruine, de charges de réassurance... Un cas particulier simplifiant énormément les calculs a été néanmoins envisagé, celui des distributions Poisson Composées.

1.3. DISTRIBUTIONS POISSON COMPOSEES¹

Définition : On appelle **distribution Poisson Composée** une distribution composée dans laquelle la v.a.r. N est de loi de Poisson.

Dans la suite, hypothèse est faite que $N \sim P(\lambda)$, loi de Poisson de paramètre $\lambda > 0$.

Propriété :

- La fonction de répartition d'une distribution Poisson Composée est :

$$F_S(x) = e^{-\lambda} \sum_{n=0}^{+\infty} \frac{\lambda^n}{n!} F_X^{*n}(x)$$

- La fonction génératrice des moments de la charge sinistre, si celle de X existe, est :

$$M_S(s) = \exp(\lambda(M_X(s) - 1))$$

- La fonction caractéristique de la distribution est :

$$\varphi_S(s) = \exp(\varphi(M_X(s) - 1))$$

D'autres résultats existent sur les distributions Poissons Composées, concernant notamment l'additivité de la famille des lois Poisson Composées, et sur l'agrégation de classes indépendantes de risques Poissons-Composées, mais ce n'est pas réellement l'objet de notre étude. Le but est de pouvoir obtenir de façon complètement déterminée la distribution de la charge de sinistre S , et c'est l'objet de la formule récursive de Panjer dans le cas particulier des distributions Composées (cf. 2.3 ci-dessous).

Cette formule se place dans le cas particulier de coûts de sinistres entiers, c'est à dire que X est à valeurs dans \mathbb{N} . Nous verrons par la suite qu'il est possible d'utiliser cette formule dans le cas où X est à valeurs dans \mathbb{R} , par le biais de méthodes d'arrondis.

¹ Besson J.L., Partrat C., « Assurance Non Vie : Modélisation, Simulation », Economica (Ouvrage à paraître).

1.4. DISTRIBUTIONS CLASSIQUES

1.4.1. Nombre de sinistres

1.4.1.1. Loi de Poisson

Définition : $(p_n)_{n \geq 0}$ est distribuée selon une loi de **Poisson** $P(\lambda)$ de paramètre $\lambda > 0$, si pour tout $n \geq 0$, n entier,

$$p_n = P(N = n) = e^{-\lambda} \frac{\lambda^n}{n!}$$

La loi de Poisson et la loi Exponentielle sont intimement liées :

Propriété : Soit V_1, V_2, \dots une suite de v.a.r. i.i.d. suivant une loi exponentielle 1. Si N est le plus petit entier k positif tel que :

$$\sum_{i=1}^{k+1} V_i > \lambda,$$

alors N suit une loi de Poisson.

Avec $\lambda > 0$ comme unique paramètre, cette loi de probabilité est facilement estimable ; elle reste la loi fondamentale de l'assurance non-vie. Ses inconvénients sont une queue de distribution légère, et l'égalité moyenne-variance :

$$E(N) = \lambda \text{ et } V(N) = \lambda$$

Inférence statistique

Si (n_1, \dots, n_v) est la réalisation d'un v -échantillon (N_1, \dots, N_v) de N , la méthode des moments et celle du maximum de vraisemblance conduisent à l'estimation par la moyenne empirique

$$\hat{\lambda} = \overline{N}$$

Cet estimateur est sans biais et converge par la loi des grands nombres. Il est de plus efficace. En effet, la matrice d'information de Fisher nous donne le résultat suivant :

$$I(\lambda) = V \left(\frac{\partial \ln L}{\partial \lambda} \right) = \frac{v}{\lambda}$$

On peut donner un indicateur de la précision d'estimation de λ à partir de la variance de cet estimateur, l'erreur standard vaut :

$$\sigma(\hat{\lambda}) = \sqrt{\frac{\lambda}{v}} \text{ qui est estimée par } \sigma(\hat{\lambda}) = \sqrt{\frac{\hat{\lambda}}{v}}$$

1.4.1.2. Loi binomiale

Définition : $(p_n)_{n \geq 0}$ est distribuée selon une loi **Binomiale** $B(a,p)$ de paramètres $a \in \mathbb{N}^*$ et $p \in [0;1]$, si pour tout n entier avec $n \in \{0, \dots, a\}$,

$$p_n = P(N = n) = C_a^n p^n q^{a-n},$$

avec $q=1-p$.

