Agrégation des risques avec CPINAR(1)

Li Zhu

École d'actuariat
Université Laval, Québec, Canada

li.zhu.1@ulaval.ca

3 février 2022

Faculté des sciences et de génie
École d'actuariat

Outline

Introduction

Description du projet

Ce document présente une grande partie du contenu de l'article [Chen and Hu, 2020], selon la compréhension de l'auteur.

Risk aggregation with dependence and overdispersion based on the compound Poisson INAR (1) process

Définition des variables

Définition $\boldsymbol{X}$: pertes des sinistres (iid)

Définition $\boldsymbol{W}$: pertes des sinistres du portefeuille à la période k avec

$$W_k = \sum_{j=1}^{N_k} X_{k,j}$$

Définition $\boldsymbol{S}$: montant cumulé des pertes pendant n périodes

$$S_n = W_1 + W_2 + \cdots + W_n$$

Définition $\boldsymbol{Z}$: montant cumulé actualisé des pertes pendant n périodes

$$Z_n = vW_1 + v^2 W_2 + \cdots + v^n W_n$$

Objectif

L'objectif de l'étude est d'appliquer le modèle CPINAR(1) sur les données des sinistres.

- Introduire la structure de dépendance pour le nombre de sinistres dans les différentes périodes via le modèle.
- **Trouver la fonciton génératrice des probabilités (PFG) pour le nombre de sinistres pendant k period (G_{L_n}).**
- Calculer la fonction de répartition et la fonction génératrice des moments (MGF) de S_n et Z_n.

Préparation mathématique

Fonction génératrice des probabilités

Soit X une variable aléatoire entière et positive, la fonction génératrice des probabilités (PGF) de X, noté par $G_x(r)$, est donnée par

$$
\begin{aligned}
G_x(r) = \mathbb{E}(r^x) &= \sum_{x=0}^{\infty} r^x \mathbb{P}(X = x) \\
&= \mathbb{P}(X = 0) + r\mathbb{P}(X = 1) + r^2\mathbb{P}(X = 2) + \cdots
\end{aligned}
$$

La fonction génératrice de probabilités est un outil utile pour traiter les variables aléatoires discrètes. Sa force particulière c'est qu'il nous donne un moyen facile de caractériser la distribution de $X + Y$ lorsque X et Y sont indépendants.

PGF de la loi Poisson

Pour $N \sim Poisson(\lambda)$,

$$
\begin{aligned}
G_N(s) &= \sum_{n=0}^{\infty} s^n \mathbb{P}(N=n) = \sum_{n=0}^{\infty} s^n \frac{\lambda^n}{n!} e^{-\lambda} \\
&= \sum_{n=0}^{\infty} e^{-\lambda} \frac{(\lambda s)^n}{n!} = e^{-\lambda} e^{\lambda s} \\
&= e^{(s-1)\lambda}
\end{aligned}
$$

PGF de la somme des variables

Soit $X_1, X_2, ... \in \mathbb{N}^+$ des v.a. i.i.d. et soit N une v.a. indépendante de X_i , et le PGF du X est noté par $G_X(s)$. Pour $W = X_1 + \cdots + X_N$, sa PGF est donc

$$\begin{aligned} G_W(s) &= \mathbb{E}(s^W) = \mathbb{E}(s^{X_1+\cdots+X_N}) \\ &= \mathbb{E}_N(\mathbb{E}(s^{X_1+\cdots+X_N} \mid N)) \\ &= \mathbb{E}_N(\mathbb{E}(s^{X_1} \cdots s^{X_N}) \mid N) \\ &= \mathbb{E}_N(\mathbb{E}(s^{X_1}) \cdots \mathbb{E}(s^{X_N})) \\ &= \mathbb{E}_N(G_X(s)^N) \\ &= G_N(G_X(s)) \end{aligned}$$

Processus autorégressif à valeurs entières de l'ordre 1 : INAR(1)

