

FRÉDÉRIC BOULANGER

Individualisation du risque en assurance automobile

Journal de la société statistique de Paris, tome 134, n° 1 (1993),
p. 31-50

http://www.numdam.org/item?id=JSFS_1993__134_1_31_0

© Société de statistique de Paris, 1993, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

INDIVIDUALISATION DU RISQUE EN ASSURANCE AUTOMOBILE

par Frédéric BOULANGER
Responsable du secteur Études Tarifs
Assurances Générales de France (AGF)

« La gestion du risque, c'est d'inclure nos manques de connaissances dans un espace mesuré. »

Daniel SIBONY

L'évolution de l'assurance automobile a été marquée ces dernières années par une individualisation du risque. Cette évolution s'est faite sous la pression des assurés qui ne veulent payer que pour leur risque et le jeu de la concurrence qui, par la recherche de segments rentables, a conduit les assureurs à introduire de nouveaux critères dans les tarifs.

Un tarif commercial repose sur une estimation, à partir de modèles, des risques individuels. Il est obtenu par optimisation de la fonction d'utilité de l'assureur (développement, rentabilité, etc.). La solution retenue doit respecter les contraintes externes liées aux assurés – ce qu'ils peuvent accepter –, à la concurrence – maîtrise des risques d'antisélection – et au mode de distribution d'un produit de masse – complexité du tarif, communication, intermédiaires. Dans la suite nous nous intéresserons à la partie concernant l'estimation des risques.

1. Introduction

Le risque individuel d'un assuré est l'espérance de sa charge sinistre, c'est-à-dire sa prime pure :

$$\text{Prime Pure} = E \left(\sum_{i=1}^N C_i \right) \quad (1.1)$$

où N est le nombre aléatoire de sinistres dans l'année,

C_i est le coût aléatoire du $i^{\text{ème}}$ sinistre.

La formule de la prime pure met en évidence deux phénomènes différents : le nombre de sinistres et le coût des sinistres. Sous l'hypothèse de l'indépendance entre N et les C_i , considérés comme indépendants et de même loi, la prime pure se met sous la forme :

$$\text{Prime Pure} = E(N) E(C) \quad (1.2)$$

INDIVIDUALISATION DU RISQUE EN ASSURANCE AUTOMOBILE

Cette décomposition permet de séparer la prime pure en un terme qui dépend du nombre de sinistres (fréquence) et un des coûts (coût moyen). Pour choisir entre une approche globale et une approche coût/fréquence, il est utile de calculer le coefficient de variation :

$$CV(\text{Prime Pure}) = [E(N) \star CV(N) + CV(C)] / E(N) \quad (1.3)$$

Le terme dépendant de N est, dans la pratique, proche de 1 – le processus de survenance des sinistres d'un assuré est proche d'un processus de Poisson – alors que celui qui dépend de C est très variable selon le type de garantie étudiée – la variation est comprise entre 1 et 10. Lorsque ces deux termes sont très différents, il peut être intéressant de prendre des segmentations différentes pour la fréquence et le coût (fiabilisation de l'estimation du coût moyen, par exemple) et donc d'étudier les deux phénomènes séparément.

Pour l'estimation de la prime pure, l'assureur dispose de trois types de variables :

- les caractéristiques du véhicule (puissance, valeur, âge...) ;
- les caractéristiques de l'assuré à un instant donné (âge, zone de circulation, CSP...) ;
- son historique « sinistres ».

Ces variables ne sont pas de même nature : les caractéristiques sont des faits objectifs ayant une valeur absolue, alors que l'historique sinistres est la réalisation d'une variable aléatoire qui n'a qu'une valeur relative. La présence d'un sinistre sur une année est en effet normale (car fort probable) pour un assuré ayant une fréquence estimée de 500 ‰, alors qu'elle peut révéler une aggravation du risque (car peu probable) pour un assuré ayant une fréquence estimée de 50 ‰. Cette différence oblige à traiter successivement les variables objectives (purement exogènes) et les variables de type historique en prenant en compte, dans la deuxième étape, les résultats de la première. Le non-respect de cette règle risque de produire des effets cumulatifs. C'est par exemple une majoration, du fait d'un sinistre, de primes qui intègrent déjà une forte sinistralité.

La tarification des risques se décompose en deux étapes : la première repose sur des critères purement exogènes – la tarification *a priori* – ; la seconde, corrige la première en fonction de l'historique – la tarification *a posteriori*. Ces deux approches sont complémentaires dans la mesure où, plus la tarification *a priori* est « bonne », moins l'historique a d'importance, et donc moins le poids de la tarification *a posteriori* est important.

2. Tarification *a priori*

L'objet de cette partie est d'estimer le risque individuel à partir des variables purement exogènes. Nous nous intéresserons successivement à l'estimation de la fréquence, du coût moyen et de la prime pure. Nous comparerons les résultats obtenus

avec l'approche fréquence/coût et ceux obtenus avec l'approche globale et nous donnerons les avantages et inconvénients des deux méthodes.

Dans la pratique, on utilise soit des modèles additifs, soit des modèles multiplicatifs. L'avantage des seconds est qu'ils sont contraints à la positivité. Ils sont de plus bien adaptés aux méthodes du maximum de vraisemblance alors que les modèles additifs sont plutôt adaptés aux méthodes des moindres carrés. Ils sont enfin simples à mettre en oeuvre dans un tarif et proches de la pratique des compagnies d'assurances en terme de structure tarifaire. Notons qu'il existe, pour des cas simples, une approche permettant de passer continûment d'une formulation additive à une formulation multiplicative (Harrington, 1986).

Dans toute cette partie, nous utiliserons des modèles multiplicatifs.

2.1. Estimation de la fréquence

Les grandeurs observables utiles sont les nombres de sinistres. Nous proposons donc une démarche reposant sur l'étude du processus de leur survenance. Nous partons d'un modèle « naturel » simple que nous enrichirons à partir de l'analyse des résultats.

2.1.1. Modèle de Poisson

Dans ce modèle, nous considérons que, pour un individu i , le processus de survenance des sinistres est un processus de Poisson de paramètre λ_i (fréquence individuelle) fonction des variables purement exogènes (après codage disjonctif) :

$$\lambda_i = f(C_i), \quad C = (c_1, \dots, c_m) \quad (2.1.1.1)$$

Du point de vue de la sinistralité, cela se traduit par :

- l'homogénéité dans le temps de la loi du nombre de sinistres (abstraction faite des variations saisonnières) ;
- l'indépendance des nombres de sinistres sur deux périodes de temps différentes ;
- la non « concomitance » des sinistres.

