Likelihood Computations for Extended Poisson Process Models

H. M. Podlich[*], M. J. Faddy[†], G. K. Smyth

Centre for Statistics

Department of Mathematics

The University of Queensland

Brisbane, Australia

Abstract

Some computational aspects of maximum likelihood estimation for extended Poisson process models are discussed, with computation of log-likelihood derivatives being of particular interest. A method is proposed for computation of these derivatives that involves extending the matrix of transition rates describing the underlying stochastic process. This scheme is designed for parametric forms of the transition rates that can include covariate dependence.

Keywords: extended Poisson process models, matrix of transition rates, expokit, maximum likelihood estimation

[*]Current address: School of Applied Mathematics and Statistics, Griffith University, Brisbane, Australia.

[†]Current address: Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand.

1 Introduction

Extended Poisson process models provide a general framework for the analysis of discrete data (Faddy, 1997a, 1997b, 1998). They involve representing a discrete distribution as the distribution of the number of events occurring in a finite time interval of a state-dependent Markov birth process. In such a process the rates at which events occur, known as the transition rates, are modelled as a function of n, the number of accumulating events. It is this state-dependency, or n-dependent form, of the rates which determines the dispersion properties of the resulting discrete distribution. The generality of the approach lies in the fact that any discrete distribution with non-negative support has such a representation (Faddy, 1997a) and therefore other models for discrete data can be seen as special cases. However, these extended Poisson process models do involve more intensive computation than other more traditional models for discrete data.

In order to form the log-likelihood, probabilities must be obtained by solving the Chapman-Kolmogorov forward equations of the underlying stochastic process, for which an exact analytical form suitable for computational purposes is not available. Numerical solutions of the Chapman-Kolmogorov equations are therefore required. There are two alternatives here: utilising a differential equation solver or computing the exponential of the Q-matrix of transition rates. Recent work (Sidje, 1998) has led us to consider the latter of these. Sidje's expokit package has been specifically developed for applications such as Markov chain models where the matrices to be exponentiated are typically large and sparse, but their matrix exponentials are large and dense. Therefore, rather than computing the matrix exponential in its entirety, expokit considers the operation of the matrix exponential on a vector.

The log-likelihood must then be maximised over the parameters specifying the transition rates. In Faddy (1997a, 1997b, 1998) the maximization is performed in MATLAB using a *simplex search* method. However for large numbers of parameters and large count sizes this search method can be computationally inefficient; larger counts mean computations involving larger matrix exponentials and hence an increase in the per iteration time, while more parameters result in an increased number of iterations. To improve the maximization process, a convenient method for computing first and second derivatives of the log-likelihood function was therefore sought.

Moreover, computation of these derivatives as a part of the maximization process means standard errors for estimates can be readily obtained and used to construct Wald tests. Similarly, first and second derivatives can be used to construct score tests.

The method proposed here for computation of derivatives involves extending the Q-matrix to be exponentiated. This is outlined is section 4. In sections 2 and 3 extended Poisson process models are described, and in section 5 an example data-set is used to demonstrate the application of the methods.

2 Extended Poisson process models

Extended Poisson process models derive their name from the fact that they are based upon generalizing the simplest Markov birth process, the *Poisson process*. This process describes a series of 'events' occurring over time t such that the probability of an event occurring in the time interval $(t, t + \delta t)$ is $\lambda \delta t + o(\delta t)$, independently of the occurrence of events up to time t (Cox and Miller, 1965). The constant λ here is the rate at which events occur, known as the *transition* rate, and under this assumption of constant transition rates, the distribution of the number of events occurring in a finite time interval of length t is *Poisson* with mean λt.

