Kurzmitteilungen

Objekttyp: **Group**

Zeitschrift: **Mitteilungen / Schweizerische Vereinigung der Versicherungsmathematiker = Bulletin / Association Suisse des Actuaires = Bulletin / Swiss Association of Actuaries**

Band (Jahr): **- (1992)**

Heft 1

PDF erstellt am: **30.04.2024**

D. Kurzmitteilungen

M.N. ISLAM and P.C. CONSUL, Calgary

A Probabilistic Model for Automobile Claims

1 Introduction

In a modern world an automobile has become a virtual necessity for almost everyone and there are some concerns in purchasing an insurance policy. The premium is usually determined by a merit-demerit rating system based on one's previous driving records. According to such a system the premium that the insured pays decreases if one does not make any claims (e.g. bodily injuries or property damage), increases if one does. Many attempts have been made to develop mathematical models for the distribution of automobile accidents or claims. Among others *Gossiaux* and *Lemaire* (1981) reported that they applied different distributions namely Poisson, inflated Poisson, Negative Binomial Distribution (NBD), and a mixed Poisson distribution to six different observed automobile accident data, obtained from five countries and studied by other researchers; and found that no single probability model provided a good fit to the data sets. Recently, *Panjer* (1987) proposed a generalized Poisson Pascal (GPP) distribution for the modelling of the number of automobile accidents or claims, and *Consul* (1990) has suggested the generalized Poisson distribution (GPD) as a plausible model for the purpose. The question of a better model becomes more interesting when one searches for an underlying explanation for appropriate factors causing the automobile accidents.

In this paper we derive the Consul distribution, introduced by *Consul* and *Shenton* (1975), as a probabilistic model for the distribution of the number of automobile accidents or claims in automobile insurance. We also obtain estimators of the parameters by three different methods. Finally, we fit the model to the same data sets as used by *Panjer* (1987) and by *Consul* (1990).

2 The Consul Model

Let us consider a population of vehicles (cars, trucks, jeeps, station wagons, trailors, or any other vehicles for which one has to buy an insurance) that may be involved in an automobile accident on a given stretch of highways in a given period. The complete and realistic mechanism which produces a single or multiple vehicle accidents is of course complicated. However, it may be approximated by a branching process. We will use the term *claim* interchangeably with *automobile accident,* so no confusion will arise.
An automobile accident always starts by a single vehicle either
(i) by collision with other vehicles either in front or on the side, or
(ii) by noncollision (e.g. skidding and running off the road, overturning, etc.). When the visibility is not clear or the road is slippery, the collision of a single vehicle may result into a multiple vehicle accident as the incoming vehicles are unable to stop quickly and keep hitting the vehicles already in collision. In a single vehicle accident the process stops right there. However, in a multiple vehicle accident, the first vehicle starting the accident can be said to belong to the 0-th generation in a Galton-Watson branching process so that $X_0 = 1$. Let this vehicle generate X_1 vehicles to join the accident. We assume that the probability generating function (pgf) of X_1 is given by $g(s) = (1 - \theta + \theta s)^m$ where (i) $m \in N^+$, $0 < \theta < 1$, or (ii) $m < 0$, $\theta < 0$. The process may continue forming different generations of branching chain of vehicles or may come to an end. Let X_n $(n \geq 0)$ denote the number of vehicles in the nth generation and then $\{X_n : n \geq 0\}$ becomes a Galton-Watson branching process. Let $Z_n = X_1 + \cdots + X_n$, and $Z = \sum_{i=0}^{\infty} X_i$. Let $F_n(s)$ be the pgf of Z_n, $n = 0, 1, 2, \ldots$, and let $F(s) = \sum_{i=1}^{\infty} P(Z = i)s^i$. Then by a standard result one may obtain the following recurrence relation

$$F_{n+1}(s) = s\, g[F_n(s)] \tag{2.1}$$

and also obtain the following functional equation

$$F(s) = s\, g[F(s)] \tag{2.2}$$

Replacing s by t and $F(t) = s$ into (2.2), we obtain

$$s = t\, g(s). \tag{2.3}$$

A Lagrange expansion of t as a function of s may be easily obtained from the above to give the probabilities for the total number of vehicles in an automobile accident generated by a single vehicle as

