An improved saddlepoint approximation

Article *in* Mathematical Biosciences · September 2007

DOI: 10.1016/j.mbs.2006.08.026 · Source: PubMed

2 authors, including:

Eric Renshaw
University of Strathclyde
114 PUBLICATIONS **2,928** CITATIONS

Available online at www.sciencedirect.com
ScienceDirect
Mathematical Biosciences 208 (2007) 359–374
Mathematical Biosciences
www.elsevier.com/locate/mbs

An improved saddlepoint approximation

Colin S. Gillespie [a], Eric Renshaw [b,*]

[a] School of Mathematics and Statistics, University of Newcastle, Newcastle upon Tyne, NE1 7RU, UK
[b] Department of Statistics and Modelling Science, Livingstone Tower, University of Strathclyde, 26 Richmond Street, Glasgow G1 1XH, UK

Received 19 May 2005; received in revised form 15 February 2006; accepted 29 August 2006
Available online 9 September 2006

Abstract

Given a set of third- or higher-order moments, not only is the saddlepoint approximation the only realistic 'family-free' technique available for constructing an associated probability distribution, but it is 'optimal' in the sense that it is based on the highly efficient numerical method of steepest descents. However, it suffers from the problem of not always yielding full support, and whilst [S. Wang, General saddlepoint approximations in the bootstrap, Prob. Stat. Lett. 27 (1992) 61.] neat scaling approach provides a solution to this hurdle, it leads to potentially inaccurate and aberrant results. We therefore propose several new ways of surmounting such difficulties, including: extending the inversion of the cumulant generating function to second-order; selecting an appropriate probability structure for higher-order cumulants (the standard moment closure procedure takes them to be zero); and, making subtle changes to the target cumulants and then optimising via the simplex algorithm.

Keywords: Cumulants; Generating functions; Moment closure; Population dynamics; Simplex; Truncation

1. Introduction

Although stochastic population models [19] have proved to be a powerful tool in the study in process generating mechanisms across a wide range of disciplines, all too often the associated

* Corresponding author. Fax: +141 552 2079.
E-mail address: eric@stams.strath.ac.uk (E. Renshaw).

0025-5564/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.mbs.2006.08.026

mathematical development involves non-linear mathematics which immediately raises difficult and challenging analytic problems if useful progress is to be made. Whilst linearising such systems will render them tractable to direct solution, ignoring non-linear characteristics of the model is highly likely to have a serious impact on long term behaviour. Invoking nonlinear approximation techniques is certainly a possibility, as illustrated by Marion et al. [13] implementation of the Tan and Hsu [24] procedure, but as this may in itself inject spurious behaviour into the system such approaches must usually be viewed with caution.

The essence of the problem lies in the general intractability of the forward (and backward) Kolmogorov partial differential equations for the moment generating function (m.g.f.)

$$M(\theta;t) \equiv \sum_{n=0}^{\infty} p_n(t)\mathrm{e}^{n\theta} \tag{1.1}$$

for the population size probabilities $p_n(t)$ at time t. In some situations, such as queueing [1] and birth–death [22] type applications, or multiple immigration-death paradigms in quantum optics [6,14,11] the underlying process is sufficiently tractable to enable the construction of exact expressions for the $p_n(t)$. However, this is the exception rather than the norm, and in many biological and ecological situations is rarely possible. However, provided the probability transition rates are polynomial in form, rather than being, for example, fractional, then limited progress in the development of moment expressions can be made by working in terms of the cumulants $\kappa_i(t)$ $(i>0)$ of the process through the cumulant generating function (c.g.f.)

$$K(\theta;t) \equiv \ln[M(\theta;t)] \equiv \sum_{i=1}^{\infty} \frac{\kappa_i(t)\theta^i}{i!}. \tag{1.2}$$

This representation carries a considerable advantage over the alternative probability, moment and factorial moment generating functions, since for small $i>3$ the $\kappa_i(t)$ require little algebraic manipulation in order to turn them into central moments, whilst $\kappa_1(t)$, $\kappa_2(t)$ and $\kappa_3(t)$ denote the mean, variance and skewness, directly.

Motivated by a desire to model both the annual catch of an invasion of muskrats in eleven Dutch provinces between 1968 and 1991 [10], and the rapid colonisation of the Africanised honey (Killer) bees of North and South America [16], Matis et al. [15] provide an excellent example of the practical implementation of this procedure. They base their analysis on the power-law logistic process with population birth and death rates

$$\lambda_N = a_1 N - b_1 N^{s+1}, \qquad \mu_N = a_2 N + b_2 N^{s+1} \tag{1.3}$$

for a_i, $b_i \geqslant 0$, integer $s \geqslant 1$ and population size $N = 0,1,2,\ldots,\mathrm{int}\{(a_1/b_1)^{1/s}\}$. For the basic logistic (i.e., $s=1$) case the forward differential equation for the c.g.f. is given by

$$\frac{\partial K}{\partial t} = [a_1(\mathrm{e}^{\theta}-1) + a_2(\mathrm{e}^{-\theta}-1)]\frac{\partial K}{\partial \theta} + [b_1(1-\mathrm{e}^{\theta}) + b_2(\mathrm{e}^{-\theta}-1)]\left[\frac{\partial^2 K}{\partial \theta^2} + \left(\frac{\partial K}{\partial \theta}\right)^2\right]. \tag{1.4}$$

Whence on successively differentiating (1.4) with respect to θ, and then placing $\theta = 0$, we obtain the system of cumulant equations

$$\begin{aligned}
\mathrm{d}\kappa_1(t)/\mathrm{d}t &= [a - b\kappa_1(t)]\kappa_1(t) - b\kappa_2(t) \\
\mathrm{d}\kappa_2(t)/\mathrm{d}t &= [c - d\kappa_1(t)]\kappa_1(t) + [2a - d - 4b\kappa_1(t)]\kappa_2(t) - 2b\kappa_3(t) \\
\mathrm{d}\kappa_3(t)/\mathrm{d}t &= [a - b\kappa_1(t)]\kappa_1(t) + [3c - b - 6d\kappa_1(t) - 6b\kappa_2(t)]\kappa_2(t) \\
&\quad + [3a - 3d - 6b\kappa_1(t)]\kappa_3(t) - 3b\kappa_4(t), \text{ etc.},
\end{aligned} \tag{1.5}$$

for $a = a_1 - a_2$, $b = b_1 + b_2$, $c = a_1 + a_2$ and $d = b_1 - b_2$. Exact solutions of these equations are not available, as the equation for any specific jth order cumulant function involves terms up to the $(j+1)$th order. For general s, terms up to the $(j+s)$th order are involved. Matis and Kiffe [16] therefore adopt a moment closure approach by solving the system for the first j cumulant equations with $\kappa_i(t) \equiv 0$ for all $i > j + s$.

