Approximate Bayesian inference in a latent Gaussian model

Adam Butler and Chris Glasbey†
Biomathematics & Statistics Scotland, Edinburgh, UK
Sarah Wanless
Centre for Ecology and Hydrology, Banchory, UK

Summary. Methods of Approximation Bayesian Computation (ABC) provide a generic simulation-based framework for statistical inference, and are suited to models for which the likelihood function is analytically or computationally intractable. The efficiency of the approach can be improved by embedding the basis ABC step within a standard Markov chain Monte Carlo or sequential Monte Carlo algorithm. Most previous applications of ABC have been in genetics or epidemiology, where they are used for the analysis of unreplicated sequences of highly dependent data, but in this paper we apply a sequential variant of ABC to replicated data that arise from a specific latent Gaussian model. Applications to simulated and real data serve to illustrate the difficulties in applying ABC methods when the prior distribution is relatively diffuse, the number of parameters is relatively large, or the set of summary statistics is not apparent from the context of the problem.

Keywords: Latent Gaussian model; Compositional data; Unit-sum constraint; Zero proportion; Approximate Bayesian Computation; Sequential Monte Carlo; Intractable likelihood; Dietary composition; Seabirds

1. Introduction

Latent Gaussian models are used in a wide range of contexts (Tobin, 1958; Allcroft and Glasbey, 2003; Oigard et al., 2006), and continue to present challenging problems for statistical inference (Rue et al., 2007). We deal here with the situation in which it is difficult to evaluate the likelihood function but straightforward to simulate realisations from the model. Methods of Approximation Bayesian Computation (ABC) provide an attractive and generic approach to likelihood-free statistical inference, and, in this paper, we explore their performance in the context of a specific latent Gaussian model.

Butler and Glasbey (2007) propose a latent Gaussian model to describe the properties of "compositional data": data on the relative proportions associated with D different components. The proportions must sum to one, and must each lie between zero and one, so compositional data are constrained to lie on either the interior or boundaries of the unit simplex. There are established methods for analysing compositional data in which all of the proportions are non-zero (Aitchison, 1986), but methods for dealing with data that contain zero proportions are less well developed (Fry et al., 2000; Martín-Fernández et al., 2003; **?**; Bull et al., 2004). Butler and Glasbey (2007) assume that compositional data y are related

†*Address for correspondence:* Adam Butler, Biomathematics & Statistics Scotland, James Clerk Maxwell Building, The King's Building, Edinburgh EH9 3JZ, United Kingdom
E-mail: adam@bioss.ac.uk

to the values of a multivariate Gaussian random variable z, with unknown mean vector μ and covariance matrix Σ, via a deterministic transformation g that performs a Euclidean projection of the latent variable onto the unit simplex. Maximum likelihood estimates for the parameters $\theta = (\mu, \Sigma)$ can easily be derived by numerical optimisation when D is small, but for general D the calculation of the likelihood requires the evaluation of difficult integrals. Gibbs sampling can also be used as a basis for inference about this specific model, since an efficient variant of rejection sampling can be used to simulate values from the conditional distribution of the latent variable z given data x and parameters θ, but could not be used for alternative choices of the transformation function g. It is of interest to explore the performance of ABC methods against that of these alternatives, because ABC provides, in principle, a generic tool for inference across a much wider class of models, including those for which alternative approaches are not available.

In Section 2 we provide an overview of ABC methods, and in Section 3 discuss the choice of an appropriate distance metric in the context of the latent Gaussian model of Butler and Glasbey (2007). In Sections 4 and 5 we explore the performance of the method using simulated data, and in Section 6 apply the method to a real dataset concerning the composition of seabird diet. We conclude with a brief discussion.

