Approximation exponentielle des gains d'une compagnie d'assurance

Projet GMM 4 : Probabilités appliquées- Statistique

Janvier 2014

Résumé Une compagnie d'assurance désire évaluer ses profits avant qu'un événement catastrophique $\mathcal{C}$ ne se produise. Sous l'hypothèse que les profits sur des périodes de temps successifs sont de variables aléatoires indépendantes, on se propose de donner une approximation de la loi du profit avant l'occurrence de l'événement catastrophique. On suppose que la probabilité θ pour que $\mathcal{C}$ se produise dans une unité de temps est petite. On modélise tout d'abord le profit P par une somme constituée d'un nombre aléatoire de variables aléatoires. On montre ensuite un théorème limite sur θP dans l'asymptotique où le paramètre θ tend vers 0. La loi limite est la loi exponentielle. On s'intéresse ensuite à l'utilisation pratique de ce théorème limite dans un cadre non asymptotique où θ prend des valeurs proches de 0.

Mots clefs : Loi exponentielle. Loi des grands nombres.
Thème applicatif : Finance et assurances.

1 Modèle probabiliste du profit

Pour construire un modèle probabiliste du profit on fait les hypothèses suivantes :

— L'échelle de temps est disrète. On considère donc des périodes de temps successives (par exemple des mois, des années ...) notées $1, \ldots, n, \ldots$.

— Pour $n \in \mathbb{N}^*$, on note X_n le profit de la compagnie d'assurance lors de la $n^{\text{ème}}$ période de temps. On suppose que (X_n) forme une suite de variables indépendantes de même espérance $\mu > 0$. On suppose de plus que toutes ces variables possèdent une variance bornée par une constante C. En d'autres termes, la loi du profit n'est pas fixée, elle peut varier au cours du temps mais sa moyenne reste constante et la moyenne de son écart à la moyenne au carré est uniformément bornée dans le temps. En outre, on suppose que la suite (X_n) est indépendante du nombre de périodes de temps avant l'occurrence de $\mathcal{C}$ (la catastrophe).

— La probabilité d'occurrence de $\mathcal{C}$ sur une période de temps donnée vaut θ. Pour $n \in \mathbb{N}$, notons A_n l'événement : *$\mathcal{C}$ se produit à la $n^{\text{ème}}$ période de temps.* On suppose que (A_n) est une suite d'événements indépendants.

On modélise alors les différentes grandeurs du modèle par les variables aléatoires suivantes :

— N est le nombre de périodes de temps avant l'occurrence de $\mathcal{C}$.

— On appelle P le profit réalisée par la compagnie d'assurance jusqu'à la période précédant l'occurrence de $\mathcal{C}$.

On a alors, d'après les hypothèses faites sur le modèle, les résultats suivants.

1.1 Premiers résultats de modélisation

a) La distribution de N est la loi géométrique de paramètre $1-\theta$:

$$\mathbb{P}(N=k) = (1-\theta)^k\theta, \quad (k \in \mathbb{N}). \tag{1}$$

b) Avec la convention que $\sum_{j=1}^{0} X_j = 0$, le profit P est donné par :

$$P = \sum_{j=1}^{N} X_j, \text{ de plus } \mathbb{E}(P) = \frac{\mu(1-\theta)}{\theta}.$$

1.2 Un exemple très particulier

On rappelle que l'on dit qu'une variable aléatoire suit la loi exponentielle d'espérance $a > 0$ si sa loi posséde la densité (par rapport à la mesure de Lebesgue) :

$$f_a(x) = \frac{1}{a}e^{-\frac{x}{a}}\mathbb{1}_{\mathbb{R}^+}(x).$$

On suppose ici que la loi des X_i est la loi exponentielle d'espérance μ. On peut alors complètement caractériser la loi de P :

— La loi de $\theta(1-\theta)^{-1}P$ est la loi exponentielle de moyenne μ.

2 Un théorème limite pour le profit

Lorsque θ est petit (θ tend vers 0), on se propose de montrer que la loi de la variable aléatoire $\theta(1-\theta)^{-1}P$ peut être approchée par la loi exponentielle de moyenne μ. Il est d'abord utile de préciser la notion de convergence que nous allons utiliser. C'est l'objet du prochain paragraphe.