Soit N une variable aléatoire à valeurs entières non négatives, pour tout $\alpha \in [0,1]$, l'opérateur $\circ$ est défini par

$$\alpha \circ N = \sum_{i=1}^{N} \xi_i$$

où ξ_i est une suite des v.a. iid, indépendante de N avec

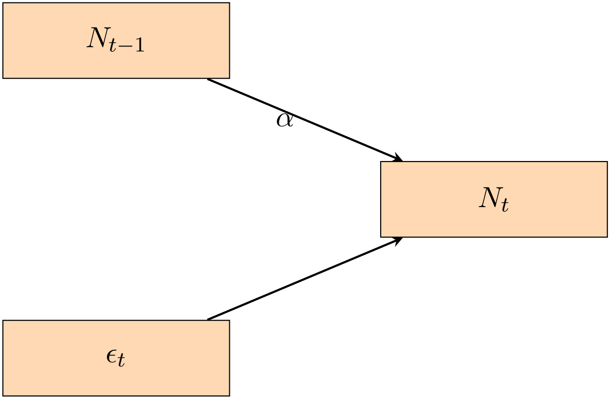
$$\mathbb{P}(\xi_i = 1) = 1 - \mathbb{P}(\xi_i = 0) = \alpha$$

Le INAR(1) processus est donné par

$$N_t = \alpha \circ N_{t-1} + \epsilon_t$$

où $\alpha \in [0,1]$ et ϵ_t est une suite des v.a. non négatives et non corrélées, avec une moyenne μ et variance finie σ^2.
La définition est donnée par [Al-Osh and Alzaid, 1987].

Schéma du processus INAR(1)



N_t est le nb de sinistres réclamés au temps t ;
N_{t-1} est le nb de sinistres réclamés au temps $t-1$;
ϵ_t est le nb de nouveaux sinistres réclamés au temps t ;
α est la probabilité qu'un sinistre réclamé au temps $t-1$ reste encore au temps t.

Fonction génératrice des probabilités pour INAR(1)

$$
\begin{aligned}
G_{N_k}(r) &= \mathbb{E}(r^{N_k}) = \mathbb{E}(r^{\alpha \circ N_{k-1} + \epsilon_k}) \\
&= \mathbb{E}(r^{\alpha \circ N_{k-1}})\mathbb{E}(r^{\epsilon_k}) \\
&= \mathbb{E}_{N_{k-1}}(r^{\sum_{i=1}^{N_{k-1}} \xi_i} \mid N_{k-1})\mathbb{E}(r^{\epsilon_k}) \\
&= \boldsymbol{G_{N_{k-1}}(G_{\xi}(r))G_{\epsilon}(r)} \\
&= G_{N_{k-1}}(1 - \alpha + \alpha r)G_{\epsilon_k}(r) \\
&= G_{N_{k-2}}[1 - \alpha + \alpha(1 - \alpha + \alpha r)]G_{\epsilon_{k-1}}(1 - \alpha + \alpha r)G_{\epsilon_k}(r) \\
&= G_{N_{k-2}}(1 - \alpha^2 + \alpha^2 r)G_{\epsilon_{k-1}}(1 - \alpha + \alpha r)G_{\epsilon_k}(r) \\
&= \cdots \\
&= G_{N_0}(1 - \alpha^k + \alpha^k r)\prod_{j=0}^{k-1} G_{\epsilon_j}(1 - \alpha^j + \alpha^j r)
\end{aligned} \tag{1}
$$

Processus de Poisson composé autorégressif à valeurs entières de l'ordre 1 : CPINAR(1)

Soit une v.a. N suit la loi Poisson avec la moyenne $\mu > 0$; Soit $U_1, U_2, ... \in \mathbb{N}^+$ des v.a. i.i.d. et indépendantes de N, mettons le PGF de U_i étant $H(r)$, sous la forme de

$$H(r) = h_1 r + \cdots + h_\nu r^\nu,$$

pour tout $h_j \geq 0, j = 1, ..., \nu$. Alors

$$\epsilon = U_1 + \cdots + U_N \sim CP_\nu(\mu, H)$$

Note : $CP_\nu(\mu, H)$ signifie la distribution Poisson composé.

