

# Optimal Sampling and Problematic Likelihood Functions in a Simple Population Model

<sup>1</sup>Pagendam, D.E. and <sup>1</sup>Pollett, P.K.

<sup>1</sup>Department of Mathematics, School of Physical Sciences  
University of Queensland, St Lucia, Qld 4072, Australia. E-Mail: pagendam@maths.uq.edu.au

*Keywords: Markov chain, optimal design, sampling, birth-death process, population model, Bayesian inference*

## ABSTRACT

Markov chains provide excellent statistical models for studying many natural phenomena that evolve with time. One particular class of continuous-time Markov chain, called birth-death processes, can be used for modelling population dynamics in fields such as ecology and microbiology. The challenge for the practitioner when fitting these models is to take measurements of a population size over time in order to estimate the model parameters, such as per capita birth and death rates. In many biological contexts, it is impractical to follow the fate of each individual in a population continuously in time, so the researcher is often limited to a fixed number of measurements of population size over the duration of the study. We show that for a simple birth-death process, with positive Malthusian growth rate, subject to common practical constraints (such as the number of samples and timeframes), there is an optimal schedule for measuring the population size that minimises the expected confidence region of the parameter estimates. This type of experimental design results in a more efficient use of experimental resources, which is often an important consideration.

Throughout our exposition of the optimal experimental design, we compare it to a simpler equidistant design, where the population is sampled at regular intervals. This is an experimental design worthy of comparison, since it can represent a much simpler design to implement in practice. We acknowledge that practitioners are likely to prefer using the simplest possible design and therefore focus upon the conditions under which the optimal design is expected to be particularly beneficial over the simpler alternative. We also examine the exact manner in which the optimal design acts to minimise the area of the confidence region compared to the alternative design.

In order to find optimal experimental designs for our population model, we make use of a combination of useful statistical machinery. Firstly, we use a Gaussian diffusion approximation of the underlying discrete-state Markov process which allows us to obtain

analytical expressions for Fisher's Information matrix (FIM), which is crucial to optimising the experimental design. We also make use of the Cross-Entropy method of stochastic optimisation for the purpose of maximising the determinant of FIM to obtain the optimal experimental designs. Our results show that the optimal schedule devised by Becker and Kersting (1983) for a simple model of population growth without death can be extended, for large populations, to the two-parameter model that incorporates both birth and death.

Population models have a history of creating problematic likelihood functions with high levels of dependence between model parameters (see Givens and Poole (2002)). For the simple birth-death process, we find that the likelihood surface is also problematic and poses serious problems for point estimation and easily defining confidence regions. There is also very high level of correlation between the estimates for the birth and death rates indicated by the likelihood surface having a long narrow ridge falling away steeply at either side.

We use simulation to examine the practical benefits of the optimal design over an equidistant design. Unless the period of time over which the population is observed is very long, the optimal design is only likely to provide a significant efficiency gain when the number of samples is relatively small. However, we find that, in general, confidence regions cannot be assumed to have elliptical contours. We therefore base our confidence regions on the asymptotic  $\chi^2$  distribution of the generalized likelihood ratio, which restricts the region to the appropriate domain. Whilst the confidence regions of this type may be poor due to the atypical nature of the likelihood surface, we utilise these regions for comparative purposes, since they reflect the contours of the surface. We find that our optimal design remains optimal even though the likelihood contours are not elliptical. It is suggested that Bayesian inference and the use of an informative prior probability distribution could overcome problems associated with the likelihood surface and parameter uncertainty for the model.

## 1 INTRODUCTION

Many natural phenomena, such as the dynamics of a population of organisms, exhibit an element of randomness which renders a purely deterministic approach to modelling inappropriate. Markovian models, such as birth-death processes, can be ideal stochastic population models for addressing many pertinent questions in population biology regarding issues such as the probability of extinction and sustainable harvesting rates. In order to make use of such models, it is necessary to monitor a population over time in order to obtain data that can be used to estimate the parameters of the underlying random process, such as per capita birth and death rates. However, in many typical experiments, the practitioner has constraints on the timeframe for the research and the number of times the individuals in the population can be counted. In such cases, there is a clear need for research to be carried out in a manner that maximises the amount of information obtained and results in as precise an estimate of these parameters as possible.