$$P(X = x) = \frac{1}{x}\binom{mx}{x-1}\theta^{x-1}(1-\theta)^{mx-x+1}, \qquad x = 1, 2, \ldots, \tag{2.4}$$

where (i) $m \in N^+$, $\quad 0 < \theta < 1$ such that $1 \leq m \leq \theta^{-1}$, or (ii) $m < 0$, $\quad \theta < 0$ such that $0 < m\theta \leq 1$. The mean and the variance of the model are given by Islam and Consul (1991) as

$$\mu = E(X) = (1 - m\theta)^{-1} \tag{2.5}$$

and

$$\sigma^2 = m\theta(1 - \theta)(1 - m\theta)^{-3}. \tag{2.6}$$

3 Estimation of Parameters

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the model (2.4) and the observed values in the sample be given by $1, 2, \ldots, k$, with frequencies f_i, $i = 1, 2, \ldots, k$ where $f_1 + f_2 + \cdots + f_k = n$. Also, let $\bar{x}$ and s^2 be the sample mean and sample variance respectively. The parameters m and θ can be estimated by the following three methods.

(1) Moment Estimators

On equating the first two moments of the Consul distribution, given by (2.5) and (2.6), to the corresponding sample moments respectively, we obtain the moment estimates θ^* and m^* as

$$\theta^* = 1 - \frac{s^2}{(\bar{x})^2(\bar{x} - 1)} \tag{3.1}$$

and

$$m^* = (\theta^*)^{-1}\left[1 - (\bar{x})^{-1}\right]. \tag{3.2}$$

(2) Estimators Based Upon Mean and the First Frequency (MFF)

By equating the first relative frequency and the sample mean $\bar{x}$ with the probability $P(X = 1) = (1 - \theta)^m$ and the population mean $\mu = (1 - m\theta)^{-1}$ respectively, the estimates θ^{**} and m^{**} of the parameters θ and m are given by the solution of the following two non-linear equations

$$\ln\left(\frac{f_1}{n}\right) - \frac{1 - (\bar{x})^{-1}}{\theta}\ln(1 - \theta) = 0 \tag{3.3}$$

and

$$m = \theta^{-1}\left[1 - (\bar{x})^{-1}\right]. \tag{3.4}$$

Since closed form solution for θ is not available from (3.3), one has to consider a numerical solution. The equation (3.3) may be solved for the estimate θ^{**} by plotting the functions

$$h_1(\theta) = \theta \ln\left(\frac{f_1}{n}\right) \tag{3.5}$$

and

$$h_2(\theta) = \left(1 - (\bar{x})^{-1}\right)\ln(1-\theta) \tag{3.6}$$

against θ on the same set of axes. Since the functions $h_1(\theta)$ and $h_2(\theta)$ are monotonically decreasing at unequal rates, they must intersect at some point. Thus θ^{**} is given by their common point of intersection. The estimate m^{**} then follows from the equation (3.4).

3 Maximum Likelihood Estimators

The likelihood function for the given sample $X_1, X_2, \ldots, X_n$ becomes

$$\begin{aligned} L =& \theta^{n\bar{x}-n}(1-\theta)^{n(m\bar{x}-\bar{x}+1)}m^{n-f_1} \times \\ & \prod_{i=3}^{k}\left(\frac{(mi-1)(mi-2)\ldots(mi-i+2)}{(i-1)!}\right)^{f_i} \end{aligned}$$

and its log-likelihood function is

$$\begin{aligned} \ln(L) = & (n\bar{x}-n)\ln(\theta) + n(m\bar{x}-\bar{x}+1)\ln(1-\theta) + (n-f_1)\ln(m) \\ & + \sum_{i=3}^{k}\sum_{r=1}^{i-2} f_i \ln(mi-r) - \sum_{i=3}^{k} f_i \ln(i-1)!. \end{aligned}$$