2.1 Quelques éléments sur la convergence en loi

Dans toute la suite, W désigne une variable aléatoire dont la loi possède une densité par rapport à la mesure de Lebesgue. Rappelons que l'on dit qu'une famille de variables aléatoires $(W_\theta)_{\theta>0}$ converge en loi vers W, quand θ tend vers 0^+, si la fonction de répartition F_{W_θ} de W_θ converge ponctuellement vers celle de W. En d'autres termes si l'on a, pour tout $t \in \mathbb{R}$,

$$\lim_{\theta\to 0^+} F_{W_\theta}(t) = \lim_{\theta\to 0^+} \mathbb{P}(W_\theta \le t) = \mathbb{P}(W \le t) = F_W(t).$$

On rappelle également que la convergence en loi est équivalente à la convergence ponctuelle des fonctions caractéristiques. C'est-à-dire que $(W_\theta)_{\theta>0}$ converge en loi, quand θ tend vers 0, vers W si, et seulement si, pour tout $t \in \mathbb{R}$, on a

$$\lim_{\theta\to 0^+} \mathbb{E}[\exp(itW_\theta)] = \mathbb{E}[\exp(itW)].$$

On admettra le Lemme suivant :

Lemme 1 *Soit $(W_\theta)_{\theta>0}$ et $(Z_\theta)_{\theta>0}$ deux familles de variables aléatoires. Soit ψ une fonction sur $\mathbb{R}^+$, continue à droite en 0, avec $\psi(0) = 1$. On suppose que :*

a) $(W_\theta)_{\theta>0}$ converge en loi, quand θ tend vers 0, vers W.

b) $(V_\theta)_{\theta>0}$ converge en probabilité, quand θ tend vers 0. C'est-à-dire, pour tout $\varepsilon > 0$, on a

$$\lim_{\theta\to 0^+} \mathbb{P}(|V_\theta| > \varepsilon) = 0.$$

Alors,

— *$(W_\theta + V_\theta)_{\theta>0}$ et $(\psi(\theta)W_\theta)$ convergent en loi, quand θ tend vers 0, vers W.*

2.2 Résultat principal pour l'approximation

Le résultat principal qui permet d'approximer la loi de P est le Théorème de convergence suivant :

Théorème 1 *$\theta(1-\theta)^{-1}P$ converge en loi vers une variable de loi exponentielle de moyenne μ.*

La preuve du Théorème est basée sur la décomposition suivante :

$$P = \mu N + \sum_{j=1}^{N}(X_j - \mu).$$

Le Lemme suivant donne la convergence en loi de $\theta(1-\theta)^{-1}\mu N$.

Lemme 2 *$\theta(1-\theta)^{-1}\mu N$ converge en loi, quand θ tend vers 0^+, vers une variable de loi exponentielle de moyenne μ.*

D'autre part la variance de $\theta(1-\theta)^{-1}\sum_{j=1}^{N}(X_j - \mu)$ tend vers 0 avec θ. L'inégalité de Markov-Tchebycheff puis le Lemme 1 permettent alors de conclure.

2.3 Formule d'approximation

Soit $t > 0$, le Théorème 1 conduit à la formule d'approximation suivante (pour $\theta > 0$ petit) :

$$\mathbb{P}(\theta P \leq t) = 1 - \exp\left(-\frac{t}{\mu}\right) + o(1).$$

3 Simulations numériques

3.1 Un algorithme simple de simulation de la loi géométrique

Les deux remarques suivantes conduisent à la construction d'un générateur de variables aléatoire de loi géométrique.

1) Il est facile de réaliser une variable de loi exponentielle à partir d'une variable de loi uniforme. En effet, si U possède une distribution uniforme sur $[0, 1]$ et $\mu > 0$, alors $-\mu \log(U)$ suit la loi exponentielle de moyenne μ.
2) Pour $x \geq 0$, on note $\lfloor x \rfloor$ sa partie entière. Soit E une variable aléatoire de loi exponentielle de moyenne $\mu > 0$. Alors $\lfloor E \rfloor$ suit la loi géométrique de paramètre $\exp(-1/\mu)$.

En conclusion, pour $0 < \theta < 1$, la variable aléatoire

$$G_\theta = \left\lfloor \frac{\log U}{\log(1-\theta)} \right\rfloor, \tag{2}$$

suit la loi géométrique de paramètre $1 - \theta$.

3.2 Un programme MATLAB de comparaison qualitative

Le programme suivant permet la simulation de réalisations indépendantes de la variable aléatoire θP. On se place dans le cas où les variables (X_n) sont indépendantes et identiquement distribuées (i.i.d.). Deux lois différentes sont utilisées :

i) La loi de Bernoulli de paramètre $p \in]0, 1[$ (`loi1.m`).
ii) La loi de densité sur $\mathbb{R}$:

$$\sqrt{\frac{2}{\pi}} \exp(-\frac{x^2}{2}) \mathbb{1}_{\mathbb{R}^+}(x), \quad (\texttt{loi2.m}).$$