We examine a simple model of population growth, known as the simple birth-death process. This model has two parameters: the per capita birth rate ( $\lambda$ ) and death rate ( $\mu$ ) and we restrict our attention to the case where the Malthusian growth rate ( $\alpha = \lambda - \mu$ ) is positive. We demonstrate how the use of an optimal design or schedule for the proposed census dates can reduce the area of our confidence regions for the estimated parameters. Such a design is useful, since it allows the practitioner to make the most of the experimental resources available. The optimal design for the simple birth-death process is identified using a diffusion approximation to the underlying Markovian population model in conjunction with the Cross-Entropy (CE) method of stochastic optimisation.

For the purposes of illustration, we compare the use of the optimal design, to the much simpler, but sub-optimal ‘equidistant’ design, where observations are scheduled at regular intervals. We outline the scenarios where the use of an optimal design is most advantageous and also where it results in almost negligible benefits. Additionally, we outline the precise manner in which the optimal design acts to reduce the area of the confidence ellipse.

## 2 THE MODEL

### 2.1 Simple birth-death processes

Let  $Y(t)$  be a birth-death process taking values in  $S := \{0, 1, \dots\}$  and let  $y(t)$  denote the observed population size at time  $t$ . In addition, we define the non-zero transition rates of the Markov process to be

$$q(y, y + 1) = \lambda_y \quad \text{and} \quad q(y, y - 1) = \mu_y,$$

where  $\mu_y$  and  $\lambda_y$  are the birth and death rates respectively when the population is of size  $y$ . For the simple birth-death process, we treat the population birth and death rates as being linearly related to the per capita birth and death rates ( $\lambda$  and  $\mu$ ), that is,  $\lambda_y = \lambda y$  and  $\mu_y = \mu y$ . This model, commonly referred to as the linear or simple birth-death process can be used to model the growth of a population of organisms in continuous time under the assumption of density-independent or weakly density-dependent population growth. From the perspective of parameter estimation, the model has received attention from a number of authors (see Keiding (1975) and Moran (1953) for example). Since our model is the stochastic counterpart to the deterministic model of exponential growth, we might use such a model to examine the bloom of a problematic algal population in an aquatic ecosystem or the initial population growth of an invasive species.

## 3 THE DESIGN PROBLEM

### 3.1 The Experiment and Constraints

Suppose we wish to estimate the parameters  $\lambda$  and  $\mu$  of a simple birth-death process by conducting a census of a population as it grows over time. We begin our experiment with the knowledge of the initial population size  $y_0$  at time  $t_0 = 0$  and, as a result of constraints on experimental resources such as budget, manpower and available time, we are limited to  $n$  opportunities to census the population over the time interval  $(0, t_{max}]$ . Our design problem amounts to finding the set of sampling times  $(t_1 < \dots < t_n)$  that will result in the most precise estimate of the parameter vector  $\theta = (\lambda, \mu)$ . We note that the same design problem was considered by Becker and Kersting (1983) for the one parameter, simple birth process, which assumes that there is no death in the population (i.e.  $\mu = 0$ ).

### 3.2 The Gaussian Diffusion Approximation and Maximum Likelihood Estimation

In order to effectively use models for addressing problems associated with real phenomena, it is necessary to obtain accurate and precise estimates of the model’s parameters. In most statistical problems, including stochastic models, the most widely accepted method of parameter estimation is maximum likelihood estimation. This involves forming the likelihood as the joint probability density of the data and then maximising this function over the parameter space. In order to formulate the likelihood, it is first necessary to obtain analytical expressions for the probability of the population making transitions from one observed state to another in the time intervals defined by our experimental design. We write the likelihood, conditional upon the initial state

of the population being known, as

$$L(\lambda, \mu; y_1, \dots, y_n | y_0) = \prod_{i=1}^n p(y_i, t_i | y_{i-1}, t_{i-1}),$$

where  $p(y_i, t_i | y_{i-1}, t_{i-1})$  is the probability of the birth-death process making a transition from a population size of  $y_{i-1}$  to  $y_i$  in the time interval formed by adjacent sample times  $t_{i-1}, t_i$ .