Le processus CPINAR(1) est un processus INAR(1), si les innovations ϵ est i.i.d. $\sim CP_\nu(\mu, H)$ avec

$$N_t = \alpha \circ N_{t-1} + \epsilon_t \tag{2}$$

La distribution Poisson composé interprétation

Comments :
Je n'ai pas trouvé l'intuition d'utiliser la loi Poisson composé pour faire la modélisation dans l'article.

Une possible interprétation d'utiliser la loi Poisson composé est que le client qui réclame des sinistres peut avoir souscrit plusieurs polices d'assurance. (Voir [Zhang et al., 2014])

Fonction génératrice des probabilités pour CPINAR(1) - I

Quand $H'(1) < \infty$, $\{N_k\}$ a une distribution marginale stationnaire unique $\Pi \sim CP_\nu(\lambda, T)$. Voir [Schweer and Weiß, 2014].

$$\begin{aligned}
G_\Pi(r) &= \lim_{k\to\infty} G_{N_0}(1-\alpha^k+\alpha^k r)\prod_{j=0}^{k-1} G_\epsilon(1-\alpha^j+\alpha^j r) \\
&= G_{N_0}(1)\prod_{j=0}^{\infty} G_\epsilon(1-\alpha^j+\alpha^j r) \\
&= \prod_{j=0}^{\infty} G_\epsilon(1-\alpha^j+\alpha^j r) \\
&= exp\{\sum_{j=0}^{\infty} \mu[H(1-\alpha^j+\alpha^j r)-1]\}
\end{aligned}$$

avec $G_\epsilon(r) = exp\{\mu[H(r)-1]\}$.

Fonction génératrice des probabilités pour CPINAR(1) - II

Donc,on peut établir la relation entre λ, T et μ, H

$$\lambda(T(r)-1)=\sum_{i=0}^{\infty}\mu[H(1-\alpha^i+\alpha^i r)-1]$$

Quand $r=0$, on obtient

$$\lambda=\mu\sum_{i=0}^{\infty}[1-H(1-\alpha^i)]$$

Agrégation des risques

Définition S_n

Quelque rappels des concepts :

Définition S : montant cumulé des sinistres pendant n périodes

$$S_n = W_1 + W_2 + \cdots + W_n$$

Réflexion :

On travaille sur la somme des v.a. ***dépendantes****. L'utlisation directe de la PGF n'est pas très intéressante. Est-ce que la somme des Poissons composés dépendants donne encore un Poisson composé ?...*

Solution de l'article :

On transforme la question et on continue à travailler avec v.a. **indépedantes**. Donc, on redéfinit S_n par S_{L_n}, avec

$$S_{L_n} = \sum_{i=1}^{L_n} Y_i$$

où $L_n = N_1 + N_2 ... + N_n$ et Y_i est i.i.d. distribué comme X.

Fonction génératrice de moment de S_{L_n}

$$
\begin{aligned}
M_{S_{L_n}}(r) &= \mathbb{E}(e^{rS_{L_n}}) = \mathbb{E}(e^{r\sum_{i=1}^{L_n} Y_i}) \\
&= \mathbb{E}_{L_n}[\mathbb{E}(e^{r\sum_{i=1}^{L_n} Y_i} \mid L_n)] \\
&= \mathbb{E}_{L_n}[\mathbb{E}(e^{rY_1}) \cdots \mathbb{E}(e^{rY_{L_n}})] \\
&= \mathbb{E}_{L_n}[M_Y(r)^{L_n}] \\
&= G_{L_n}[M_X(r)]
\end{aligned}
$$

