A model for distributions of injuries in auto-accidents

P.C. Consul, Calgary

A Model for Distributions of Injuries in Auto-Accidents

1 Introduction

Auto insurance for collision and liability is one of the most important businesses in the modern world and it concerns every person because almost everybody possesses an automobile and pays the insurance premium which is based upon the injuries inflicted in auto-accidents and the amounts paid for claims. In view of this many attempts have been made to find a mathematical model for the distributions of injuries in auto-accidents. *Thyrion* (1960), *Derron* (1962), *Bichsel* (1964), *Johnson/Hey* (1971), *Lemaire* (1975, '77) and *Justens* (1979) have done important work in that direction by studying the actual data on injuries inflicted by auto-accidents in different countries. In an effort to find if there exists one probability model applicable to such distributions *Gossiaux/Lemaire* (1981) used six observed claims distributions from third party liability portfolios, obtained from five countries and studied by other researchers, and fitted the Poisson distribution, inflated geometric distribution (generalized geometric distribution), negative binomial distribution and a mixed Poisson distribution to each one of them by different methods. They concluded that no single probability law provided a good fit to all of them.
In this paper we apply the generalized Poisson distribution (GPD), introduced by *Consul/Jain* (1973) and fully described in a recent book by *Consul* (1989), to the same six observed data sets and it seems to be a plausible model.

2 Generalized Poisson Distribution (GPD)

A discrete random variable Y is said to have a generalized Poisson distribution (GPD) if its probability mass is given by

$$P(Y=k)=p_k = \begin{cases} \theta(\theta+k\lambda)^{k-1}\frac{e^{-\theta-k\lambda}}{k!}, & \text{for } k=0,1,2,\dots \\ 0, & \text{for } k>m \quad \text{when} \quad \lambda<0, \end{cases} \tag{2.1}$$

and zero otherwise, where $\theta>0$, $\max(-1,-\theta/4)\leq\lambda<1$ and m is the largest positive integer for which $\theta+\lambda m>0$ when λ is negative. The GPD reduce to the Poisson distribution when $\lambda = 0$ and it possesses the twin properties of over-dispersion and under-dispersion according as $\lambda > 0$ or $\lambda < 0$. This property makes the GPD a very useful model for numerous applications. The GPD gets truncated for negative values of λ but the truncation error is always less than 0.07 %. Its mean and variance are given by

$$\text{mean} \quad \mu = \theta(1-\lambda)^{-1}$$
$$\text{variance} \quad \sigma^2 = \theta(1-\lambda)^{-3}.$$

Though *Consul* (1989) has given many methods for the estimation of θ and λ, we shall use the maximum likelihood (ML) method. If the observed distribution is given by $(k, n_k;\ k = 0, 1, 2, \ldots, r)$ where n_k denotes the number of vehicles each of which had k injured persons in an accident, then the ML estimate $\hat{\lambda}$ is given by the unique root of λ (in its domain) given by the equation

$$\sum_{k=0}^{r} n_k \frac{k(k-1)}{\bar{x} + (k-\bar{x})\lambda} - n\bar{x} = 0, \tag{2.2}$$

where $n = n_0 + n_1 + \cdots + n_r$ and $\bar{x}$ is the mean of the observed distribution. The ML estimate $\hat{\theta}$ is then given by

$$\hat{\theta} = \bar{x}(1-\hat{\lambda}). \tag{2.3}$$

A computer program for estimation of $\hat{\theta}$ and $\hat{\lambda}$ and for fitting the model is given by *Consul* (1989).

3 Application

We apply the GPD model (2.1) to all the six examples considered by *Gossiaux/Lemaire* (1981). They used the Poisson distribution, negative binomial distribution, inflated geometric distribution and mixed Poisson distribution as the null hypothesis against its negation in all six examples. The Poisson distribution does not fit any example and there is at least one example where each model gets rejected. For comparison sake we provide the expected frequencies for those two models which were found to be better for each example together with the expected GPD frequencies by maximum likelihood method. Somehow an error was made in computing the degrees of freedom. We give the correct degrees of freedom and the corresponding probabilities $\mathbb{P}(\chi^2 \geq \chi^2_{obs})$.