In the case of maximum likelihood estimation for Markov processes, obtaining analytical expressions for the transition probabilities is not always simple and, furthermore, these expressions may not allow the likelihood to be maximised easily. This is certainly true of the simple birth-death process and we therefore rely on a Gaussian diffusion approximation of the model to form an approximate likelihood. Gaussian diffusion approximations have been used with great success in modelling population models and we refer the reader in particular to Ross *et al.* (2006) and Pollett (2001) for more information on their application in population biology. The likelihood from the resulting diffusion approximation is given by

$$L(\lambda, \mu; y_1, \dots, y_n | y_0) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{m})\Sigma^{-1}(\mathbf{y} - \mathbf{m})'\right),$$

where  $\mathbf{y} = (y_1, \dots, y_n)$  and  $\mathbf{m} = (y_0 \exp((\lambda - \mu)t_1), \dots, y_0 \exp((\lambda - \mu)t_n))$  are the vectors of the observed and expected population sizes respectively for the sampling times  $\mathbf{t} = (t_1, \dots, t_n)$  and  $\Sigma$  is the  $n$  by  $n$  covariance matrix, with elements

$$\Sigma_{i,j} = \begin{cases} y_0 \text{Var}(Z(t_i)) & i = j, \\ y_0 \text{Var}(Z(\min(t_i, t_j))) e^{((\lambda - \mu)|t_i - t_j|)} & i \neq j, \end{cases}$$

where

$$\text{Var}(Z(t)) = \begin{cases} \frac{\lambda + \mu}{\lambda - \mu} e^{(\lambda - \mu)t} (e^{(\lambda - \mu)t} - 1) & \lambda \neq \mu \\ 2\lambda t & \lambda = \mu. \end{cases}$$

We note that our diffusion approximation is only a good approximation when the initial population size  $y_0$  is reasonably large (say  $\geq 100$ ) and improves as  $y_0 \rightarrow \infty$ .

### 3.3 The D-Optimal Design via the Cross-Entropy Method

The Gaussian diffusion approximation used for maximum likelihood estimation is also particularly useful in that it allows us to tackle experimental design problems of the type described above for the simple birth-death process. An optimal experimental design is one that minimises the expected variability of the parameter estimates in some way. Formally, a design  $g^*$  in the set of possible designs,  $G$ , is considered to be D-optimal if it maximises the determinant ( $\det(\cdot)$ ) of FIM,  $\mathcal{I}$ , which is a function of  $g$ , that is,

$$g^* = \underset{g \in G}{\operatorname{argmax}} \det(\mathcal{I}(g)),$$

where, for our model,

$$\mathcal{I} = \begin{bmatrix} \frac{\partial^2 \log L}{\partial \lambda^2} & \frac{\partial^2 \log L}{\partial \lambda \partial \mu} \\ \frac{\partial^2 \log L}{\partial \lambda \partial \mu} & \frac{\partial^2 \log L}{\partial \mu^2} \end{bmatrix},$$

whose elements have a particularly simple form for the multivariate Gaussian distribution (see Porat, 1995), namely

$$\mathcal{I}_{i,j} = \left( \frac{\partial \mathbf{m}}{\partial \theta_i} \right) \Sigma^{-1} \left( \frac{\partial \mathbf{m}}{\partial \theta_j} \right)' + \frac{1}{2} \operatorname{tr} \left( \Sigma^{-1} \left( \frac{\partial \Sigma}{\partial \theta_i} \right) \Sigma^{-1} \left( \frac{\partial \Sigma}{\partial \theta_j} \right) \right).$$

One notable and somewhat counter-intuitive aspect of optimal designs in general is that, in order to design the best experiment for estimating the parameters of interest, the practitioner must already know the parameter values. This arises, since the optimal design is dependent upon the parameters. In practice however, we make some educated guess, use this to design the experiment and then perform the experiment to obtain better estimates.