Dans la pratique cela se traduit par : « l'assuré a un comportement **constant** (à variables exogènes données) et **connu** (fonction des variables exogènes) », ce qui est une hypothèse forte.

Dans notre présentation nous nous limitons à :

$$\lambda_i = \exp(C_i^T B), \quad B = (b_1, \dots, b_n) \quad (2.1.1.2)$$

Cette forme de f permet, en plus des avantages cités ci-dessus, de réduire le nombre de paramètres à estimer et donc de fiabiliser l'estimation. Toutefois, elle induit une relation forte sur les fréquences estimées : un changement de modalité entraîne une modification uniforme des fréquences en pourcentage.

INDIVIDUALISATION DU RISQUE EN ASSURANCE AUTOMOBILE

Pour l'estimation de B , nous allons nous placer dans le cas où les durées d'observation sont l'année (unité de temps). Pour l'individu i la loi du nombre de sinistres N_i est une loi de Poisson de paramètre λ_i :

$$P(N_i = n/C_i) = \lambda(C_i)^n / n! \exp(-\lambda(C_i)) \quad (2.1.1.3)$$

d'où l'expression de la vraisemblance :

$$L(n_1, \dots, n_m / C_1, \dots, C_m) = \prod \lambda(C_i)^{n_i} / (n_i)! \exp(-\lambda(C_i)) \quad (2.1.1.4)$$

Par dérivation du logarithme de la vraisemblance, on obtient le système :

$$\forall j, \quad 1 \leq j \leq m \quad \sum_i c_{ij} \exp(C_i^T B) = \sum_i c_{ij} n_i \quad (2.1.1.5)$$

soit encore :

$$\forall j, \quad 1 \leq j \leq m \quad \sum_i c_{ij} \lambda_i = \sum_i c_{ij} n_i \quad (2.1.1.5')$$

Cette expression traduit le respect des marges par le modèle, c'est-à-dire que, pour toute population associée à une modalité, le nombre de sinistres estimés avec le modèle est égal au nombre de sinistres observés (grandeur observable utile). Inversement, cette adéquation des marges peut, du fait de son importance pour l'assureur, conduire à prendre le système (2.1.1.5) comme définition des estimateurs des paramètres d'un modèle multiplicatif. On retrouve un cas particulier de la méthode des marges appliquée à la fréquence. Cette approche peut être étendue à l'estimation directe de la prime pure (Bailey, 1963 ; Jung, 1968 ; Van Eeghen, 1984).

L'unicité de la solution (à un coefficient multiplicatif près) est assurée dans la mesure où les dérivées secondes sont négatives.

$$\forall j, k, \quad 1 \leq j, k \leq m \quad \partial^2 \ln(L) / \partial b_j \partial b_k = - \sum_i c_{ij} c_{ik} \exp(C_i^T B) \quad (2.1.1.6)$$

Notons enfin que la part de variance expliquée par un tel modèle, même s'il est bien adapté, est toujours très faible par rapport à la variance endogène du phénomène observé. Par exemple, pour un mélange de deux processus de Poisson de paramètres respectifs $\lambda_1 = 50 \text{ ‰}$ et $\lambda_2 = 70 \text{ ‰}$ en proportion égale, la variance du nombre de sinistres pour une année est de 60.1 ‰ et la variance inter-classe de 0.1 ‰ , soit une part de variance expliquée de 1.66 ‰ . Cette mesure, prise dans l'absolu, ne permet pas de mesurer la qualité d'un modèle.

Nous avons testé ce modèle sur la garantie RC matérielle à partir d'un échantillon de 679 950 assurés. La fréquence est bien estimée. Toutefois, la répartition du portefeuille en fonction de nombre de sinistres annuel (*Tableau 1*) n'est pas bien reproduite (la queue de la distribution est sous-estimée) : la valeur du test du Chi-2 (à 2 degrés de liberté) est 447. Le modèle est rejeté.

INDIVIDUALISATION DU RISQUE EN ASSURANCE AUTOMOBILE

*Tableau 1. Fréquence et distribution des assurés
en fonction du nombre de sinistres*

	<i>Réelle</i>	<i>Estimée</i>
Fréquence ‰	47,9	47,9
Nombre de sinistres	répartition	répartition
0	648 785	648 227
1	29 835	30 898
2	1 266	809
3 et plus	64	15

Au niveau des tests statistiques nous avons considéré :

- le pourcentage de la population appartenant à une classe dans laquelle le test du Chi-2 est vérifié ;
- un test asymptotique D_1 comparant le nombre de sinistres survenu et le nombre théorique estimé :

$$D_1 = \sum_{k \leq K} (n_k - \lambda_k * m_k)^2 / \sigma_k^2 \quad (2.1.1.7)$$

où n_k est le nombre de sinistres de la classe k ,

m_k est le nombre d'assurés de la classe k ,

λ_k est la fréquence estimée.

D_1 suit, sous l'hypothèse d'adéquation de la loi de Poisson et de l'expression donnant λ , un Chi-2 à $K-p$ (p étant le nombre de paramètres à estimer) degrés de liberté ;

- un test D_2 basé sur le logarithme du rapport entre le maximum de vraisemblance du modèle étudié et celui du sur-modèle général (la répartition du nombre de sinistres est la répartition empirique) :

$$D_2 = -2 \ln (RMV)$$

D_2 suit un Chi-2 à $K * n_m - p$, où est n_m le nombre maximal de sinistres que peut avoir un assuré.

Nous avons obtenu :

- 53 % de bien classé,
- $D_1 = 1\,454$ avec 1 252 degrés de liberté,
- $D_2 = 3\,822$ avec 3 010 degrés de liberté.

INDIVIDUALISATION DU RISQUE EN ASSURANCE AUTOMOBILE

Les valeurs de D_1 et D_2 doivent être, vu la taille de l'échantillon, considérées plus en tant que distances qu'en tant que tests statistiques ; elles permettent de comparer des modèles entre eux.

Au niveau de l'estimation des paramètres (*Tableau 2*), on constate une différence importante entre la fréquence d'une modalité et la valeur du paramètre estimé. Cet écart traduit le pouvoir de « décorrélation » du modèle par rapport au tableau de contingences sous-jacent, et ce, malgré les contraintes fortes liées à l'hypothèse multiplicative.