Propriété : Semi-additivité

Soient deux v.a.r indépendantes B_1 et B_2 , B_1 suivant $B(a_1,p)$ et B_2 suivant $B(a_2,p)$. Alors la var $B=B_1+B_2$ est distribuée selon la loi $B(a_1+a_2,p)$.

1.4.1.3. Loi binomiale négative

Définition : $(p_n)_{n \geq 0}$ est distribuée selon une loi **Binomiale Négative** $B(r,p)$ de paramètres $r > 0$ et $p \in [0;1]$, si pour tout n entier avec $n \in \mathbb{N}$,

$$p_n = P(N = n) = \frac{\Gamma(r+n)}{\Gamma(r)n!} p^r q^n,$$

avec $q=1-p$, et

$$\frac{\Gamma(r+n)}{\Gamma(r)} = \begin{cases} (r+n-1) \dots r & \text{si } n \geq 1 \\ 1 & \text{si } n=0 \end{cases}$$

La loi Binomiale Négative est en réalité une distribution Poisson-mélange. Cette propriété est à la base de la méthode de simulation de la distribution Binomiale Négative.

Les deux premiers moments sont donnés par :

$$E(N) = r \times \left(\frac{1-p}{p} \right) \text{ et } V(N) = r \times \left(\frac{1-p}{p^2} \right) > E(N)$$

Inférence statistique

Soit (n_1, \dots, n_v) la réalisation d'un v -échantillon (N_1, \dots, N_v) de N , la méthode des moments et celle du maximum de vraisemblance ne conduisent plus à la même estimation. La méthode du *maximum de vraisemblance* nécessite en plus d'un développement analytique compliqué, une résolution numérique pour déterminer l'estimateur. Dans un souci de simplification, on choisira d'utiliser la méthode des moments.

On remplace donc dans les formules précédentes l'espérance et la variance, par \bar{N} et \bar{s} (qui sont respectivement la moyenne et la variance empirique débiaisées). Ces deux estimateurs sont convergents.

L'inconvénient principal de la méthode des moments est que les estimateurs qu'elle fournit sont en général assez peu précis, et qu'il est difficile d'étudier leur loi autrement que par simulation.

On estime donc les paramètres \mathbf{p} et \mathbf{r} par

$$\hat{\mathbf{p}} = \frac{\bar{N}}{\bar{s}} \quad \text{et} \quad \hat{\mathbf{r}} = \frac{(\bar{N})^2}{\bar{s} - \bar{N}}$$

1.4.1.4. Distribution de Poisson-mélange

Définition : Distributions Poisson-mélange

Soit un couple (Λ, N) de v.a.r., avec $\Lambda > 0$, pour lequel sont connus :

- ✓ La distribution conditionnelle de N / Λ : loi de Poisson
- ✓ La distribution marginale de Λ , une fonction de répartition F .

La loi marginale de N a pour expression :

$$p_n = p_n(F) = P(N = n) = \int_0^{+\infty} P(N = n / \Lambda = \lambda) dF(\lambda) = \int_0^{+\infty} e^{-\lambda} \frac{\lambda^n}{n!} dF(\lambda)$$

avec $p_n > 0$ pour tout $n \in \mathbb{N}$. La distribution entière ainsi obtenue est une distribution Poisson F -mélange.

Propriété : Une distribution Binomiale Négative (r, p) est une distribution Poisson-Gamma $(r, \frac{p}{q})$.

Preuve : d'après la définition précédente, on a :

$$\begin{aligned} P(N = n) &= \int_0^{+\infty} P(N = n / \Lambda = \lambda) f_{\Lambda}(\lambda) d\lambda \quad \text{en notant } f_{\Lambda} \text{ la fonction de densité de } \Lambda \\ &= \frac{(p/q)^r}{\Gamma(r)} \int_0^{+\infty} \frac{e^{-\lambda} \lambda^n}{n!} e^{-\frac{p}{q}\lambda} \lambda^{r-1} d\lambda \\ &= \frac{(p/q)^r}{\Gamma(r)n!} \int_0^{+\infty} e^{-(1+\frac{p}{q})\lambda} \lambda^{n+r-1} d\lambda \end{aligned}$$

$$= \frac{(p/q)^r \Gamma(r+n)}{\Gamma(r)n! (1/q)^{r+n}} p^r q^n, \text{ où l'on reconnaît une distribution BN}(r,p).$$