The generalization of the Poisson process considered here is to have the transition rates dependent on the number of accumulating events. That is, the probability of an event occurring in the time interval $(t, t+\delta t)$ is $\lambda(n)\delta t + o(\delta t)$, where n is the number of events that have occurred by time t. The distribution of the number of events occurring in any finite time interval of such a state-dependent Markov process is no longer Poisson. For example, any increasing transition rate sequence $\lambda(n)$ gives rise to a distribution that is over-dispersed relative to the Poisson distribution, while any decreasing transition rate sequence $\lambda(n)$ gives rise to a distribution that is under-dispersed relative to the Poisson distribution (Ball, 1995). A linear increasing transition rate sequence gives rise to the negative binomial distribution, while a linear decreasing transition rate sequence gives rise to the binomial distribution. Further, it has been conjectured that a concave increasing sequence (negative second differences) corresponds to over-disperison relative to the Poisson distribution, but under-dispersion relative to the negative binomial distribution, while a convex increasing sequence (positive second differences) corresponds to over-dispersion relative to the negative binomial distribution. Similarly, a concave decreasing sequence corresponds to over-dispersion relative to the binomial distribution, but under-dispersion relative to the Poisson distribution, while a convex decreasing sequence corresponds to under-dispersion relative to the binomial distribution (Faddy, 1997b). In fact, any discrete distribution with non-negative support has a *unique* representation as the distribution of the number of events occurring in a finite time interval of a state-dependent Markov birth process (Faddy, 1997a), and hence the generality of this modelling approach is apparent.

In transition rate models such as those proposed by Faddy (1997a, 1997b, 1998), the n-dependence form is specified parametrically. One or more parameters control the variation in the model and other parameters may be modelled as some function of covariates to allow for the assessment of these effects. For example, in a later section the model

$$\lambda(n) = \begin{cases} \lambda_0 & n = 0 \\ \lambda_1 \, n^c & n \geq 1. \end{cases}$$

will be considered. The c parameter here controls the variation. The $c = 0$ case is similar to a Poisson distribution with modified probability at zero, while $c > 0$ and $c < 0$ correspond to more and less variation, respectively, than the $c = 0$ model (Faddy, 1998). The other parameters, λ_0 and λ_1, may be modelled as log-linear functions of covariates,

$$\begin{aligned} \lambda_0 &= \exp(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0) \\ \lambda_1 &= \exp(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_1), \end{aligned}$$

where $\mathbf{x}^{\mathrm{T}}$ is a vector of covariates and $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ are vectors of covariate coefficients. The use of a log-linear form here will ensure that the transition rates remain positive.

3 The Q-matrix

To construct the probability distribution $p_n(t), n \geq 0$, arising from a Markov birth process with birth rates, $\lambda(n)$ $n \geq 0$, it is necessary to solve a system of differential equations from the underlying stochastic process, the *Chapman-Kolmogorov* forward equations (Cox and Miller, 1965),

$$\begin{aligned} p_0'(t) &= -\lambda(0)p_0(t) \quad \text{with } p_0(0) = 1 \\ p_n'(t) &= \lambda(n-1)p_{n-1}(t) - \lambda(n)p_n(t) \quad \text{with } p_n(0) = 0, \quad n \geq 1. \end{aligned} \tag{1}$$

In constructing discrete probability distributions from the solution of (1), the time t may be taken to be one without any loss of generality.

Suppose it is required to calculate the probability of obtaining a count of size N. The Chapman-Kolmogorov equations, (1), may be re-written in the matrix-vector form,

$$\frac{d\mathbf{p}(t)}{dt} = \mathbf{p}(t)\mathbf{Q}, \tag{2}$$

where $\mathbf{p}(t)$ is the vector of probabilities, $(p_0(t)\ p_1(t)\ \ldots\ p_N(t))$, and $\mathbf{Q}$ is the Q-matrix of transition rates,

$$\mathbf{Q} = \begin{pmatrix} -\lambda(0) & \lambda(0) & 0 & \cdots & 0 & 0 \\ 0 & -\lambda(1) & \lambda(1) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\lambda(N-1) & \lambda(N-1) \\ 0 & 0 & 0 & \cdots & 0 & -\lambda(N) \end{pmatrix}.$$

The solution of (1) may therefore be expressed in terms of the matrix exponential function,

$$\mathbf{p}(t) = (1\ 0\ \ldots\ 0\ 0)\exp(\mathbf{Q}t),$$

or, with t taken to be one,

$$(p_0\ p_1\ \ldots\ p_{N-1}\ p_N) = (1\ 0\ \ldots\ 0\ 0)\exp(\mathbf{Q}). \tag{3}$$

The probability of obtaining a count of size N, p_N, may therefore be computed by taking the $(N+1)$th, or last, entry of the resulting matrix exponential vector operation (3). The expokit software developed by Sidje (1998), for example, may be used to perform this operation.

4 The extended Q-matrix

To allow for computation of first and second derivatives of probabilities for iterative maximization of a log-likelihood function, the Q-matrix of transition rates may be extended as follows. Recall that throughout t will be taken to be one.