Considering the values of χ^2_{obs} and $\mathbb{P}(\chi^2 \geq \chi^2_{obs})$ for the observed data in Belgium (1975–76) [Example 1] the mixed Poisson distribution does not fit at all but the inflated geometric distribution, the negative binomial distribution and the GPD fit very well, the last two giving better fits than the first one. Also, in example 2 [for Zaire (1974) data] all these three models fit very well; however the model GPD seems to be the best as it gives the largest value of $\mathbb{P}(\chi^2 \geq \chi^2_{obs}) = 0.9992$.

For the observed data in Belgium (1958) [Example 3] the inflated geometric distribution does not fit while the null hypotheses of negative binomial distribution and the GPD cannot be rejected. Again, the GPD gives slightly a better fit. Similarly, both the negative binomial distribution and the inflated geometric distribution get rejected for the 1961 observed data on Switzerland [Example 4] but the GPD and the mixed Poisson distribution fit reasonably well; however the mixed Poisson distribution seems to be somewhat better.

For the observed data set on Germany (1960), given in Example 5, the inflated geometric distribution, the mixed Poisson distribution and the GPD are three models which fit very well as the P-values are 0.54, 0.64 and 0.44 respectively. Among these the mixed Poisson distribution seems to be the best. For the observed data on Great Britain (1968), given in Example 6, the same three models provide a good fit, though the inflated geometric distribution seems to apply best as it has a high P-value of 0.886.

It is clear from the six examples that the GPD is the only model which does not get rejected as a null hypothesis in any one of the above six data sets. Possibly, one has to apply this model on some more observed data sets and test its applicability to them. Also, one has to examine the prevalent conditions in each country and see if the GPD can be developed as a stochastic model for the Auto-injuries under those conditions.

We have also given the maximum likelihood estimates of the two GPD parameters θ and λ in each example. We shall now try to give some meaning to these estimated values of the parameters for the six examples. It has been shown by *Consul* (1989) that the GPD represents the cumulative effect of two stochastic processes or a queueing process where θ represents the average effect of one process and λ represents the average effect of the other process.

Example 1 [9]

Belgium (1975–76): Total 106974; $\overline{x} = 0.10108, s^2 = 0.10745$

k	Observations	Neg. Binomial distribution	Inf. Geometric distribution	GPD
0	96978	96980.8	96978.0	96980.5
1	9240	9230.9	9239.0	9232.8
2	704	708.6	699.7	706.0
3	43	50.1	53.0	50.8
4	9	3.4	4.0	3.6
5	0	0.2	0.3	0.3
Number of Classes after regrouping		4	4	4
χ^2_{obs}		0.09	0.53	0.145
degrees of freedom		2	2	2
$\mathbb{P}(\chi^2 \geq \chi^2_{obs})$		0.956	0.766	0.93

ML estimates for GPD parameters: $\widehat{\theta} = 0.098075, \widehat{\lambda} = 0.029731$

Example 2 [8]

Zaire (1974): Total 4000; $\overline{x} = 0.08650$, $s^2 = 0.12255$

k	Observations	Neg. Binomial distribution	Inf. Geometric distribution	GPD
0	3719	3719.2	3719.0	3719.1
1	232	229.9	228.2	231.2
2	38	39.9	42.9	38.4
3	7	8.4	8.1	8.4
4	3	1.9	1.5	2.1
5	1	0.5	0.3	0.6
6	0	0.1	0.1	0.2
Number of Classes after regrouping		4	4	4
χ^2_{obs}		0.36	0.87	0.015
degrees of freedom		2	2	2
$\mathbb{P}(\chi^2 \geq \chi^2_{obs})$		0.87	0.64	0.992

ML estimates for GPD parameters: $\widehat{\theta} = 0.072808, \widehat{\lambda} = 0.158290$

Example 3 [10]