Tableau 2. Garantie Vol : comparaison entre les fréquences des marges et les paramètres estimés

<i>Fréquence</i>	<i>Poisson</i>	<i>Fréquence</i>	<i>Poisson</i>	<i>Fréquence</i>	<i>Poisson</i>
1,00	1,00	1,00	1,00	1,00	1,00
1,41	1,37	1,03	0,93	0,74	0,77
1,69	1,58	1,21	1,05	0,57	0,64
2,17	1,94	1,49	1,20	0,46	0,53
3,06	2,74	1,82	1,31	0,27	0,33
4,99	4,06	2,29	1,46	0,14	0,18
4,35	3,70	2,12	1,32		
5,93	4,68	3,02	1,79		
5,36	4,49	3,66	1,82		
6,87	5,80	2,93	1,60		
		3,35	2,85		
		4,86	3,16		
		6,31	3,89		
		9,29	4,69		
		10,79	6,06		

La non-adéquation de la répartition des assurés en fonction du nombre de sinistres annuel ne doit pas nous surprendre dans la mesure où les hypothèses du modèle de Poisson ne sont pas compatibles avec les observations. En effet :

- la fréquence individuelle ne dépend pas que des variables purement exogènes dans la mesure où l'on constate qu'un assuré ayant eu un sinistre a plus de chance que les autres assurés de sa classe d'en avoir d'autres ;
- la variance intra-classe est supérieure à la fréquence (49,9 ‰ contre 47,9 ‰).

Afin de tenir compte de ces phénomènes importants pour une bonne individualisation du risque nous allons généraliser le modèle de Poisson.

2.1.2. Le modèle de Poisson randomisé

Les grandeurs observables utiles sont les nombres de sinistres et la distribution des assurés en fonction du nombre de sinistres. Dans le modèle de Poisson randomisé la fréquence individuelle λ_i n'est plus une constante mais une variable aléatoire Λ_i qui suit une loi Gamma (Gouriéroux, 1989 b). La moyenne de la loi est λ_i et ne dépend que des variables purement exogènes, la variance relative est $1/a$, valeur constante pour toutes les classes.

D'un point de vue pratique, cela revient à supposer que « le comportement de l'assuré est **constant** dans le temps (à variables exogènes données), comme pour le modèle de Poisson, mais qu'il est **inconnu** » ; ce n'est pas la fréquence individuelle qui change au cours du temps, mais l'estimateur de cette fréquence, c'est-à-dire la connaissance qu'on en a.

D'après les résultats classiques sur la randomisation d'une loi de Poisson par une loi Gamma, la loi du nombre de sinistres suit une binomiale négative de moyenne λ_i et de variance $\sigma_i^2 = \lambda_i (a + \lambda_i)/a$:

$$P(N_i = n/C_i) = (\lambda_i^n / n!) * (\Gamma(n+a)/\Gamma(a)) * (a^a / (a + \lambda_i)^{n+a}) \quad (2.1.2.1)$$

On vérifie que :

- la variance intra-classe est supérieure à la fréquence ;
- le modèle de Poisson est un cas limite du modèle de Poisson randomisé quand a tend vers l'infini.

Si l'on met λ_i sous la forme $\exp(C_i^T B)$ (hypothèse multiplicative), la solution obtenue par le maximum de vraisemblance est donnée par le système :

$$\forall j, \quad 1 \leq j \leq m \quad \sum_i c_{ij} (\lambda_i - n_i) (a/(a + \lambda_i)) = 0 \quad (2.1.2.2.a)$$

$$\sum_{k>0} m_k / (a + k - 1) + \sum_i \ln(a/(a + \lambda_i)) + (\lambda_i - n_i)/(a + \lambda_i) = 0 \quad (2.1.2.2.b)$$

où m_k est le nombre d'assurés ayant eu au moins k sinistres.

Le système (2.1.2.2.a) est une généralisation du système des marges (2.1.1.5') en pondérant les observations par $p_i = a/(a + \lambda_i) = \lambda_i/\sigma_i^2$, poids dépendants des paramètres à estimer et accordant une plus grande importance aux bonnes classes (λ_i petit). Le système (2.1.2.2.a) est équivalent à :

$$\forall j, \quad 1 \leq j \leq m \quad \sum_i (\partial \lambda_i / \partial b_j) (\lambda_i - n_i) / \sigma_i^2 = 0 \quad (2.1.2.2.a')$$

qui est le système obtenu par la méthode des moindres carrés généralisés lorsque les variances intra-classe sont connues. Bien que cette généralisation des équations de la méthode des marges ne respecte plus les marges, les estimateurs obtenus sont « meilleurs » car ils tiennent compte des variances des observations dans les différentes classes.

INDIVIDUALISATION DU RISQUE EN ASSURANCE AUTOMOBILE

Par dérivation seconde du système (2.1.2.2.a) on obtient :

$$\forall j, k, \quad 1 \leq j, k \leq m \quad \frac{\partial^2 \ln(L)}{\partial b_j \partial b_k} = - \sum_i c_{ij} c_{ik} a \lambda_i \frac{(n_i + a)}{(\lambda_i + a)^2} \leq 0 \quad (2.1.2.3.a)$$

d'où l'existence, pour a fixé, d'une solution (à un coefficient multiplicatif près).

Par dérivation seconde de l'équation (2.1.2.2.b) on obtient :

$$\frac{\partial^2 \ln(L)}{(\partial a)^2} = - \sum_{k>0} \frac{m_k}{(a+k-1)^2} + \sum_i \frac{\lambda_i}{a(\lambda_i + a)} - \frac{\lambda_i - n_i}{(a + \lambda_i)^2} \quad (2.1.2.3.b)$$

On n'est pas assuré de l'existence d'une solution. Toutefois, nous avons une condition suffisante d'existence :

Propriété 2.1.2.1 : Si la variance intra-classe expérimentale est supérieure à la moyenne expérimentale $\left(\sum_i (\lambda_i - n_i)^2 > \sum_i n_i \right)$ alors l'équation (2.1.2.2.b) admet au moins une solution.

Démonstration :

– Pour a au voisinage de 0 nous avons :

$$\partial \ln(L) / \partial a = m_1 / a + \sum_i \ln(a) + 0(1)$$

$$\text{d'où :} \quad \lim_{a \rightarrow 0} \partial \ln(L) / \partial a = +\infty$$

– pour a au voisinage de $+\infty$ nous avons :

$$\partial \ln(L) / \partial a = \sum_{k>0} m_k / a - \sum_i n_i / a + \sum_i (\lambda_i / a)^2 / 2 - \lambda_i (\lambda_i - n_i) / a^2 + o(1/a^2)$$

$$\text{d'où :} \quad \lim_{a \rightarrow +\infty} \partial \ln(L) / \partial a = 1/2a^2 \left[\sum_i (n_i - (\lambda_i - n_i)^2) \right]$$

qui est négative si la condition suffisante est satisfaite. $\partial \ln(L) / \partial a$ s'annule donc en au moins un point, le premier zéro étant un maximum relatif de L .

