Anne-Marie Gossiaux et Jean Lemaire, Bruxelles

Méthodes d'ajustement de distributions de sinistres

§1 Introduction

Lorsqu'il s'agit d'ajuster les observations faites dans un portefeuille automobile sur le nombre de sinistres occasionnés par un véhicule pendant une année, il est presque toujours fait appel à la loi binomiale négative; ce modèle présente des avantages théoriques non négligeables, mais il semble qu'une des raisons principales qui motivent son choix soit, d'après les auteurs qui l'utilisent, la qualité des ajustements. Nous nous proposons de vérifier, à l'aide d'exemples, si ce dernier argument mérite autant de poids: nous allons ajuster 6 distributions prélevées dans la littérature actuarielle de 6 manières différentes, et comparer les résultats.

§2 Ajustements

Soit une distribution observée $(k, n_k\,;\, k=0,1,\ldots,r)$, où n_k représente le nombre de véhicules touchés par k sinistres en un an. Soit $\bar{x}$ la moyenne, s^2 la variance, et n l'effectif de cette distribution. On peut envisager un ajustement par l'une des distributions théoriques $(k, p_k\,;\, k=0,1,\ldots,r)$ suivantes.

1 *Distribution de Poisson*

$$p_k = \frac{e^{-\lambda}\lambda^k}{k\,!} \qquad \lambda > 0$$

moyenne $m = \lambda$

variance $\sigma^2 = \lambda$

Estimateur (méthode des moments): $\hat{\lambda} = \bar{x}$

(méthode du maximum de vraisemblance): $\hat{\lambda} = \bar{x}$.

2 *Distribution binomiale négative*

$$p_k = \frac{\Gamma(a+k)}{k\,!\,\Gamma(a)}\left(\frac{1}{1+\tau}\right)^k\left(\frac{\tau}{1+\tau}\right)^a \qquad a,\ \tau > 0$$

moyenne $m = \frac{a}{\tau}$

variance $\sigma^2 = \frac{a}{\tau}\left(1 + \frac{1}{\tau}\right)$

Estimateurs (méthode des moments): $\hat{\tau} = \frac{\bar{x}}{s^2 - \bar{x}}$

$\hat{a} = \bar{x}\hat{\tau}$

(méthode du maximum de vraisemblance): $\hat{\tau} = \frac{\hat{a}}{\bar{x}}$

$\hat{a}$ est solution de l'équation transcendante

$$\sum_{k=0}^{r} n_k\left(\frac{1}{a} + \ldots + \frac{1}{a+k-1}\right) = \sum_{k=0}^{r} n_k \operatorname{Log}\left(1 + \frac{\bar{x}}{a}\right)$$

3 *Distribution géométrique généralisée* [4]

$$p_0 = 1 - a\theta \qquad 0 \leqq \theta \leqq 1$$

$$p_k = a\theta^k(1-\theta) \qquad k \geqq 1 \quad 0 \leqq a \leqq \frac{1}{\theta}$$

moyenne $m = \frac{a\theta}{1-\theta}$

variance $\sigma^2 = \frac{a\theta(1+\theta-a\theta)}{(1-\theta)^2}$

Estimateurs (méthode des moments): $\hat{\theta} = \frac{s^2 - \bar{x} + \bar{x}^2}{s^2 + \bar{x} + \bar{x}^2}$

$\hat{a} = \frac{2\bar{x}^2}{s^2 + \bar{x} + \bar{x}^2}$

(méthode du maximum de vraisemblance): $\hat{\theta} = 1 - \frac{n - n_0}{n\bar{x}}$

$\hat{a} = \frac{n - n_0}{n\hat{\theta}}$

4 *Loi mixte de Poisson* [8]

$$p_k = a_1 \frac{e^{-\lambda_1}\lambda_1^k}{k!} + a_2 \frac{e^{-\lambda_2}\lambda_2^k}{k!} \qquad a_1, a_2, \lambda_1, \lambda_2 > 0, \; a_1 + a_2 = 1$$

moyenne $\quad m = a_1\lambda_1 + a_2\lambda_2$

variance $\quad \sigma^2 = \alpha_2 - m^2$, où

$$\alpha_2 = a_1\lambda_1^2 + a_1\lambda_1 + a_2\lambda_2^2 + a_2\lambda_2$$

dissymétrie $\quad \mu_3 = \alpha_3 - 3m\alpha_2 + 2m^3$, où

$$\alpha_3 = a_1\lambda_1^3 + a_2\lambda_2^3 + 3(a_1\lambda_1^2 + a_2\lambda_2^2) + a_1\lambda_1 + a_2\lambda_2$$

Estimateurs (méthode des moments): $\hat{a}_1 = \dfrac{a - \hat{\lambda}_2}{\hat{\lambda}_1 - \hat{\lambda}_2}$

$$\hat{\lambda}_1, \hat{\lambda}_2 = \frac{S \pm \sqrt{S^2 - 4P}}{2}$$

avec $S = \dfrac{c - ab}{b - a^2}$, $P = \dfrac{ac - b^2}{b - a^2}$, $a = \bar{x}$, $b = \alpha_2^* - \bar{x}$, $c = \alpha_3^* - 3\alpha_2^* + 2\bar{x}$, α_2^* et α_3^* étant, respectivement, les moments d'ordre 2 et 3 par rapport à l'origine de la distribution observée.