Belgium (1958): Total 9461; $\bar{x} = 0.21435$, $s^2 = 0.28893$

k	Observations	Neg. Binomial distribution	Inf. Geometric distribution	GPD
0	7840	7847.0	7840.0	7847.0
1	1317	1288.4	1295.7	1292.0
2	239	256.5	260.0	251.3
3	42	54.1	52.2	54.1
4	14	11.7	10.5	12.4
5	4	2.6	2.1	3.0
6	4	0.6	0.4	0.7
7	1	0.1	0.1	0.2
8	0	0	0	0.1
Number of Classes after regrouping		5	5	5
χ^2_{obs}		7.61	12.38	6.45
degrees of freedom		3	3	3
$\mathbb{P}(\chi^2 \geq \chi^2_{obs})$		0.055	0.007	0.094

ML estimates for GPD parameters: $\widehat{\theta} = 0.187049$, $\widehat{\lambda} = 0.127380$

Example 4 [1]

Switzerland (1961): Total 119853; $\bar{x} = 0.15513$, $s^2 = 0.17931$

k	Observations	Inf. Geometric distribution	Mixed Poisson distribution	GPD
0	103704	103704.0	103692.7	103723.1
1	14075	14025.4	14116.0	14002.9
2	1766	1844.3	1714.4	1838.1
3	255	242.5	278.3	248.5
4	45	31.9	44.8	34.6
5	6	4.2	6.1	5.0
6	2	0.6	0.7	0.7
7	0	0.1	0.1	0.1
Number of Classes after regrouping		6	6	6
χ^2_{obs}		11.48	3.80	8.333
degrees of freedom		4	3	4
$\mathbb{P}(\chi^2 \geq \chi^2_{obs})$		0.022	0.28	0.125

ML estimates for GPD parameters: $\widehat{\theta} = 0.144541$, $\widehat{\lambda} = 0.068274$

Example 5 [4]

Germany (1960): Total 23589; $\bar{x} = 0.14422$, $s^2 = 0.16387$

k	Observations	Inf. Geometric distribution	Mixed Poisson distribution	GPD
0	20592	20592.0	20558.7	20596.8
1	2651	2640.2	2662.2	2632.4
2	297	314.3	285.0	315.8
3	41	37.4	44.5	38.4
4	7	4.5	7.5	4.8
5	0	0.5	1.1	0.6
6	1	0.1	0.1	0.1
0	0	0	0	0.1
Number of Classes after regrouping		5	5	5
χ^2_{obs}		2.21	0.90	2.455
degrees of freedom		3	2	3
$\mathbb{P}(\chi^2 \geq \chi^2_{\text{obs}})$		0.54	0.62	0.443

ML estimates for GPD parameters: $\widehat{\theta} = 0.135641$, $\widehat{\lambda} = 0.059482$

Example 6 [6]

Great Britain (1968): Total 421240; $\bar{x} = 0.13174$, $s^2 = 0.13852$

k	Observations	Inf. Geometric distribution	Mixed Poisson distribution	GPD
0	370412	370412.2	370408.9	370390.5
1	46545	46563.5	46557.5	46509.7
2	3935	3906.8	3916.8	4016.6
3	317	327.8	327.5	299.1
4	28	27.4	27.1	22.6
5	3	2.1	2.0	1.4
6	0	0	0	0.1
Number of Classes after regrouping		5	5	5
χ^2_{obs}		0.64	1.05	3.39
degrees of freedom		3	3	3
$\mathbb{P}(\chi^2 \geq \chi^2_{\text{obs}})$		0.886	0.60	0.34

ML estimates for GPD parameters: $\widehat{\theta} = 0.128645$, $\widehat{\lambda} = 0.024198$

In the case of auto-injuries the value of θ may represent the cumulative effect of the conditions of the roads in the country, the amount of traffic on the roads, the sense of road discipline in the users and the safety features provided in the automobiles. Accordingly, the values $\widehat{\theta} = 0.187049$ for Belgium (1958) and $\widehat{\theta} = 0.098075$ for Belgium (1974–75) can be very well justified by the above criteria. The reduction on the value of θ possibly implies that the mean average rate of injuries declined on account of the improvement in the conditions of the roads, better roads signs, better sense of road discipline and some effect of the use of seat-belts etc. Similarly, the slightly higher value of $\widehat{\theta} = 0.144541$ for Switzerland (1961) in comparison to $\widehat{\theta} = 0.135641$ for Germany (1960) is possibly due to the fact that the driving on the mountain roads of Switzerland is somewhat more hazardous. Also, the small value of $\widehat{\theta} = 0.072808$ for Zaire (1974) may be due to a much smaller intensity of traffic.