2. Approximate Bayesian Computation

Methods of Approximate Bayesian Computation provide a generic approach for drawing inferences about the parameters of complex statistical and probabilistic models. ABC methods do not require us to compute the likelihood associated with the model; instead, they require only (a) that we are prepared to put a prior distribution $\pi(\theta)$ on the parameters θ, and (b) that we are able to repeatedly generate a large number of simulated datasets $y(\theta)$ using the model $f(x|\theta)$. ABC algorithms accept parameter values which simulate data with similar properties to those of the real data, x. The degree of similarity between the actual data and data simulated from the model using a particular parameter set (θ) is quantified by

$$\rho(y(\theta), x),$$

where ρ is a distance metric with the property that $\rho(y, x) = 0$ if $y = x$. In practise it will often be difficult to select an appropriate distance metric (Section 3).

2.1. Vanilla ABC

The most straightforward implementation of ABC - which is sometimes called the "vanilla" algorithm - involves repeatedly generating candidate values of θ from the prior, and accepting only those parameter sets for which the distance between x and $y(\theta)$ is less than or equal to a pre-specified threshold ϵ (Fu and Li, 1997; Weiss and von Haeseler, 1998; Pritchard et al., 1999). The algorithm generates N values from $\theta|\rho(x, y(\theta)) \leq \epsilon$, where the value of N is specified by the user. For sufficiently small ϵ this distribution should provide a good approximation to the posterior; if $\epsilon = 0$ then the algorithm generates values from the exact posterior. In formal terms, the algorithm involves:

V1 Setting $i = 1$

  V2 Generating $\theta \sim \pi(\theta)$;

  V3 If $\pi(\theta) > 0$ then generating $y(\theta) \sim f(y|\theta)$, else returning to step V1;

V4 If $\rho(y(\theta), x) < \epsilon$ then setting $\bar{\theta}_i = \theta$, else returning to step V1;

V5 If $i < N$ then setting $i = i + 1$ and returning to step V2.

Properties of the sample $\bar{\theta}_1, \ldots, \bar{\theta}_N$ can be used to derive approximate properties of the posterior distribution $\theta|x$.

2.2. ABC-MCMC

The vanilla algorithm will be highly inefficient if $\pi(\theta)$ is diffuse, if θ is of moderately large dimension, or if a very accurate approximation to the posterior is required. To overcome these difficulties, Marjoram et al. (2003) proposed embedding the ABC step - i.e. accepting θ if and only if $\rho(x, y(\theta)) < \epsilon$ - within a Metropolis-Hastings algorithm. This approach involves generating a sequence of dependent realisations, $\theta_1, \ldots, \theta_m$, from a Markov chain whose equilibrium distribution is equal to $\theta|(\rho(y(\theta), x) \leq \epsilon)$. Bortot et al. (2007) note that this algorithm will often exhibit very poor mixing, because excursions into the tails of the distribution are associated with a severe reduction in the acceptance rate. They propose an alternative version of the algorithm in which the parameter set θ is augmented by the threshold ϵ, leading to improved mixing; they assign a pseudo-prior to ϵ, and this choice controls the accuracy of the approximation generated by the algorithm.

2.3. Sequential ABC

Sequential Monte Carlo algorithms provide a powerful set of tools for Bayesian inference (Del Moral et al., 2006), and the development of new sequential algorithms is an active area of statistical research (Robert, 2004). Sisson et al. (2007) embed the ABC step within a sequential importance sampler, and show that the resulting algorithm performs well on simulated and real data; in particular, they demonstrate that the efficiency of the algorithm is substantially greater than that of the vanilla or unaugmented MCMC variants of ABC.