We utilise the CE method of stochastic optimisation to obtain the optimal design for the process. Our method works by generating random designs over the time interval  $[0, t_{max}]$  by using  $n$  beta probability density functions, each with parameter vector  $\nu_i = (\alpha_i, \beta_i)$ . The probability distributions are initiated with parameter vector  $\nu_i = (1, 1)$ , so that the densities are uniform on the time interval under consideration. Our algorithm for the updating of these beta probability distributions is outlined below and is based upon recommendations in Rubinstein and Kroese (2004).

#### CE algorithm for the D-optimal design

1. Define values for  $t_{max}$ ,  $n$ ,  $y_0$ ,  $\lambda$  and  $\mu$ .
2. Set  $\gamma = 0.9$ .
3. Set the maximum variance threshold, used to end the optimisation routine, to be  $\epsilon = 10^{-5}$ .
4. Set the timestep  $k = 0$  and the initial parameter vectors  $\theta_i$  for each of  $n$  beta probability density functions  $f_i(t_i; \nu_i)$  as  $(1, 1)$ .
5. Create  $m$  inspection schedules,  $\mathbf{T}_{1,k} \dots, \mathbf{T}_{m,k}$ , such that the  $i$ th element of each schedule is generated from the beta probability density function  $f_i(t_i; \alpha_{i,k}, \beta_{i,k})$ , where  $\mathbf{T}_j = (T_{1,j,k}, \dots, T_{n,j,k})$  and the inspection times satisfy  $0 < T_1 < \dots < T_n \leq t_{max}$ .
6. Calculate the performance (D-optimality)  $p_{1,k}, \dots, p_{m,k}$  of each of the  $m$  inspection schedules by calculating the  $\det(\text{FIM})$ .
7. Calculate the  $\gamma$ -quantile of  $p_{1,k}, \dots, p_{m,k}$  and record this as  $q_k$ .

8. Calculate the mean and variance for each of the  $n$  beta densities (indexed by  $i$ ) using the best performing observation points .
9. Update the time step:  $k = k + 1$ .
10. Update the parameter vectors for each of the beta probability densities, such that the density giving rise to the  $i$ th inspection point has the mean and variance calculated in step 8 for the  $i$ th inspection point.
11. If  $\max(\text{Var}(T_{1,k}), \dots, \text{Var}(T_{n,k})) \leq \epsilon$ , then return the vector  $(\text{Mean}(T_{1,k}, \dots, T_{n,k}))$  as the optimal design and terminate the routine; otherwise go to step 5.

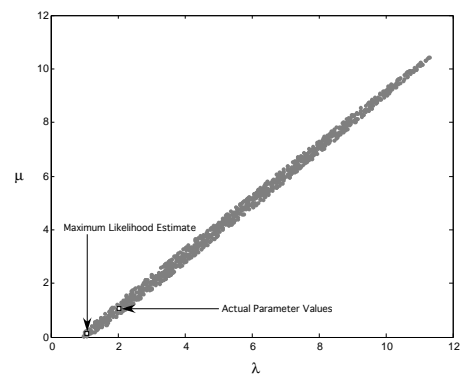
### 3.4 Checking Optimal Designs via Simulation

For the purpose of illustrating our methods, we simulated the growth of a population of organisms according to birth-death processes with per capita birth rate,  $\lambda = 2$  and death rate  $\mu = 1$  and initial populations of 100 individuals. We simulated the growth of populations according to a number of scenarios (various time intervals with various numbers of samples) to assess the benefits of the optimal design, using 10 simulated trajectories for each scenario. For each simulation, the areas ( $A_{opt}$  and  $A_{eq}$ ) of 95% confidence regions were calculated for data collected using the optimal design and the equidistant design respectively. For each trajectory the relative areas of the confidence regions were calculated as  $A_{relative} = A_{opt}/A_{eq}$ , which can be thought of as being analogous to the efficiency of the equidistant design relative to the optimal design.