1.4.2. Coût des sinistres

1.4.2.1. Loi Normale

Définition : Fonction de densité de la loi Normale

La fonction de densité f de la loi Normale est donnée par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\left(\frac{x-\mu}{\sigma}\right)^2}{2}}$$

1.4.2.2. Loi Log Normale LN($\mu; \sigma^2$)

Cette loi possède deux paramètres μ réel, et $\sigma > 0$. Cette loi est très utilisée en assurance non-vie. C'est la transformée exponentielle de la loi $N(\mu; \sigma^2)$. En effet X suit $LN(\mu; \sigma^2)$ si, et seulement si, $\ln(X)$ est distribuée selon $N(\mu; \sigma^2)$.

Les principales propriétés de cette loi sont les suivantes :

$$\text{Densité : } f(x) = \frac{1}{\sigma x \sqrt{2\pi}} \times e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} \quad \text{où } x > 0$$

$$\text{Fonction de répartition}^1 : F(x) = \Phi\left(\frac{\ln(x)-\mu}{\sigma}\right)$$

$$E(\ln(X)) = \mu \quad \text{ou} \quad E(X) = e^{\mu + \frac{\sigma^2}{2}}$$

$$V(\ln(X)) = \sigma^2 \quad \text{ou} \quad V(X) = e^{2\mu + \sigma^2} \times (e^{\sigma^2} - 1)$$

Remarque : On peut aussi utiliser une loi Lognormale dite « translatée » de t , où $t > 0$. On obtient alors l'équivalence suivante :

$$X \sim \text{LN}(t, \mu, \sigma) \Leftrightarrow \frac{\ln(X-t)-\mu}{\sigma} \sim N(0,1)$$

Inférence statistique

¹ Φ est la fonction de répartition de la loi $N(0,1)$

Si (x_1, \dots, x_v) est la réalisation d'un v -échantillon (X_1, \dots, X_v) de X , la méthode du maximum de vraisemblance permet d'estimer les paramètres.

$$\hat{\mu} = \frac{1}{v} \times \sum_{i=1}^v \ln(x_i)$$

$$\hat{\sigma}^2 = \frac{1}{v} \times \sum_{i=1}^v \left(\ln(x_i) - \hat{\mu} \right)^2$$

Ces deux estimateurs sont convergents, et celui de σ est asymptotiquement sans biais.

Le couple $\left(\hat{\mu}, \frac{v}{v-1} \hat{\sigma} \right)$ est un estimateur sans biais de (μ, σ) .

De plus, l'erreur asymptotique de chaque estimateur est donnée par la suite :

$$\sigma_{as}(\hat{\mu}) = \frac{\sigma}{\sqrt{v}}$$

$$\sigma_{as}(\hat{\sigma}) = \sigma^2 \sqrt{\frac{2}{v}}$$

$$\frac{\sigma_{as}(\hat{\mu})}{\hat{\mu}} = \frac{\hat{\sigma}}{\sqrt{v}} \times \left(1 + \frac{\hat{\sigma}^2}{2} \right)^{\frac{1}{2}}$$

Remarque : Pour le cas de la Lognormale translatée on a les estimations suivantes

$$\hat{t} = \min_i(x_i)$$

$$\hat{\mu} = \frac{1}{v} \times \sum_{i=1}^v \ln \left(\frac{x_i}{\hat{t}} \right)$$

$$\hat{\sigma} = \left[\frac{1}{v} \times \sum_{i=1}^v \left(\ln \left(\frac{x_i}{\hat{t}} \right) \right)^2 - \left(\frac{1}{v} \times \sum_{i=1}^v \ln \left(\frac{x_i}{\hat{t}} \right) \right)^2 \right]^{1/2}$$

1.4.2.3. Loi de Pareto $P(a, a)$

Cette loi très couramment utilisée en réassurance, dépend d'un paramètre $\alpha > 0$, de forme, et d'un paramètre $a > 0$ de position (le seuil de prise en compte des montants de sinistres).