A special case of the sequential ABC algorithm arises if the prior distribution has uniform probability across a set Θ and zero probability elsewhere, and if we use a symmetric proposal distribution, $q(\theta^*|\theta)$. Assume that we have pre-specified a sequence of decreasing thresholds, $e_0 > e_1 > \ldots > e_{T-1} > e_T = \epsilon$. The algorithm is:

S1 Set $i = 1$;

S2 Generate $\theta^* \sim \pi(\theta)$;

S3 Generate $y \sim f(y|\theta^*)$;

S4 If $\rho(y(\theta^*), x) < e_0$ then set $\bar{\theta}_{0,i} = \theta^*$, else return to S2;

S5 If $i < N$ then set $i = i + 1$ and return to S2;

S6 Set $t = 1$ and $i = 1$;

S7 Sample θ^* at random from the sequence $\{\theta_{(t-1),i} \ldots, \theta_{(t-1),N}\}$;

S8 Generate $\theta^{**} \sim q(\theta|\theta^*)$;

S9 If $\theta^{**} \in \Theta$ then generate $y(\theta^{**}) \sim f(\mathbf{y}|\theta^{**})$, else return to S7;

S10 If $\rho(y(\theta^{**}), x) < e_t$ then set $\bar{\theta}_{t,i} = \theta^{**}$, else return to S7;

S11 If $i < N$ then set $i = i + 1$ and return to S7;

S12 If $t < T$ then set $t = t + 1$ and return to S7.

The sample $\bar{\theta}_{T,1}, \ldots, \bar{\theta}_{T,N}$ provides an iid set of N realisations from the target distribution, $\theta|\rho(x, y(\theta)) \leq \epsilon$.

We have found that the acceptance rate can typically be improved substantially by the inclusion of two additional steps,

S7a generate $y^* \sim f(\mathbf{y}|\theta^*)$, and

S7b if $\rho(y(\theta^*), x) \geq e_t$ then return to S7,

between steps **S7** and **S8** of the original algorithm, and this is the version which we use in our simulation study. There is, however, no theoretical basis for the inclusion of these extra steps into the algorithm.

2.4. Post-hoc adjustment

Beaumont et al. (2002) show that a post-hoc adjustment can be used to improve the approximation of the ABC output to the posterior, by using a local regression of θ on $\rho(y(\theta), x)$ to estimate the conditional distribution of $\theta|(\rho(y(\theta), x) = 0)$. This adjustment may be applied to output from the vanilla, sequential or augmented ABC algorithms.

3. Selection of a distance measure ρ

The function $\rho(y(\theta), x)$ quantifies the distance between $y(\theta)$ and x. This function must be sensitive to changes in the value of θ, so that $|\partial\rho(y(\theta), x)/\partial\theta|$ is relatively large across a wide range of values of θ. It must, however, also be relatively insensitive to sampling variability that arises from the generation of the simulated data, $y(\theta)$.

The best balance between these two requirements will typically be achieved by identifying a (fairly small) set of summary statistics, S, which capture the key properties of the observed and simulated datasets, and then comparing the distance $\rho(S(y(\theta)), S(x))$ between the two sets of summary statistics. No general procedure for the formal identification of a set of summary statistics has yet been developed, and the usual approach is to proceed using a mixture of prior information and trial and error.

For the latent Gaussian model of Butler and Glasbey (2007) we take the elements of $S(x)$ to be the marginal means and variances of the components of x, the means of the pairwise differences between components of x, and the proportion of zero and one values in each component of x. The elements are scaled in such a way that they are all constrained to lie on the range $[0, 1]$: marginal variances are multiplied by two, pairwise differences by dividing by two, and the remaining elements already lie on this range. We then take the distance measure ρ to be the mean of the absolute differences between the elements of $S(x)$ and $S(y(\theta))$.

These choices were made on the basis of trial and error. We found, in particular, that it is essential to include the proportion of zeroes and ones within $S(x)$, or the ABC algorithm will generate a poor approximation to the posterior even for very small values of ϵ. Qualitatively similar results were obtained by making obvious modifications to the set of summary statistics, or to the distance measure (e.g. using the mean sum of squares rather than the mean of the absolute differences), but we have made no attempt to select S or ρ using any kind of formal or systematic procedure.

