A LOGLINEAR LAGRANGIAN POISSON MODEL

By PETER TER BERG

SFB Verzekeringen, Amsterdam, Netherlands

ABSTRACT

Maximum likelihood estimation is derived for the Lagrangian Poisson distribution for a simple and a loglinear model and illustrated with real data

KEYWORDS

Loglinear; Lagrangian Poisson; Maximum likelihood, Newton-Raphson.

1 INTRODUCTION

The monograph by CONSUL (1989) is on a nice counting distribution known as Lagrangian Poisson[1]

$$(1.1)\qquad p(n \mid \theta, \zeta) = \theta(\theta + n\zeta)^{n-1} \exp\{-(\theta + n\zeta)\}/n!$$

$$n = 0, 1, 2, 3, 4, \dots \quad \theta > 0 \quad 0 \leq \zeta \leq 1$$

The mean and variance of this probability distribution are finite when $\zeta < 1$ and are given by

$$E[N] = \theta(1-\zeta)^{-1} \qquad \operatorname{var}(N) = \theta(1-\zeta)^{-3}$$

When $\zeta = 0$ this distribution reduces to the well-known Poisson distribution

CONSUL and SHENTON (1972) derive (1.1) using Lagrange's expansion A probabilistic derivation of the Lagrangian Poisson distribution can be found in GOOVAERTS and KAAS (1991). Once we know this distribution it is possible to recognize it in the literature Browsing in COX and MILLER (1965), I encountered it as a queueing exercise on page 250. They refer to MCMILLAN and RIORDAN (1957)

Clearly, the Lagrangian Poisson is a useful distribution which belongs to the statistical toolkit

A generalization of the loglinear Poisson model using the Lagrangian Poisson may be useful This allows the variance to exceed the mean Maximum likelihood

[1] This adheres to the terminology in CONSUL and SHENTON (1972) The distribution is also known as Generalized Poisson The adjective *generalized*, being rather uninformative anyway, applies also to arbitrary mixed and compound Poisson distributions

analysis for a simple model is given in the next section, followed with the exposition for a loglinear model

Evaluation of Information matrices needs the following moments, which can be found on page 157 in CONSUL (1989)

$$(1\ 2)\qquad E\left[\frac{N-1}{(\theta+N\zeta)^2}\right] = \frac{\theta(1-\zeta)+2\zeta}{\theta(\theta+2\zeta)} - \theta^{-2}$$

$$E\left[\frac{N(N-1)}{(\theta+N\zeta)^2}\right] = \frac{\theta}{\theta+2\zeta}$$

$$E\left[\frac{N^2(N-1)}{(\theta+N\zeta)^2}\right] = \frac{\theta(\theta+2)}{(1-\zeta)(\theta+2\zeta)}$$

A numerical illustration and some final remarks complete the paper.

2 A SIMPLE MODEL

The convolution of two independent Lagrangian Poisson random variables with parameters (θ_1, ζ) and (θ_2, ζ) is Lagrangian Poisson with parameters $(\theta_1+\theta_2, \zeta)$. This justifies the replacement of θ in (1 1) by $m\theta$, where m is a known size factor. This situation may be appropriate for a statistical agency, which collects economy-wide data on the number of policies and number of claims Let there be R reporting units numbered $r = 1, \quad , R$ Minus the logarithm of the likelihood function can be written as:

$$(2\ 1)\qquad f = c + \theta\sum m_r + \zeta\sum n_r + \sum(n_r-1)\ln a_r - R\ln\theta$$

$$a_r = (m_r\theta + n_r\zeta)^{-1} \qquad c = \sum[\ln n_r! - \ln m_r]$$

where $\sum$ denotes summation over r The elements of the gradient of f result as

$$(2\ 2)\qquad \frac{\partial f}{\partial\theta} = \sum m_r - R\theta^{-1} - \sum m_r(n_r-1)a_r$$

$$(2\ 3)\qquad \frac{\partial f}{\partial\zeta} = \sum n_r - \sum n_r(n_r-1)a_r$$