We found that confidence ellipses frequently extended into illegal parts of the parameter space and the classical assumption that  $\hat{\theta}$  was normally distributed was therefore flawed. This was largely because the maximum likelihood estimate frequently occurred very close to the boundary of the parameter space and the likelihood surface had an unusual shape. It was apparent that the confidence regions were not elliptical and we instead generated confidence regions by basing them on the asymptotic  $\chi^2$  distribution of the generalised likelihood ratio test, giving an approximate  $(1-\alpha)\%$  confidence region for  $\theta$  as  $\{\theta \in \Theta : 2[l(\hat{\theta})] - l(\theta) \leq \chi_{2;1-\alpha}^2\}$ . The maximum likelihood estimate was identified using a simple CE algorithm similar to that presented in Ross *et al.* (2006) and, although convergence of the routine was achieved, estimates were poor since the greatest likelihood appeared to consistently occur very close to the edge of the parameter space. As noted by Poole and Givens (2002) in their discussion of problematic likelihood surfaces of population models, confidence regions based on the likelihood ratio are likely to be unreliable. However, they also noted that since such

confidence regions are based on the contours of the likelihood surface, they still have a convenient use in comparing the shapes and contours of different surfaces. We find such regions to be convenient herein for examining the change in the shape of the contours of the likelihood surface under different experimental designs.

Since the D-optimal design is formed under the assumption of elliptical confidence regions, we examined the differences in areas of simulated (non-elliptical) confidence regions, approximating their area via Importance Sampling (IS), using a bivariate Gaussian sampling density with parameters obtained using the CE method. The CE algorithm for the IS was based on an example given on page 40 of Rubinstein and Kroese (2004). For each of the 10 simulations, we wished to generate random points over the square region  $R$  in the parameter space, where  $R := \{[0, 15] \times [0, 15]\}$ . Since the confidence regions were very narrow, the probability of random points being generated in these regions was very low. IS allowed us to simulate these rare events in a much more efficient manner. For each simulation, 50 000 points were generated according to the IS density, rendering a high probability of the points being contained in the confidence region. We were then able to estimate the probability,  $p$ , that a single point was contained in the confidence region, had it been generated uniformly at random over  $R$ . The area of the 95% confidence region was estimated by multiplying  $p$  by 225 (the area of  $R$ ). Figure 1 demonstrates how the random points were used to obtain a confidence region for an example simulation. The shape of the confidence region and the placement of the maximum likelihood estimate highlight the problematic nature of likelihood surfaces obtained from this model.



**Figure 1.** Plot of randomly generated points in the parameter space falling within the 95% confidence region for a simulated birth-death process with a 2 sample optimal design over the time interval  $[0, 1]$ .

## 4 RESULTS

### 4.1 The Optimal Design

The Gaussian diffusion approximation in conjunction with the CE method allowed us to quickly establish D-optimal designs for the estimation of the parameter vector  $\theta = (\lambda, \mu)$ . We found that for the simple birth-death process (when  $y_0$  is sufficiently large), the optimal design is, for all practical purposes, identical to that obtained by Becker and Kersting (1983) for the simple birth process, where the design is now dependent upon the Malthusian growth rate  $\alpha = \lambda - \mu$ , rather than the just birth rate  $\lambda$ . The D-optimal design is therefore defined by the relation

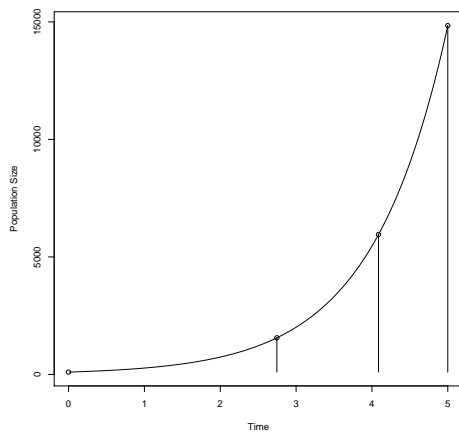
$$\phi_1(\alpha(t_i - t_{i-1})) = \phi_2(\alpha(t_{i+1} - t_i)), \quad i = 1, \dots, n-1,$$