Truncating the equations in this manner raises two non-trivial questions. The first [15] relates to the accuracy of the resulting cumulant estimates. The second [20,21] is that assuming that only the first $(j+s)$ cumulants are known accurately, what error is induced in the underlying probability structure by taking all higher-order cumulants to be zero? This is addressed by considering the truncated c.g.f.

$$K_n(\theta) \equiv \sum_{i=1}^{n} \frac{\kappa_i \theta^i}{i!} \tag{1.6}$$

in tandem with the associated saddlepoint approximation [17]. Daniels [2] provides a superb account of its derivation; the approach essentially involves taking the dominant term in the contour-integration formula for the inversion of the c.g.f. $K(\theta)$ corresponding to the p.d.f. $h(x)$. The key result is that for θ_0 an appropriate root of the saddlepoint equation

$$x = K'(\theta_0) \tag{1.7}$$

we have the approximation

$$h(x) \simeq f(x) = [2\pi K''(\theta_0)]^{-1/2} \exp\{K(\theta_0) - \theta_0 x\}. \tag{1.8}$$

The power of this approach can be seen immediately on noting that it not only reproduces the Normal p.d.f. exactly, but that the saddlepoint approximation for the gamma p.d.f. differs only from the exact result in that $\Gamma(\alpha)$ is replaced by Stirling's approximation in the normalising factor. Furthermore, the relative error is of order $O(n^{-1})$. This error cannot be obtained by using techniques such as Edgeworth expansions.

Easton and Ronchetti [3] propose a corresponding truncated approximation $f_n(x)$. This approximation is obtained by replacing $K(\theta)$ by $K_n(\theta)$ in (1.7) and (1.8), thereby yielding

$$f_n(x) = \left(2\pi \sum_{i=0}^{n-2} \frac{\kappa_{i+2}\theta_0^i}{i!}\right)^{-1/2} \exp\left\{\sum_{i=1}^{n} \frac{\kappa_i \theta_0^i}{i!} - \theta_0 x\right\}, \tag{1.9}$$

where

$$x = \sum_{i=0}^{n-1} \frac{\kappa_{i+1}\theta_0^i}{i!}. \tag{1.10}$$

This is a completely general representation and is extremely useful when we wish to examine the structure of a p.d.f. corresponding to a given finite set of cumulants. However, it does suffer from a possible lack of support in the tail regions [20]. This paper therefore explores possible methods for overcoming this problem, and also investigates whether the basic saddlepoint approximation can be improved still further.

2. Second-order improvements

One obvious improvement is to include the second term in the contour integration for the inversion of $K(\theta)$. This yields [2] the second-order approximation

$$g(x) = f(x)\left(1 + \frac{1}{8}\frac{K^{iv}(\theta_0)}{K''(\theta_0)^2} - \frac{5}{24}\frac{K'''(\theta_0)^2}{K''(\theta_0)^3}\right). \tag{2.1}$$

To investigate the merits of this representation, let us consider the Poisson (α) and negative binomial (r,p) distributions. First, as the Poisson c.g.f.

$$K^{\mathrm{P}}(\theta) = \alpha(\mathrm{e}^{\theta} - 1) = \alpha\sum_{i=1}^{\infty}\frac{\theta^i}{i!} \tag{2.2}$$

we see that all the cumulants $\kappa_i \equiv \alpha$, and so $f(x)$ and $g(x)$ take the form

$$f^{\mathrm{P}}(x) = \frac{\alpha^x \mathrm{e}^{x-\alpha}}{\sqrt{2\pi x}x^x} \quad \text{and} \quad g^{\mathrm{P}}(x) = f^{\mathrm{P}}(x)\left(1 - \frac{1}{12x}\right) \quad \text{for } x = 1, 2, \ldots . \tag{2.3}$$

These approximations clearly fail at $x = 0$, though this is not surprising. For Daniels [2] proves that, for a distribution with support on (a,b), where $a < b$ and a, b or both may be $\pm\infty$, when $x \to a,b$ then $K'(\theta_0) \to a,b$ and so $K''(\theta_0) \to 0$. Thus as $K(\theta_0) - \theta_0 x$ is bounded, it follows from (1.8) that $f(x) \to \infty$.

A way of overcoming this problem is to define

$$S \equiv \sum_{x=1}^{\infty} f^{\mathrm{P}}(x). \tag{2.4}$$

For since S will, in general, not equal one, we may simply take:

$$\begin{aligned} &f^{\mathrm{P}}(0) = 1 - S \quad \text{for } S \leqslant 1; \\ &f^{\mathrm{P}}(0) = 0, \quad \text{and replace} \quad f^{\mathrm{P}}(x) \text{ by } f^{\mathrm{P}}(x)/S \text{ otherwise.} \end{aligned} \tag{2.5}$$

A similar approach can be applied to $g^{\mathrm{P}}(x)$. For example, when $\alpha = 1$ the values $f^{\mathrm{P}}(0) = 0.32695$, and especially $g^{\mathrm{P}}(0) = 0.37032$ (i.e., $S < 1$ in both cases), compare well with the exact probability 0.36788. Whilst when $\alpha = 10$, $f^{\mathrm{P}}(0) = 0$ ($S > 1$) and $g^{\mathrm{P}}(0) = 0.0001069$ ($S < 1$) lie equidistant from the true probability 0.000045.

In the case of the negative binomial example, we note that the c.g.f.

$$K(\theta) = r\ln\left(\frac{p\mathrm{e}^{\theta}}{1 + (p-1)\mathrm{e}^{\theta}}\right), \tag{2.6}$$

which, for $x = r+1, r+2, \ldots$, gives rise to the first-order saddlepoint approximation

$$f^{\mathrm{NB}}(x) = \frac{p^r}{\sqrt{2\pi(x-r)}}\left(\frac{x}{r}\right)^{r/2}\left(\frac{x-r}{x(1-p)}\right)^{r-x} \tag{2.7}$$

and the second-order approximation

$$g^{\mathrm{NB}}(x) = f^{\mathrm{NB}}(x)\left(1+\frac{(r-x)^2+rx}{12rx(r-x)}\right). \tag{2.8}$$

These approximations again fail at the ends of the support region, since $f^{\mathrm{NB}}(x)$ and $g^{\mathrm{NB}}(x) \to \infty$ as $x \downarrow r$. Though this problem is easily avoided if we implement the same procedure (2.5) as used for the Poisson example.