Following page 102 in CONSUL (1989), we multiply (2 2) by θ and (2 3) by ζ Adding together and equating to 0 results in:

$$(2\ 4)\qquad \theta\sum m_r = (1-\zeta)\sum n_r$$

So, the implied maximum likelihood estimator for the population mean is given by the sample mean. The elements of the Hessian of f can be displayed in partitioned matrix form as:

$$(2.5)\qquad \mathbf{H} = \frac{\partial^2 f}{\partial\begin{bmatrix}\theta\\ \varsigma\end{bmatrix}\partial\begin{bmatrix}\theta\\ \varsigma\end{bmatrix}'} = R\theta^{-2}\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} + \sum (n_r - 1)a_r^2\begin{bmatrix}m_r\\ n_r\end{bmatrix}\begin{bmatrix}m_r\\ n_r\end{bmatrix}'$$

If we have a pathological sample for which all $n_r = 0$, the Hessian will be a zero-matrix In case also $n_r = 1$ are observed, the Hessian will be positive semi-definite. In all other cases the Hessian is positive definite, which implies f to be convex with a unique stationary point

Searching along the line (2.4) we derive that this stationary point will have $\zeta < 1$ and $\zeta > 0$ depending on the sample. To this end we substitute (2.4) in (2 1) and get a convex function in ζ.

$$(2\,6)\ \tilde{f} = (c + \sum n_r - R\ln\hat{\lambda}) - R\ln(1-\zeta) - \sum (n_r - 1)\ln[m_r\hat{\lambda}(1-\zeta) + n_r\zeta]$$

where $\hat{\lambda} = \sum n_r/\sum m_r$, the sample mean. Numerically, $\tilde{f}$ is defined on the open interval $(z, 1)$ where $z<0$ is given as:

$$z = 1 - \min_r [n_r u_r^{-1} \mid u_r > 0] \qquad u_r = n_r - m_r\hat{\lambda}$$

At the boundaries of this interval $\tilde{f}$ approaches $+\infty$ Probabilisticaly, the stationary point of $\tilde{f}$ should be in the interval [0, 1]. Differentiation of (2 6) results in:

$$(2\,7)\qquad \frac{d\tilde{f}}{d\zeta} = R(1-\zeta)^{-1} - \sum (n_r - 1)u_r(m_r\hat{\lambda} + \zeta u_r)^{-1}$$

Taking the limit of (2 7) for $\zeta \to 1$ results in $+\infty$ So, the search for a stationary point, starting in this limit point, will be in the direction of $\zeta < 1$. Next we evaluate (2.7) for $\zeta = 0$. When this value is negative, the stationary point will have $\zeta > 0$. This condition can be simplified to:

$$\left(\sum n_r\right)^2 < \left(\sum m_r\right)\left[\sum n_r(n_r - 1)/m_r\right]$$

When all m_r are equal, this condition amounts to the statement that the sample variance exceeds the sample mean. This agrees with the findings on page 102 in CONSUL (1989)

So, there is a possibility that the stationary point will have $\zeta < 0$. In such cases we should decide to use the Poisson model by setting $\zeta = 0$

The Information matrix results as the expectation of (2 5) Using the expectations in (1.2) and adjusting for m, we get

$$(2\,8)\ E[\mathbf{H}] = \sum m_r(m_r\theta + 2\zeta)^{-1}\begin{bmatrix} m_r(1-\zeta)+2\zeta\theta^{-1} & m_r\theta \\ m_r\theta & (1-\zeta)^{-1}\theta(m_r\theta+2) \end{bmatrix}$$

Both the inverse of (2.5) or (2 8) give an estimator for the covariance matrix of the maximum likelihood estimator.

Whenever all m_r are equal we may normalize $m_r = 1$ and the sample data are conveniently stored in a frequency table Such a situation arises when considering the number of claims per policy In Table 1 we find data from BICHSEL (1964) together with the maximum likelihood fit

TABLE 1

LIABILITY CLAIMS PER AUTOMOBILE POLICY, SWITZERLAND 1961

Number of claims	Number of policies	Lagrangian Poisson fit
0	103 704	103 722 2
1	14 075	14 003 7
2	1 766	1 838 2
3	255	248 5
4	45	34 6
5	6	4 9
6	2	0 7
≥7	0	0 1

The maximum likelihood estimates are $\hat{\theta} = 0.14455$ and $\hat{\zeta} = 0\,06826$ with standard deviations 0.0011 and 0 0028

3 A LOGLINEAR MODEL

We model θ in the following loglinear way, which reconciles with the loglinear Poisson model as presented in TER BERG (1980)

$$(3.1)\qquad \theta_r = \lambda_r(1-\zeta)$$
$$\lambda_r = \exp[\mathbf{x}_r'\boldsymbol{\beta}]$$

where $\mathbf{x}_r$ is a vector of explanatory variables and $\boldsymbol{\beta}$ a parameter vector with K elements.