§3 Exemples

Les 6 distributions observées suivantes ont été ajustées par les 6 méthodes proposées. Pour chaque ajustement, les tableaux qui suivent fournissent les effectifs théoriques, la valeur observée χ^2 ainsi que la probabilité pour que cette valeur soit dépassée.

Il convient cependant d'émettre quelques réserves en ce qui concerne la précision des tests χ^2 effectués; il est bien connu que, lorsqu'on utilise des estimateurs asymptotiquement normaux et efficaces, l'écart

$$\sum_{k=0}^{r} \frac{(n_k - np_k)^2}{np_k}$$

converge, sous l'hypothèse nulle, vers une loi χ^2 à $r - s$ degrés de liberté où s est le nombre de paramètres estimés. Mais les estimateurs obtenus par la méthode des moments n'étant pas asymptotiquement normaux et efficaces, cette distribution limite n'est pas applicable. La seule manière d'effectuer une comparaison valable entre les différentes méthodes consiste alors à supposer les estimateurs donnés, et à tester la qualité des ajustements au moyen d'une χ^2 à r degrés de liberté.

Remarquons également que, dans la littérature statistique, les opinions sont fort divergentes en ce qui concerne la rapidité de la convergence; certains auteurs exigent de regrouper toutes les classes dont l'effectif théorique est inférieur à 20; d'autres sont plus tolérants et estiment que l'approximation est bonne dès que tous les effectifs théoriques sont supérieurs à 1. La décision à prendre n'est malheureusement pas sans importance sur le résultat du test: il est fréquent qu'un ajustement, accepté lorsqu'on suit les recommandations d'un auteur, ne l'est plus lorsque les regroupements sont effectués d'une autre manière. Ceci démontre que la distribution limite est loin d'être atteinte, malgré les effectifs considérés, et illustre le fait que le test χ^2 (comme d'ailleurs tout test basé sur une distribution asymptotique!) ne constitue qu'une approximation.

Dans ce qui suit, nous avons regroupé les classes extérieures de manière à obtenir des effectifs théoriques au moins égaux à 5.

Exemple 1 [6]:

Belgique (1975–1976). Effectif 106974 $\bar{x}=0{,}1011$ $s^2=0{,}1074$

k	Observations	Distribution de Poisson	Binomiale nég. Méth. des mom.	Binomiale nég. Max. de vrais.	Dist. Géom. Méth. des mom.	Dist. Géom. Max. de vrais.	Dist. mixte de Poisson
0	96978	96689,6	96985,4	96980,8	96978,0	96978,0	96975,0
1	9240	9773,5	9222,5	9230,9	9239,0	9240,7	9252,1
2	704	493,9	711,7	708,6	699,7	698,2	685,0
3	43	16,6	50,7	50,1	53,0	52,7	57,0
4	9	0,4	3,6	3,4	4,0	4,0	4,6
5	0	0	0	0,2	0,3	0,3	0,3
Nombre de classes après regroupement		4	4	4	4	4	5
χ^2_{obs}		191,33	0,21	0,09	0,53	0,49	9,17
degrés de liberté		3	3	3	3	3	4
$\mathbb{P}(\chi^2 > \chi^2_{obs})$		0	0,9760	0,9930	0,9123	0,9211	0,0570

Exemple 2 [5]:
Zaïre (1974). Effectif 4000 $\bar{x}=0{,}0865$ $s^2=0{,}1225$

k	Observations	Distribution de Poisson	Binomiale nég. Méth. des mom.	Binomiale nég. Max. de vrais.	Dist. Géom. Méth. des mom.	Dist. Géom. Max. de vrais.	Dist. mixte de Poisson
0	3719	3668,5	3720,9	3719,2	3720,6	3719,0	3717,9
1	232	317,3	227,2	229,9	225,6	228,2	235,2
2	38	13,7	40,3	39,9	43,4	42,9	33,2
3	7	0,4	8,7	8,4	8,3	8,1	10,2
4	3	0,0	2,1	1,9	1,6	1,5	2,7
5	1	0,0	0,5	0,5	0,3	0,3	0,6
6	0	0,0	0,1	0,1	0,1	0,1	0,1
Nombre de classes après regroupement		3	4	4	4	4	4
χ^2_{obs}		110	0,59	0,36	0,90	0,87	1,24
degrés de liberté		2	3	3	3	3	3
$\mathbb{P}(\chi^2>\chi^2_{obs})$		0	0,8987	0,9484	0,8254	0,8327	0,7434

