

SUJET B

LES MATHÉMATIQUES DE L'ASSURANCE
AUTOMOBILE

Introduction de la discussion (résumé)

P. J. DELAPORTE

Plusieurs communications sont présentées au Colloque sur les Mathématiques de l'Assurance Automobile. Afin de pouvoir mieux situer chacune d'elles dans ce vaste sujet, il peut être bon de schématiser l'ensemble de la question. Il comprend :

1. la survenance des accidents
2. le tarif et la mesure de son efficacité par rapport à l'estimation optimale.

1. *La Survenance des accidents*

Groupons dans une classe les propriétaires de véhicules ayant un même ensemble de caractéristiques extérieures de risques: même modèle de voiture, même zone de garage habituel, même usage ou profession du propriétaire, etc. Les risques s de survenance d'accident ne sont pas identiques pour tous les propriétaires de la classe, mais sont distribués selon une fonction de répartition $F(s)$.

Désignons par $P(s' | s)$ la probabilité de survenance de s' sinistres pour un assuré dont le risque individuel est s (des études antérieures ont montré que $P(s' | s)$ est représentable par une loi de probabilité de Poisson). La probabilité qu'un assuré, pris au hasard dans la classe, ait s' sinistres est :

$$P(s') = \int_{s_0}^{+\infty} P(s' | s) \cdot dF(s) \quad (1)$$

s_0 étant le risque minimum de la classe.

Désignons par $g(X)$ la densité de probabilité de la distribution des coûts X des sinistres. On peut admettre qu'à l'intérieur de la classe, au moins en première approximation, $g(X)$ est indépendant de s , d'où la probabilité de survenance d'un sinistre de coût X est :

$$P(s' = 1; X) = g(X) \cdot P(s' = 1) \quad (2)$$

La probabilité de survenance d'un sinistre de coût supérieur à une franchise X_e est:

$$P(s' = 1; X > X_e) = P(s' = 1) \cdot \left[\int_{x_e}^{+\infty} dG(X) \right] \quad (3)$$

et la probabilité de survenance de s'_e sinistres de coûts supérieurs à une franchise est:

$$P(S'_e; {}_1X > X_e, \dots, s'_e X > X_e) = P(s'_e) \cdot \left[\int_{x_e}^{+\infty} dG(X) \right]^{s'_e} \quad (4)$$

Dans sa communication: „L'influence de la franchise sur le nombre de sinistres dans l'Assurance Automobile”, Monsieur Mogens Muff étudie les distributions empiriques de ces sinistres, et après ajustement sur une distribution binomiale négative, il observe que l'un des paramètres dépend de la clause de bonus et l'autre varie avec la franchise.

Le réassureur excess pourra utiliser ce même modèle en prenant pour s non pas le risque probable d'un assuré, mais le risque probable d'une compagnie réassurée. Il remplacera la franchise X_e par un montant d'excess $X_e(1 + d)^t$ décroissant avec le temps écoulé depuis le début de l'année de survenance jusqu'à l'époque moyenne de règlement t , pour tenir compte d'une dépréciation monétaire dont le taux annuel est de d pour les règlements de sinistres.

Dans sa communication: „Une approche d'analyse des sinistres observés dans la réassurance excess de responsabilité civile”, Monsieur H. G. Verbeek étudie le tableau à double entrée des sinistres par année de survenance et par année de déclaration au réassureur au fur et à mesure que chacun d'eux dépasse le montant de l'excess.

M. Verbeek admet que la distribution des nombres de sinistres déclarés par la compagnie cédante au réassureur est représentable par des distributions de Poisson et qu'il y a indépendance entre le nombre et le coût des sinistres. Une estimation des paramètres par la méthode du Maximum de Vraisemblance lui permet d'extrapoler, pour les années futures, la valeur probable des sinistres qui seront déclarés au réassureur, ce qui peut servir de base à la cotation d'un traité de réassurance excess.