where,

$$\phi_1(x) = x(e^x - 1)^{-2}(2e^x - x - 2), \quad (1)$$

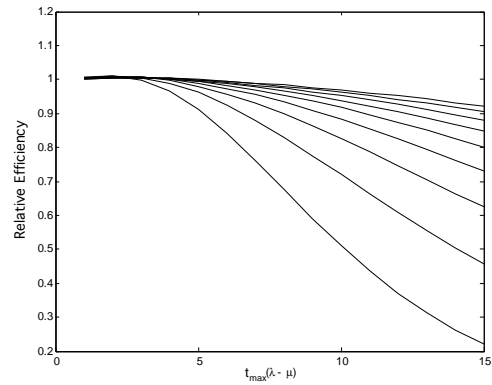
$$\phi_2(x) = xe^x(e^x - 1)^{-2}(2e^x - 2 - xe^x), \quad (2)$$

and where  $t_n = t_{max}$ . Figure 2 illustrates the optimal sampling times for a population with initial population size of 100 individuals, observed over the interval  $[0, 5]$ . Figure 3 was generated by obtaining



**Figure 2.** An example of a 3 sample optimal design, showing the expected trajectory of a population and the times at which observations should be made.

optimal designs for various numbers of observations and time intervals and comparing the determinant of FIM under this design to the determinant under equidistant sampling. We define the relative efficiency of designs using the D-efficiency criterion  $D_{\text{eff}} = \left(\frac{|I_{\text{eq}}|}{|I_{\text{opt}}|}\right)^{1/2}$  (see Atkinson and Donev, 1992), which (for the two parameter model) gives the relative areas of the theoretical confidence ellipses expected under the two experimental designs. Averages of the relative confidence region areas and the standard errors obtained by simulation are shown in Table 1.



**Figure 3.** Relative efficiency of the equidistant design compared with the D-optimal design for different numbers of samples (labels).

**Table 1.** Relative areas of 95% confidence regions obtained from simulated data.

$n$	$t_{max}$	$\bar{A}_{relative}$	std. err.
2	2	0.9993	0.0058
	5	0.9039	0.0131
	10	0.5372	0.0179
5	2	0.9924	0.0034
	5	0.9879	0.0121
	10	0.9072	0.0443

We note that  $\bar{A}_{relative}$  can be considered analogous to the D-efficiency of the equidistant design relative to the optimal design.

## 5 DISCUSSION

### 5.1 D-Optimal Versus Equidistant Design

It was noted in the previous section that once the D-optimal design was applied to the analysis of simulated data, we found that the likelihood surface did not have elliptical contours and, furthermore, presented serious problems for point estimation. This cast some doubt upon whether maximising the determinant of FIM would necessarily provide an optimal design in practice, since FIM relies upon the asymptotic normality of the estimator and associated elliptical contours of the likelihood surface. However, upon further investigation of the optimal design, we found that we could also consider the D-optimal design as maximising objective functions equal to (i) the smallest eigenvalue of FIM, and (ii) the trace of FIM (the so called A-optimal design), because the maximum eigenvalue of FIM is relatively insensitive to the sampling times. This indicated that the major axis of the confidence region was unlikely to be reduced significantly by changing sampling times,

as long as  $t_n = t_{max}$ . However, it also highlighted that the optimal design operated by narrowing the confidence region (i.e. minimising the minor axis of the region). We note that since the profile of the log-likelihood surface in the direction of the minor axis generally appears to be approximated well by tangent parabolas (a characteristic typical of a normally distributed parameter estimate) and since it is along this axis that the optimal design was found to minimise the confidence region, we conclude that our optimal design based on FIM will remain optimal (or very close to optimal) even though the confidence regions are not necessarily elliptical. The relative areas of the simulated non-elliptical confidence regions displayed in Table 1 support this finding and appear to be strongly consistent with the theoretical relative efficiencies depicted in Figure 3.

