An Improved Saddlepoint Approximation Based on the Negative Binomial Distribution for the General Birth Process

Gordon K. Smyth and Heather M. Podlich

Department of Mathematics and School of Land and Food, University of Queensland, Australia 4072

Summary

Daniels (*JAP* 1982) gave a saddlepoint approximation to the probabilities of a general birth process. This paper gives an improved approximation which is only slightly more complex than Daniels' approximation and which has considerably reduced relative error in most cases. The new approximation has the characteristic that it is exact whenever the birth rates can be reordered into a linear increasing sequence.

Keywords: saddlepoint approximation, birth process, negative binomial distribution, binomial distribution, exponential tilting.

1 Introduction

Daniels (1982) gave a saddlepoint approximation to the probabilities of a general birth process. This paper gives an improved approximation which is only slightly more complex than Daniels' approximation and which has considerably reduced relative error in most cases. The new approximation has the characteristic that it is exact whenever the birth rates can be reordered into a linear increasing sequence.

A birth process is an example of a Markov chain and there is a large and active literature on the numerical computation of transient probabilities for Markov chains. Stewart (1994, Chapter 8) gives a detailed and readable survey of the available methods and Stewart (1995) is a recent collection of papers. Popular methods include matrix exponentiation by uniformization (Grassman, 1977), matrix exponention by Krylov space methods (Sidje, 1998) and direct solving of the Chapman-Kolmogorov differential equations. Sidje and Stewart (1999) compare a range of methods in a recent numerical study. Podlich et al (1999a) discuss numerical computation of derivatives of probabilities. Despite the existance of these numerical methods there is still considerable interest in accurate analytic approximations. This is partly because of the complexity of the numerical methods but also because it diffi-cult to guarantee the accuracy for all probabilities in all cases. The numerical methods are matrix orientated and generally return all probabilities starting from $n = 0$ even if only one probability is required. Error bounds are usu-ally dominated by the larger probabilities and very small probabilities may not be computed to good relative accuracy. The saddlepoint approximation solves these problems and also can be manipulated analytically, for example differentiated with respect to unknown parameters.

The ordinary saddlepoint approximation is accurate for a wide range of birth rates to two or more significant figures. This accuracy is satisfactory for descriptive purposes, but greater accuracy is desirable for example when using the probabilities for maximum likelihood fitting of parametric models. See Podlich et al (1999b) for a recent application requiring accurate analytic approximations for birth rate probabilities.

Daniels (1982) derived his approximation using a method from asymptotic analysis known as the saddlepoint technique or the method of steepest descent applied to the inversion integral of the Laplace transform of the probabili-ties. Reid (1988) pointed out that the saddlepoint technique has a statistical interpretation in terms of exponentially tilted densities. The distribution of interest is imbedded in a linear exponential family and an Edgeworth expan-sion is applied to the exponential family density at an optimal choice of the canonical parameter. The ordinary saddlepoint approximation corresponds to a single term Edgeworth expansion, i.e., to a normal approximation to the exponential family density.

Several authors have investigated using a non-normal density in place of the normal as the first term of the Edgeworth expansion. For example Em-brechts et al (1985) and Jensen (1991, 1995) proposed the use of gamma distributions. The motivation for using the gamma distribution was to more accurately approximate distributions which are asymptotically gamma in the extreme tail. McCullagh (1987, Chapter 5) gives general formulae for Edge-worth type expansions with an arbitrary smooth density as the first term.

In this paper we note that the count distribution of a birth process follows a negative binomial distribution when the birth rates are linear increasing.

This allows us to derive the waiting time distribution for any birth rate sequence which is a permutation of a linear increasing sequence. This distribution is then used as the leading term for the saddlepoint approximation for a general birth rate process. The result is exact not only in the negative binomial case, but also in the Poisson case (constant birth rates), binomial case (linear decreasing rates) and any case when the birth rates can be reordered in a linear sequence. Even when the birth rates are not evenly spaced, the negative binomial based saddlepoint approximation is often considerably more accurate than the usual saddlepoint approximation.