Une solution (a^*, B^*) du système (2.1.2.2) peut être calculée de manière itérative. Le processus est initialisé à partir, d'une part, des paramètres estimés avec le modèle de Poisson, et d'autre part, de l'estimation de a obtenue par la méthode des moments :

$$a = \frac{\sum_i \lambda_i}{\sum_i ((\lambda_i - n_i)^2 - n_i)}$$

qui, si la condition suffisante est satisfaite, est bien un nombre positif.

INDIVIDUALISATION DU RISQUE EN ASSURANCE AUTOMOBILE

Nous avons appliqué ce modèle au cas étudié dans le paragraphe précédent. Comme pour le modèle de Poisson la fréquence est bien estimée. De plus, si l'on compare les nombres d'assurés ayant eu 0, 1, 2 et 3 et plus (*Tableau 3*), on constate une assez bonne adéquation avec une légère sous-estimation de la queue de distribution. La valeur du test du Chi-2 (à 1 degré de liberté) est 0,56 : le modèle est accepté.

Tableau 3. Fréquence et distribution des assurées en fonction du nombre de sinistres

	<i>Réelle</i>	<i>Estimée</i>
Fréquence ‰	47,9	47,9
Nombre de sinistres	répartition	répartition
0	648 785	648 788
1	29 835	29 823
2	1 266	1 280
3 et plus	64	59

Pour les autres tests statistiques nous avons :

- 70 % de bien classé, à la place de 53 % avec le modèle de Poisson – les distributions par classe sont nettement mieux reproduites ;
- $D_1 = 1\,396$ avec 1 251 degrés de liberté, à la place de 1 454 – les fréquences par classe ne sont que légèrement mieux estimées ;
- $D_2 = 2\,693$ avec 3 821 degrés, à la place de 3 010 – le modèle de Poisson randomisé est un sur modèle du modèle de Poisson.

Pour les quatre mesures retenues, le modèle de Poisson randomisé est meilleur. Toutefois, au niveau de B , les estimateurs obtenus ne diffèrent que très faiblement des paramètres estimés avec le modèle de Poisson. Cette généralisation (*Tableau 4*) n'apporte pas d'éléments neufs pour l'estimation du risque *a priori* : les fréquences sont bien les seules grandeurs observables utiles pour le risque *a priori*.

Si ce modèle n'améliore pas l'estimation des fréquences, le calcul du paramètre a permet d'avoir une mesure de la qualité de la segmentation puisqu'il est inversement proportionnel à la variance des fréquences individuelles dans les classes. En effet, plus la segmentation est fine, plus les classes sont homogènes, et donc plus a est grand.

Dans notre exemple, a vaut :

- 0.76 avant segmentation,
- 1.41 après segmentation (trois critères),

la segmentation permet d'expliquer 46 % de la variance relative.

INDIVIDUALISATION DU RISQUE EN ASSURANCE AUTOMOBILE

*Tableau 4. Garantie Vol : comparaison entre le modèle de Poisson
et le modèle de Poisson randomisé*

<i>Poisson</i>	<i>P. Randomisé</i>	<i>Poisson</i>	<i>P. randomisé</i>	<i>Poisson</i>	<i>P. randomisé</i>
1,00	1,00	1,00	1,00	1,00	1,00
1,37	1,37	0,93	0,93	0,77	0,77
1,58	1,58	1,05	1,05	0,64	0,64
1,94	1,94	1,20	1,20	0,53	0,53
2,74	2,74	1,31	1,31	0,33	0,33
4,06	4,07	1,46	1,46	0,18	0,18
3,70	3,70	1,32	1,32		
4,68	4,69	1,79	1,80		
4,49	4,50	1,82	1,82		
5,80	5,80	1,60	1,61		
		2,85	2,86		
		3,16	3,17		
		3,89	3,91		
		4,69	4,70		
		6,06	6,12		

Cette propriété permet de définir une famille d'estimateurs basée sur la minimisation de la variance intra-classe des lois Gamma (fréquences individuelles). On montre que les estimateurs ainsi obtenus sont équivalents à ceux obtenus par les méthodes classiques de minimisation de la variance intra-classe utilisées en analyse de variance (Rubinstein, 1992). L'intérêt de cette approche est de faire la part entre la variance intra-classe endogène (le processus de survenance des sinistres d'un assuré suit un processus de Poisson) qui est la justification de l'assureur, et la variance intra-classe exogène (distribution des fréquences individuelles à l'intérieure des classes, loi de Gamma) qui mesure la qualité de la segmentation.

2.1.3. Conclusion

Outre la bonne adéquation entre les grandeurs observables utiles et les estimateurs obtenus, l'intérêt de cette approche réside dans l'utilisation de modèles génétiques (c'est-à-dire définis à partir de la genèse des phénomènes étudiés). Cette propriété permet :

INDIVIDUALISATION DU RISQUE EN ASSURANCE AUTOMOBILE

- d’avoir une mesure de la qualité de la segmentation et donc de permettre la mise au point de procédures de sélection de variables par analogie avec les méthodes utilisées en analyse de variance ;
- de généraliser les estimateurs au cas d’assurés qui ne sont présents que sur une partie de l’année (Rubinstein, 1992) ;
- de construire des généralisations incrémentales ;
- de fournir un mode opératoire de simulation.

2.2. Estimation du coût moyen

Les grandeurs observables utiles sont les coûts moyens par classe. Préalablement à toute étude, il est important de bien définir le coût d’un sinistre.

- Doit-on prendre les frais ou les traiter de manière annexe ?
- Comment doit-on traiter le coût des sinistres réglés dans le cadre de la convention IDA ?
- Comment analyser les sinistres ouverts (non réglés), les sinistres « sans suite » ?

Les réponses à toutes ces questions ne sont pas uniques et ont un impact important sur les estimations des coûts. De plus, pour que le produit de la fréquence par le coût ait bien un sens, il est nécessaire que les choix soient cohérents avec ceux faits pour l’estimation des fréquences.