4.1 First derivatives

In general, $\mathbf{Q}$ and $\mathbf{p}$ will be functions of parameters specifying the transition rates and therefore the matrix-vector form of the Chapman-Kolmogorov equations (2) may be written as

$$\frac{\partial \mathbf{p}}{\partial t} = \mathbf{p}\mathbf{Q}. \tag{4}$$

Suppose computation of $\partial \mathbf{p}/\partial a$ is required, where a is such a parameter. Then differentiating (4) with respect to a,

$$\frac{\partial}{\partial a}\left(\frac{\partial \mathbf{p}}{\partial t}\right) = \frac{\partial \mathbf{p}}{\partial a}\mathbf{Q} + \mathbf{p}\frac{\partial \mathbf{Q}}{\partial a}.$$

On reversing the order of differentiation,

$$\frac{\partial}{\partial t}\left(\frac{\partial \mathbf{p}}{\partial a}\right) = \frac{\partial \mathbf{p}}{\partial a}\mathbf{Q} + \mathbf{p}\frac{\partial \mathbf{Q}}{\partial a}. \tag{5}$$

Combining (5) with (4) gives,

$$\frac{\partial}{\partial t}\begin{pmatrix}\mathbf{p} & \frac{\partial \mathbf{p}}{\partial a}\end{pmatrix} = \begin{pmatrix}\mathbf{p} & \frac{\partial \mathbf{p}}{\partial a}\end{pmatrix}\begin{pmatrix}\mathbf{Q} & \frac{\partial \mathbf{Q}}{\partial a} \\ \mathbf{0} & \mathbf{Q}\end{pmatrix}. \tag{6}$$

Let $\mathbf{Q}^* = \begin{pmatrix} \mathbf{Q} & \frac{\partial \mathbf{Q}}{\partial a} \\ \mathbf{0} & \mathbf{Q} \end{pmatrix}$, the extended blocked Q-matrix. Then the solution to (6) is given by,

$$\begin{pmatrix} \mathbf{p} & \frac{\partial \mathbf{p}}{\partial a} \end{pmatrix} = \mathbf{e}_1^{\mathrm{T}} \exp(\mathbf{Q}^* t),$$

where $\mathbf{e}_1^{\mathrm{T}}$ is the initial vector for the left-hand side of (6), $(1\ 0\ \ldots 0\ 0\ 0\ \ldots\ 0)$.

Therefore, in order to compute the vector of first derivatives, the matrix exponential vector operation involves components of dimension *twice* the order of the original Q-matrix. Note that the vector of probabilities is also computed in this operation.

To allow for computation of second derivatives as well, such a $\mathbf{Q}^*$ can be further extended.

4.2 Second derivatives

Differentiating (5) again with respect to another parameter b, say, gives

$$\frac{\partial^2}{\partial a \partial b}\left(\frac{\partial \mathbf{p}}{\partial t}\right) = \frac{\partial^2 \mathbf{p}}{\partial a \partial b}\mathbf{Q} + \frac{\partial \mathbf{p}}{\partial a}\frac{\partial \mathbf{Q}}{\partial b} + \frac{\partial \mathbf{p}}{\partial b}\frac{\partial \mathbf{Q}}{\partial a} + \mathbf{p}\frac{\partial^2 \mathbf{Q}}{\partial a \partial b}.$$

So that combining the above with (5) and (6) together with,

$$\frac{\partial}{\partial t}\left(\frac{\partial \mathbf{p}}{\partial b}\right) = \frac{\partial \mathbf{p}}{\partial b}\mathbf{Q} + \mathbf{p}\frac{\partial \mathbf{Q}}{\partial b},$$

results in

$$\frac{\partial}{\partial t}\begin{pmatrix} \mathbf{p} \\ \frac{\partial \mathbf{p}}{\partial a} \\ \frac{\partial \mathbf{p}}{\partial b} \\ \frac{\partial^2 \mathbf{p}}{\partial a \partial b} \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} \mathbf{p} \\ \frac{\partial \mathbf{p}}{\partial a} \\ \frac{\partial \mathbf{p}}{\partial b} \\ \frac{\partial^2 \mathbf{p}}{\partial a \partial b} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \mathbf{Q} & \frac{\partial \mathbf{Q}}{\partial a} & \frac{\partial \mathbf{Q}}{\partial b} & \frac{\partial^2 \mathbf{Q}}{\partial a \partial b} \\ \mathbf{0} & \mathbf{Q} & \mathbf{0} & \frac{\partial \mathbf{Q}}{\partial b} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q} & \frac{\partial \mathbf{Q}}{\partial a} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q} \end{pmatrix}. \tag{7}$$