2. *Le tarif et la mesure de son efficacité par rapport à l'estimation optimale*

Le tarif d'assurance automobile comprend :

1. *une information à priori* sur l'assuré et l'usage qu'il fera du véhicule, son lieu de garage habituel et, d'autre part, sur la voiture, son modèle, son état d'entretien, etc. ;
2. *une information à posteriori* sur les sinistres déclarés et sur la manière dont le conducteur se comporte (entretien du véhicule, infractions au Code de la Route), relative à la période écoulée depuis la souscription du contrat.

L'information à priori permet un découpage de l'ensemble des risques en *classes de tarif* dans lesquelles on groupe les assurés ayant à priori des risques identiques.

Désignons par :

- $1s'_i, 2s'_i, \dots$ les nombres de sinistres déclarés par l'assuré i pendant les années d'assurance 1, 2, etc. Les probabilités de survenance d'accidents ont été définies ci-dessus (formule 1).
- $1c'_i, 2c'_i, \dots$ la manière dont le conducteur s'est comporté pendant les années 1, 2, etc. (par exemple le nombre des infractions graves à la loi sur la circulation routière).
- $H(jc'_i | s)$ la fonction de distribution des comportements des conducteurs de risques s pendant l'année j .

et admettons que les classes de tarif ont été constituées de manière suffisamment homogènes pour que le coût moyen des sinistres \bar{X} soit, dans la classe, indépendant du risque individuel d'accident. La prime pure de l'année $k + 1$ connaissant les antécédents des k premières années d'assurance est :

$$\begin{aligned} & \Pi_{(k+1s'_i | 1s'_i, \dots, ks'_i; 1c'_i, \dots, kc'_i)} = \\ & = \bar{X} \frac{\int_0^{+\infty} s \cdot P(1s'_i | s) \dots P(ks'_i | s) \cdot H(1c'_i | s) \dots H(kc'_i | s) \cdot dF(s)}{\int_0^{+\infty} P(1s'_i | s) \dots P(ks'_i | s) \cdot H(1c'_i | s) \dots H(kc'_i | s) \cdot dF(s)} \end{aligned} \tag{5}$$

$P(s' | s)$ ayant une loi de distribution de Poisson, ceci se simplifie en

$$= \bar{X} \frac{\int_{s_0}^{+\infty} s^{1s'_i + \dots + ks'_i + 1} \cdot e^{-ks} \cdot H({}_1c'_i | s) \dots H({}_kc'_i | s) \cdot dF(s)}{\int_{s_0}^{+\infty} s^{1s'_i + \dots + ks'_i} \cdot e^{-ks} H({}_1c'_i | s) \dots H({}_kc'_i | s) \cdot dF(s)} \quad (6)$$

A notre connaissance, la fonction H n'a pas encore été étudiée, elle a donc été supposée constante dans tous les travaux antérieurs.

Au contraire, l'étude des observations a montré que la fonction de distribution des risques individuels $F(s)$ est voisine de l'eulérienne de seconde espèce:

$$dF(s) = \frac{a^b}{\Gamma(b)} \cdot e^{-a(s-s_0)} \cdot (s - s_0)^{b-1} \cdot ds \quad (7)$$

et que S_0 , qui est le risque minimum, est voisin de zéro; a et b sont des paramètres à estimer par le Maximum de Vraisemblance pour que les voitures aient eu les nombres de sinistres observés.

Le risque d'un assuré i connaissant les sinistres qu'il a eus pendant les k années antérieures peut être obtenu d'après l'espérance mathématique de s liée par les sinistres observés, c'est-à-dire:

$$\bar{X} \cdot E[{}_{k+1}s_i | {}_1s'_i, \dots, {}_ks'_i] = \bar{X} \frac{b + {}_1s'_i + \dots + {}_ks'_i}{a + k} \quad (8)$$

Notre Président, Monsieur P. Thyron, présente une communication intitulée: „Quelques observations statistiques sur la variable „nombre de sinistres” en assurance automobile”, dans laquelle il a étudié le comportement de 1.410 polices pendant 4 années et de 1.094 polices pendant 7 années.