PFG de L_n - I

Rappel que $L_n = N_1 + N_2 ... + N_n$, pour $t = 2, ..., n$, on a

$$N_t = \alpha \circ N_{t-1} + \epsilon_t$$

Pour $n = 1$, on a

$$G_{N_1} = \mathbb{E}(r^{N_1})$$

Pour $n = 2$, on a

$$\begin{aligned} G_{N_1+N_2} &= \mathbb{E}(r^{N_1+N_2}) = \mathbb{E}(r^{N_1+\alpha\circ N_1+\epsilon_2}) \\ &= \mathbb{E}(r^{N_1+\alpha\circ N_1})\mathbb{E}(r^{\epsilon_2}) \\ &= \mathbb{E}(r^{N_1+\sum_{i=1}^{N_1}\xi_i})G_\epsilon(r) \\ &= \mathbb{E}_{N_1}[r^{N_1+\sum_{i=1}^{N_1}\xi_i} \mid N_1]G_\epsilon(r) \\ &= \mathbb{E}_{N_1}\{[rG_\xi(r)]^{N_1}\}G_\epsilon(r) \\ &= \boldsymbol{G_{N_1}(rG_\xi(r))G_\epsilon(r)} \\ &= G_{N_1}[r(1-\alpha+\alpha r)]G_\epsilon(r) \end{aligned}$$

PFG de L_n - II

Pour $n = 3$, on a

$$\begin{aligned} G_{N_1+N_2+N_3} &= \mathbb{E}(r^{N_1+N_2+N_3}) \\ &= \mathbb{E}(r^{N_1+\alpha\circ N_1+\epsilon_2+\alpha\circ N_2+\epsilon_3}) \\ &= \mathbb{E}(r^{N_1+\alpha\circ N_1+\epsilon_2+\alpha\circ(\alpha\circ N_1+\epsilon_2)+\epsilon_3}) \\ &= \mathbb{E}[r^{N_1+\alpha\circ N_1+\alpha\circ\alpha\circ N_1}]\mathbb{E}(r^{\epsilon_2+\alpha\circ\epsilon_2})\mathbb{E}(r^{\epsilon_3}) \end{aligned}$$

Montré dans l'équation (1),

$$\mathbb{E}(r^{\epsilon_2+\alpha\circ\epsilon_2})\mathbb{E}(r^{\epsilon_3}) = G_{\epsilon_2}(1-\alpha+\alpha r)G_{\epsilon_3}(r)$$

PFG de L_n - III

$$
\begin{aligned}
\mathbb{E}[r^{N_1+\alpha\circ N_1+\alpha\circ\alpha\circ N_1}] &= \mathbb{E}(r^{N_1} r^{\sum_{i=1}^{N_1}\xi_i} r^{\sum_{j=1}^{\sum_{N_1}\xi_i}\xi_j}) \\
&= \mathbb{E}(r^{N_1} r^{\sum_{i=1}^{N_1}\xi_i} r^{\sum_{j=1}^{\sum_{N_1}\xi_i}\xi_j}) \\
&= \mathbb{E}[r^{N_1} r^{\sum_{i=1}^{N_1}\xi_i} \mathbb{E}(r^{\sum_{j=1}^{\sum_{N_1}\xi_i}\xi_j} \mid \sum_{N_1}\xi_i)] \\
&= \mathbb{E}(r^{N_1} r^{\sum_{i=1}^{N_1}\xi_i} G_{\xi_j}(r)^{\sum_{i=1}^{N_1}\xi_i}) \\
&= \mathbb{E}(r^{N_1} (rG_{\xi_j}(r))^{\sum_{i=1}^{N_1}\xi_i}) \\
&= \mathbb{E}(r^{N_1} G_{\xi_i}(rG_{\xi_j}(r))^{N_1}) \\
&= G_{N_1}(rG_{\xi_i}(rG_{\xi_j}(r)))
\end{aligned}
$$