Définition : Fonction de répartition de la loi de Pareto

La fonction de répartition F de la loi de Pareto est donnée par :

$$F(x) = \begin{cases} 1 - \left(\frac{a}{x}\right)^\alpha, & x > a \\ 0, & x \leq a \end{cases}$$

De même que pour la loi Exponentielle, il est possible de calculer analytiquement la fonction réciproque de la fonction de répartition, sur l'ensemble $x \in]a; +\infty[$:

$$F^{-1}(u) = \frac{a}{\sqrt[\alpha]{1-u}} \text{ si } u > 0.$$

Preuve : par définition de la fonction réciproque, on a :

$F \circ F^{-1}(u) = u$. On se limite à $x > a$, c'est à dire à $F(x) > F(a) = 0$, c'est à dire $u > 0$.

$$\Leftrightarrow 1 - \left(\frac{a}{F^{-1}(u)}\right)^\alpha = u \text{ (en remplaçant } F \text{ par son expression)}$$

$$\Leftrightarrow \left(\frac{a}{F^{-1}(u)}\right)^\alpha = 1 - u$$

$$\Leftrightarrow F^{-1}(u) = \frac{a}{\sqrt[\alpha]{1-u}}$$

Les deux premiers moments sont donnés par :

$$E(X) = \frac{\alpha t}{(\alpha-1)} \text{ avec } \alpha > 1 \quad V(X) = \frac{\alpha t^2}{(\alpha-1)^2(\alpha-2)} \text{ avec } \alpha > 2$$

Remarque : On a la relation d'équivalence suivante :

$$X \sim P(t, \alpha) \Leftrightarrow \ln\left(\frac{X}{t}\right) \sim \mathcal{E}(\alpha)$$

Inférence statistique

Si (x_1, \dots, x_n) est la réalisation d'un n -échantillon (X_1, \dots, X_n) de X , l'existence des moments n'étant pas assurée, la méthode des moments est à proscrire. On estime que le paramètre α par la méthode de maximum de vraisemblance, car on suppose t donné :

$$\hat{\alpha} = \frac{n}{\ln\left(\frac{G}{t}\right)} \text{ avec } G = \left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}} \text{ la moyenne géométrique de l'échantillon}$$

Comme $2n \frac{\alpha}{\hat{\alpha}} = 2n \times \alpha \times \ln\left(\frac{G}{t}\right)$ suit une loi du χ_{2n}^2 , on $E(\hat{\alpha}) = \frac{n}{(n-1)}\alpha$. Cet estimateur est donc biaisé, asymptotiquement sans biais.

La matrice d'information de Fisher nous donne $I(\hat{\alpha}) = \frac{n}{\alpha^2}$. On en déduit que l'estimateur est asymptotiquement efficace. On a alors :

$$\sigma_{as}(\hat{\alpha}) = \frac{\alpha}{\sqrt{n}}$$

Cela permet de déterminer un intervalle de confiance de niveau asymptotique 95% pour α .

Remarque : Dans le cas où t n'est pas donné, on a alors pour estimateur de maximum de vraisemblance de t :

$$\hat{t} = \min_i(x_i)$$

Nota Bene : Rytgaard¹ a fourni un estimateur sans biais $\hat{\alpha}^* = \frac{n-1}{n} \hat{\alpha}$

1.4.2.4. Loi de Weibull $W(\tau, \alpha)$

Cette loi dépend d'un paramètre de forme $\tau > 0$, et d'un paramètre d'échelle $\alpha > 0$.

Définition : Fonction de répartition de la loi de Weibull

La fonction de répartition F de la loi de Weibull est donnée par : $F(x) = 1 - e^{-\left(\frac{x}{\alpha}\right)^\tau}$

De même que pour la loi Exponentielle, et la loi de Pareto, il est possible de trouver la fonction réciproque de la fonction de répartition :

$$F^{-1}(u) = \alpha \times (-\ln(1-u))^{\frac{1}{\tau}}$$

Preuve : par définition de la fonction réciproque, on a :

$$F \circ F^{-1}(u) = u$$

$$\Leftrightarrow 1 - \exp\left(-\left(\frac{F^{-1}(u)}{\alpha}\right)^\tau\right) = u \text{ en remplaçant } F \text{ par son expression.}$$

$$\Leftrightarrow -\left(\frac{F^{-1}(u)}{\alpha}\right)^\tau = \ln(1-u)$$

$$\Leftrightarrow F^{-1}(u) = \alpha \times (-\ln(1-u))^{\frac{1}{\tau}}$$

1.4.2.5. Loi Gamma $\gamma(v; \beta)$

¹ METTE RYTGAARD (1989) Estimation of the Pareto Distribution, (XXIème Colloque Astin à New York)

ν est un paramètre de forme et β un paramètre d'échelle ; ils doivent être strictement positifs.