Although the saddlepoint method has been very successful in practice, it is difficult to compute analytic error bounds for the approximations. For the same reasons it is difficult to make a general analytic comparison between the usual saddlepoint approximation and that based on the negative binomial distribution. In this paper we instead demonstrate the superiority of the new approximation empirically using the same examples as given by Daniels (1982). One special case in which superiority can be proved analytically is that in which the birth process being examined lies in a neighbourhood of one of the exact cases referred to above. Such applications arise naturally for example when using birth processes to extend standard regression models for categorical data as in Faddy (1997a, b), Faddy and Bosch (1999) or Smyth and Podlich (2000).

In Section 2 we derive a generalized form of the saddlepoint approximation for birth processes. Although not unexpected, the general approximation with correction term that we give at the end of Section 2 appears to be new. The approximation based on the negative binomial distribution is developed in Section 3. Numerical comparisons with the standard saddlepoint approximation are given in Section 4.

2 Saddlepoint Approximations

Consider a pure birth process with birth rates λ_i, $i = 0, 1, \ldots$. Here λ_i is the birth rate when i individuals are present, and we assume no individuals present at time zero. Let $p_n(t)$ be the probability mass function for the number of births by time t, $n = 0, 1, \ldots$.

Let $f_n(t)$ be the density function of the waiting time until the nth birth. Daniels (1982) pointed out that

$$f_{n+1}(t) = \lambda_n p_n(t).$$

Daniels (1982) used the saddlepoint technique to invert the Laplace transformation of $p_n(t)$. The same approximation can be derived by inverting the moment generating function of $f_{n+1}(t)$,

$$M(s) = \prod_{i=0}^{n} \frac{\lambda_i}{\lambda_i - s}.$$

Let $f(t,\theta)$ be the exponential family density generated by $f_{n+1}(t)$,

$$f(t,\theta) = f_{n+1}(t)\exp\{t\theta - K(\theta)\}$$

where $K(\theta) = \log M(\theta)$ is the cumulant generating function of $f_{n+1}(t)$. The exponential family distribution has cumulants

$$\kappa_r(\theta) = K^{(r)}(\theta) = (r-1)!\sum_{i=0}^{n}\frac{1}{(\lambda_i - \theta)^r}.$$

Let $\tilde{\theta}$ be the value of θ which solves $\kappa_1(\theta) = t$. We can write

$$f_{n+1}(t) = f(t,\tilde{\theta})\exp\{K(\tilde{\theta}) - t\tilde{\theta}\}.$$

The ordinary saddlepoint approximation consists of replacing $f(t,\tilde{\theta})$ by its normal approximation, i.e.,

$$\tilde{f}_{n+1}(t) = \{2\pi\kappa_2(\tilde{\theta})\}^{-1/2}\exp\{K(\tilde{\theta}) - t\tilde{\theta}\}.$$

This approximation has high relative accuracy because the normal approximation is used only at the mean of the approximated distribution. The approximation for $p_n(t)$ is finally

$$\tilde{p}_n(t) = \tilde{f}_{n+1}(t)/\lambda_n = \{2\pi\kappa_2(\tilde{\theta})\}^{-1/2}\exp\{H(t,\boldsymbol{\lambda})\}$$

with

$$H(t,\boldsymbol{\lambda}) = -t\tilde{\theta} + \sum_{i=0}^{n-1}\log\lambda_i - \sum_{i=0}^{n}\log(\lambda_i - \tilde{\theta}).$$

Note that this final approximation can be used even for $\lambda_n = 0$.