Exemple 3 [7]:
Belgique (1958). Effectif 9461 $\bar{x}=0{,}2144$ $s^2=0{,}2889$

k	Observations	Distribution de Poisson	Binomiale nég. Méth. des mom.	Binomiale nég. Max. de vrais.	Dist. Géom. Méth. des mom.	Dist. Géom. Max. de vrais.	Dist. mixte de Poisson
0	7840	7635,6	7871,3	7847,0	7878,1	7840,0	7825,6
1	1317	1636,7	1251,9	1288,4	1235,7	1295,7	1364,7
2	239	175,4	261,1	256,5	271,1	260,0	189,0
3	42	12,5	58,8	54,1	59,5	52,2	53,2
4	14	0,7	13,7	11,7	13,0	10,5	19,8
5	4	0,0	3,3	2,6	2,9	2,1	6,4
6	4	0,0	0,8	0,6	0,6	0,4	1,8
7	1	0	0	0,1	0,1	0,1	0,4
8	0	0	0	0	0	0	0,1
Nombre de classes après regroupement		4	5	5	5	5	6
χ^2_{obs}		294,26	11,17	7,32	16,95	12,38	18,98
degrés de liberté		3	4	4	4	4	5
$\mathbb{P}(\chi^2>\chi^2_{obs})$		0	0,0247	0,1199	0,0020	0,0147	0,0019

Exemple 4 [1]:
Suisse (1961). Effectif 119853 $\bar{x}=0,1551$ $s^2=0,1793$

k	Observations	Distribution de Poisson	Binomiale nég. Méth. des mom.	Binomiale nég. Max. de vrais.	Dist. Géom. Méth. des mom.	Dist. Géom. Max. de vrais.	Dist. mixte de Poisson
0	103704	102629,6	103760,8	103723,6	103760,9	103704	103692,7
1	14075	15922,0	13927,3	13989,9	13926,8	14025,4	14116,0
2	1766	1235,1	1873,5	1857,1	1874,0	1844,3	1714,4
3	255	63,9	252,2	245,2	252,2	242,5	278,3
4	45	2,5	34,0	32,3	33,9	31,9	44,8
5	6	0,1	4,6	4,2	4,6	4,2	6,1
6	2	0	0,6	0,6	0,6	0,6	0,7
7	0	0	0,1	0,1	0,1	0,1	0,1
Nombre de classes après regroupement		4	6	6	6	6	6
χ^2_{obs}		2002	12,73	12,34	12,87	11,48	3,80
degrés de liberté		3	5	5	5	5	5
$\mathbb{P}(\chi^2>\chi^2_{\text{obs}})$		0	0,0260	0,0304	0,0246	0,0427	0,5786

Exemple 5 [2]:
Allemagne (1960). Effectif 23589 $\bar{x}=0,1442$ $s^2=0,1639$

k	Observations	Distribution de Poisson	Binomiale nég. Méth. des mom.	Binomiale nég. Max. de vrais.	Dist. Géom. Méth. des mom.	Dist. Géom. Max. de vrais.	Dist. mixte de Poisson
0	20592	20420,9	20605,8	20596,8	20605,3	20592,0	20588,7
1	2651	2945,1	2615,5	2631,0	2616,8	2640,2	2662,2
2	297	212,4	322,8	318,4	321,8	314,3	285,0
3	41	10,2	39,5	37,8	39,6	37,4	44,5
4	7	0,4	4,8	4,5	4,9	4,5	7,5
5	0	0,0	0,6	0,6	0,6	0,5	1,1
6	1	0,0	0,1	0,1	0,1	0,1	0,1
7	0	0	0	0	0	0	0
Nombre de classes après regroupement		4	5	5	5	5	5
χ^2_{obs}		296,61	4,26	3,37	2,73	2,21	0,90
degrés de liberté		3	4	4	4	4	4
$\mathbb{P}(\chi^2>\chi^2_{\text{obs}})$		0	0,3720	0,4979	0,6040	0,6972	0,9246