Let

$$\mathbf{Q}^* = \begin{pmatrix} \mathbf{Q} & \frac{\partial \mathbf{Q}}{\partial a} & \frac{\partial \mathbf{Q}}{\partial b} & \frac{\partial^2 \mathbf{Q}}{\partial a \partial b} \\ \mathbf{0} & \mathbf{Q} & \mathbf{0} & \frac{\partial \mathbf{Q}}{\partial b} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q} & \frac{\partial \mathbf{Q}}{\partial a} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q} \end{pmatrix},$$

then the solution to (7) is given by,

$$\begin{pmatrix} \mathbf{p} \\ \frac{\partial \mathbf{p}}{\partial a} \\ \frac{\partial \mathbf{p}}{\partial b} \\ \frac{\partial^2 \mathbf{p}}{\partial a \partial b} \end{pmatrix}^{\mathrm{T}} = \mathbf{e}_1^{\mathrm{T}} \exp\left(\mathbf{Q}^* t\right).$$

Computation of mixed derivatives therefore requires considering a matrix exponential vector operation of dimension *four* times the order of the original Q-matrix. For the second derivative, $\partial^2 \mathbf{p}/\partial a^2$, a matrix *three* times the order of the Q-matrix is required,

$$\mathbf{Q}^* = \begin{pmatrix} \mathbf{Q} & \frac{\partial \mathbf{Q}}{\partial a} & \frac{\partial^2 \mathbf{Q}}{\partial a^2} \\ \mathbf{0} & \mathbf{Q} & 2\frac{\partial \mathbf{Q}}{\partial a} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q} \end{pmatrix}. \tag{8}$$

For computation of all the first and second derivatives with respect to two parameters a and b, the single extended Q-matrix which allows for computation of the probability vector together with these derivatives is given by,

$$\mathbf{Q}^* = \begin{pmatrix} \mathbf{Q} & \frac{\partial \mathbf{Q}}{\partial a} & \frac{\partial \mathbf{Q}}{\partial b} & \frac{\partial^2 \mathbf{Q}}{\partial a^2} & \frac{\partial^2 \mathbf{Q}}{\partial b^2} & \frac{\partial^2 \mathbf{Q}}{\partial a \partial b} \\ \mathbf{0} & \mathbf{Q} & \mathbf{0} & 2\frac{\partial \mathbf{Q}}{\partial a} & \mathbf{0} & \frac{\partial \mathbf{Q}}{\partial b} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q} & \mathbf{0} & 2\frac{\partial \mathbf{Q}}{\partial b} & \frac{\partial \mathbf{Q}}{\partial a} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q} \end{pmatrix}.$$

All derivative vectors and the probability vector may then be computed at once since,

$$\begin{pmatrix} \mathbf{p} \\ \frac{\partial \mathbf{p}}{\partial a} \\ \frac{\partial \mathbf{p}}{\partial b} \\ \frac{\partial^2 \mathbf{p}}{\partial a^2} \\ \frac{\partial^2 \mathbf{p}}{\partial b^2} \\ \frac{\partial^2 \mathbf{p}}{\partial a \partial b} \end{pmatrix}^{\mathrm{T}} = \mathbf{e}_1^{\mathrm{T}} \exp(\mathbf{Q}^* t), \tag{9}$$

where $\mathbf{e}_1^{\mathrm{T}} = (1\ 0\ \ldots\ 0\ 0)$ is the initial vector for the left-hand side of (9). Such extensions of the Q-matrix may be continued in a natural way to deal with derivatives with respect to three or more parameters.

However, it may be more economical for large counts to construct a number of smaller matrices and compute the derivatives required in stages, rather than computing all of the derivatives at once. A matrix such as (8) may be formed for each parameter and all first derivatives and second derivatives obtained by successively exponentiating each of these matrices. For computation of mixed derivatives, a matrix such as (7) may be constructed for each distinct pair of parameters. Several matrix exponential vector operations will be required, but the matrices and vectors will be of smaller dimension.

5 Example: Leadbeater's possum counts

To illustrate the method of extending the Q-matrix to allow for computation of likelihood derivatives, the analysis of some species abundance data by Faddy (1998) will be re-visited.