If the λ_i have bounded range then the Central Limit Theorem applies to the waiting time $f_{n+1}(t)$ and hence also to $f(t,\theta)$ and the relative error of the saddlepoint approximation is $O(n^{-1})$,

$$p_n(t) = \tilde{p}_n(t)\left\{1 + O(n^{-1})\right\}.$$

A higher order approximation can be obtained by including adjustments for the third and fourth cumulants of $f(t,\theta)$ (Daniels, 1982; McCullagh, 1987). If the Central Limit Theorem applies then $\kappa_r(\theta)/\kappa_2(\theta)^{1/2} = O(n^{-r/2-1})$, $r \geq 3$, for any θ. The second and generally more accurate approximation is

$$p_n(t) = \tilde{p}_{n+1}(t)\left\{1 + \frac{1}{8}\frac{\kappa_4}{\kappa_2^2} - \frac{5}{24}\frac{\kappa_3^2}{\kappa_2^3} + O(n^{-2})\right\}$$

where here and in the following $\kappa_r = \kappa_r(\tilde{\theta})$. On taking logarithms we also have

$$\log p_n(t) = -\frac{1}{2}\log\{2\pi\kappa_2\} + H(t,\boldsymbol{\lambda}) + \frac{1}{8}\frac{\kappa_4}{\kappa_2^2} - \frac{5}{24}\frac{\kappa_3^2}{\kappa_2^3} + O(n^{-2}) \qquad (1)$$

and this form of the expansion is usually slightly preferable because it in principle preserves positivity of the density.

Although the use of the normal distribution as the leading term in the saddlepoint approximation is usual, any reasonable approximation can in principle be used for $f(t, \tilde{\theta})$. If $f(t, \tilde{\theta})$ follows a known distributional form other than the normal in some limit then an improved approximation can often be obtained by using the alternative distribution. We know for example that $f(t, \theta)$ converges to a gamma density as the range of the λ_i tends to zero. Let g be an arbitrary density with mean t and variance κ_2. An alternative saddlepoint approximation is

$$\tilde{f}_{n+1}(t) = g(t) \exp\{K(\tilde{\theta}) - t\tilde{\theta}\}.$$

Write the cumulants of g as k_r. If $g(t)$ itself can be approximated to $O(n^{-2})$ using the second order saddlepoint approximation based on the normal distribution, then we obtain the generalized saddlepoint approximation

$$\log p_n(t) = \log g(t) + H(t, \boldsymbol{\lambda}) + \frac{1}{8}\frac{\kappa_4 - k_4}{\kappa_2^2} - \frac{5}{24}\frac{\kappa_3^2 - k_3^2}{\kappa_2^3} + O(n^{-2}) \qquad (2)$$

where again $\kappa_r = \kappa_r(\tilde{\theta})$.

3 Linear Birth Rate Sequences

Suppose that the birth rates form a linear increasing sequence, $\lambda_i = \lambda + ai$ for $i = 0, \ldots, n$ and $\lambda > 0$. If λ/a is a positive integer then the number of births follows a negative binomial distribution with probability mass function

$$p_n(t) = \binom{\lambda/a + n - 1}{n} e^{-\lambda t}(1 - e^{-at})^n = \frac{e^{-\lambda t}}{n!}\left(\frac{1 - e^{-at}}{a}\right)^n \prod_{i=0}^{n-1}(\lambda + ai).$$

It follows that

$$f_{n+1}(t) = \frac{e^{-\lambda t}}{n!}\left(\frac{1 - e^{-at}}{a}\right)^n \prod_{i=0}^{n}(\lambda + ai).$$

Now define

$$h(a, t) = \begin{cases} (1 - e^{-at})/a & a \neq 0 \\ t & a = 0 \end{cases}$$

Then $h(a, t)$ is continuous in both a and t and the expression

$$f_{n+1}(t) = \frac{e^{-\lambda t}}{n!} h(a, t)^n \prod_{i=0}^{n}(\lambda + ai) \qquad (3)$$

is valid for general a. As $a \to 0$, $f_{n+1}(t)$ converges to a gamma density and hence we recover the Poisson distribution with mean λ for $p_n(t)$. For λ/a a

negative integer we recover the binomial distribution for $p_n(t)$ with size $-\lambda/a$ and probability $1 - e^{-at}$. Although the negative binomial and binomial laws for the count distributions require that λ/a be an integer, the expression (3) for the waiting time density remains valid for any a and λ such that $\lambda + ai \geq 0$ for $i = 0, \ldots, n$.