Although the Poisson *probabilities* can be accurately represented by the saddlepoint approximation [20,8], an equally relevant question concerns the accuracy to which the *cumulants* of the saddlepoint approximation reflect those of the original distribution? To answer this, we begin by defining the relative difference

$$\widetilde{\kappa}_i = 100(\hat{\kappa}_i - \kappa_i)/\kappa_i, \tag{2.9}$$

where κ_i denotes the ith cumulant of the target distribution and $\hat{\kappa}_i$ the ith cumulant of the approximation. Now we see from the Poisson entries in Table 1 that for $\alpha = 1$, and hence $\kappa_i \equiv 1$, the $\hat{\kappa}_i$ do not match κ_i particularly well, with large errors of 25%, 63% and 502% for the fifth, sixth and eighth cumulants, respectively. In contrast, when $\alpha = 10$, as the Poisson distribution can now be approximated quite well by the Normal distribution (i.e., the saddlepoint approximation based on κ_1 and κ_2), the estimated cumulants $\hat{\kappa}_1, \ldots, \hat{\kappa}_6$ are more accurate, though the same cannot be said for higher-order cumulants. Moreover, including the extra term, $g^{\mathrm{P}}(x)$, yields an improvement in the accuracy of only the first two cumulants. This demonstrates that using a higher-order saddlepoint procedure does not necessary improve the accuracy of the estimated distribution. Note that both $f^{\mathrm{P}}(x)$ and $g^{\mathrm{P}}(x)$ 'explode' once we reach the eighth cumulant, which highlights

Table 1
Relative differences (2.9) between the cumulants of the Poisson and Negative Binomial distributions and their first-order, $f(x)$, and second-order, $g(x)$, saddlepoint approximations

	Poisson				Negative Binomial (geometric)			
	$\alpha = 1$		$\alpha = 10$		$r = 1$ and $p = 0.4$		$r = 1$ and $p = 0.8$	
	$f^{\mathrm{P}}(x)$	$g^{\mathrm{P}}(x)$	$f^{\mathrm{P}}(x)$	$g^{\mathrm{P}}(x)$	$f^{\mathrm{NB}}(x)$	$g^{\mathrm{NB}}(x)$	$f^{\mathrm{NB}}(x)$	$g^{\mathrm{NB}}(x)$
$\widetilde{\kappa_1}$	5.33	−0.274	−0.104	−0.004	5.81	−0.453	2.34	−0.214
$\widetilde{\kappa_2}$	−2.52	0.204	0.103	0.049	2.24	−0.152	8.21	−0.712
$\widetilde{\kappa_3}$	−8.06	0.518	0.108	−0.42	−0.359	0.0649	3.89	−0.268
$\widetilde{\kappa_4}$	−2.41	−0.006	−1.960	2.82	−0.519	0.0691	0.283	0.098
$\widetilde{\kappa_5}$	25.50	−2.28	12.09	−14.41	−0.322	0.0409	−1.45	0.263
$\widetilde{\kappa_6}$	63.50	−5.04	−38.61	39.45	−0.201	0.0233	−1.69	0.266
$\widetilde{\kappa_7}$	−2.32	4.01	−79.64	126.39	−0.137	0.0146	−1.35	0.206
$\widetilde{\kappa_8}$	502.10	62.04	−1573	−1888	−0.099	0.010	−0.99	0.14

the danger in thinking that 'more cumulants necessarily equals better' (see [4] , for a similar example).

Further inspection of Table 1 shows that in terms of cumulant accuracy the approximation to the heavy-tailed negative binomial distribution with $r = 1$ (i.e., geometric) fares substantially better than that for the Poisson distribution. Moreover, we also see from Fig. 1 that it also captures the probabilities quite satisfactorily, with only slight errors apparent for low population numbers. However, in spite of these similarities, some of the differences exhibited between the binomial and saddlepoint cumulants are disturbing. For in both the $p = 0.4$ and 0.8 cases $f^{\mathrm{NB}}(x)$ exhibits substantial errors in the mean, variance and eighth cumulant; though κ_3 to κ_8 are estimated with an error rate of less then 2%. Including the additional term in the saddlepoint approximation, to form $g^{\mathrm{NB}}(x)$, improves the accuracy of all the tabulated cumulants. The reason the saddlepoint approximation fails to capture all the cumulants well can be illustrated by noting that the equation $K'(\theta_0) = \kappa_1$ always takes the solution $\theta_0 \equiv 0$. So it follows from (1.8) that at $x = \kappa_1$

$$f(\kappa_1) = \frac{1}{\sqrt{2\pi\kappa_2}}, \tag{2.10}$$

which means that irrespective of the number of cumulants employed the saddlepoint approximation at $x = \kappa_1$ can never be improved on that given by the Normal approximation. However, on considering expression (2.1), we see that the second-order value

$$g(\kappa_1) = f(\kappa_1)\left(1 + \frac{\kappa_4}{8\kappa_2^2} - \frac{5\kappa_3^2}{24\kappa_2^3}\right) = \frac{1}{\sqrt{2\pi\kappa_2}}\left(1 + \frac{\kappa_4}{8\kappa_2^2} - \frac{5\kappa_3^2}{24\kappa_2^3}\right). \tag{2.11}$$

This suggests that inclusion of higher-order terms in the inversion of $K(\theta)$ should (generally) result in a better approximation, since these utilise more information. Indeed, since the saddlepoint approximation at x is always expressed in terms of the leading term (i.e., the Normal approximation) of the tilted Edgeworth expansion at its mean, it always expresses the accuracy of the Edgeworth expansion at the mean which is why the saddlepoint approximation achieves such high accuracy.

It must be stressed that the higher-order cumulants are *very* sensitive to the mode of calculation. First, there is extreme sensitivity in regard to whether the scaling procedure (2.5) is used. For example, $f^{\mathrm{P}}(x)$ for $\alpha = 10$ requires scaling since $S = 1.0094 \neq 1$. If we did not employ rescaling

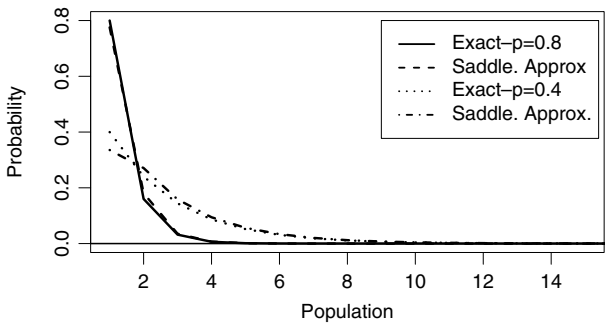

Fig. 1. Comparing the exact p.d.f.'s for the Negative Binomial distribution with $r = 1$, $p = 0.4$ and *0.8* against its saddlepoint approximation, $f^{\mathrm{NB}}(x)$.

