

ON THE LOGLINEAR POISSON AND GAMMA MODEL

PETER TER BERG

Maximum likelihood estimation in case of a Poisson or Gamma distribution with loglinear parametrization for the mean is quite akin. The asymptotic variance-covariance matrix for the maximum likelihood estimator is derived as well as a linear estimator, which can serve as a starting value for the nonlinear search procedure.

1. INTRODUCTION

The loglinear Poisson model is very useful for the analysis of frequency data. The mathematical statistics for this model can be found in HABERMAN (1974) as well as ANDERSEN (1977), who refers to actuarial applications by BAILY and SIMON (1960), JUNG (1968) and AJNE (1975). The procedure to calculate the maximum likelihood estimates of this model remains valid when the discrete data are replaced by continuous data, as has been done for the analysis of claim costs. However, this procedure does not correspond with maximum likelihood estimation of the parameters of a well-defined stochastic process, whereas claim frequency analysis under the Poisson assumption does.

The more interesting is the fact that maximum likelihood estimation of the parameters of a Gamma distribution with loglinear mean can be performed with the very same nonlinear search procedure as for the loglinear Poisson model. The Gamma distribution may be a good description for the fluctuation of mean claim costs. For such a claim costs analysis, the only modification we have to make is to replace the exposure by the total claim costs, maintaining the position of the number of claims.

In the next paragraph, the loglinear Poisson model will be outlined and an easy to calculate estimator to start the search procedure will be derived. Then, the loglinear Gamma model will be specified and convenient properties as regards estimation and aggregation will be demonstrated.

2. THE LOGLINEAR POISSON MODEL

For claim frequency analysis, this model can be formulated as follows:

$$(2.1) \quad f(n_r | m_r \lambda_r) = (m_r \lambda_r)^{n_r} \exp(-m_r \lambda_r) / n_r! \quad n_r = 0, 1, 2, 3, \dots$$

$$(2.2) \quad \log(\lambda_r) = \mathbf{x}'_r \boldsymbol{\beta} \quad r = 1, \dots, R$$

where  $r$  indexes the riskgroup,  $n_r$  stands for the number of claims and  $m_r$  equals exposure. The riskgroup  $r$  is characterized by the column vector  $\mathbf{x}_r$  consisting of the dummies 0 and 1. The parameters to estimate are the  $k$  elements of the column vector  $\boldsymbol{\beta}$ . What matters for this model is the linearity

in  $\beta$  for  $\log(\lambda_r)$ . It is assumed that the dummy variable trap has been taken care of such that the  $R \times K$  matrix  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_R)'$  has full column rank implying  $\mathbf{X}'\mathbf{X}$  to be invertible. This avoids a basic indeterminacy on the elements of  $\beta$ .

For instance, with two factors, having 2 and 3 levels,  $\mathbf{X}$  amounts to a  $6 \times 4$  matrix given by:

$$(2.3) \quad \mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

The natural logarithm of the likelihood function associated with (2.1) and (2.2) amounts to <sup>1</sup>:

$$(2.4) \quad \log L = \text{const} + (\sum n_r \mathbf{x}_r)' \beta - \sum m_r \exp(\mathbf{x}_r' \beta).$$

Differentiating with respect to the elements of  $\beta$  gives:

$$(2.5) \quad \frac{\partial \log L}{\partial \beta} = \sum n_r \mathbf{x}_r - \sum m_r \exp(\mathbf{x}_r' \beta) \mathbf{x}_r.$$

Putting (2.5) equal to  $\mathbf{0}$  defines the maximum likelihood estimator  $\hat{\beta}$  whose elements will be finite as soon as  $\sum n_r \mathbf{x}_r$  does not have elements equal to 0. From (2.5) we can see that the maximum likelihood estimate implies predicted marginal totals equal to the observed marginal totals.

The loglikelihood function is concave in  $\beta$  implying a unique solution for the maximum likelihood estimate.

The concavity can be demonstrated with the help of the matrix of second order derivatives:

$$(2.6) \quad \frac{\partial^2 \log L}{\partial \beta \partial \beta'} = - \sum m_r \exp(\mathbf{x}_r' \beta) \mathbf{x}_r \mathbf{x}_r' = -\mathbf{X}' \mathbf{M} \mathbf{X}$$

where  $\mathbf{M}$  and  $\mathbf{X}$  are  $R \times R$  diagonal matrices given by  $\text{diag}(m_1, \dots, m_R)$  and  $\text{diag}(\lambda_1, \dots, \lambda_R)$ .