Définition : Fonction de densité de la loi Gamma

La fonction de densité f de la loi Gamma est donnée par :

$$f(x) = \frac{\beta^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\beta x}$$

Définition : Fonction de répartition de la loi Gamma

La fonction de répartition F de la loi Gamma est donnée par :

$$F(x) = \Gamma(\nu, \beta x)$$

où $\Gamma(\nu; x)$ représente la fonction gamma incomplète.

Définition : Fonction gamma incomplète

Elle est définie sur $\mathbb{R}_+^* \times \mathbb{R}_+^*$ par :

$$\Gamma(\nu, x) = \frac{1}{\Gamma(\nu)} \int_0^x t^{\nu-1} e^{-t} dt$$

où $\Gamma(\nu)$ représente la fonction gamma appliquée à ν .

Comme le montrent les formules ci-dessus, il paraît difficile d'appliquer ici une méthode type inversion de la fonction de répartition. La procédure de simulation va donc se déduire de quelques propriétés de la loi, présentées ici.

Proposition : Une loi $E(\beta)$ est une loi $\gamma(1; \beta)$.

Preuve : Par calcul de la densité de $\gamma(1; \beta)$ qui est celle d'une loi $E(\beta)$.

Propriété : Semi-additivité de la loi Gamma.

Si X_1 et X_2 sont des v.a.r. i.i.d. distribuées selon $\gamma(\nu_1; \beta)$ et $\gamma(\nu_2; \beta)$, respectivement, alors la v.a.r. $X = X_1 + X_2$ suit $\gamma(\nu_1 + \nu_2; \beta)$.

Lemme : Si Y et Z sont des v.a.r. indépendantes distribuées selon $\beta f(\nu, 1-\nu)$, et $E(1)$, alors la

v.a.r. $X = \frac{1}{\beta} YZ$ suit $\gamma(\nu; \beta)$.

1.4.2.6. Loi Beta $\beta_I(a, b)$

Cette loi dépend de deux paramètres $a, b > 0$. Elle nécessite l'utilisation de la fonction Beta incomplète qui elle-même utilise la fonction Beta. Définissons tout d'abord ces deux fonctions.

Définition : Fonction Beta

Elle est définie sur $R_+^* \times R_+^*$ par :

$$\beta(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Définition : Fonction Beta incomplète

Elle est définie sur $R_+^* \times R_+^* \times [0;1]$ par :

$$\beta(a, b, x) = \frac{1}{\beta(a, b)} \int_0^x u^{a-1} (1-u)^{b-1} du$$

Définition : Fonction de répartition de la loi Beta

La fonction de répartition F de la loi Beta est donnée par :

$$F(x) = \beta(a, b, x)$$

où β représente la fonction Beta incomplète.

Définition : Fonction de densité de la loi Beta

La fonction de densité f de la loi Beta est donnée par :

$$f(x) = \frac{1}{\beta(a, b)} x^{a-1} (1-x)^{b-1} 1_{]0;1[}(x)$$

où β représente la fonction Beta et $1_{]0;1[}(x)$ représente la fonction indicatrice sur $]0;1[$ appliquée à x .

2. DETERMINATION DE LA LOI DE S

2.1. FONCTION DE REPARTITION DE S, AVEC X A VALEURS DANS \mathfrak{R}

La v.a.r. parente X des coûts de sinistres est considérée dans ce paragraphe à valeurs dans \mathfrak{R} .

Définition : Convoluée d'une fonction de répartition.

Soit F la fonction de répartition de X , et $n \in N$.

La $n^{\text{ième}}$ convoluée de F , F^{*n} , est définie par : $F^{*0}(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases}$ si $n=0$

Et $F^{*n}(x) = P(X_1 + \dots + X_n < x)$ avec X_1, \dots, X_n v.a.r. i.i.d. distribuées selon la loi de X , si $n > 0$.

Propriété : Soit x réel.