The negative binomial saddlepoint approximation is defined as follows. Let the λ_i be a general, not necessarily linear, sequence of birth rates. Let $\tilde{\theta}$ satisfy

$$\sum_{i=1}^{n} \frac{1}{\lambda_i - \tilde{\theta}} = 1.$$

Find $\lambda > 0$ and $a \geq 0$ such that

$$\sum_{i=0}^{n} \frac{1}{\lambda + ai} = 1$$

and

$$\sum_{i=0}^{n} \frac{1}{(\lambda + ai)^2} = \sum_{i=0}^{n} \frac{1}{(\lambda_i - \tilde{\theta})^2}.$$

It is shown in the Appendix that there exists a unique solution for λ and a for any sequence of non-negative λ_i. The approximation is then given by (2) with

$$g(t) = \frac{e^{-\lambda t}}{n!} h(a, t)^n \prod_{i=0}^{n} (\lambda + ai),$$

$$k_3 = 2! \sum_{i=0}^{n} \frac{1}{(\lambda + ai)^3}$$

and

$$k_4 = 3! \sum_{i=0}^{n} \frac{1}{(\lambda + ai)^4}.$$

Note that λ/a is not restricted to be integer in this approximation.

In order to avoid numerical instability in limited precision arithmetic the function $h(a, t)$ can be computed as

$$h(a, t) \approx \begin{cases} t - at^2/2 + a^2t^3/6 & at < 10^{-5} \\ (1 - e^{-at})/a & at \geq 10^{-5} \end{cases}$$

This yields $h(a, t)$ to nearly full accuracy in standard 64-bit double precision arithmetic.

Table 1: Exact probabilities and percentage errors for saddlepoint approximations for the simple epidemic with $\lambda_n = (n+1)(5-n)$ and $t = 0.1$.

	Prob.	Normal		Gamma		Neg. Binomial	
n	$\times 100$	First	Second	First	Second	First	Second
0	60.6531	8.4438	-0.2269	0.0000	0.0000	0.0000	0.0000
1	26.2003	4.1258	0.0100	-0.1137	0.0431	0.0000	0.0000
2	9.6413	2.7611	0.0062	-0.0534	0.0163	-0.0019	0.0022
3	2.8407	2.0846	0.0003	-0.0190	0.0046	-0.0009	0.0008
4	0.5973	1.6686	-0.0002	-0.0115	0.0020	-0.0001	0.0001
5	0.0674	1.3752	0.0032	-0.0256	0.0045	-0.0019	0.0011

4 Numerical Calculations

Daniels (1982) gave numerical calculations for a simple epidemic with $\lambda_n = (n+1)(N-n)$. We consider the same epidemics with $N = 5$ (Table 1 and 2) and $N = 10$ (Tables 3 and 4). Results are given for the usual saddlepoint approximation based on the normal distribution and for the negative binomial saddlepoint described in the previous section. Results are also given for the saddlepoint approximation with a gamma density as leading term. The approximations are based on the expansions for the log-densities without renormalisation. In each case the tables give the percentage errors for the approximations with and without the $O(n^{-1})$ correction term. Exact probabilities were computed numerically using the method of Sidje (1998).

The results show that the first term of the negative binomial approximation is considerably more accurate than the first term of the normal approximation. The gamma approximation is also more accurate than the normal when the range of $\lambda_n t$ is not large. The correction term improves the normal approximation very much more than it improves the gamma or negative binomial approximations. This is to be expected as the correction is obtained from Hermite polynomials and the normal density. Nevertheless the negative binomial approximation remains more accurate than the normal except in isolated cases.

With $N = 5$ and $t = 0.1$ the maximum relative error is reduced from 0.2% for the usual second-order saddlepoint approximation to 0.002% for the negative binomial saddlepoint approximation with correction term. With $t = 1$ the maximum relative error is reduced from 1% to 0.45%. With $N = 10$ and $t = 0.1$ the maximum relative error is reduced from 0.6% to 0.02%. With $t = 1$ the maximum relative error is reduced from 2% to 0.8%. The minimum relative error is reduced to zero in each case, because the negative binomial based approximation is always exact for $n = 0$ and $n = 1$.