It may be noted that the value λ does not affect the zero-injury frequency, reduces the 1-injury frequency and increases all the frequencies of higher injuries. Accordingly, λ may represent the average effect of the number of passengers in the autos on the number of injuries. The large values $\widehat{\lambda} = 0.158290$ for Zaire (1974) and $\widehat{\lambda} = 0.127380$ for Belgium (1958) do possibly signify more passengers in the cars in those countries at that time. The value of $\widehat{\lambda} = 0.029731$ for Belgium (1975–76) signifies the new trend of single drivers in most automobiles. Other values of $\widehat{\lambda}$ for the other countries possibly demonstrate similar trends.

4 Conclusion

The GPD appears to be a very plausible model for the distributions of auto-injuries but it needs a more detailed study on the observed data for a number of normal years in many different countries. Possibly, some insurance companies will make such a collaborative study in future.

P.C. Consul
Department of Mathematics and Statistics
University of Calgary
Calgary, Alberta
Canada

Bibliography

[1] *Bichsel, F.* (1964): Erfahrungs-Tarifierung in der Motorfahrzeughaftpflicht-Versicherung. MVSV *64*, 119–143.

[2] *Consul, P.C.* (1989): Generalized Poisson Distributions: Properties and Applications. Marcel Dekker Inc., New York/Basel.

[3] *Consul, P.C./Jain, G.C.* (1973): A generalization of Poisson distribution. Technometrics *15*, 791–799.

[4] *Derron, M.* (1962): Mathematische Probleme der Automobilversicherung. MVSV *62*, 103–123.

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

[6] *Johnson, P./Hey, G.* (1971): A review of the scope for using claims histories of individual policies in risk assessment. Bulletin A.R.A.B. *66*, 159–195.

[7] *Justens, D.* (1979): Le bonus-malus en assurance automobile: une nouvella approache analytique des probabilites de transition. Memoire de license. U.L.B.

[8] *Lemaire, J.* (1975): Sur les critéres de détermination d'un tarif bonus-malus optimal. Application á l'assurance automobile á kinshasa. Ann. Fac. Sci. UNAZA, Vol. *1*, 338–366.

[9] *Lemaire, J.* (1977): Selection procedures of regression analysis applied to automobile insurance. MVSV *77*, 143–160.

[10] *Thyrion, P.* (1960): Contribution á l'étude due bonus pour non-sinistre en assurance auto. ASTIN *I*, 143–162.

Summary

The generalized Poisson distribution (GPD), introduced by *Consul/Jain* (1973) and fully described in the book by *Consul* (1989) is defined. The GPD is applied to six observed data sets on injuries in auto-accidents as a plausible model and some possible interpretation has been given to the values of the two parameters in different cases.

Zusammenfassung

Die verallgemeinerte Poissonverteilung (GPD) wurde von *Consul/Jain* (1973) eingeführt und ist im Buch von *Consul* (1989) beschrieben. Die GPD wird auf sechserlei Datenmaterial im Zusammenhang mit Körperverletzungen bei Autounfällen angewendet. Sie erweist sich dabei als plausibles Modell, und für verschiedene Fälle werden mögliche Interpretationen der beiden Parameter angegeben.

Résumé

La distribution de Poisson généralisée (GPD) est présentée dans cette article. Cette distribution, introduite par *Consul/Jain* (1973), est étudiée dans le livre de *Consul* (1989). Cette distribution est appliquée à six échantillons de données concernant des blessures corporelles lors d'accidents de la circulation. La GPD apparait dans cette analyse comme un modèle plausible et des interprétations des paramètres peuvent être décrites dans différents cas.