PFG de L_n - IV

Pour $n > 3$,

$$\begin{aligned} G_{L_n}(r) &= \mathbb{E}(r^{N_1+\ldots+N_n}) \\ &= \mathbb{E}(r^{N_1+\alpha\circ N_1+\cdots+\alpha^{n-1}\circ N_1}) \\ &\times \mathbb{E}(r^{\epsilon_2+\alpha\circ\epsilon_2+\cdots+\alpha^{n-2}\circ\epsilon_2})\cdots\mathbb{E}(r^{\epsilon_{n-1}+\alpha\circ\epsilon_{n-1}})\mathbb{E}(r^{\epsilon_n}) \end{aligned}$$

Afin de présenter les résultats, on définit les terms suivant :

$$\begin{aligned} f_1(r) &= r \\ &\cdots \\ f_n(r) &= r(\alpha f_{n-1}(r) + 1 - \alpha) \end{aligned}$$

On obtient

$$G_{L_n}(r) = G_{N_1}(f_n(r)) \prod_{k=1}^{n-1} G_\epsilon(f_k(r))$$

où $N_1 \sim CP(\lambda, T)$ et $\epsilon \sim CP(\mu, H)$

PFG de L_n - V

Rappel des PFG pour N_1 et ϵ :

$$G_{N_1}(r) = exp\{\lambda[T(r) - 1]\}$$
$$G_{\epsilon}(r) = exp\{\mu[H(r) - 1]\}$$

On obtient donc,

$$G_{L_n}(r) = \boldsymbol{G_{N_1}(f_n(r)) \prod_{k=1}^{n-1} G_{\epsilon}(f_k(r))}$$
$$= exp\{\lambda[T(f_n(r)) - 1] + \mu \sum_{k=1}^{n-1} [H(f_k(r)) - 1]\}$$
$$= exp\left\{\lambda_{S_n}\left[\frac{\lambda}{\lambda_{S_n}} T(f_n(r)) + \frac{\mu}{\lambda_{S_n}} \sum_{k=1}^{n-1} H(f_k(r)) - 1\right]\right\}$$

où $\lambda_{S_n} = \lambda + (n-1)\mu$.
si on remplace r par $M_X(r)$, on obtient $M_{S_n}(r)$.

$Z \sim CPINAR(1)$

Z est le montant cumulé actualisé des sinistres pendant n périodes

$$Z_n = vW_1 + v^2 W_2 + \cdots + v^n W_n$$

De la même façon, on peut obtenir $M_{Z_n}(r)$.

$$exp\left\{\lambda_{S_n}\left[\frac{\lambda}{\lambda_{S_n}} T(f_{(n)}(r_1, ..., r_n)) + \frac{\mu}{\lambda_{S_n}} \sum_{k=1}^{n-1} H(f_{(k)}(r_{n-j}, ..., r_n)) - 1\right]\right\}$$

avec

$$f_{(1)}(r_1) = r_1$$
$$f_{(n)}(r_1, ..., r_n) = r_1(\alpha f_{(n-1)}(r_2, ..., r_n) + 1 - \alpha)$$
$$\lambda_{S_n} = \lambda + (n-1)\mu$$

où $(r_1, ..., r_n) = M_X(vr), ..., M_X(v^n r)$.

Cas Spécial : F_{S_n} quand $X \sim exp(\beta)$

Rappel des concepts :
Soient $X_1, ...X_{L_n}$ des v.a. iid et $S_{L_n} = X_1 + \cdots + X_{L_n}$. Par la loi de la probabilité totale, la CDF du S est donnée par

$$F_S(s) = \sum_{n=0}^{\infty} F_{L_n}(S)\mathbb{P}(L_n = n)$$

où $F_{L_n}(S)$ est la CDF est des v.a. indépendantes $X_1 + \cdots + X_{L_n}$.
$F_{L_n}(S)$ suit donc une CDF de loi Erlang.

Le problème est donc de trouver $\mathbb{P}(L_n = n)$, sachant la PFG de L_n.