then the cumulant errors rise dramatically to $\widetilde{\kappa_5} = 3100$, $\widetilde{\kappa_6} = -18007$, $\widetilde{\kappa_7} = 150,000$ and $\widetilde{\kappa_8} = -1,900,000$. Second, the calculations were performed using MAPLE with 50-digit accuracy; using only 12-digit accuracy produces widely different values for higher-order cumulants. Third, we have illustrated that even though the saddlepoint approximation might accurately reproduce only the first few cumulants of a distribution, it is still able to provide an accurate approximation to the probability structure. Even the exceedingly large errors of κ_8 do not appear to have any noticeable effect. Finally, we note that saddlepoint approximations are low-order approximations recentred at the point of interest. This contrasts with other types of approximations such as Edgeworth or Taylor expansions which are high-order approximations based around a fixed point for all values of the distribution's support. It is with these features in mind that we now proceed to examine the situation where we possess only limited knowledge of the cumulant structure.

3. Approximating a distribution using just two or three cumulants

Given the first three cumulants, the truncated c.g.f. (1.6) becomes

$$K_3(\theta) = \kappa_1\theta + \kappa_2\theta^2/2 + \kappa_3\theta^3/6. \tag{3.1}$$

Thus on solving (1.10), namely $K_3'(\theta_0) = x$, we obtain

$$\theta_0 = (-\kappa_2 + \sqrt{\psi})/\kappa_3 \quad \text{where } \psi(x) = \kappa_2^2 + 2\kappa_3(x - \kappa_1), \tag{3.2}$$

whence the third-order approximation (1.9) takes the simple form

$$f_3(x) = (4\pi^2\psi(x))^{-1/4}\exp\{-(1/6\kappa_3^2)[\kappa_2^3 - 3\kappa_2\psi(x) + 2\psi(x)^{3/2}]\} \tag{3.3}$$

[20]. This is a completely general and algebraically tractable result which provides a significant improvement over the standard Normal approximation if we know the skewness. However, both Easton and Ronchetti [3] and Renshaw [20] note that this approximation can collapse in the lower tails of a distribution, due to the necessity of having $\psi(x) > 0$ in (3.2). This condition is equivalent to

$$x \geqslant \kappa_1 - \frac{\kappa_2^2}{2\kappa_3}, \tag{3.4}$$

and so for large $|\kappa_3|$ the approximation will fail for $x < \kappa_1$.

To obtain full support when only three cumulants are known, Wang [25] proposes multiplying κ_3 by

$$h(\theta) = \exp(-\kappa_2 b^2\theta^2/2) \tag{3.5}$$

thereby yielding

$$K^{3W}(\theta) = \kappa_1\theta + \kappa_2\frac{\theta^2}{2} + \kappa_3\frac{\theta^3}{6}h(\theta), \tag{3.6}$$

where $b > 0$ is an 'appropriately chosen' constant and the subscript 3W denotes third-order truncation under Wang's improvement. The idea is to retain full support by controlling the effect of κ_3 as θ increases. Here $h(\theta)$ serves this purpose well: for $h(0) = 1$, and so the κ_3-term remains virtually

unchanged for small θ; whilst $h(\theta) \to 0$ as θ or $b \to \infty$, whence the saddlepoint approximation reverts to the Normal distribution, for which full support is guaranteed. It is therefore always possible to find a b large enough to allow Wang's approximation to exist. Wang suggests taking $b \geqslant 1/2$, stating that if $b < 1/2$ then the modification has little effect. However, this approach is not easy to defend, since it can be argued that by controlling κ_3 we are simply reducing the skewness to a smaller level. So our saddlepoint approximation will capture this altered κ_3 rather than the target value. Also, as $b \to \infty$ the Normal approximation will be obtained, so we are simply ignoring the κ_3 value altogether!

The approximation (3.6) has been implemented on a number of occasions, for example by Harvill and Newton [9], Gatto [5], Good [7] and Lieberman [12]. However, the values of κ_3 used have generally been small in comparison with the mean and variance. Wang [25], for example, uses $\kappa_1 = 2/7$, $\kappa_2 = 0.005102$ and $\kappa_3 = 9.718 \times 10^{-5}$. We will now demonstrate that when κ_3 is large in comparison to the mean and variance, Wang's approximation presents problems.

As an example of a distribution which has a large third cumulant, suppose we consider the distribution of the simple birth–death process with parameters $\lambda = 1.5$ and $\mu = 1.0$ and initial population size $n_0 = 10$, at time $t = 0.65$. This produces target cumulants of

$$\text{Example A}: \quad \kappa_1 = 13.84, \quad \kappa_2 = 26.57 \quad \text{and} \quad \kappa_3 = 82.87.$$

Let $f^{3W}(x)$ denote the saddlepoint approximation using the first three cumulants and Wang's correction term, and $g^{3W}(x)$ the corresponding approximation based on the second-order expression (2.1), with $b = 0.5$. Then we see from Table 2 that $f^{3W}(x)$ estimates the target mean and variance to within 3.5%, but has an error rate of over 32% for κ_3. Whilst although $g^{3W}(x)$ involves a small loss of accuracy in the variance, this is more than compensated for by large improvements in κ_1 and κ_3. If we consider a more extreme situation, with target cumulants

$$\text{Example B}: \kappa_1 = 12, \quad \kappa_2 = 18 \quad \text{and} \quad \kappa_3 = 100,$$

then to achieve full support we have to increase b to 0.6 because κ_3 is relatively larger. On inspecting Table 2 we see that this results in a substantial loss of accuracy in comparison with Example A. Moreover, incorporating the extra saddlepoint term, $g^{3W}(x)$, results in a further decrease of accuracy. Fig. 2 exposes yet more difficulties; not only does $f^{3W}(x)$ exhibit unwelcome 'cusps' at $x = 10$ and $x = 25$, but these cusps become even more prominent when $g^{3W}(x)$ is considered.