D’un point de vue méthodologique il y a deux approches possibles : l’une paramétrique, l’autre non paramétrique.

2.2.1. Modèles paramétriques

Dans ce type d’étude, il est indispensable de découper les garanties de base en sous-garanties homogènes pour les coûts afin de faciliter l’ajustement de la loi expérimentale à une loi simple. Par exemple, il faut séparer les vols totaux des vols partiels pour que les lois expérimentales soient uni-modales.

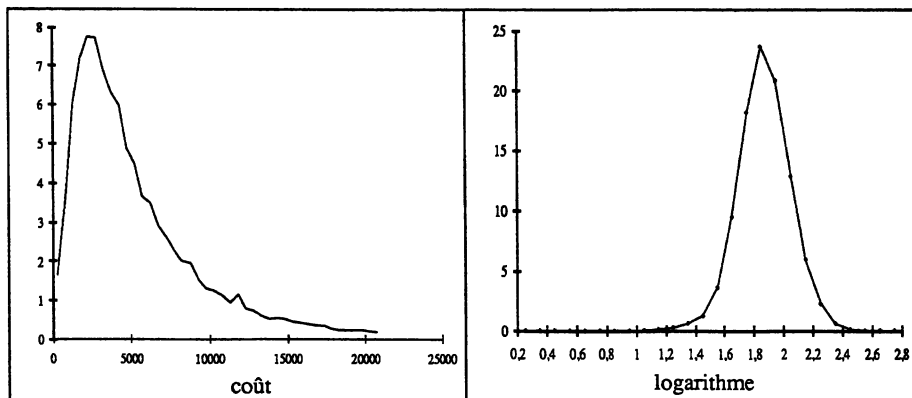
Une fois cette décomposition réalisée, il faut :

- choisir une famille de lois ;
- définir la fonction f reliant les paramètres de cette famille aux variables exogènes ;
- estimer les paramètres du modèle.

Contrairement au cas des fréquences, il n’existe pas de lois « naturelles » pour les coûts. Pour choisir la loi, nous sommes donc partis de la loi expérimentale des coûts (et/ou logarithmes des coûts) de l’ensemble de l’échantillon. A cette étape, du fait de l’absence de segmentation, les tests d’adéquation à une loi ne sont pas vérifiés : la loi expérimentale est un mélange de lois expérimentales qui dépendent des classes. Néanmoins, cette pré-étude permet de réduire le nombre de lois à analyser. Dans notre exemple, c’est la loi log Gamma qui semble la mieux adaptée (*figure 1*).

INDIVIDUALISATION DU RISQUE EN ASSURANCE AUTOMOBILE

Figure 1. Loi expérimentale des coûts, des logarithmes



Les grandeurs observables utiles sont les coûts moyens. De plus, afin que le tarif reste multiplicatif, nous nous limitons à la famille des estimateurs qui s'expriment comme un produit fonction des variables exogènes. Cette contrainte nous conduit à abandonner la loi log Gamma au profit de la loi log Normale.

La loi Normale a une moyenne par classe l_i fonction des variables exogènes ($l_i = C_i^T B$) et une variance σ^2 indépendante des classes. La solution obtenue par le maximum de vraisemblance est donnée par le système :

$$\forall j, \quad 1 \leq j \leq n \quad \sum_i c_{ij} C_i^T B = \sum_i c_{ij} \ln(x_i) \quad (2.2.1.1.a)$$

$$\sigma^2 = 1/m \sum_i (C_i^T B - \ln(x_i))^2 \quad (2.2.1.1.b)$$

où x_i est le coût du $i^{\text{ème}}$ sinistre

Le système (2.2.1.1.a) est équivalent au système obtenu par la méthode des marges, sous contrainte additive, appliquée aux logarithmes des coûts : le modèle respecte donc les logarithmes sur les marges. Toutefois, cela ne nous assure pas, lors du passage des logarithmes aux coûts, une bonne restitution des coûts :

$$m_i = \exp(l_i + \sigma^2/2) \quad (2.2.1.2)$$

Après segmentation, 60 % des assurés appartiennent à une classe où le test d'adéquation à la loi Normale est accepté (ce chiffre passe à 68 % pour la loi Gamma). Pour ces classes, le test de la moyenne est accepté dans 96 % des cas. La limite de cette approche réside dans l'instabilité vis-à-vis de la variance (supposée indépendante des classes). Cela se traduit, lors du retour aux coûts moyens (équation 2.2.1.2), par une mauvaise estimation de la moyenne globale (+ 1,8 %).

INDIVIDUALISATION DU RISQUE EN ASSURANCE AUTOMOBILE

Tableau 5. Garantie RC matérielle : comparaison entre les coûts moyens
des marges et les paramètres estimés

<i>Fréquence</i>	<i>Log normal</i>	<i>Fréquence</i>	<i>Log normal</i>	<i>Fréquence</i>	<i>Log normal</i>
1,00	1,00	1,00	1,00	1,00	1,00
1,06	1,05	1,09	1,04	0,95	0,97
1,07	1,06	1,14	1,11	0,96	0,98
1,13	1,11	1,16	1,11	0,94	0,97
1,14	1,11	1,16	1,11	0,95	0,97
1,23	1,19	1,21	1,17	0,95	0,97
1,17	1,13	1,19	1,14	0,90	0,93
1,20	1,19	1,26	1,19	1,14	1,11
1,18	1,13	1,32	1,22	0,83	0,88
1,03	1,01	1,38	1,21	1,01	1,00
		1,43	1,29	0,92	0,96
		1,21	1,17		

Comme dans le cas de l'estimation des fréquences, on constate que le modèle a un pouvoir de « décorrélation » par rapport au tableau de contingences sous-jacent (*Tableau 5*). L'amplitude des paramètres est moins importante que pour la fréquence, et ce, bien que le coefficient de variation des coûts soit plus grand. Le poids relatif de l'aléa endogène aux coûts est plus important ; cela va dans le sens d'une segmentation moins fine pour l'estimation des coûts.

2.2.2. Modèle non paramétrique

Dans cette approche, la seule hypothèse porte sur la décomposition des coûts en un terme ne dépendant que des variables exogènes et en un résidu u_i de moyenne nulle :

$$X_i = f(C_i, B) + u_i$$

Dans notre application, f est une fonction multiplicative :

$$f(C, B) = \exp(C^T B)$$

La solution est alors obtenue par minimisation de :

$$\sum_i (X_i - \exp(C_i^T B))^2$$

L'intérêt de cette approche est de travailler directement sur les coûts (grandeurs observables utiles) et non sur leurs logarithmes. De plus, elle ne nécessite pas la recherche d'une loi sous-jacente. Toutefois, cette abstraction n'est qu'en partie apparente, dans la mesure où la solution ainsi trouvée est équivalente à la solution du pseudo-maximum de vraisemblance à l'ordre 1 appliquée à la famille des lois Normales (Gouriéroux, 1989 a).