Further experiments show that, although there are isolated values of n for which the normal saddlepoint is more accurate that the negative binomial

Table 2: Exact probabilities and percentage errors for saddlepoint approximations for the simple epidemic with $\lambda_n = (n+1)(5-n)$ and $t = 1$.

n	Prob. $\times 100$	Normal		Gamma		Neg. Binomial	
		First	Second	First	Second	First	Second
0	0.6738	8.4438	−0.2269	0.0000	0.0000	0.0000	0.0000
1	1.0671	3.0869	−1.1499	−2.2276	−1.0836	0.0000	0.0000
2	1.9221	0.5093	−0.8416	−3.1423	−0.8183	−0.2822	−0.1687
3	4.2854	−0.1543	0.1389	−2.6627	0.1467	−0.5359	0.0199
4	13.8699	0.9616	0.0061	−0.8884	0.0091	0.1471	−0.1589
5	78.1817	−2.0778	−0.6616	−4.1672	−0.6569	−1.0578	−0.4482

Table 3: Exact probabilities and percentage errors for saddlepoint approximations for the simple epidemic with $\lambda_n = (n+1)(10-n)$ and $t = 0.1$.

n	Prob. $\times 100$	Normal		Gamma		Neg. Binomial	
		First	Second	First	Second	First	Second
0	36.7879	8.4438	−0.2269	0.0000	0.0000	0.0000	0.0000
1	25.3226	3.6496	0.1279	−0.6979	0.1641	0.0000	0.0000
2	16.5760	2.3281	0.1347	−0.5692	0.1459	−0.0193	0.0169
3	10.2136	1.7554	0.0942	−0.4044	0.0989	−0.0285	0.0223
4	5.8450	1.4430	0.0593	−0.2724	0.0616	−0.0288	0.0197
5	3.0508	1.2450	0.0349	−0.1749	0.0363	−0.0238	0.0143
6	1.4157	1.1039	0.0191	−0.1062	0.0199	−0.0165	0.0087
7	0.5623	0.9928	0.0094	−0.0614	0.0100	−0.0092	0.0043
8	0.1797	0.8972	0.0044	−0.0376	0.0048	−0.0039	0.0016
9	0.0412	0.8083	0.0032	−0.0328	0.0035	−0.0021	0.0008
10	0.0051	0.7187	0.0042	−0.0475	0.0045	−0.0064	0.0012

Table 4: Exact probabilities and percentage errors for saddlepoint approximations for the simple epidemic with $\lambda_n = (n+1)(10-n)$ and $t = 1$.

	Prob.	Normal		Gamma		Neg. Binomial	
n	$\times 100$	First	Second	First	Second	First	Second
0	0.0045	8.4438	-0.2269	0.0000	0.0000	0.0000	0.0000
1	0.0057	6.7814	-0.5969	-0.0199	-0.4720	0.0000	0.0000
2	0.0073	5.9009	-0.7889	0.0020	-0.7051	0.0639	0.0270
3	0.0097	5.1804	-0.9517	-0.0505	-0.8921	0.1755	0.0681
4	0.0136	4.4635	-1.1187	-0.1985	-1.0756	0.3256	0.1121
5	0.0204	3.6457	-1.3128	-0.4836	-1.2822	0.4955	0.1369
6	0.0338	2.5798	-1.5661	-1.0064	-1.5455	0.6236	0.0849
7	0.0671	0.9542	-1.9275	-2.0227	-1.9152	0.4915	-0.1996
8	0.1923	-1.7380	-2.0998	-3.9428	-2.0944	-0.4789	-0.8457
9	1.4904	-0.1964	-0.4272	-1.6350	-0.4258	1.4494	-0.4677
10	98.1552	0.1840	-2.0674	-2.2964	-2.0601	0.4759	-0.3659

based saddlepoint, the latter is overall substantially more accurate.