Although we took $b = 0.5$ in the birth–death example A, to achieve full support we only require $b \simeq 0.2$. However, whilst this lower value results in improved cumulant estimates, the resulting distribution exhibits strange characteristics, similar to those shown in Fig. 2. Likewise, in Example

Table 2
Relative differences (2.9) between the target cumulants of Examples A and B, and the cumulants obtained using Wang's approach, and the saddlepoint approximations based on a Poisson (with $b = 0.5$) and Negative binomial (with $b = 0.6$) moment structure

	Example A				Example B			
	$f^{3W}(x)$	$g^{3W}(x)$	$f^{3P}(x)$	$g^{3P}(x)$	$f^{3W}(x)$	$g^{3W}(x)$	$f^{3NB}(x)$	$g^{3NB}(x)$
$\widetilde{\kappa_1}$	3.45	1.42	1.70	0.43	7.60	−0.106	3.35	1.17
$\widetilde{\kappa_2}$	−2.42	−3.63	−1.77	−1.7	−1.62	−21.08	8.34	2.66
$\widetilde{\kappa_3}$	−32.34	−13.28	−9.90	−2.38	−53.54	−34.22	−14.19	−9.05

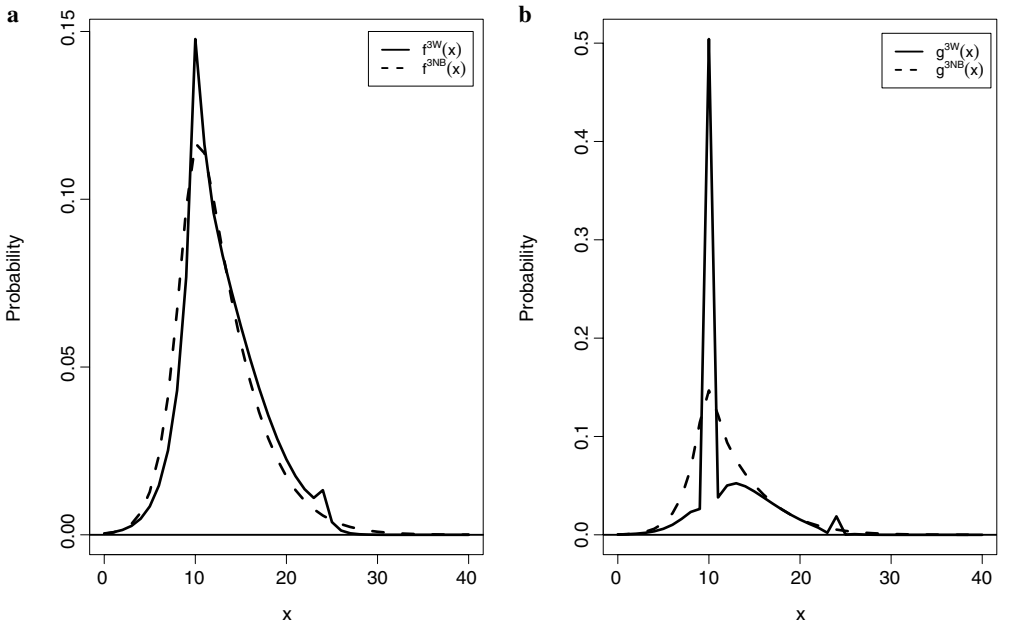

Fig. 2. (a) Distributions obtained using cumulants $\kappa_1 = 12$, $\kappa_2 = 18$ and $\kappa_3 = 100$ for Wang's approximation $f^{3W}(x)$, with $b = 0.6$, together with the saddlepoint approximation under a negative binomial structure with $p = 0.4$ and $r = 10$; (b) as (a), but including the second term of the saddlepoint approximation.

B we see that b can be reduced, but this also exacerbates the situation. Conversely, although increasing b leads to a cusp-free distribution, it does so at the expense of substantially increasing the error surrounding the third cumulant. This raises the question of whether the distributions shown in Fig. 2 are aberrant results, or whether they form part of a larger problem?

To answer this, consider the corresponding moment generating equation of the c.g.f. (3.6), namely

$$M^{3W}(\theta) = \exp\{K^{3W}(\theta)\} = \exp\left\{\kappa_1\theta + \kappa_2\frac{\theta^2}{2} + \kappa_3\frac{\theta^3}{3!}e^{-\kappa_2 b^2\theta^2/2}\right\}. \tag{3.7}$$

This easily yields the raw moments via

$$\mu'_k = \int_{-\infty}^{\infty} x^k p(x)dx = \left.\frac{\partial^k M(\theta)}{\partial\theta^k}\right|_{\theta=0}. \tag{3.8}$$

If we calculate the eighth raw moment for Expression (3.7), since this is where Table 1 'explodes', on taking $\kappa_1 = 0$ for simplicity we obtain

$$\mu'_8 = 35\kappa_2[3\kappa_2^3 + 8\kappa_3^2(1 - 2b^2)]. \tag{3.9}$$

So μ'_8 may become negative when $b > 1$. However, since even raw moments are by definition positive over $x \geqslant 0$, it follows that for $\kappa_1 = 0$ and $b > \sqrt{(1/2 + 3\kappa_2^3/16\kappa_3^2)}$ we have generated moments which cannot correspond to any proper distribution. So it is by no means clear that using Wang's saddlepoint approximation based on only the first three cumulants will lead to a 'sensible' distribution.

Even for the mild case of Example A, we discover that some even raw moments of order above 52 are negative. Whilst for the more extreme Example B, complications arise for orders above 32. Indeed, when we include the second-order saddlepoint approximation, $g(x)$, we are estimating higher-order moments of an 'improper' distribution with a greater degree of accuracy, which most likely explains the strange shapes exhibited in Fig. 2. Such difficulties, of course, arise because the approximations take full support over $(-\infty, \infty)$, rather than operating over the restricted range $[0, \infty)$.

4. An improved method

Given that we have shown that Wang's correction term does not provide good estimates of the target cumulants as κ_3 increases, it is clearly necessary to find a alternative approach for obtaining full support when using only three cumulants. We begin by acknowledging that when we just use the first three cumulants, it is not that the remainder are zero but that we take them to be zero simply because we do not possess any information on them. Thus a possible solution to this problem is to assume some form of non-zero structure for the unknown cumulants. This is particularly apposite, since the presence of skewness automatically implies a non-Normal distribution, yet having third- and higher-order cumulants equal to zero characterises the Normal distribution! For example, suppose we take advantage of the simple structure of the Poisson process, for which $K(\theta) = \alpha(e^{\theta} - 1)$ since all cumulants equal the Poisson mean α, by considering

$$K^{3P}(\theta) = (\kappa_1 - \alpha)\frac{\theta}{1!} + (\kappa_2 - \alpha)\frac{\theta^2}{2!} + (\kappa_3 - \alpha)\frac{\theta^3}{3!} + \alpha(e^{\theta} - 1). \tag{4.1}$$

Note that target cumulants κ_i are unaffected by the choice of α. Moreover, on noting that the saddlepoint approximation only really captures the first few moments (see Table 1), our choice of α should not exert an overt influence on the approximation.