Following the interpretation of the R reporting units with different size given by m_r this model is capable of incorporating differences between reporting units and differences between time periods by introducing appropriate dummies A more common interpretation is the modelling of claim frequencies as a step towards a multiplicative rating structure

The $R \times K$ matrix $\mathbf{X}$ is defined as.

$$\mathbf{X} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \dots \quad \mathbf{x}_R]'$$

and is assumed to have full column rank. The first column of $\mathbf{X}$ equals $\mathbf{1}$, a vector containing the elements 1.

Minus the logarithm of the likelihood function is a function of $\boldsymbol{\beta}$ and ζ and can be written as:

$$(3.2)\quad f = c + (1-\zeta)\sum m_r\lambda_r + \zeta\sum n_r + \sum(n_r - 1)\ln a_r - R\ln(1-\zeta) - \mathbf{1}'\mathbf{X}\boldsymbol{\beta}$$

$$a_r = (m_r\theta_r + n_r\zeta)^{-1}$$

Differentiation of f with respect to $\boldsymbol{\beta}$ and ζ results in the elements of the gradent $\mathbf{g}$

$$(3.3)\quad \frac{\partial f}{\partial \boldsymbol{\beta}} = (1-\zeta)\sum[1-(n_r-1)a_r]m_r\lambda_r\mathbf{x}_r - \mathbf{X}'\mathbf{1}$$

$$\frac{\partial f}{\partial \zeta} = R(1-\zeta)^{-1} + \sum[1-(n_r-1)a_r]u_r \qquad u_r = n_r - m_r\lambda_r$$

Differentiation of $\mathbf{g}$ with respect to $\boldsymbol{\beta}$ and ζ results in the elements of the Hessian $\mathbf{H}$:

$$(3.4)\quad \frac{\partial^2 f}{\partial\boldsymbol{\beta}\,\partial\boldsymbol{\beta}'} = \sum[1 - n_r(n_r-1)a_r^2\zeta]m_r\theta_r\mathbf{x}_r\mathbf{x}_r'$$

$$\frac{\partial^2 f}{\partial\boldsymbol{\beta}\,\partial\zeta} = -\sum[1 - n_r(n_r-1)a_r^2]m_r\lambda_r\mathbf{x}_r$$

$$\frac{\partial^2 f}{\partial\zeta^2} = R(1-\zeta)^{-2} + \sum(n_r-1)[a_r u_r]^2 > 0$$

The Information matrix is given as the expected value of the Hessian

$$E[\mathbf{H}] = (1-\zeta)\sum m_r\theta_r\begin{bmatrix}\mathbf{x}_r\\ 0\end{bmatrix}\begin{bmatrix}\mathbf{x}_r\\ 0\end{bmatrix}'$$

$$+\,2\sum m_r\theta_r(m_r\theta_r + 2\zeta)^{-1}\begin{bmatrix}\zeta\mathbf{x}_r\\ (\zeta-1)^{-1}\end{bmatrix}\begin{bmatrix}\zeta\mathbf{x}_r\\ (\zeta-1)^{-1}\end{bmatrix}'$$

The Newton-Raphson search direction is given by $\mathbf{H}^{-1}\mathbf{g}$. Whenever the Hessian $\mathbf{H}$ is not positive definite, we replace the Hessian by the Information matrix Occasionally, an iteration may imply $\zeta \geq 1$, an $a_r \leq 0$ or an increase of f. In such cases, we halve the stepsize

Taking the loglinear Poisson model as a starting value, convergence is quick and swift

4 NUMERICAL ILLUSTRATION

In Table 2 we find data from BAILEY and SIMON (1960) on $R = 20$ risk groups defined by 2 risk indicators with 5 and 4 levels.