This matrix is negative semi-definite and will be negative definite if the diagonal matrix  $\mathbf{M}$  contains  $K$  or more positive elements, that is: the presence of enough different exposure. If this is the case, the loglikelihood function will be concave in  $\beta$ .

Minus the expectation of this matrix, which does not contain any  $n_r$ , gives the information matrix of  $\beta$ . The inverse of the information matrix gives the Rao-Cramér lower bound for the asymptotic variance-covariance matrix of  $\hat{\beta}$ .

<sup>1</sup> All summations run from  $r = 1$  to  $r = R$ , unless explicitly otherwise stated.

To be complete: the maximum likelihood estimator  $\hat{\beta}$  follows asymptotically a multivariate normal distribution with mean  $\beta$  and variance-covariance matrix  $(\mathbf{X}'\mathbf{M}\mathbf{A}\mathbf{X})^{-1}$ . This matrix can be estimated by substituting  $\hat{\beta}$  for  $\beta$  in the diagonal elements of  $\mathbf{A}$ .

It is possible to reduce the dimensionality of the problem by maximizing (2.4) with respect to the parameters of a single factor conditionally on the parameters of the other factors. It is of course advisable to choose that factor with the maximum number of levels. In case of the matrix  $\mathbf{X}$ , given by (2.3) this results in a concentrated loglikelihood function with one unknown parameter.

In case that  $\kappa$  is large, a good starting value for the nonlinear search procedure may be important. In (2.2), a system of  $\mathbf{r}$  linear equations in  $\kappa$  unknowns is defined. However,  $\log(\lambda_r)$  is unknown and should be replaced by an estimate. Dropping the index  $r$  and adopting a Bayesian point of view, utilizing an improper prior density for  $\lambda$  given by  $p(\lambda) \propto 1/\lambda$  with (2.1) as likelihood function results in a posterior density for  $\lambda$  which is a Gamma density function:

$$(2.7) \quad p(\lambda|n, m) = \lambda^{n-1} \exp(-m\lambda) m^n / \Gamma(n)$$

In order this posterior density function to be proper,  $n$  must be positive. Calculating the posterior mean and variance of  $\log(\lambda)$  gives:

$$(2.8) \quad E(\log \lambda) = \psi(n) - \log(m) = \sum_{i=1}^{n-1} i^{-1} - \log(m) - \gamma$$

$$\doteq \log\{(n - \frac{1}{2})/m\}$$

$$V(\log \lambda) = \psi'(n) = \pi^2/6 - \sum_{i=1}^{n-1} i^{-2} \doteq (n - \frac{1}{2})^{-1}$$

where  $\psi$  and  $\psi'$  denote the digamma and trigamma function and  $\gamma = 0.5772156649\dots$ ; see ABRAMOWITZ and STEGUN (1970, ch. 6) for the above results associated with the gamma function.

Now we can write:

$$(2.9) \quad E(\log \lambda) = \mathbf{x}'\beta + \{E(\log \lambda) - \log \lambda\}$$

where  $\{E(\log \lambda) - \log \lambda\}$  is unobservable, having mean 0 and finite variance given by  $V(\log \lambda)$ . Utilizing the approximations for  $E(\log \lambda)$  and  $V(\log \lambda)$ , adding the index  $r$  again, assuming  $n_r \geq 1$  for simplicity of notation, enables us to apply weighted least squares to (2.9).

This results in a linear estimator for  $\beta$  given by:

$$(2.10) \quad \hat{\beta} = [\mathbf{X}'(\mathbf{N} - \frac{1}{2}\mathbf{I})\mathbf{X}]^{-1} \mathbf{X}'(\mathbf{N} - \frac{1}{2}\mathbf{I})\mathbf{v}$$

where  $\mathbf{N} = \text{diag}(n_1, \dots, n_{\mathbf{R}})$  and  $\mathbf{v}$  contains the  $\mathbf{r}$  values  $\log\{(n_r - \frac{1}{2})/m_r\}$ .

3. THE LOGLINEAR GAMMA MODEL

The analysis of claim costs is hampered by the fact that it is difficult to specify a probability distribution which gives a good description of empirical reality. However, the total claim costs for  $n$  claims will tend to have a smooth probability distribution and will approach the Normal distribution.

The Gamma distribution approaches the Normal distribution too and assigns no probability mass to the negative axis. So, it may be fruitful to adopt a Gamma distribution for the total claim costs and act as if the individual claim costs follow a Gamma distribution too.