La $n^{\text{ième}}$ convoluée de F en x , $F^{*n}(x)$, vérifie la relation de récurrence suivante :

$$F^{*n}(x) = \int_0^x F^{*(n-1)}(x-z) dF(z), \quad n \geq 1.$$

L'objet de ces définitions est de connaître la distribution de S , ici par l'intermédiaire de sa fonction de répartition.

Propriété : Soit x réel strictement positif. Soient S , N et F^{*n} définies comme précédemment. Alors on a :

$$F_S(x) = \sum_{n=0}^{+\infty} P(N=n) F^{*n}(x)$$

2.2. FONCTION DE REPARTITION DE S, AVEC X A VALEURS DANS N

La v.a.r. parente X des coûts de sinistres est considérée dans ce paragraphe à valeurs dans N .

Définition : Convoluée de la loi de probabilité d'une distribution entière

Soit $P_X = P(X=x)$ la loi de probabilité de X , et $n \in \mathbb{N}$.

La $n^{\text{ième}}$ convoluée de P_X , P_X^{*n} , est définie par : $P_x^{*n}(x) = \begin{cases} 1, & x=0 \\ 0, & x \geq 1 \end{cases}$ si $n=0$

Et $P_X^{*n}(x) = P(X_1 + \dots + X_n = x)$ avec X_1, \dots, X_n v.a.r. i.i.d. distribuées selon la loi de X , si $n > 0$.

Propriété : Soit x entier.

La $n^{\text{ième}}$ convoluée de P_X en x , $P_X^{*n}(x)$, vérifie la relation de récurrence suivante :

$$P_x^{*n}(x) = \sum_{z=1}^x P_{x-z}^{*(n-1)} P_z, \quad n \geq 2.$$

Propriété : Soit x entier strictement positif. Soient S , N et P_x^{*n} définies comme précédemment. Alors on a :

$$P(S=x) = \sum_{n=0}^{+\infty} P(N=n) P_x^{*n}(x)$$

2.3. LA FORMULE RECURSIVE DE PANJER

Comme nous l'avons évoqué dans la présentation du modèle collectif, la formule récursive de Panjer permet d'obtenir de façon complètement déterminée la distribution de la charge de sinistre S dans le cas particulier des distributions Composées avec coûts de sinistres entiers.

Notons $\pi_i = P(X=i)$ avec $i \in \mathbb{N}$. S est à valeurs entières, et sa loi dépend donc des probabilités $P(S=x)$, avec $x \in \mathbb{N}$.

On note $P(N=n) = p_n$ pour tout $n \in \mathbb{N}$

Hypothèse: (H): $\exists a, b \in \mathbb{R}$ tels que $p_n = (a + \frac{b}{n}) p_{n-1}$

Les seules distributions vérifiant (H) sont dégénérées, Binomiale, Poisson ou Binomiale Négative.

Proposition : **Formule récursive de Panjer**

Sous (H) , on a :

$$P(S=0) = \sum_{n=0}^{+\infty} p_n \pi_0^n$$

$$P(S=x) = \frac{\sum_{y=1}^x \left(a + b \frac{y}{x} \right) \pi_y P(S=x-y)}{1 - a \pi_0}, x \geq 1$$

Cette formule peut être étendue au cas où l'hypothèse (H) n'est vérifiée qu'à partir d'un certain rang. De même, en utilisant des méthodes d'arrondis ou d'égalisation locale des moments, il est possible de rendre "entières" des distributions de coûts de sinistres. Nous ne présenterons pas ces méthodes, mais il semblait ici important de signaler leur existence.

L'objet de la présentation de cet algorithme était de montrer qu'il est réellement difficile d'obtenir la loi de S , par l'intermédiaire de sa fonction de répartition par exemple, dans le cas général des distributions composées. C'est justement dans cet objectif qu'ont été mises en place les méthodes stochastiques.

L'utilisation de méthodes stochastiques permet de réaliser le calcul de la distribution de S . Ainsi, il est possible d'utiliser de nombreux indicateurs de risque (probabilité de risque...), et de les calculer sur un portefeuille pour lequel on connaît les distributions de fréquences et de coûts de sinistres.

La connaissance de telles lois est la résultante de tests statistiques effectués sur des données d'expérience du passé du portefeuille, sur plusieurs années.