TABLE 2

EXPOSURE AND NUMBER OF CLAIMS, AUTOMOBILE LIABILITY, CANADA 1957-1958

X-matrix	Earned car years	Number of claims		
		Observed	Poisson	Lagrangian Poisson
1 0000 000	2757 520	217 151	219 950 1	219 868 7
1 1000 000	130 535	14 506	14 052 3	14 083 6
1 0100 000	247 424	31 964	31 546 8	31 590 9
1 0010 000	156 871	22 884	21 170 2	21 085 6
1 0001 000	64 130	6 560	6 345 7	6 394 7
1 0000 100	130 706	13 792	13 688 2	13 761 4
1 1000 100	7 233	1 001	1 022 3	1 030 5
1 0100 100	15 868	2 695	2 656 3	2 675 3
1 0010 100	17 707	3 054	3 137 4	3 142 8
1 0001 100	4 039	487	524 7	531 8
1 0000 010	163 544	19 346	18 607 9	18 631 6
1 1000 010	9 726	1 430	1 493 5	1 499 3
1 0100 010	20 369	3 546	3 704 6	3 715 9
1 0010 010	21 089	3 618	4 059 7	4 050 1
1 0001 010	4 869	613	687 3	693 7
1 0000 001	273 944	37 730	35 772 8	35 715 3
1 1000 001	21 504	3 421	3 789 9	3 793 6
1 0100 001	37 666	7 565	7 862 3	7 863 5
1 0010 001	56 730	11 345	12 533 7	12 468 2
1 0001 001	8 601	1 291	1 393 3	1 402 4

Applying the maximum likelihood model of the previous section, we have $K = 1 + (5-1) + (4-1) = 8$ elements for $\boldsymbol{\beta}$. The maximum likelihood results are presented in Table 3

TABLE 3

MAXIMUM LIKELIHOOD ESTIMATES AND STANDARD DEVIATIONS

Parameter	Loglinear Poisson	Loglinear Lagrangian Poisson		
	Maximum likelihood estimate	Maximum likelihood estimate	Standard deviation based on	
			Hessian	Information matrix
β_1	- 2 5287	- 2 5291	0110	0111
β_2	2998	3024	0392	0392
β_3	4691	4708	0273	0272
β_4	5259	5222	0294	0291
β_5	2156	2236	0575	0575
β_6	2723	2780	0385	0385
β_7	3552	3568	0337	0336
β_8	4930	4917	0247	0244
ζ	0	8154	0294	0294

We see close agreement between the parameter estimates for β in the loglinear Poisson as well as the loglinear Lagrangian Poisson model This agreement is also evident in the fitted claim numbers in Table 2. The relative large value for ζ increases the variances in the Lagrangian Poisson model

The similarity between the standard deviations based on the inverse of the Hessian and Information matrix is reassuring

If, in this illustration, we shrink the **X**-matrix to the first column, the maximum likelihood estimate for β_1 changes to -2.3295 and that for ζ increases to 0.9738

5 FINAL REMARKS

The use of the Lagrangian Poisson distribution with loglinear mean contains, through the parameter ζ, a diagnostic tool to infer the presence of omitted explanatory variables in the loglinear specification. As such, the loglinear Lagrangian Poisson model is a possible starting point from which to model the loglinear Poisson model.

When the maximum likelihood estimate for ζ is clearly different from 0, the variance of the Poisson distribution is too small, whereas the Lagrangian Poisson distribution implies the appropriate larger variances

As shown by GOOVAERTS and KAAS (1991) the Lagrangian Poisson distribution also allows a recursive evaluation of a compound Lagrangian Poisson distribution. So, from an applied point of view, there is little reason to object to the use of the Lagrangian Poisson distribution.

REFERENCES

BAILEY, R A and SIMON, L J (1960) Two studies in automobile insurance ratemaking *ASTIN Bulletin* **1**, 192-217

BICHSEL, F (1964) Erfahrungs-Tarifierung in der Motorfahrzeughaftpflicht-Versicherung *Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker* **64**, 119-129

CONSUL, P C (1989) *Generalized Poisson Distributions* Marcel Dekker, New York

CONSUL, P C and SHENTON, L R (1972) Use of Lagrange expansion for generating discrete generalized probability distributions *SIAM Journal of Applied Mathematics* **23**, 239-248

COX, D R and MILLER, H D (1965) *The Theory of Stochastic Processes* Chapman & Hal, London

GOOVAERTS, M and KAAS, R (1991) Evaluating compound generalized Poisson distributions recursively *ASTIN Bulletin* **21**, 193-198

MCMILLAN, B and RIORDAN, J (1957) A moving single server problem *Annals of Mathematical Statistics* **28**, 471-478

TER BERG, P (1980) On the loglinear Poisson and Gamma model *ASTIN Bulletin* **11**, 35-40