Acknowledgement

Appendix

Suppose that $\lambda_0, \ldots, \lambda_n$, $n \geq 1$, are non-negative and that $\sum_{i=0}^{n} 1/(\lambda_i - \tilde{\theta}) = 1$. Let $\kappa_2 = \sum_{i=0}^{n} 1/(\lambda_i - \tilde{\theta})^2$. Then $1/(n+1) \leq \kappa_2 \leq 1$. We seek λ and a such that $\sum_{i=0}^{n} 1/(\lambda + ai) = 1$ and $\sum_{i=0}^{n} 1/(\lambda + ai)^2 = \kappa_2$. If the λ_i are all equal then the solution is $a = 0$ and $\lambda = \lambda_0$. If the λ_i are not all equal then $a \neq 0$ and $b = \lambda/a$ is finite. Then a can be solved in terms of b as

$$a = \sum_{i=0}^{n} 1/(b+i)$$

and b solves

$$\frac{\sum_{i=1}^{n} 1/(b+i)^2}{\left\{\sum_{i=0}^{n} 1/(b+i)\right\}^2} = \kappa_2. \tag{4}$$

The left-hand side of (4) tends to 1 as $b \to 0$ and to $1/(n+1)$ as $b \to \infty$ and is monotonic decreasing between these two values. There is therefore a unique non-negative solution for b and hence for a. The left-hand of (4) is

in fact convex in b so that Newton's method for finding b is monotonically convergent from a suitable starting value.

References

Daniels, H. E. (1954). Saddlepoint approximations in statistics. *Annals of Mathematical Statistics* **25**, 631–650.

Daniels, H. E. (1982). The saddlepoint approximation for a general birth process. *Journal of Applied Probability* **19**, 20–28.

Embrechts, P., Jensen, J. L., Maejima, M., and Teugels, J. L. (1985). Approximations for compound Poisson and Polya processes. *Advances of Applied Probability* **17**, 623–637.

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. *Math. Scientist.* **22**, 77–82.

Faddy, M. J. and Bosch, R.J. (1999). Likelihood based modelling and analysis of data under-dispersed relative to the Poisson distribution. *Biometrics.* To appear.

Grassman, W. (1977). Transient solutions in Markovian queuing systmes. *Computers and Operations Research* **4**, 47–66.

Jensen, J. L. (1991). Uniform saddlepoint approximations and log-concave densities. *Journal Royal Statistical Society* B **53**, 157–172.

Jensen, J. L. (1991). A large deviation-type approximation for the "Box class' of likelihood ratio criteria. *Journal of the American Statistical Society* **86**, 437–440.

Jensen, J. L. (1995). *Saddlepoint Approximations.* Oxford University Press, New York.

McCullagh, P. (1987). *Tensor Methods in Statistics.* Chapman and Hall, London.

Podlich, H. M., Faddy, M. J., and Smyth, G. K. (1999a). Likelihood computations for extended Poisson process models. *InterStat*, September #1.

Podlich, H. M., Faddy, M. J., and Smyth, G. K. (1999b). Semi-parametric extended Poisson process models. Research Report 23, Department of Statistics and Demography, University of Southern Denmark, Odense.

Reid, N. (1988). Saddlepoint methods and statistical inference. With comments and a rejoinder by the author. *Statistical Science* **3**, 213–238.

Sidje, R. B. (1998). EXPOKIT software package for computing matrix exponentials. *ACM Transactions on Mathematical Software* **24**, 130–156.

Smyth, G. K., and Podlich, H. M. (2000). Score tests for Poisson variation against general alternatives. *Computing Science and Statistics* **33**. To appear.

Stewart, W. J. (1994). *Introduction to the Numerical Solutions of Markov chains.* Princeton University Press, Princeton, New Jersey.

Stewart, W. J. (ed.) (1995). *Computations with Markov Chains: Proceedings of the 2nd International Workshop on the Numerical Solution of Markov Chains.* Kluwer Academic Publishers, Norwell, Massachusetts.