The maximum likelihood (ML) estimate of m, given by $\widehat{m}$, is the unique root of m (in its domain) given by the equation,

$$n - f_1 + nm\bar{x}\ln\left(1 - \frac{1}{m} + \frac{1}{m\bar{x}}\right) + m\sum_{i=3}^{k}\sum_{r=1}^{i-2}\frac{if_i}{(mi-r)} = 0$$

or

$$1 - \frac{1}{m} + \frac{1}{m\bar{x}} = \exp\left[\frac{1}{nm\bar{x}}\left(- (n - f_1) - m \sum_{i=3}^{k} \sum_{r=1}^{i-2} \frac{if_i}{(mi - r)}\right)\right] \tag{3.7}$$

and then the ML estimate $\widehat{\theta}$, of θ, becomes

$$\widehat{\theta} = \widehat{m}^{-1} - (\widehat{m}\bar{x})^{-1}. \tag{3.8}$$

The equation (3.7) cannot be solved for m explicitly and the Newton-Raphson method cannot also be used in numerical methods. However, one may consider the graphs of the $H_1(m)$ and $H_2(m)$ given below:

$$H_1(m) = 1 - \left(\frac{1}{m}\right) + \left(\frac{1}{m\bar{x}}\right) = 1 - \frac{1 - \frac{1}{\bar{x}}}{m} \tag{3.9}$$

and

$$H_2(m) = \exp\left[\frac{-1}{n\bar{x}}\left(\frac{-1}{m}(n - f_1) + \sum_{i=3}^{k} \sum_{r=1}^{i-2} \frac{if_i}{(mi - r)}\right)\right]. \tag{3.10}$$

Since the functions $H_1(m)$ and $H_2(m)$ are both monotonically increasing at different rates, they must intersect at some common value $\widehat{m}$, which is the ML estimate.

4 Applications

Panjer (1987) considered three sets of automobile accidents frequency data for the GPP model and *Consul* (1990) used six sets for his GP model. All these included the zero-class frequency. Here, we are excluding the zero-class frequency i.e. the number of policies for which no claims were made or those policies for which there were no accidents because the Consul model deals only with autos in accident.
The parameters m and θ of the Consul distribution were estimated by all three methods discussed in section 3 and the distribution was fitted to each of the above data sets. Though the estimated values of m and θ, obtained by the three methods, were different yet they were close to each other. The ML method seemed to give a slight edge over the other two methods. The observed frequencies and the expected frequencies (by ML method) are given in the following tables (*Table 4.1 to 4.3*) for all the data sets used by *Panjer* (1987) and *Consul* (1990). For a comparison the corresponding χ^2-values as well as the degrees of freedom ($d.f.$), obtained by them, are also shown in parentheses.

Table 4.1: Observed and fitted Consul distribution for the number x of automobile accidents.

x	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency
	California (1964)		*Bühlmann*		*Hossak*	
1	21,350	$21,352.16$	14,075	$14,074.57$	68,714	$68,719.64$
2	3,425	$3,415.19$	1,766	$1,765.88$	5,177	$5,160.87$
3	530	543.35	255	258.80	365	376.53
4	89	86.14	45	41.30	24	26.92
5+	19	16.16	8	8.45	6	2.05
Total	25,413	$25,413.00$	16,149	$16,149.00$	74,286	$74,286.00$
Parameter		$\widehat{\theta}=0.16134$		$\widehat{\theta}=0.08723$		$\widehat{\theta}=0.07903$
estimates		$\widehat{m}=0.98951$		$\widehat{m}=1.50640$		$\widehat{m}=0.94602$
χ^2-values	(1.0)	0.95	$(0,5)$	0.41	(0.3)	0.43
$d.f.$	(2)	2	(2)	2	(1)	1
p-values	(0.6070)	0.6242	(0.7790)	0.8166	(0.6535)	0.5407

Source: *Panjer* (1987)

Table 4.2: Observed and fitted Consul distribution for the number x of automobile accidents.

x	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency
	Belgium (1975 – 76)		Zaire (1974)		Belgium (1958)	
1	9240	9242.31	232	232.07	1317	1319.96
2	704	695.37	38	37.34	239	226.09
3	43	53.73	7	8.44	42	53.66
4	9	4.59	3	2.22	14	14.77
5			1	0.93	4	4.43
6					4	1.40
7					1	0.69
Total	9996	9996.00	281	281.00	1621	1621.00
Parameter		$\widehat{\theta} = 0.07147$		$\widehat{\theta} = 0.03578$		$\widehat{\theta} = 0.04592$
estimates		$\widehat{m} = 1.05714$		$\widehat{m} = 5.25000$		$\widehat{m} = 4.37000$
χ^2-values	(7.88)	6.48	(0.66)	0.54	(10.45)	7.41
d.f.	(2)	1	(2)	2	(3)	3
p-values	(.0194)	0.0111	(0.7240)	0.7652	(0.0163)	0.0613