Détaillons l'utilisation de ces méthodes de simulation, dans le modèle composé. Tout d'abord on se place sous les hypothèses du modèle composé. La charge sinistre S , annuelle par exemple, est donnée par :

$$S = \sum_{i=0}^N X_i,$$

avec N v.a. à valeurs entières, représentant la fréquence de sinistres sur une durée donnée, annuellement par exemple, et $(X_i)_{i \geq 0}$ suite des variables aléatoires réelles représentant les montants individuels de sinistre dans l'ordre de survenance, de v.a.r. parente X , avec $X_0=0$, pour envisager simplement le cas où N réalise 0.

Une réalisation de S est donc obtenue à partir :

- d'une réalisation n de N ,

- de n réalisations $(x_i)_{1 \leq i \leq n}$, indépendantes et identiquement distribuées, de loi la loi de X .

Les hypothèses du modèle collectif proposent de prendre N et $(X_i)_{i \geq 0}$ indépendantes, idem à l'intérieur même des $(X_i)_{i \geq 0}$.

2.4. LA FFT (FAST FOURIER TRANSFORM)^{2,3}

2.4.1. La transformée de Fourier inverse

Soit φ_Y la fonction caractéristique d'une variable aléatoire réelle Y . Si Y est une v.a.r. arithmétique à valeurs dans $\{kh, k \in \mathbb{N}\}$ avec $h > 0$ fixé, on a :

$$P(Y=kh) = \frac{1}{2\pi} \int_0^{2\pi} e^{-itk} \varphi_Y\left(\frac{t}{h}\right) dt$$

Sauf rares exceptions, le calcul des intégrales fait appel à des techniques numériques. Il en existe 2 qui sont très utilisées parce qu'elles ont été développées pour le calcul d'intégrales de cette forme. Il s'agit de la FFT et de l'algorithme de Heckman-Meyers

2.4.2. L'algorithme de la FFT

La FFT se déroule en 4 étapes :

- Si nécessaire, on commence par « arithmétiser » la loi de Y (c'est à dire ramener les valeurs de Y dans $\{k \cdot h, k \in \mathbb{N}\}$). On partage ensuite l'intervalle $[0, 2\pi]$ en m sous-intervalles de longueur $2\pi/m$. Si on note φ_Y la fonction caractéristique de la v.a.r. donnant le coût d'un sinistre, on calcule le vecteur ϕ des ϕ_k (pour k de 0 à $m-1$) de la manière suivante :

$$\phi_k = \frac{1}{m} \sum_{j=0}^{m-1} e^{-\frac{2\pi}{m}ijk} \varphi_Y\left(\frac{2\pi j}{mk}\right)$$

- On calcule ensuite le vecteur φ en multipliant V , la matrice des $v_{jk} = e^{2\pi ijk/m}$ (pour j et k de 0 à $m-1$) par le vecteur ϕ :

$$\varphi = V^* \phi, \text{ noté } \varphi = \text{FFT}^+(\phi)$$

² Besson J.L., Partrat C., « Assurance Non Vie : Modélisation, Simulation », Economica (Ouvrage à paraître).

³ Robertson J.P., « The computation of aggregate loss distributions », Proceedings of the Casualty Actuarial Society, Vol. LXXIX, p. 57-133, 1992.

- Puis on calcule $\varphi' = g_N(\varphi)$ où g_N est la fonction génératrice de N le nombre de sinistres de la classe C .
- Enfin, on obtient la loi approchée de S en multipliant W , la matrice des $w_{jk} = e^{-2\pi i j k / m}$ (pour j et k de 0 à $m-1$) par le vecteur φ' :

$$\phi' = 1/m * W * \varphi', \text{ noté } \phi' = 1/m * \text{FFT}^-(\varphi')$$

L'ensemble de cette procédure se note : $1/m * \text{FFT}^-(g_N(\text{FFT}^+(\phi)))$

2.5. ALGORITHME DE HECKMAN-MEYERS⁴

Pour utiliser cet algorithme, on fait l'hypothèse que la distribution du coût des sinistres est linéaire par morceaux. Sa forme mathématique est la suivante :

- ✓ les « morceaux » sont délimités par les montants a_k , k de 0 à $m+1$, $a_0 = 0$
- ✓ les p_k sont les probabilités que le montant du sinistre soit entre a_k et a_{k+1}
- ✓ on définit les $d_k = p_k / (a_{k+1} - a_k)$