In general, the likelihood surface for the parameters  $\lambda$  and  $\mu$  of the simple birth-death process was a problematic one from the point of view of point estimation and identifying confidence regions. The surface had two unusual features: (i) a very long ridge running along a line with an approximate gradient of 1 and intercept on the  $\lambda$  axis approximately equal to  $\alpha$  ( $= \lambda - \mu$ ); and (ii) non-elliptical contours that appeared to be maximal very near (possibly on) the boundary of the parameter space. Observation of long thin confidence regions from population models was also made by Ross *et al.* (2006) and Poole and Givens (2002), and is related to the fact that the individual parameter estimates are highly correlated. However, in the case of Ross *et al.* (2006), the likelihood surface was well behaved and allowed the evaluation of confidence ellipses.

Observations of high levels of dependency between parameters in population models has been noted by previous authors and led to discussion of the problematic likelihood functions that can arise in these situations (see Givens and Poole (2002)). Indeed, the shape of likelihood based confidence regions for the simple birth-death process makes it very difficult to visually compare the areas unless the observation period is quite short (as used for Figure 1). The use of the CE method for rare-event simulation did allow us to compare the areas however and indicated that the optimal design was more efficient than the equidistant design, narrowing the ridge of the likelihood surface so that a confidence region more closely resembled a straight line.

It seemed that the optimal design allowed us to obtain a more precise estimate of the Malthusian growth rate ( $\alpha$ ), but did practically nothing to improve our certainty about the reproductive ratio  $\lambda/\mu$ . Point estimates based on the likelihood surface are likely to be poor and grossly underestimate the death rate since the net change in the population is unlikely to

be negative between successive observations and the likelihood function therefore favours the model where the death rate is negligible. This appears to be the case even when the period between successive observations is very short and a decline in the population size is observed over some of the time intervals.

Figure 3 demonstrates the conditions under which the optimal design greatly outperforms the equidistant design. It is apparent that the differences are greatest when the sample size is small. When sample sizes are relatively large, the benefits are only likely to be realised for much longer observation periods (i.e. large values of  $t_{max}$ ). Even though the relative efficiencies in Figure 3 are based on FIM, our simulations suggest we can rely upon these for the non-elliptical contours observed. In practice, the optimal design is therefore likely to be most advantageous in situations where the number of samples is a more serious constraint than the length of time available for the sampling. This is likely to be the case in the ecological context, where sampling the population is often labour intensive and contributes greatly to the overall experimental cost, or, data is required in a relatively short time-frame in order to aid urgent management decisions.

We have seen that the likelihood surface presents some significant problems for point estimation and quantifying parameter uncertainty, even when the optimal sampling design is used. The practitioner might overcome these deficiencies by employing a Bayesian approach, which treats model parameters as random variables rather than as fixed but unknown constants in the parameter space. This approach to statistical inference can be particularly useful when the researcher has some prior ‘expert knowledge’ of the distribution of parameters across the parameter space. In simple terms, the practitioner uses available information to propose a joint probability density for the model parameters ( $p(\theta)$ ) prior to performing the experiment (commonly called the prior probability density). The data are then collected and the likelihood  $L(y|\theta)$  is calculated. Bayes rule is then used to calculate the posterior probability density of the parameters in light of the data as

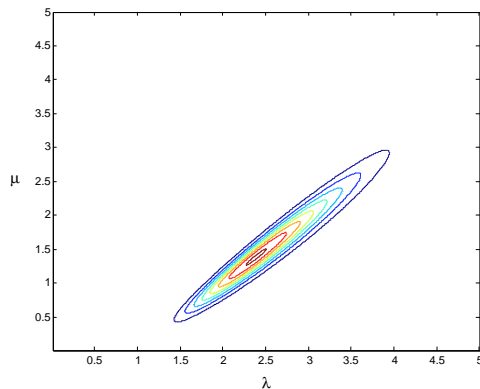
$$p(\theta|y) = \frac{L(y|\theta)p(\theta)}{\int L(y|\theta)p(\theta)d\theta}$$

For the simple birth-death process, the use of an informative prior for the per capita birth and death rates seems logical since we know that the likelihood surface has a very long ridge spanning a wide range of parameter values, many of which are likely to be absurd in a biological context. Under the Bayesian approach, posterior credible regions (similar to confidence regions) can be determined via Markov Chain Monte Carlo (MCMC) methods. It should be noted, however, that MCMC methods may suffer from poor mixing because of the high degree of correlation

between parameters. For this reason, it may be more desirable to work in the parameter space formed by  $(\lambda - \mu, \frac{\lambda}{\mu})$ .