INDIVIDUALISATION DU RISQUE EN ASSURANCE AUTOMOBILE

Tableau 6. Garantie RC matérielle : comparaison entre méthode paramétrique (log normale) et méthode non paramétrique

<i>Log normal</i>	<i>MC non lin.</i>	<i>Log normal</i>	<i>MC non lin.</i>	<i>Log normal</i>	<i>MC non lin.</i>
1,00	1,00	1,00	1,00	1,00	1,00
1,05	1,05	1,04	1,08	0,97	0,98
1,06	1,06	1,11	1,13	0,98	0,99
1,11	1,09	1,11	1,15	0,97	1,03
1,11	1,08	1,11	1,14	0,97	0,96
1,19	1,17	1,17	1,20	0,97	0,98
1,13	1,10	1,14	1,18	0,93	0,92
1,19	1,15	1,19	1,25	1,11	1,10
1,13	1,10	1,22	1,29	0,88	0,85
1,01	0,98	1,21	1,34	1,00	0,98
		1,29	1,36	0,96	0,96
		1,17	1,18		

Les valeurs estimées par les deux méthodes (*tableau 6*) sont proches pour les modalités fortement représentées dans l'échantillon. Des différences apparaissent dans les cas peu représentés.

2.2.3. Conclusion

Les grandeurs observables utiles étant les coûts moyens, les méthodes non paramétriques suffisent. Elles sont plus simples à mettre en oeuvre, elles s'appliquent de la même manière à toutes les garanties et s'appuient directement sur les coûts. Toutefois, même en l'absence d'hypothèse sur la loi sous-jacente, il est important d'analyser les lois expérimentales. En effet, l'estimateur du pseudo-maximum de vraisemblance est d'autant plus précis que la famille sur laquelle on projette est proche de la loi expérimentale.

Les méthodes paramétriques restent utiles lorsque les coûts moyens ne sont pas les seules grandeurs utiles (loi de Pareto en réassurance par exemple).

Notons enfin qu'il existe des approches semi-paramétriques basées sur la méthode de l'échantillon saturé. Dans cette méthode, la famille de lois utilisée est déduite de la loi expérimentale globale par des transformations simples.

2.3. Estimation de la prime pure

Les grandeurs observables utiles sont les charges sinistres par classe. Toutefois, afin de fiabiliser les estimateurs, on se limite aux marges (Bailey, 1963 ; Jung, 1968 ; Van Eeghen, 1984). Les contraintes sont :

INDIVIDUALISATION DU RISQUE EN ASSURANCE AUTOMOBILE

- une expression multiplicative de l'estimateur de la prime pure par classe ;
- l'égalité, au niveau des marges, entre les charges sinistres estimées et les charges sinistres observées.

Ces contraintes se traduisent par le système

$$s(C_i) = \exp(C_i^T B) \quad (2.3.1.a)$$

$$\forall j, \quad 1 \leq j \leq n \quad \sum_i c_{ij} \exp(C_i^T B) = \sum_i c_{ij} s_i \quad (2.3.1.b)$$

La résolution algorithmique de ce système est simple et l'existence d'une solution vérifiée expérimentalement. De manière plus précise, nous avons la propriété suivante :

Propriété 2.3.1 :

a) Le système (2.3.1.b) admet une et une unique solution si et seulement si :

$$\sum_i s_i C_i \text{ appartient au cône polyédrique } \sum_i \mathbb{R}^+ C_i$$

b) L'estimateur ainsi obtenu est convergent.

Démonstration :

a) Se reporter à la démonstration dans « Tarification *a priori* et *a posteriori* des risques en assurance automobile » (Montocchio, 1992).

b) Prenons le modèle de régression à l'ordre 1 :

$$S_i = \exp(C_i^T B) + u_i$$

de loi : $f(x, C_i, B) = (\exp(C_i^T B))^x / \Gamma(x+1) \exp(-\exp(C_i^T B))$

L'expression de f est telle que nous pouvons appliquer la méthode du pseudo-maximum de vraisemblance (Gouriéroux, 1989a). Il en résulte que l'estimateur B^* obtenu par maximisation de :

$$\sum_i \ln(f(s_i, C_i, B))$$

est convergent. Par dérivation on obtient :

$$\forall j, \quad 1 \leq j \leq n \quad \partial \left(\sum_i \ln(f(s_i, C_i, B)) \right) / \partial b_j = \sum_i c_{ij} (s_i - \exp(C_i^T B)) = 0$$

qui est le système obtenu par la méthode des marges.

2.4. Comparaison entre l'approche globale et l'approche fréquence/coût

Nous avons appliqué les deux approches, d'une part, à la garantie RC matérielle en prenant la même segmentation pour les coûts et les fréquences, et, d'autre part, à la garantie RC corporelle en prenant des segmentations différentes (le coefficient de variation des coûts des sinistres corporels est égal à 9).

INDIVIDUALISATION DU RISQUE EN ASSURANCE AUTOMOBILE

Dans le premier cas (*tableau 7*), on constate qu'il y a peu de différence entre les estimateurs obtenus : l'écart maximal, après combinaison des trois critères, est de 5% et correspond à une classe très peu représentée dans le portefeuille.

Tableau 7. Garantie RC matérielle : comparaison entre une approche globale et une approche fréquence/coût

<i>Marges</i>	<i>Fréq. + Coût</i>	<i>Marges</i>	<i>Fréq. + Coût</i>	<i>Marges</i>	<i>Fréq. + Coût</i>
1,00	1,00	1,00	1,00	1,00	1,00
1,13	1,14	1,58	1,59	0,77	0,76
1,24	1,24	1,94	1,98	1,16	1,16
1,36	1,36	2,13	2,16	1,08	1,11
1,42	1,42	2,23	2,27	0,96	0,95
1,90	1,90	2,35	2,39	0,80	0,80
1,71	1,71	2,14	2,18	0,72	0,72
2,10	2,11	2,44	2,48	1,77	1,77
1,88	1,88	2,54	2,59	1,15	1,16
1,56	1,57	2,56	2,56	1,15	1,15
		2,49	2,51	1,36	1,37
		1,95	1,94		

Dans le second cas (*tableau 8*), on observe des résultats très différents, et ce, même dans des classes très représentées.