The probability density function for the Gamma distribution can be written as:

$$(3.1) \quad g(q|\mu, \varphi) = q^{\varphi-1} \exp(-\varphi q/\mu) \varphi^\varphi \mu^{-\varphi} / \Gamma(\varphi)$$

implying a mean and variance given by  $u$  and  $u^2/\varphi$ .

The sum of  $n$  of such distributed random variables will follow a Gamma distribution with parameters  $n\mu$  and  $n\varphi$ . The sample mean will follow a Gamma distribution with parameters  $\mu$  and  $n\varphi$ . The sample mean happens to be the maximum likelihood estimator for the population mean  $\mu$ .

The parametrization:

$$(3.2) \quad \log(\mu_r) = \mathbf{x}'_r \boldsymbol{\theta} \quad r = 1, \dots, R$$

results in a loglikelihood function:

$$(3.3) \quad \log L = \text{const} + \varphi \sum n_r \log(y_r) - \sum \log \{ \Gamma(\varphi n_r) \} + \varphi \log(\varphi) \sum n_r + \varphi \{ -(\sum n_r \mathbf{x}_r)' \boldsymbol{\theta} - \sum y_r \exp(-\mathbf{x}'_r \boldsymbol{\theta}) \}$$

where  $y_r$  stands for the total claim costs for the  $n_r$  claims in riskgroup  $r$ . The form between curls in (3.3) is equivalent with the crucial part of (2.4) if we substitute  $\boldsymbol{\beta}$  for  $-\boldsymbol{\theta}$  and  $m_r$  for  $y_r$ . This gives the possibility to calculate the maximum likelihood estimator for  $\boldsymbol{\theta}$  in the same way as in case of the loglinear Poisson model.

It is interesting to observe that aggregation of the individual claim costs to total claim costs does not have any implications for the maximum likelihood estimate of  $\boldsymbol{\theta}$ . This remains valid for arbitrary specifications for  $\mu_r$ , as long as these do not depend on  $\varphi$ . The aggregation does have implications for the estimation of  $\varphi$ , however.

Differentiating (3.3) with respect to  $\boldsymbol{\theta}$  and  $\varphi$  gives:

$$(3.4) \quad \frac{\partial \log L}{\partial \boldsymbol{\theta}} = \varphi \{ -\sum n_r \mathbf{x}_r + \sum y_r \exp(-\mathbf{x}'_r \boldsymbol{\theta}) \mathbf{x}_r \}$$
  
$$\frac{\partial \log L}{\partial \varphi} = \sum \{ n_r \log(y_r) - n_r \mathbf{x}'_r \boldsymbol{\theta} - y_r \exp(-\mathbf{x}'_r \boldsymbol{\theta}) + n_r \} + \log(\varphi) \sum n_r - \sum n_r \psi(\varphi n_r).$$

The second order derivatives amount to:

$$\begin{aligned}
 (3.5) \quad \frac{\partial^2 \log L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} &= -\varphi \sum y_r \exp(-\mathbf{x}_r' \boldsymbol{\theta}) \mathbf{x}_r \mathbf{x}_r' \\
 \frac{\partial^2 \log L}{\partial \boldsymbol{\theta} \partial \varphi} &= -\sum n_r \mathbf{x}_r + \sum y_r \exp(-\mathbf{x}_r' \boldsymbol{\theta}) \mathbf{x}_r \\
 \frac{\partial^2 \log L}{\partial \varphi^2} &= \varphi^{-1} \sum n_r - \sum n_r^2 \psi'(\varphi n_r).
 \end{aligned}$$

Minus the expectation of these second order derivatives gives:

$$(3.6) \quad -E \left[ \frac{\partial^2 \log L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] = \varphi \mathbf{X}' \mathbf{N} \mathbf{X}$$

$$(3.7) \quad -E \left[ \frac{\partial^2 \log L}{\partial \boldsymbol{\theta} \partial \varphi} \right] = \mathbf{0} .$$

The vanishing of (3.7) implies a blockdiagonal information matrix for  $\boldsymbol{\theta}$  and  $\varphi$ . This implies that the maximum likelihood estimator for  $\boldsymbol{\theta}$  follows asymptotically a multivariate normal distribution with mean  $\boldsymbol{\theta}$  and variance-covariance matrix given by  $\varphi^{-1} (\mathbf{X}' \mathbf{N} \mathbf{X})^{-1}$ . This matrix can be estimated by substituting the maximum likelihood estimate for  $\varphi$ ,

$$\begin{aligned}
 (3.8) \quad -E \left[ \frac{\partial^2 \log L}{\partial \varphi^2} \right] &= \sum n_r^2 [\psi'(\varphi n_r) - (\varphi n_r)^{-1}] \\
 &\sim \sum \frac{1}{2} \varphi^{-2} = \frac{1}{2} \mathbf{R} \varphi^{-2}, \text{ for large } n_r \varphi.
 \end{aligned}$$