Denote $f^{3P}(x)$ and $g^{3P}(x)$ to be the first- and second-order saddlepoint approximation of the c.g.f. (4.1). Then as the saddlepoint approximation only captures the first few cumulants successfully (see Table 1), to ensure that the choice of α does not exert an undue influence on the approximation, we base it on three criteria. First, it must lend full support over the required range. Second, it should match our target cumulants κ_i to a 'reasonable' degree of accuracy; and third, the resulting approximation should not have 'unnatural' characteristics as demonstrated, for example, by the cusps in Fig. 2. In practice, setting $\alpha = c\kappa_3$, where c takes the trial values $c = 1, 2, \ldots, 10$, produces excellent results. For example, in the birth–death scenario of Example A, we found that taking $\alpha = 4 \times \kappa_3 = 331.48$ produced a good approximation. Indeed, Table 2 shows that not only does $f^{3P}(x)$ produce significantly better cumulant approximations than Wang's approach, but that these estimates may be refined still further through $g^{3P}(x)$.

It is satisfying to observe that by invoking a specific cumulant structure we not only obtain better cumulant estimates, but that these also correspond to an improved approximation for the target distribution. Fig. 3 shows the absolute error for five different approximation schemes used to estimate the birth–death process (Example A), where the absolute error

$$\text{err} = 100\left|\frac{p(x) - f(x)}{p(x)}\right|;$$

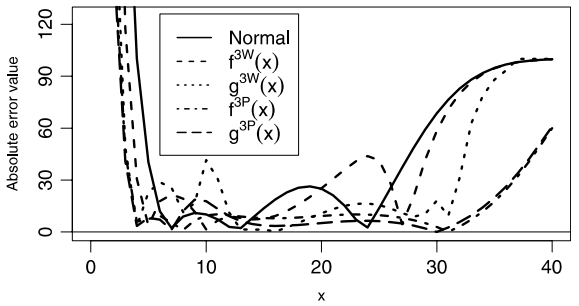

Fig. 3. A plot of the absolute error values for five different approximations of a birth–death process at time $t = 0.65$, having parameters $\lambda = 1.5$, $\mu = 1.0$ and $n_0 = 10$.

here $p(x)$ denotes the exact distribution and $f(x)$ its approximation. Overall, $f^{3P}(x)$ and $g^{3P}(x)$ outperform both Wang's $f^{3W}(x)$ and $g^{3W}(x)$ approximations, and the Normal approximation, most notably in the centre of the distribution. Over $8 < x < 32$ the error estimates for $f^{3P}(x)$ are about 10%, compared to a maximum of more than 40% for $f^{3W}(x)$. As to whether $f^{3P}(x)$ is superior to $g^{3P}(x)$, we could argue either way; for although g^{3P} possesses a lower error rate for much of the distribution, it has a significantly poorer error rate near $x = 10$. Whilst no approximation performs well in the tails of the distribution, it should be recalled that we are only using three cumulants for a particularly long-tailed distribution.

When we examine Example B, where the target cumulants are $\kappa_1 = 12$, $\kappa_2 = 18$ and $\kappa_3 = 100$, we discover that an assumed Poisson structure for the remaining cumulants does not produce satisfactory results. Instead, using a negative binomial cumulant structure provides a much better approximation. Denote $f^{3NB}(x)$ and $g^{3NB}(x)$ to be the first- and second-order saddlepoint approximations using the associated c.g.f.

$$K^{NB}(\theta) = \left(\kappa_1 - \frac{r}{p}\right)\frac{\theta}{1!} + \left(\kappa_2 - \frac{r(1-p)}{p^2}\right)\frac{\theta^2}{2!} + \left(\kappa_3 - \frac{r(1-p)(2-p)}{p^3}\right)\frac{\theta^3}{3!} + r \times \ln\left(\frac{p e^{\theta}}{1 + e^{\theta}(p-1)}\right). \tag{4.2}$$

Then on parallelling the approach used when choosing α in the Poisson c.g.f. (4.1), we select the parameters p and r in (4.2) to ensure that the associated cumulants correspond closely to the target saddlepoint cumulants, whilst at the same time allowing the approximation to generate a realistic 'p.d.f. shape'. Here, values of $p = 0.4$ and $r = 10$ achieve this goal; note that these parameters correspond to a negative binomial distribution with $\kappa_1 = 25$, $\kappa_2 = 37.5$ and $\kappa_3 = 150$, which are similar to our target values.

We observe from Table 2 that $f^{3NB}(x)$ produces better estimates for our target cumulants than Wang's expansion, with $g^{3NB}(x)$ providing yet further improvement. Moreover, Fig. 2 shows that $f^{3NB}(x)$ and $g^{3NB}(x)$ do not exhibit the cusps produced via Wang's approximation. Now it seems highly likely that yet further improvement can be gained by extending the Poisson and negative binomial expressions (4.1) and (4.2) into a more general framework. For consider the representation

$$K^{\mathrm{Gen}}(\theta) = (\kappa_1 - \kappa_1^G)\frac{\theta}{1!} + (\kappa_2 - \kappa_2^G)\frac{\theta^2}{2!} + (\kappa_3 - \kappa_3^G)\frac{\theta^3}{3!} + K^G(\theta), \tag{4.3}$$

where $K^G(\theta)$ is the c.g.f. of any distribution whose first three cumulants are κ_1^G, κ_2^G and κ_3^G. For example, the uniform distribution $p(x) = 1/(n+1)$ for $x = 0, 1, \ldots, n$ takes the relatively simple c.g.f.

$$K^G(\theta) = \ln\left(\frac{\mathrm{e}^{(n+1)\theta} - 1}{(\mathrm{e}^{\theta} - 1)(n+1)}\right) \tag{4.4}$$

which is highly amenable to manipulation. Study of the negative binomial, Poisson and uniform examples shows that in practise the Poisson distribution performs particularly well when we only have knowledge of the first three cumulants, as α can always be chosen to ensure that the target cumulants are well-matched, and that the associated approximating p.d.f. (1.8), namely

$$f(x) \simeq \{2\pi[K^{\mathrm{Gen}}(\theta_0)]''\}^{-1/2} \exp\{K^{\mathrm{Gen}}(\theta_0) - \theta_0 x\} \tag{4.5}$$

has full support and exhibits smooth behaviour. Even so, much more detailed examination of these and further candidate c.g.f.'s would be highly desirable. Note that we can only write down a closed expression for the saddlepoint approximation, such as (1.8), if the saddlepoint equation $K(\theta_0) = x$ is a polynomial of degree less than five. So in essence, this means that solutions obtained via $K^{\mathrm{Gen}}(\theta)$ will not be explicitly expressible purely in terms of x, but only in terms of θ_0. This, however, is a small price to pay for acquiring a substantial improvement in the quality of the approximation.