The data consist of the numbers of Leadbeater's possums (*Gymnobelideus leadbeateri*) sampled over 151 $3ha$. sites across Central Victoria, Australia, with the main interest being the assessment of significance of covariates describing various habitat characteristics. The covariates which were measured on each site and believed to be relevant to the abundance of this species (Welsh et al., 1996), are given below,

lstags: $\log_e$(no. of trees with hollows +1),
age: forest age,
baa: basal area of *Acacia species on site*,
slope: slope of the site,
aspect: aspect of the site
bark: score for degree of *decorticating* or peeling bark,
no_s: score for number of shrubs on the site.

All of these covariates are quantitative with the exception of the *aspect* covariate which is a factor variable at four levels.

To allow for the high proportion of zero counts in the data, the transition rate model introduced earlier in section 2 was used,

$$\lambda(n) = \begin{cases} \lambda_0 & n = 0 \\ \lambda_1 \, n^c & n \geq 1, \end{cases}$$

where λ_0 and λ_1 are log-linear functions of the covariates,

$$\begin{aligned} \lambda_0 &= \exp(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_0) \\ \lambda_1 &= \exp(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_1). \end{aligned}$$

The extended Q-matrix for such a model, which will allow for the probability vector and all of its first and second derivatives to be computed at once, is given by,

$$
\mathbf{Q}^* = \begin{pmatrix}
\mathbf{Q} & \frac{\partial \mathbf{Q}}{\partial \lambda_0} & \frac{\partial \mathbf{Q}}{\partial \lambda_1} & \frac{\partial \mathbf{Q}}{\partial c} & \frac{\partial^2 \mathbf{Q}}{\partial \lambda_0^2} & \frac{\partial^2 \mathbf{Q}}{\partial \lambda_1^2} & \frac{\partial^2 \mathbf{Q}}{\partial c^2} & \frac{\partial^2 \mathbf{Q}}{\partial \lambda_0 \partial \lambda_1} & \frac{\partial^2 \mathbf{Q}}{\partial \lambda_0 \partial c} & \frac{\partial^2 \mathbf{Q}}{\partial \lambda_1 \partial c} \\
\mathbf{0} & \mathbf{Q} & \mathbf{0} & \mathbf{0} & 2\frac{\partial \mathbf{Q}}{\partial \lambda_0} & \mathbf{0} & \mathbf{0} & \frac{\partial \mathbf{Q}}{\partial \lambda_1} & \frac{\partial \mathbf{Q}}{\partial c} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{Q} & \mathbf{0} & \mathbf{0} & 2\frac{\partial \mathbf{Q}}{\partial \lambda_1} & \mathbf{0} & \frac{\partial \mathbf{Q}}{\partial \lambda_0} & \mathbf{0} & \frac{\partial \mathbf{Q}}{\partial c} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q} & \mathbf{0} & \mathbf{0} & 2\frac{\partial \mathbf{Q}}{\partial c} & \mathbf{0} & \frac{\partial \mathbf{Q}}{\partial \lambda_0} & \frac{\partial \mathbf{Q}}{\partial \lambda_1} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{Q}
\end{pmatrix}.
$$

Derivatives with respect to the components of vectors of covariate coefficients, $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$, can be obtained in terms of those of λ_0 and λ_1 by applying the usual chain rule. For example,

$$
\begin{aligned}
\frac{\partial p_i}{\partial \boldsymbol{\beta}_{0j}} &= \frac{\partial p_i}{\partial \lambda_0} \frac{\partial \lambda_0}{\partial \boldsymbol{\beta}_{0j}} \\
&= \lambda_0 \, x_j \frac{\partial p_i}{\partial \lambda_0},
\end{aligned}
$$

where p_i is the probability of the ith observation and x_j is the jth covariate value for this observation. Therefore, the actual number of covariates is of no consequence to the complexity of the proposed scheme. Only those factors determining the size of the matrix to be exponentiated, namely the number of parameters used when specifying the n-dependence of the transition rates (three here) and the size of the counts, will affect the computation required.

Given below are the resulting covariate coefficient estimates, together with their corresponding asymptotic standard errors computed from the observed information matrix. The corner point constraint ($asp1$= 0) has been imposed for the *aspect* variable.