Dans les deux cas, on génère un grand nombre de réalisations ($N = 10000$). On compare ensuite, graphiquement, les distributions empiriques aux lois limites données par le Théorème 1.

```
function x = loi1(n,m,p)
%Génération d'une matrice nxm dont les composantes
%sont iid de loi de Bernoulli de paramètre p
y=rand(n,m);
x=(y<p);

function x = loi2(n,m)
%Génération d'une matrice nxm dont les composantes
%sont iid de loi valeur absolue d'une gaussienne standard
y=randn(n,m);
x=abs(y);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Programme de simulation d'une somme aléatoire géometrique
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Initialisation des constantes:
%
% - teta probabilité de l'événement catastrophique
%
```

```
% - N nombre de réalisations de la somme
% aléatoire
%
% - p paramètre de la loi de Bernoulli (loi1)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
clear;
teta=0.1;
N=10000;
p=0.5;
%unN est un vecteur ligne de taille N contenant des 1 utile par la suite
unN=ones(1,N);
%pa est le pas pour tracer les densités théoriques
pa=0.1;
%Nombre de classes pour l'affichage des histogrammes
mclas=floor(sqrt(N));
%Simulation de N réalisations de la variable teta*P
%
%
%Simulation de N réalisations de loi Geo(1-teta)
g=geo(1,N,teta);
%Calcul de la plus grande réalisation observée
gmax=max(g);
%Création du masque pour le calcul sans boucle des sommes
ungmax=ones(gmax,1);
masqg=ungmax*g;
masq=(1:gmax)'*unN;
masq=(masq<=masqg);
%Simulation de gmax*N variables iid de loi1
x=loi1(gmax,N,p);
%Utilisation du masque pour la sommation
x=x.*masq;
%Calcul des réalisations des sommes géometriques (teta*P)
P1=teta*sum(x);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% On recommence les mêmes operations pour la deuxième loi
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
clear g gmax ungmax masqg masq x
%Simulation de N réalisations de la variable teta*P
%
%
%Simulation de N réalisations de loi Geo(1-teta)
g=geo(1,N,teta);
%Calcul de la plus grande réalisation observée
gmax=max(g);
```

```
%Création du masque pour le calcul sans boucle des sommes
ungmax=ones(gmax,1);
masqg=ungmax*g;
masq=(1:gmax)'*unN;
masq=(masq<=masqg);
%Simulation de gmax*N variables iid de loi2
y=loi2(gmax,N);
%Utilisation du masque pour la sommation
y=y.*masq;
%Calcul des réalisations des sommes géometriques (teta*P)
P2=teta*sum(y);
%Comparaison graphique avec les lois limites
%
%Calcul des densités exponentielles de moyennes mu1 et mu2
mu1m1=1/p;
mu2m1=sqrt(pi/2);
P1max=max(P1);
P2max=max(P2);
ex1=0:pa:P1max;
ex2=0:pa:P2max;
ey1=mu1m1*exp(-mu1m1*ex1);
ey2=mu2m1*exp(-mu2m1*ex2);
%Affichage des graphiques
subplot(2,2,1);
hist(P1,mclas);
title('loi empirique de teta*P (Bernoulli)');
ylabel('effectifs');
subplot(2,2,2);
plot(ex1,ey1);
title('loi asymptotique de teta*P');
subplot(2,2,3);
hist(P2,mclas);
title('loi empirique de teta*P (absolue gauss)');
ylabel('effectifs');
subplot(2,2,4);
plot(ex2,ey2);
title('loi asymptotique de teta*P');
```

Après exécution du programme, on obtient les graphiques donnés en dernière page.

Suggestions de développement

Soulignons qu'il s'agit d'un menu à la carte et que vous pouvez choisir d'étudier certains points, pas tous, pas nécessairement dans l'ordre, et de façon plus ou moins fouillée. Vous pouvez aussi vous poser d'autres questions que celles indiquées plus bas. Il est très vivement souhaité que vos investigations comportent une partie traitée sur ordinateur et, si possible, des représentations graphiques de vos résultats.

- Justifier comment la traduction des hypothèses conduit aux résultats du paragraphe 1.1.
- Développer l'exemple du paragraphe 1.2 (montrer par le calcul que l'on trouve bien la loi exponentielle).
- Montrer le Lemme 2 (par exemple en utilisant la fonction caractéristique).
- Se placer dans le cas particulier où les variables (X_n) sont indépendantes et identiquement distribuées et montrer le Théorème 1.
- Justifier la formule d'approximation du paragraphe 2.3.
- Pour $\theta > 0$, montrer que la variable G_θ définie dans (2) suit la loi géométrique de paramètre $1 - \theta$.
- Commenter les figures obtenues après exécution du programme.
- Compléter le programme pour donner une évaluation quantitative de la *distance* entre les répartitions empiriques et la loi limite. On pourra par exemple construire une *distance* en considérant la différence des fonctions de répartition.
- Etudier numériquement plusieurs valeurs de θ.
  Attention, pour θ très petit, le nombre de simulations pour calculer la valeur d'une somme explose.