Tableau 8. Garantie RC corporelle : comparaison entre une approche globale et une approche fréquence/coût

<i>Marges</i>	<i>Fréq. + Coût</i>	<i>Marges</i>	<i>Fréq. + Coût</i>	<i>Marges</i>	<i>Fréq. + Coût</i>
1.00	1.00	1.00	1.00	1.00	1.00
1.70	1.15	2.27	2.03	0.61	0.81
0.97	1.16	3.34	3.04	1.11	1.17
0.95	1.31	3.87	3.44	0.88	1.03
1.14	1.30	3.96	3.73	1.06	1.23
1.32	1.50	7.07	4.80	0.77	0.85
1.40	1.56	4.98	4.00	0.85	0.90
1.31	1.48	5.31	5.30	1.81	2.17
1.50	2.12	7.12	4.81	0.44	0.84
2.21	2.86	3.48	5.57	2.94	1.15
		6.91	6.37	1.37	1.56
		2.44	1.55		

2.5. Conclusion

La méthode des marges peut, vu ses propriétés (existence, unicité et convergence), être utilisée lorsque le coefficient de variation des coûts n'est pas trop éloigné de 1 (la tolérance dépend de la taille de l'échantillon étudié). Cette méthode présente, en outre, l'avantage de ne pas nécessiter un système d'information complexe. Elle ne permet toutefois pas de simuler les évolutions d'un portefeuille.

A l'opposé, une approche fréquence/coût est bien adaptée aux cas où une segmentation différente est nécessaire. De plus, elle fournit des méthodes explicites de simulations. Enfin, grâce au modèle de Poisson randomisé, elle permet de traiter le problème de la tarification *a posteriori* comme le prolongement de la tarification *a priori*.

3. Tarification *a posteriori*

Nous avons vu, lors de l'étude du modèle de Poisson que les seules variables exogènes ne permettaient pas d'expliquer la queue de distribution, c'est-à-dire les risques aggravés. L'objet de la tarification *a posteriori* est de prendre en compte la sinistralité passée afin de mesurer l'aggravation du risque individuel en faisant la part entre le hasard et le comportement dans la survenance d'un sinistre.

3.1. Tarification *a posteriori* et modèle de Poisson randomisé

Dans le cadre du modèle de Poisson randomisé, le calcul de la loi conditionnelle de la fréquence d'un assuré en fonction des variables exogènes, de ses sinistres, de ses risques passés (i.e. classe de tarification) et de sa durée d'observation permet d'inclure l'historique dans l'estimation du risque individuel. Pour traiter ce problème, il est nécessaire de généraliser la notion de comportement constant au cas du changement de classe. Nous faisons l'hypothèse que :

$$\Lambda_i(t)/\lambda(C_i(t)) = U_i \quad (3.1.1)$$

où U_i est une variable aléatoire indépendante des variables exogènes. En d'autres termes, le rapport entre la fréquence de l'assuré et la fréquence de sa classe de tarification est constant dans le temps.

Disposant de cette nouvelle hypothèse, nous pouvons calculer la loi de $\Lambda_i(t)$ conditionnellement à $(n_i, \lambda(C_i(u)) \ u \leq t)$, où n_i est le nombre de sinistres sur la période $[0, t]$.

Propriété 3.1 : la loi de $\Lambda_i(t)/(n_i, \lambda(C_i(u)), u \leq t)$ est une loi Gamma de moyenne :

$$\lambda_i'(t) = \lambda_i(t) (a + n_i) / (a + \int_{[0, t]} \lambda_i(u) du) \quad (3.1.2)$$

et de variance relative :

$$\sigma_i'(t)^2 / \lambda_i'(t)^2 = 1 / (a + n_i)$$

INDIVIDUALISATION DU RISQUE EN ASSURANCE AUTOMOBILE

Définition 3.2 : On appelle exposition au risque sur une période $[t, t']$ la quantité définie par :

$$I_{[t, t']} = \int_{[t, t']} \lambda_t(u) du \quad (3.1.3)$$

Démonstration :

Notons $f_{n_t, \lambda(C_i(u), u \leq t)}$ la densité de la loi conditionnelle. En appliquant la formule de Bayes on obtient :

$$f_{n_t, \lambda(C(u), u \leq t)}(x) = \frac{P_{\Lambda(t)=x}(N_{[0, t]} = n_t) f(x)}{\int_{x > 0} P_{\Lambda(t)=x}(N_{[0, t]} = n_t) f(x) dx}$$

conditionnellement à $\Lambda(t) = x$, $N_{[0, t]}$ suit une loi de Poisson de paramètre :

$$(x/\lambda_t(t)) \int_{[0, t]} \lambda_t(u) du = x (I_{[0, t]}/\lambda_t(t))$$

d'où : $P_{\Lambda(t)=x}(N_{[0, t]} = n_t) = ((x^{n_t} (I_{[0, t]}/\lambda_t(t))^{n_t}/n_t!) \exp(-x (I_{[0, t]}/\lambda_t(t)))$

En introduisant la fonction f (loi non conditionnelle de Λ), on obtient :

$$f_{n_t, \lambda(C(u), u \leq t)}(x) = K x^{a+n_t-1} \exp(-x (a + I_{[0, t]}/\lambda_t(t)))$$

K étant une constante de normalisation. Si on pose :

$$\tau' = (a + I_{[0, t]})$$

et $a' = a + n_t$, on obtient :

$$f_{n_t, \lambda(C_i(u), u \leq t)}(x) = \tau'^{a'}/\Gamma(a') x^{(a'-1)} \exp(-\tau'x) \quad (3.1.4)$$

qui est la densité d'une loi Gamma.

L'estimateur optimal de la fréquence individuelle est, connaissant la classe tarifaire et l'historique de l'assuré, la moyenne de la loi conditionnelle. Il est proportionnel au risque *a priori* – moyenne de la loi non conditionnelle – dans un rapport égal à l'inverse de la variance relative des classes *a* plus le nombre de sinistres constaté sur $[0, t]$, divisés par *a* plus l'exposition aux risques sur $[0, t]$ – c'est-à-dire le nombre de sinistres prévu *a priori*.

$$\lambda'_t(t)/\lambda_t(t) = (a + n_t)/(a + I_{[0, t]}) \quad (3.1.5)$$

Le risque individuel converge en $+\infty$ vers la fréquence expérimentale multipliée par le rapport entre le risque *a priori* à l'instant t et le risque *a priori* moyen sur l'intervalle $[0, t]$. Dans le cas où l'assuré ne change pas de classe, le risque individuel est égal à la fréquence expérimentale ; il est indépendant de la tarification *a priori*.