Further research clearly needs to be undertaken to develop a general methodology for constructing 'optimal' k-order c.g.f.'s

$$K^{\mathrm{Gen}}(\theta) = \sum_{i=1}^{k} (\kappa_i - \kappa_i^G)\frac{\theta^i}{i!} + K^G(\theta), \tag{4.6}$$

where κ_i^G is the ith cumulant of the c.g.f. $K^G(\theta)$. This, however, is a difficult problem, since it would entail the development of least squares, minimum entropy, etc., criteria subject to the constraints that $f(x)$ is non-negative, has full support over the range of interest, and $|f''(x)| < A$ for some appropriately small A to avoid the cusp-like behaviour exhibited in Fig. 2.

5. A further refinement

A further, albeit intuitive, variation for achieving an 'optimal' saddlepoint approximation would be not only to alter say α in the Poisson-based c.g.f. (4.1), or $\kappa_1^G, \ldots, \kappa_k^G$ in the general representation (4.6), but additionally to choose our initial cumulants more carefully. For example, we observe from Table 2 that in the birth–death case A we have a relative error of 1.70% for $f^{\mathrm{3P}}(x)$ when estimating κ_1. So suppose we input a slightly lower target value of κ_1, say 13.5 instead of 13.84, and then proceed to calculate *overall* relative error, namely

$$S = a\left|\frac{\kappa_1 - \hat{\kappa_1}}{\kappa_1}\right| + b\left|\frac{\kappa_2 - \hat{\kappa_2}}{\kappa_2}\right| + c\left|\frac{\kappa_3 - \hat{\kappa_3}}{\kappa_3}\right| \tag{5.1}$$

for positive constants a, b and c, and cumulants $\hat{\kappa}_i$ corresponding to the chosen saddlepoint approximation. Then if S is reduced we accept the proposed value, otherwise we do not. As (here) three cumulants are involved, the full approach is to alter κ_1, κ_2, κ_3 and α, to $\overline{\kappa_1}$, $\overline{\kappa_2}$, $\overline{\kappa_3}$ and $\overline{\alpha}$, in order that we achieve the smallest possible value of S.

Now the simplex technique provides a fast and efficient method of determining the minimum of the function S. We define a simplex to be a set of $n+1$ points in n-dimensional space. So in two-dimensions the corners of a triangle form a simplex, while in three dimensions it would be the corners of a tetrahedron. In the simplex approach, originally developed by Spendley et al. [23] and then extended by Nelder and Mead [18], the function is first minimised at the corners of a regular simplex (which have sides of equal length), and then a new simplex is formed by deleting the corner at which the function value is largest. A replacement vertex is then formed by reflecting the rejected vertex through the centre of all of the remaining vertices, except for the rejected one. We then proceed by reducing the size of the simplex (known as contraction). Nelder and Mead extended the model to allow expansion of the simplex as well as contraction, in order to reduce the chance of the procedure becoming trapped in a local minimum. In general, the following procedure works well.

Simplex Algorithm:

1. set the initial saddlepoint cumulants $\overline{\kappa}_i$ to the target values κ_i and choose an appropriate value for α;
2. evaluate the saddlepoint probabilities over the required region of support Ω, and re-scale to ensure that $\sum_{x\in\Omega} f^{3\mathrm{P}}(x) = 1$;
3. evaluate the cumulants, $\hat{\kappa}_i$, corresponding to the saddlepoint approximation $f^{3\mathrm{P}}(x)$;
4. evaluate S, i.e., Expression (5.1), and use the simplex NAG routine E04CCF to construct $\overline{\kappa_1}$, $\overline{\kappa_2}, \overline{\kappa_3}$ and $\overline{\alpha}$;
5. print $\overline{\kappa_1}$, $\overline{\kappa_2}, \overline{\kappa_3}$ and $\overline{\alpha}$;
6. if S reaches say 10^{-6}, then stop; else return to 2.

Whilst in general the target cumulants appear to be good starting values, it is recommended that different initial values of α be used and the results compared [21]. Furthermore, experimentation with the coefficients a, b and c in S is helpful in finding an optimal solution (especially if the simplex routine takes a long time to converge). In practise we used the values $a = 3/2$, $b = 1$ and $c = 1/2$. This means that more emphasis is placed on lower-order cumulants. Note that S may not always be the most efficient function to implement, especially for small target cumulants. An alternate measure could, for example, involve the sum-of-squares

$$\tilde{S} = \sum_{i=1}^{3} (\kappa_i - \hat{\kappa}_i)^2, \tag{5.2}$$

which is less susceptible to small κ_i's; whilst other possibilities include minimising χ^2 or entropy.

In Example A, minimising $\tilde{S}$ (i.e., Expression (5.2) where $\hat{\kappa}_i$ are the cumulants from $f^{3P}(x)$), involves five-dimensional space, since we are changing the parameters $\overline{\kappa_1}, \overline{\kappa_2}, \overline{\kappa_3}$ and α. Now implementing the above algorithm yields the minimising values $\overline{\kappa_1} = 13.603, \overline{\kappa_2} = 27.378, \overline{\kappa_3} = 90.557$ and $\alpha = 324.987$. Whence on denoting $f^{3\tilde{\mathrm{S}}\mathrm{P}}(x)$ to be the saddlepoint approximation based on these

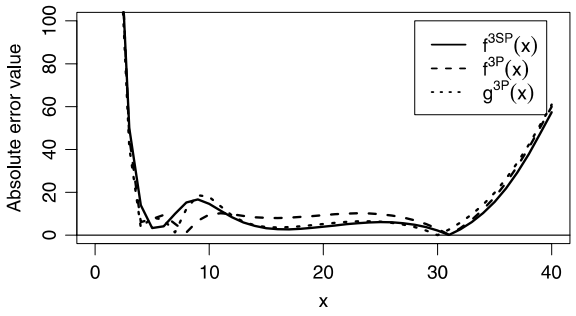

Fig. 4. A plot contrasting the relative error values (5.1) of three approximations using the cumulants in Example A.

values, we see that the three relative error rates between the cumulants of $f^{3\tilde{S}P}(x)$ and the target cumulants are -0.0002%, 0.73% and 0.00002%. This is clearly a dramatic improvement when compared with the other approximations in Table 2. Furthermore, we observe from Fig. 4 that this procedure also yields an improved approximation of the target distribution. For $x > 12$ we have a lower error rate than $f^{3P}(x)$ and, in general, a slightly lower rate than $g^{3P}(x)$. Additionally, there is a reduced error rate in the right-hand tail.