Exemple 6 [3]:
Grande-Bretagne (1968). Effectif 421 240 $\bar{x}=0{,}1317$ $s^2=0{,}1385$

k	Observations	Distribution de Poisson	Binomiale nég. Méth. des mom.	Binomiale nég. Max. de vrais.	Dist. Géom. Méth. des mom.	Dist. Géom. Max. de vrais.	Dist. mixte de Poisson
0	370412	369246,9	370460,0	370438,9	370404,8	370412,2	370408,9
1	46545	48643,6	46413,2	46451,3	46567,6	46563,5	46557,5
2	3935	3204,1	4044,0	4030,5	3908,7	3906,8	3916,8
3	317	140,7	300,9	297,8	328,2	327,8	327,5
4	28	4,6	20,5	20,1	27,8	27,4	27,1
5	3	0,0	1,4	1,3	2,5	2,1	2,0
6	0	0	0	0,1	0,4	0	0
Nombre de classes après regroupement		4	5	5	5	5	5
χ^2_{obs}		543,72	7,96	7,89	0,57	0,64	1,05
degrés de liberté		3	4	4	4	4	4
$\mathbb{P}(\chi^2 > \chi^2_{\text{obs}})$		0	0,0931	0,0957	0,9663	0,9585	0,9021

§4 Conclusions

1. Comme il a déjà été remarqué à maintes reprises, la distribution de Poisson ne convient en aucun cas, même lorsque la variance observée est proche de la moyenne.
2. A l'exception de l'exemple 6, la méthode du maximum de vraisemblance fournit de meilleurs ajustements que la méthode des moments. L'écart est cependant si faible que l'on peut se demander si la plus grande simplicité des calculs pour la méthode des moments ne compense pas la précision un peu supérieure de l'autre méthode.
3. La comparaison des différents exemples ne permet pas de dégager une loi particulière. En effet, si la binomiale négative réalise le meilleur ajustement pour les exemples 1 à 3 (encore faudrait-il préciser que pour l'exemple 3 il conviendrait de parler de «moins mauvais ajustement»), elle arrive en dernier lieu pour les exemples 4 à 6. Dans l'exemple 4, seule la loi mixte de Poisson est satisfaisante. Dans l'exemple 6, la distribution géométrique l'emporte de peu devant la loi mixte de Poisson, tandis que l'ajustement par une binomiale négative est rejeté.

Ces exemples nous paraissent affaiblir un peu les arguments d'ordre pratique en faveur de la loi binomiale négative. En fait, comme les distributions à ajuster comportent fort peu de classes, un grand nombre de lois théoriques peuvent fournir des approximations satisfaisantes, et il convient de se montrer circonspect avant d'adopter l'une d'entre elles.

Anne-Marie Gossiaux
Jean Lemaire
Université Libre de Bruxelles
Campus de la Plaine, C.P. 210
50, bd du Triomphe
B-1050 Bruxelles

Bibliographie

[1] F. Bichsel: Erfahrungs-Tarifierung in der Motorfahrzeughaftpflicht-Versicherung. MVSV 64 (1964), pp. 119–143.
[2] M. Derron: Mathematische Probleme der Automobilversicherung. MVSV 62 (1962), pp. 103–123.
[3] P. Johnson et G. Hey: A review of the scope for using claims histories of individual policies in risk assessment. Bulletin A.R.A.B. 66 (1971), pp. 159–195.
[4] D. Justens: Le bonus-malus en assurance automobile: une nouvelle approche analytique des probabilités de transition. Mémoire de licence. U.L.B. 1979.
[5] J. Lemaire: Sur les critères de détermination d'un tarif bonus-malus optimal. Application à l'assurance automobile à Kinshasa. Ann. Fac. Sci. UNAZA. Vol. I (1975), pp. 338–366.
[6] J. Lemaire: Selection procedures of regression analysis applied to automobile insurance. MVSV 77 (1977), pp. 143–160.
[7] P. Thyrion: Contribution à l'étude du bonus pour non-sinistre en assurance auto. ASTIN I (1960), pp. 143–162.
[8] A. Tröbliger: Mathematische Untersuchungen zur Beitragsermässigung in der Kraftfahrversicherung. Blätter DGVM, V (1961), pp. 327–348.

Résumé

6 distributions observées de sinistres dans un portefeuille RC auto ont été ajustées de 6 manières différentes. Aucune loi de probabilité ne paraît émerger comme fournissant uniformément le meilleur ajustement.

Zusammenfassung

6 beobachtete Schadenverteilungen in der Motorfahrzeughaftpflicht-Versicherung werden auf 6 verschiedene Arten ausgeglichen. Keines der Verfahren scheint sich dadurch auszuzeichnen, dass es eine beste Ausgleichung liefert.

Summary

6 observed claims distributions in an automobile third party liability portfolio were fitted in 6 different ways. No single probability law seems to emerge as providing "the" best fit.