L'équation (3.1.5) est du même type que les formules obtenues en appliquant la théorie de la crédibilité (Buhlmann, 1964 ; 1970). En effet, si l'assuré n'a pas changé de classe on a :

$$\lambda'_t(t') = (n_t/t) z_t + \lambda_t(1 - z_t) \quad (3.1.5')$$

où $z_t = t \lambda_t/(a + t \lambda_t)$ est le coefficient de crédibilité, il dépend de la classe.

INDIVIDUALISATION DU RISQUE EN ASSURANCE AUTOMOBILE

Toutefois, la crédibilité ne permet pas de :

- traiter simultanément la tarification *a priori* et la tarification *a posteriori* sauf à prendre des modèles de crédibilité hiérarchique ;
- de travailler en continu ;
- de traiter les problèmes de changement de classe tarifaire et/ou de garantie.

La formule (3.1.5) permet de calculer un système de bonus-malus optimal. Pour une période $[t, t']$ sans sinistre (et sans changement de classe) nous avons :

$$(\lambda'_i(t') - \lambda'_i(t)) / \lambda'_i(t) = -I_{[t, t']} / (a + I_{[0, t']}) \quad (3.1.6)$$

Le coefficient de réduction est tel que :

- plus l'exposition sur $[t, t']$ est importante, plus la baisse est forte (la qualité reflétée par l'absence de sinistre est d'autant plus importante que l'assuré avait un risque *a priori* important) ;
- plus a est grand, moins la baisse est importante (c'est-à-dire que plus la segmentation *a priori* est « bonne » moins le poids de la tarification *a posteriori* est important) ;
- plus l'historique est important, moins le poids d'une année est important.

Pour une période $[t, t']$ avec un sinistre nous avons :

$$\frac{\lambda'_i(t') - \lambda'_i(t)}{\lambda'_i(t)} = \frac{a + I_{[0, t']} - I_{[t, t']}(a + n_{[0, t']})}{(a + I_{[0, t']})(a + n_{[0, t']})} \quad (3.1.7)$$

Le coefficient de majoration (le premier terme est en général positif en assurance automobile) est tel que :

- plus l'exposition sur $[t, t']$ est importante, moins la hausse est forte (la prime *a priori* incluait, pour une part plus importante, la sinistralité observée) ;
- plus a est grand, moins la majoration est importante (c'est-à-dire que plus la segmentation *a priori* est « bonne », moins le poids de la tarification *a posteriori* est important) ;
- plus l'historique est important, moins la majoration est importante.

L'expression des coefficients de réduction-majoration montre qu'il est important de segmenter le tarif tant pour le calcul des paramètres – la valeur de a calculée sans segmentation conduit à accorder un poids trop important au tarif *a posteriori* –, que pour le calcul des coefficients – les poids dépendent du tarif *a priori*.

3.2. Conclusion

Les limites du modèle résident dans :

- La non reproduction de la queue de distribution. Néanmoins, il est possible de définir des profils de sinistralité nécessitant un traitement spécial (méthode de scoring post-sinistres par exemple).
- Le caractère mono-variable de l'approche, bien que des solutions commencent à apparaître (Larsen, 1991 ; Partrat, 1992).

INDIVIDUALISATION DU RISQUE EN ASSURANCE AUTOMOBILE

L'intérêt de ce modèle est :

- de fournir une approche homogène de la tarification ;
- de permettre de simuler les impacts en terme de résultats d'un système de bonus-malus commercial ou d'une politique de surveillance des risques.

4. Conclusion

La démarche que nous avons présentée s'applique de manière plus générale à la tarification de tous les produits de masse. La concurrence a contraint les assureurs à développer cette approche pour les tarifs automobile et maintenant, il est probable que, fort de cette expérience, elle se généralise à d'autres types de risques : systèmes de bonus-malus en assurance habitation par exemple.

La partie traitée ci-dessus n'est qu'une des étapes dans le calcul d'un tarif commercial. Aujourd'hui, les compagnies disposent peu ou prou d'outils performants permettant de passer du calcul du coût de la « matière première » (la charge sinistre prévisible) au prix de vente. Si ce manque n'était pas important hier, l'individualisation de plus en plus poussée des risques rend le développement de tels outils indispensable.

BIBLIOGRAPHIE

- BAILEY R.A. (1963) "Insurance Rates with Minimum Bias", *Proceedings of the Casualty Actuarial Society*, vol. 50, 4-11.
- BUHLMANN H. (1964) "A Distributive Free Method for General Risk Problems", *ASTIN Bulletin*, vol. 3.
- BUHLMANN H. (1970) *Mathematical Methods in Risk Theory*, Springer-Verlag.
- HARRINGTON S.E. (1986) "Estimation and Testing for Functional Form in Pure Premium Regression Models", *ASTIN Bulletin*, vol. 16, 31-43.
- GOURIÉROUX Ch., MONFORT A. (1989a) *Statistique et modèles économétriques*. Economica.
- GOURIEROUX Ch. (1989b) *Econométrie des variables qualitatives*, Economica.
- JUNG J. (1968) "On Automobile Insurance Rate Making", *Astin Bulletin*, vol. 5, 41-48.
- LARSEN C.R., DENGSOE Ch., EGEBO J.H. et HASEN J.E. (1991) "Risk Analysis of Customers not Products", *SCOR, International Prize in Actuarial Science*, 71-119.
- MONTOCCHIO L., PESTRE G., PINQUET J. et ROBERT J.C. (1992) *Tarification a priori et a posteriori des risques en assurance automobile*, Mémoire du Centre d'Études Actuarielles, Cycle 1990-1991.
- PARTRAT Ch. (1992) "Risque et statistique", *Bulletin de l'IAf*, n° 356-357, 11-24.
- RUBINSTEIN J.C. (1992) *Études comparées de modèles de tarification*, ISUP, Mémoire de 3^e année.
- SIBONY D. (1990) "Psychanalyse du risque", *Risque*, vol. 1, 52-55.
- VAN EEGHEN J., GREUP E.K. et NUSSEN J.A. (1984) "Rate Making", *Surveys of Actuarial Studie*, National-Nederlanden N.V., Rotterdam : The Netherlands.