However, extending this simplex approach to the far more extreme Example B does not give rise to satisfactory results. For although we do obtain better cumulants in relation to other target values, the resulting shape of the distribution is not realistic since the spike shown in Fig. 2 becomes even more pronounced. However, it should be stressed that when κ_3 is reduced to approximately three times κ_2, this cusp no longer features and the method once again works extremely well.

Finally, we note that when extending the simplex algorithm to allow even more cumulants to be changed, i.e., fourth- and higher-order cumulants, caution needs to be exercised. For if more than six parameters are used, then the simplex routine is more likely to find a local minimum than a global one.

6. Summary and conclusions

The general intractability of nonlinear Kolmogorov equations poses a major problem in the analysis of applied stochastic processes. For although many standard 'statistical' procedures employ the assumption of an underlying Normal distribution, with the Physics paradigm being the adoption of 'mean-field' theory, it is becoming increasingly acknowledged that Gaussian processes have little relevance in many real-life situations. Whilst in general, the standard saddlepoint approximation arguably provides the 'optimal' approximating distribution, since it is based on the method of fastest descents, the fact that it can easily give rise to limited regions of support is clearly a major problem. Moreover, including the second term in the underlying contour integration (see Expression (2.1)) does not lead to a uniformly better improvement. Consideration of the Poisson and negative binomial distributions highlights the accuracy to which both probabilities and cumulants can be reproduced under both first- and second-order saddlepoint approximations. The key point is that at $x = \kappa_1$ the basic saddlepoint approximation can never be improved

on that given by the Normal approximation, which suggest that inclusion of higher-order terms in the inversion of $K(\theta)$ should result in an improved approximation since more information is utilised.

If we are given just the first three cumulants, then although the resulting saddlepoint approximation $f_3(x)$ has the closed form (3.3), it may not take full support. Wang's scaling of κ_3 does reclaim full support, but the price paid involves both reduced accuracy and the chance of producing aberrant (e.g. cusped) distributions. Development of an improved method which yields good accuracy, full support and also 'sensible' approximating p.d.f.s therefore requires an alternative approach. To construct this we dispense with the 'usual' moment closure technique based on replacing all higher-order cumulants κ_i for $i > k$ by zero, and adopt, instead, a specific non-zero structure. Here we consider the Poisson and negative binomial cases, and through them show that this procedure yields substantially better approximations than those delivered through Wang's approach. Progress can be advanced still further by choosing our 'start' cumulants more carefully. For example, we show that making a subtle change to κ_1, and then minimising overall relative error (weighted with respect to cumulant order) via the simplex algorithm, has the potential for yielding a dramatic improvement in cumulant accuracy. Moreover, using second-order inversion provides yet further improvement. So not only do these new techniques for producing the 'best' cumulant structure in a given situation prove to be extremely useful and beneficial, but the extension of these ideas into the construction of a general kth-order methodology could well prove to be a highly profitable field of future research development.

Acknowledgement

This work was supported by an EPSRC Research Studentship.

References

[1] A. Chen, E. Renshaw, The M/M/1 queue with mass exodus and mass arrivals when empty, J. Appl. Prob. 34 (1997) 192.
[2] H.E. Daniels, Saddlepoint approximations in statistics, Ann. Math. Stat. 25 (1954) 631.
[3] G.S. Easton, E. Ronchetti, General saddlepoint approximations with applications to L statistics, J. Am. Stat. Assoc. 81 (1986) 420.
[4] C.A. Field, E. Ronchetti, Small Sample Asymptotics. Lecture Notes and Monograph Series, 13, Institute of Mathematical Statistics, Hayward, 1990, p. 4.
[5] R. Gatto, Saddlepoint approximations of marginal densities and confidence intervals in the logistic regression measurement error model, Biometrics 52 (1996) 1096.
[6] C.S. Gillespie, E. Renshaw, The evolution of a batch-immigration death process subject to counts, Proc. R. Soc. Lond. A 461 (2005) 1563.
[7] I.J. Good, The multivariate saddlepoint method and chi-squared for the multinomial, Ann. Math. Stat. 32 (1961) 535.
[8] C. Goutis, G. Casella, Explaining the saddlepoint approximation, Am. Stat. 53 (1999) 216.
[9] J.L. Harvill, H.J. Newton, Saddlepoint approximations for the difference of order statistics, Biometrika 82 (1995) 226.
[10] R. Hengeveld, Dynamics of Biological Invasions, Chapman & Hall, London, 1984.

[11] E. Jakeman, K.I. Hopcraft, J.O. Matthews, Distinguishing population processes by external monitoring, Proc. R. Soc. Lond. A 459 (2003) 623.
[12] O. Lieberman, The effect of nonnormality, Econ. Th. 13 (1997) 52.
[13] G. Marion, E. Renshaw, G. Gibson, Stochastic effects in a model of nematode infection in ruminants, IMA J. Math. Appl. Med. Biol. 15 (1998) 97.
[14] J.O. Matthews, K.I. Hopcraft, E. Jakeman, Generation and monitoring of discrete stable random processes using multiple immigration population models, J. Phys. A 36 (2003) 11585.
[15] J.H. Matis, T.R. Kiffe, P.R. Parthasarathy, Density-dependent birth–death-migration models, in: XVIII International Biometric Conference, Amsterdam, I5, 1996, p. 79.
[16] J.H. Matis, T.R. Kiffe, On approximating the moments of the equilibrium distributions of a stochastic logistic model, Biometrics 52 (1996) 980.
[17] J.H. Matis, T.R. Kiffe, E. Renshaw, J. Hassan, A simple saddlepoint approximation for the equilibrium distribution of the stochastic logistic model of population growth, Ecol. Model. 161 (2003) 239.
[18] J.A. Nelder, R. Mead, A simplex method for function minimization, Comp. J. 7 (1965) 308.
[19] E. Renshaw, Modelling Biological Populations in Space and Time, Cambridge University Press, Cambridge, MA, 1991.
[20] E. Renshaw, Saddlepoint approximations for stochastic processes with truncated cumulant generating functions, IMA J. Math. Appl. Med. Biol. 15 (1998) 41.
[21] E. Renshaw, Applying the saddlepoint approximation to bivariate stochastic processes, Math. Biosci. 168 (2000) 57.
[22] E. Renshaw, A. Chen, Birth–death processes with mass annihilation and state-dependent immigration, Stoch. Models 13 (1997) 239.
[23] W. Spendley, G.R. Hext, F.R. Himsworth, Sequential applications of the simplex designs on optimization and evolutionary operation, Technometrics 4 (1962) 441.
[24] W.Y. Tan, H. Hsu, Some stochastic models of AIDS spread, Stat. Med. 8 (1989) 121.
[25] S. Wang, General saddlepoint approximations in the bootstrap, Prob. Stat. Lett. 27 (1992) 61.