Sur cet ensemble, il a d'abord examiné la distribution empirique du nombre des sinistres d'une année en fonction du nombre total des sinistres des années antérieures. Il a montré que les observations vérifient les conditions pour être représentables par une loi de Poisson composée et que la „fonction de structure” $U(\lambda)$ (qui est la fonction $F(s)$ dans les notations précédentes, formule [1]) reste stable au cours du temps. Les résultats montrent que cette fonction est plus concentrée que la loi exponentielle et a un seul mode situé près de l'origine.

Ensuite, il a calculé, d'après les observations, pour chacune des

années, la valeur du risque liée par les nombres de sinistres observés pendant les années précédentes et a donné, pour plus de précision, les valeurs médianes du rapport du risque lié divisé par le risque moyen de l'ensemble des polices, dans le dernier tableau de sa communication.

On peut noter que les résultats entièrement empiriques de P. Thyron, obtenus sur la Belgique, sont en excellent accord avec ceux que nous avons obtenus, sur la France, il y a plus de 10 ans, par un modèle théorique de loi de Poisson composée avec une eulérienne de seconde espèce (formules 1, 7 et 8 ci-dessus) pour une même fréquence moyenne $\bar{s} = 0,193$ et pour la valeur du paramètre $b = 1,6$, alors que les résultats de P. Thyron donnent $b = 1,32$

n	Valeur	S'il y a eu n sinistres pendant les années antérieures, quotient de la fréquence liée à la fréquence moyenne après :					
		1 an	2 ans	3 ans	4 ans	5 ans	6 ans
0 sin.	empirique	0,85	0,79	0,71	0,66	0,64	0,59
	théorique	0,89	0,80	0,73	0,67	0,62	0,58
1 sin.	empirique	1,48	1,25	1,21	1,01	0,96	0,79
	théorique	1,45	1,31	1,19	1,09	1,01	0,94

P. Thyron indique qu'avec une vraisemblance suffisante pour la pratique on peut conclure que l'hypothèse de l'hétérogénéité des classes de risques se montre apte à prévoir l'influence du comportement passé d'une sous-classe de risques sur son comportement futur en ce qui concerne les nombres de sinistres, ce qui est une des bases pratiques d'une tarification de l'assurance automobile utilisant un système de bonus-malus.

Dans sa communication : „Etude sur la survenance des sinistres en assurance automobile”, M. Bricler propose de poser $b = 1$ et a égal à l'inverse de la fréquence moyenne du groupe. Cette solution correspond à un cas limite de la représentation par l'eulérienne de seconde espèce réduite ae^{-as} . Ceci semble valable lorsque la classe de risques est très dissymétrique et l'espérance mathématique du nombre de sinistres très faible, mais conviendrait moins bien à des classes plus hétérogènes pour lesquelles on trouve b variant de 1,5 à 2.

D'autre part, M. Brichler a essayé, empiriquement, de trouver une représentation des fréquences des années successives par des formules plus générales, notamment en donnant un poids plus fort aux sinistres des années récentes et plus faible aux sinistres anciens.

Nous ne parlerons pas du problème de jugement d'un tarif, car aucune communication n'est présentée sur ce sujet. Je voudrais cependant rappeler que l'analyse de la variance permet de décomposer la variance totale des sinistres en variance liée par le tarif et variance non expliquée par le tarif. La variance du risque probable divisée par la variance liée par le tarif donne la mesure de l'efficacité de ce tarif. Cette méthode a encore peu été utilisée.

En terminant, je voudrais faire remarquer que, pour établir les modèles représentant la survenance des accidents et l'estimation des fréquences individuelles des assurés, nous avons fait appel seulement à quatre distributions :

- $F(s)$ fonction de répartition des risques individuels
- $P(s' | s)$ probabilité de survenance d'un sinistre
- $G(X)$ fonction de répartition des coûts de sinistres
- $H(c' | s)$ fonction de répartition du comportement des conducteurs

qui nous permettent d'obtenir toute l'information désirée. Un effort très grand devrait être fait pour obtenir de manière très précise ces distributions.