So, the asymptotic variance of the maximum likelihood estimator for  $\varphi$  equals  $2\varphi^2/\mathbf{R}$ , which goes to 0 when  $\mathbf{R}$  goes to infinity.  $\mathbf{R}$  can be large, but is fixed, contrary to the number of claims. In order to improve the estimation of  $\varphi$ , aggregation of claim costs should not take place blindly. Risk groups with many claims should form separate totals such that  $\mathbf{R}$  increases.

As the sample mean  $y_r/n_r$  follows a Gamma distribution with mean  $\mu_r$  and variance  $\mu_r^2/n_r \varphi$ , the mean and variance of  $\log(y_r/n_r)$  are given and approximated by:

$$\begin{aligned}
 (3.9) \quad E[\log(y_r/n_r)] &= \mathbf{x}_r' \boldsymbol{\theta} + \psi(n_r \varphi) - \log(n_r \varphi) \\
 &\doteq \mathbf{x}_r' \boldsymbol{\theta} - (2n_r \varphi)^{-1} \\
 V[\log(y_r/n_r)] &= \psi'(n_r \varphi) \doteq (n_r \varphi)^{-1}
 \end{aligned} \left. \vphantom{\begin{aligned} E[\log(y_r/n_r)] \\ \doteq \mathbf{x}_r' \boldsymbol{\theta} - (2n_r \varphi)^{-1} \\ V[\log(y_r/n_r)] = \psi'(n_r \varphi) \doteq (n_r \varphi)^{-1} \end{aligned}} \right\} r = 1, \dots, \mathbf{R}$$

Utilizing the approximations and adding the normality assumption, enables us to apply the maximum likelihood method to (3.9), resulting in closed form estimators for  $\boldsymbol{\theta}$  and  $\varphi$  given by:

$$(3.10) \quad \tilde{\theta} = (\mathbf{X}'\mathbf{N}\mathbf{X})^{-1} [\mathbf{X}'\mathbf{N}\mathbf{w} + (2\tilde{\varphi})^{-1} \mathbf{X}'\mathbf{1}]$$

$$(3.11) \quad \tilde{\varphi} = \frac{\mathbb{R} + \{\mathbb{R}^2 + \mathbf{w}'[\mathbf{N} - \mathbf{N}\mathbf{X}(\mathbf{X}'\mathbf{N}\mathbf{X})^{-1}\mathbf{X}'\mathbf{N}]\mathbf{w} \cdot \mathbf{1}'[\mathbf{N}^{-1} - \mathbf{X}(\mathbf{X}'\mathbf{N}\mathbf{X})^{-1}\mathbf{X}']\mathbf{1}\}^{1/2}}{2\mathbf{w}'[\mathbf{N} - \mathbf{N}\mathbf{X}(\mathbf{X}'\mathbf{N}\mathbf{X})^{-1}\mathbf{X}'\mathbf{N}]\mathbf{w}}$$

where  $\mathbf{1}$  is a vector with all element equal to 1 and  $\mathbf{w}$  contains the  $\mathbb{R}$  values  $\log(y_r/n_r)$ . When the diagonal elements of  $\mathbf{N}$  grow large (3.10) will approach  $(\mathbf{X}'\mathbf{N}\mathbf{X})^{-1}\mathbf{X}'\mathbf{N}\mathbf{w}$ , being independent of  $\varphi$ . This furnishes a convenient starting value for  $\theta$ .

#### REFERENCES

- ABRAMOWITZ, M. and I. A. STEGUN (Eds.). (1970). *Handbook of Mathematical Functions*, New York: Dover Publications, Inc.
- AJNE, B. (1975). A Note on the Multiplicative Ratemaking Model, *Astin Bulletin* **8**, 144-153.
- ANDERSEN, E. B. (1977). Multiplicative Poisson Models with Unequal Cell Rates, *Scandinavian Journal of Statistics* **4**, 153-158.
- BAILY, R. A. and L. J. SIMON (1960). Two Models in Automobile Insurance Ratemaking, *Astin Bulletin* **1**, 912-217.
- HABERMAN, S. (1974). *The Analysis of Frequency Data*, Chicago: The University of Chicago Press.
- JUNG, J. (1968). On Automobile Insurance Ratemaking, *Astin Bulletin* **5**, 41-48.