La simulation de S selon l'algorithme de Heckman-Meyers se déroule en 5 étapes :

- Etape 1 : on se donne les paramètres des distributions du nombre de sinistres et du montant des sinistres. Les paramètres de cette dernière sont les montants a_k , les probabilités p_k et les d_k . Les paramètres de la distribution du nombre de sinistres sont les différents nombres de sinistres envisageables. On se donne également deux paramètres, l'un dit de contagion, c , et un autre dit de mixage, b . Si $|c| < 10^{-7}$, alors on se donne $c = 10^{-7}$. Il en va de même pour b . Soient N et Z les lois du nombre de sinistres et du montant des sinistres. Appelons n l'espérance de N et z celle de Z .
- Etape 2 : on calcule l'espérance et la variance de S de la manière suivante :
 - $\mu = n * z$
 - $\sigma^2 = n E[Z^2] (1 + b) + n^2 E[Z]^2 (b + c + bc)$
- Etape 3 : on discrétise l'intervalle de calcul de l'intégrale. On partage cet intervalle en sous-intervalles de longueur $h = 2\pi\sigma / (\text{charge de sinistres maximale})$.
- Etape 4 : Pour chacun des intervalles d'intégration $[l, r]$ définis à l'étape 3, on considère les 5 points suivants :

⁴ Heckman P.E., Meyers G.G., « *The calculation of aggregate loss distributions from claim severity and claim count distributions* », Proceedings of the Casualty Actuarial Society, Vol. LXX, p. 22-61, 1983.

- $t_1 = (-0,90617985 (r - l) + (r + l)) / 2$
- $t_2 = (-0,53846931 (r - l) + (r + l)) / 2$
- $t_3 = (r + l) / 2$
- $t_4 = (0,53846931 (r - l) + (r + l)) / 2$
- $t_5 = (0,90617985 (r - l) + (r + l)) / 2$

On évalue alors $f(t_j)$ et $g(t_j)$ pour j de 1 à 5, où f et g sont construites de la manière suivante :

- soit $H(t) = \frac{1}{t} \sum_{k=1}^m d_k (\sin(t a_{k+1}) - \sin(t a_k)) + (1 - \sum_{k=1}^m p_k) \cos(t a_{m+1})$
- soit $K(t) = \frac{1}{t} \sum_{k=1}^m d_k (\cos(t a_k) - \cos(t a_{k+1})) + (1 - \sum_{k=1}^m p_k) \sin(t a_{m+1})$
- soit $h(t) = H(t/\sigma) - 1$
- soit $k(t) = K(t/\sigma)$
- alors $f(t) = [(1 - c * n * h(t))^2 + (c * n * k(t))^2]^{-1/2c}$
- alors $g(t) = \text{argument} [1 - c * n * (h(t) + i * k(t))]^{-1/c}$

On poursuit ces évaluations jusqu'à ce que $f(t_j) / t_j < 0.00002$ ou bien jusqu'à ce que l'on ait parcouru 256 intervalles.

➤ Etape 5 : on calcule pour chaque intervalle :

$$I = \left(\sum_{j=1}^5 W_j P(t_j) \right) * \text{longueurIntervalle} / 2$$

$$\text{où } P(t) = f(t) / t * \sin(tx/\sigma - g(t))$$

$$\text{et où } W_1 = W_5 = 0.23692689, W_2 = W_4 = 0.47862867, W_3 = 0.56888889.$$

$$\text{Alors } F(x) = 0.5 + (\text{somme de tous les } I) / \pi$$

2.6. LES METHODES DE MONTE-CARLO

Comme cela a été dit en introduction de cette section, dans le cadre du modèle collectif, la méthode de Monte-Carlo se décompose selon les étapes suivantes :

- ✓ On commence par tirer, dans la loi du nombre de sinistres, un nombre de sinistres N .
- ✓ Puis, N fois de suite, on tire, dans la loi du montant des sinistres, un montant de sinistre z .
- ✓ Pour obtenir une réalisation de la loi S de la charge de sinistres, il suffit de faire la somme des N nombres z tirés.

Pour obtenir la distribution de S , il faut réitérer ce processus autant de fois que cela est nécessaire à l'obtention de la distribution de la loi du nombre de sinistres.