Source: *Consul* (1990)

Table 4.3: Observed and fitted Consul distribution for the number x of automobile accidents.

x	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency
	Switzerland (1961)		Germany (1960)		Great Britain (1958)	
1	14,075	$14,075.65$	2,651	$2,650.82$	46,545	$46,547.17$
2	1,766	$1,762.86$	297	297.48	3,935	3927.85
3	255	259.92	41	41.12	317	324.30
4	45	41.88	7	6.33	28	26.38
5	6	7.14	1	1.25	3	2.30
6	2	1.55				
Total	16,149	$16,149.00$	2,997	$2,997.00$	50,828	$50,828.00$
Parameter		$\widehat{\theta} = 0.08488$		$\widehat{\theta} = 0.06374$		$\widehat{\theta} = 0.08769$
estimates		$\widehat{m} = 1.54920$		$\widehat{m} = 1.86360$		$\widehat{m} = 0.95864$
χ^2-values	(7.33)	0.64	(2.48)	0.12	(5.54)	0.49
d.f.	(3)	3	(3)	2	(3)	2
p-values	$(.0634)$	0.8860	(0.4798)	0.9498	(0.1396)	0.7832

Source: *Consul* (1990)

By a comparison of the expected frequencies in *Table 4.1* with the observed ones and also by their χ^2-values and p-values, it is quite evident that the model *(2.4)* provides an excellent fit to each of the data sets. One can also say that, on the basis of the data and on comparing the χ^2-values, that the Consul model is as good as the GPP model. However, one of the disadvantages in using the GPP is that the computation of the probabilities is tedious because the successive probabilities depend on all the preceding ones. Their dependence also leads to the accumulation of errors which may seriously affect the later probabilities. Furthermore, the estimation of its three parameters is a relatively more complex job.

The only data set in *Table 4.2 (Belgium 1975 – 76)* is not well explained by the model although the χ^2-value would not lead to rejection of the model at one percent level of significance. It may also be noted that the expected frequencies for $x = 1, 2,$ and 3 are very close to the observed ones, while the fit is poor when $x = 4$ which gives a substantial contribution of 4.24 to the χ^2-value. This may be due to some unknown factors in the process or for the small number of classes in the χ^2-test. Apart from this, in all other five data sets given in the *Table 4.2* and *Table 4.3,* one can see a very close agreement of the expected frequencies with the corresponding observed ones, as indicated by the corresponding χ^2-values and the p-values. Once again, by

comparing the χ^2-values between the Consul model and GPP model, one can say that the Consul model gives a much better fit than the GPP model.
Concluding, the Consul distribution may be used as a model for the distribution of automobile accidents or claims and it may be useful to the insurance companies.

M. N. Islam and P. C. Consul
Dept. of Mathematics and Statistics
University of Calgary
Calgary, Alberta, Canada T2N 1N4

References

Consul, P.C. (1990): A model for distributions of injuries in auto-accidents. Journal of the Swiss Association of Actuaries, 161 – 168.

Consul, P. C. and *Shenton, L. R.* (1975): On the probabilistic structure and properties of discrete Lagrangian distributions. A modern course on statistical distributions in scientific work. Vol. 1, 41 – 57. Boston, Massachusetts, D. Reidel publishing company.

Gossiaux, A. and *Lemaire, J.* (1981): Methodes d'ajustement de distributions de sinistres. MVSV *81*, 87 – 95.

Islam, M. N. and *Consul, P. C.* (1991): On some properties of the Consul distribution. (Submitted).

Panjer, H. H. (1987): Models of Claim frequency. Advances in the statistical sciences, Actuarial Science. Vol. V1, 115 – 125. Edited by I. B. MacNeill and G. J. Umphrey.