Figure 4 illustrates how the Bayesian approach can aid our analysis and reduce parameter uncertainty. For our example we created a prior using independent Gamma densities for each of the parameters, such that  $\lambda \sim \text{Gam}(9, 1/3)$  and  $\mu \sim \text{Gam}(4, 1/2)$ . We see that even though the prior consisted of two independent gamma densities, the posterior density exhibits the correlation between the parameters that was noted as a characteristic of the likelihood surface and which is apparent in Figure 1. We note that the posterior mode in Figure 4 is heavily influenced by the prior probability density and highlights that, when using Bayesian inference, the prior should be based on the best available information.

Bayesian inference may be an avenue for future research in optimal design for population models.



**Figure 4.** Contours of the posterior density of the parameters of a simulated birth-death process observed using a 2 sample optimal design over the time interval  $[0, 1]$ .

## 6 CONCLUSION

We have presented an optimal design for the linear birth-death process, which is both D-optimal and approximately A-optimal for large initial population sizes. The optimal sampling times appear to be identical to those obtained by Becker and Kersting (1983) for the linear birth process and can therefore be formulated according to equations (1) and (2). The CE algorithm, used to find the optimal design, can easily be utilised for other population models where a Gaussian diffusion approximation is appropriate.

For the linear birth-death process, we have shown that the likelihood surface has some unusual features which hinder point estimation and the definition of confidence regions. We have suggested the use of Bayesian inference for overcoming the issues associated with the likelihood surface. Our exploration of

the optimal design and its use in reducing parameter uncertainty has shown that our optimal design should remain optimal even when the confidence region is based on contours of the likelihood surface, as is the case in Bayesian inference. Using numerical methods we have demonstrated that the optimal design for our model remains optimal even though the contours of the likelihood function are not elliptical. We have shown that for observations of a population where research constraints allow only a limited number of observations of a population over a fixed interval of time, the optimal design results in greater experimental efficiency than a scheme where sampling is undertaken at regular intervals, and this is most pronounced when the number of samples is small, but time is in relative abundance.

## 7 REFERENCES

- Atkinson, A.C. and A.N. Donev (1992), *Optimum Experimental Design*, Oxford University Press, Oxford.
- Becker, G. and G. Kersting (1983), Design problems for the pure birth process, *Advances in Applied Probability*, **15**(2), 255–273.
- Givens, G.H. and D. Poole (2002), Problematic likelihood functions from sensible population dynamics models: a case study, *Ecological Modelling*, **151**, 109–124.
- Keiding, N. (1975), Maximum likelihood estimation in the birth-and-death process, *The Annals of Statistics*, **3**(2), 363–372.
- Moran, P.A.P. (1953), The estimation of the parameters of a birth and death process, *Journal of the Royal Statistical Society, Series B*, **15**(2), 241–245.
- Pollett, P.K. (2001), In (Ed. Fred Ghassemi) *Proceedings of the International Congress on Modelling and Simulation Vol.2*, Modelling and Simulation Society of Australia and New Zealand, Australia, pp. 843–848.
- Porat, B. (1995), On the Fisher Information for the mean of a Gaussian process, *IEEE transactions on Signal Processing*, **43**(8), 2033–2035.
- Rubinstein, R.Y. and D.P. Kroese (2002), *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning*, Springer-Verlag, New York, USA.
- Ross, J.V., Taimre, T. and P.K. Pollett (2006), On parameter estimation in population models, *Theoretical Population Biology*, **70**, 498–510.

## 8 ACKNOWLEDGEMENT

We gratefully acknowledge the financial support of Australian Research Council Centre of Excellence for Mathematics and Statistics of Complex Systems.