	Parameter		Parameter	
Coef	$\log(\lambda_0)$	s.e.	$\log(\lambda_1)$	s.e.
Const	−5.0715	(1.1941)	1.6803	(0.8509)
lstags	0.8902	(0.2500)	0.3890	(0.1973)
age	0.1304	(0.0913)	−0.0490	(0.0705)
baa	0.0670	(0.0179)	0.0094	(0.0185)
slope	0.0073	(0.0206)	−0.0360	(0.0220)
asp2	0.5117	(0.3952)	0.3788	(0.3551)
asp3	0.2629	(0.4064)	0.2758	(0.4055)
asp4	−0.0967	(0.0333)	0.0021	(0.4601)
bark	0.0634	(0.0333)	0.0392	(0.0246)
no_s	0.0101	(0.0437)	−0.1486	(0.0447)

The corresponding log-likelihood is $\ell\ell = -181.3683$, and the estimated c parameter is $c = -0.5509\,(0.2833)$. The log-likelihood maximization here was performed using the 'nlminb' routine (derivatives supplied), with matrix exponentiation carried out in outer fortran routines using the expokit software package. On the Pentium 300MHz machine the maximization was completed in under ten seconds, fast enough for interactive modelling. The original computation described by Faddy (1998) using the *simplex search* took several hours.

The log-likelihood and its derivatives computed as a part of the maximization process can be used to construct tests for model parameters. The log-likelihood ratio test for $c = 0$, based on the chi-squared approximation to twice the change in log-likelihood between the restricted $c = 0$ and full maximum likelihood model fits, returns p-value ≈ 0.0327. An approximate score test for $c = 0$, based on

$$u^T J^{-1} u \sim \chi^2_1$$

evaluated at $c = 0$, with u being the score vector and J the observed information matrix, gives p-value ≈ 0.0448. Both results suggest that the negative value for c should be retained. Such a negative value indicates the presence of less dispersion than the $c = 0$ model can allow for. An approximate Wald test for $c = 0$, based on

$$\frac{\hat{c}}{s.e.(\hat{c})} \sim N(0, 1),$$

where $\hat{c}$ denotes the maximum likelihood estimate of c, returns p-value ≈ 0.0518. However the Wald test might be considered less reliable than the two previous tests in this context because of its lack of invariance under reparameterization.

Comparing regression coefficients with their corresponding standard errors in the table above suggests that the significant variables are *stags*, *baa*, and *bark* for λ_0, and *stags*, *slope*, *bark* and *no_s* for λ_1. Forward selection using score tests, or backward elimination using likelihood ratio tests as in Faddy (1998), confirms this selection of significant variables.

6 Conclusion

It has been shown in this paper how probabilities and derivatives used in model fitting of extended Poisson process models can be computed by exponentiating an extended Q-matrix of transition rates. Such calculations can be done using the expokit package (Sidje, 1998) since only a matrix exponential vector operation is required, and not the entire matrix exponential.

The method makes it practical for extended Poisson process modelling to be used as an interactive data analysis tool, and gives users access to the full range of likelihood based inference. S-Plus functions and compiled object code for Windows which implement the methods described in this paper are available from `http://www.maths.uq.edu.au/~hmp/`.

References

Ball, F. (1995) A Note on Variation in Birth Processes. *Mathematical Scientist* **20**, 50-55.

Cox, D.R. and Miller, H.D. (1965) *The Theory of Stochastic Processes.* Methuen, London.

Faddy, M.J. (1997a) Extended Poisson Process Modelling and Analysis of Count Data. *Biometrical Journal* **39**, 431-440.

Faddy, M.J. (1997b) On Extending the Negative Binomial Distribution, and the Number of Weekly Winners of the UK National Lottery. *Mathematical Scientist* **22**, 77-82.

Faddy, M.J. (1998) Stochastic Models for Analysis of Species Abundance Data in *Statistics in Ecology and Environmental Monitoring 2*, 33-40. Fletcher, D.J. Kavalieris, L. and Manly, B.F.J. (Eds.) University of Otago Press.

Sidje, R.B. (1998) EXPOKIT: Software Package for Computing Matrix Exponentials. *ACM Transactions on Mathematical Software* **24**, 130-156.

Welsh, A.H., Cunningham, R.B., Donnelly, C.F. and Lindenmayer, D.B. (1996) Modelling the Abundance of Rare Species: Statistical Models for Counts with Extra Zeros. *Ecological Modelling* **88**, 297-308.