Cas Spécial : $\mathbb{P}(L_n = n)$ - I

$$G_{L_n}(r) = exp\left\{\lambda_{S_n}\left[\frac{\lambda}{\lambda_{S_n}}T(f_n(r)) + \frac{\mu}{\lambda_{S_n}}\sum_{k=1}^{n-1}H(f_k(r)) - 1\right]\right\}$$
$$= exp\{\lambda_{S_n}[\Phi_n(r) - 1]\}$$

On note $\phi_i = \frac{\lambda}{\lambda_{S_n}}T(f_i(r)) + \frac{\mu}{\lambda_{S_n}}\sum_{k=1}^{i-1}H(f_k(r))$.

Cas Spécial : $\mathbb{P}(L_n = n)$ - II

Si on regarde la fonction $\Phi_n(r)$, on a

$$\Phi_n(0) = 0$$
$$\Phi_n(1) = 1$$

Posons $\Phi_n(r)$ est une PGF, on peut écrire la fonction $\Phi_n(r)$ sous la forme suivante :

$$\Phi_n(r) = \phi_1 r + \cdots + \phi_n r^n$$

où $\sum_{j=1}^{n} \phi_j = 1$, avec $\phi_j \geq 0$.

Les dérivées de la fonction $\Phi_n(r)$ se donnent par

$$\Phi_n'(r) = \phi_1$$
$$\Phi_n''(r) = 2\phi_2$$
$$\cdots$$
$$\Phi^n{}_n(r) = n\phi_n.$$

Cas Spécial : $\mathbb{P}(L_n = n)$ - III

$$\mathbb{P}(L_n = 0) = G_{L_n}(0) = exp\{\lambda_{S_n}[\Phi_{L_n}(0) - 1]\} = e^{-\lambda_{S_n}}$$

$$\begin{aligned}\mathbb{P}(L_n = 1) &= G'_{L_n}(0) = exp\{\lambda_{S_n}[\Phi_{L_n}(0) - 1]\}\lambda_{S_n}\Phi'_{L_n}(0) \\ &= \lambda_{S_n}\phi_1\mathbb{P}(L_n = 0)\end{aligned}$$

$$\mathbb{P}(L_n = 2) = \frac{1}{2}G''_{L_n}(0) = \frac{\lambda_{S_n}}{2}[2\phi_2\mathbb{P}(L_n = 0) + \phi_1\mathbb{P}(L_n = 1)]$$

$$\cdots$$

$$\begin{aligned}\mathbb{P}(L_n = i) &= \frac{1}{i!}G^{(i)}{}_{L_n}(0) \\ &= \frac{\lambda_{S_n}}{i}[i\phi_n\mathbb{P}(L_n = 0) + (i-1)\phi_{i-1}\mathbb{P}(L_n = 1) \\ &+ \cdots + \phi_1\mathbb{P}(L_n = i-1)]\end{aligned}$$

Conclusion

Conclusion

Transformation des problèmes :

- Quand on veut travailler sur S_n, on a tranformé la question pour travailler sur S_{L_n}, où $L_n = N_1 + \cdots + N_n$.
- Quand on veut travailler sur MGF et CDF de S_{L_n}, on a fait des efforts pour trouver PGF de L_n.

Quelques techniques au travail :

- Quand on travaille sur les variables indépendantes, PGF est très intéressant ;
- Quand on travaille sur les variables dépendantes, l'espérance conditionnelle est très utile.

Bibliography

Al-Osh, M. A. and Alzaid, A. A. (1987).
First-Order Integer-Valued Autoregressive (Inar(1)) Process.
Journal of Time Series Analysis, 8(3) :261–275.

Chen, M. and Hu, X. (2020).
Risk aggregation with dependence and overdispersion based on the compound poisson inar(1) process.
Communications in Statistics - Theory and Methods, 49(16) :3985–4001.

Schweer, S. and Weiß, C. H. (2014).
Compound poisson inar(1) processes : Stochastic properties and testing for overdispersion.
Computational Statistics Data Analysis, 77(C) :267–284.

Zhang, H., Liu, Y., and Li, B. (2014).
Notes on discrete compound Poisson model with applications to risk theory.
Insurance : Mathematics and Economics, 59(C) :325–336.