4. Application to simulated data with two components

We use a simulation study to explore the performance of the sequence Monte Carlo ABC algorithm in the two component case, for which the likelihood is tractable and it is therefore possible to compare estimates obtained using ABC against those obtained using numerical maximum likelihood. When $D = 2$ the model contains just two unknown parameters: the mean μ and variance σ^2 associated with the first component of y. We simulate five datasets from the model, each of size $n = 200$ and each generated using the same seed for random number generation. The datasets are generated using a range of different parameter values:

2a: $\mu = 0.1$, $\sigma = 0.1$; **2b**: $\mu = 0$, $\sigma = 0.1$; **2c**: $\mu = -0.1$, $\sigma = 0.1$;

2d: $\mu = 0.5$, $\sigma = 0.5$; **2e**: $\mu = 0.5$, $\sigma = 1$.

The first three sets of parameters are associated with increasingly large probabilities of obtaining a zero proportion in component one (0.159 for 2a, 0.5 for 2b and 0.841 for 2c) but have a negligible probability (less than 0.001) of obtaining a zero proportion in component two. The last two set of parameters are associated with non-negligible probabilities of obtaining zero proportions in either of the components (0.159 in each component for 2d, 0.309 in each component for 2e).

Preliminary simulations suggest that the vanilla and unaugmented MCMC variants of ABC are highly inefficient in the context of the proposed latent Gaussian model model, and that they do not provide a viable basis for inference in this context. We therefore concentrate on the sequential variant of ABC, using the modified algorithm in which steps 7a and 7b are included. Prior distributions are taken to be relatively diffuse, so that $\pi(\mu) \sim U(-5, 5)$ and $\pi(\log \sigma) \sim U(-10, 2)$. The proposal distribution $q(\theta|\theta^*)$ is taken to be a bivariate normal distribution with mean θ^* and covariance matrix $\tau^2 I$, where $\tau = 0.05$. We take $\epsilon = 1/500$, $N = 1000$, $e_0 = 1/10$, and intermediate thresholds to be of the form $e_t = 2e_0/(t + 2)$ for $t \leq 98$.

In Figure 1 we compare results obtained using the sequential ABC algorithm against those obtained using numerical maximum likelihood; note that the results obtained using maximum likelihood are virtually identical to those obtained using a standard Metropolis-Hastings algorithm. Plots are shown for three of the datasets: results for 2b are qualitatively similar to those for 2a, and results for 2e are similar to those for 2d. The plots for all of the parameters and datasets suggest that the means and medians of the ABC samples do converge towards the maximum likelihood estimate as e_t tends towards zero, and that convergence is relatively rapid (with results changing minimally for e_t smaller than 1/300). The more extreme quantiles of the ABC samples (2.5%, 25%, 75%, 97.5%) become increasingly close to the corresponding quantiles of the distribution of the MLE as e_t becomes small, but for datasets 2d and 2e, and to a lesser extent 2a, they continue to systematically underestimate the uncertainty within the estimator even when the threshold ϵ is very small. For dataset 2a we investigated the impact of reducing the range of the (uniform) prior distributions, of changing the seed used for random number generation, and of reducing the number of particles N (from 1000 down to 200), but we found that these modifications all had a negligible impact on the results obtained.

We measure the efficiency of the algorithm using $\Gamma(\epsilon)$, which we define to be the ratio of N, the number of iid realisations that we generate from the target distribution $\theta|\rho(y(\theta), x) < \epsilon$, to the total number of simulated datasets required to attain this sample (Table 1). We see that the algorithm is much less efficient for the datasets that contain both zero and one

Fig. 1. Results from fitting the latent Gaussian model to simulated datasets 2a, 2c and 2d. We plot 2.5% (solid), 25% (dotted), 50% (thick solid), 75% (dotted) and 97.5% (solid) quantiles for the parameters μ and $\log \sigma$, as obtained using the sequential ABC algorithm of Section 2.3. Corresponding quantiles of the maximum likelihood estimators for μ and $\log \sigma$ are also shown (grey).

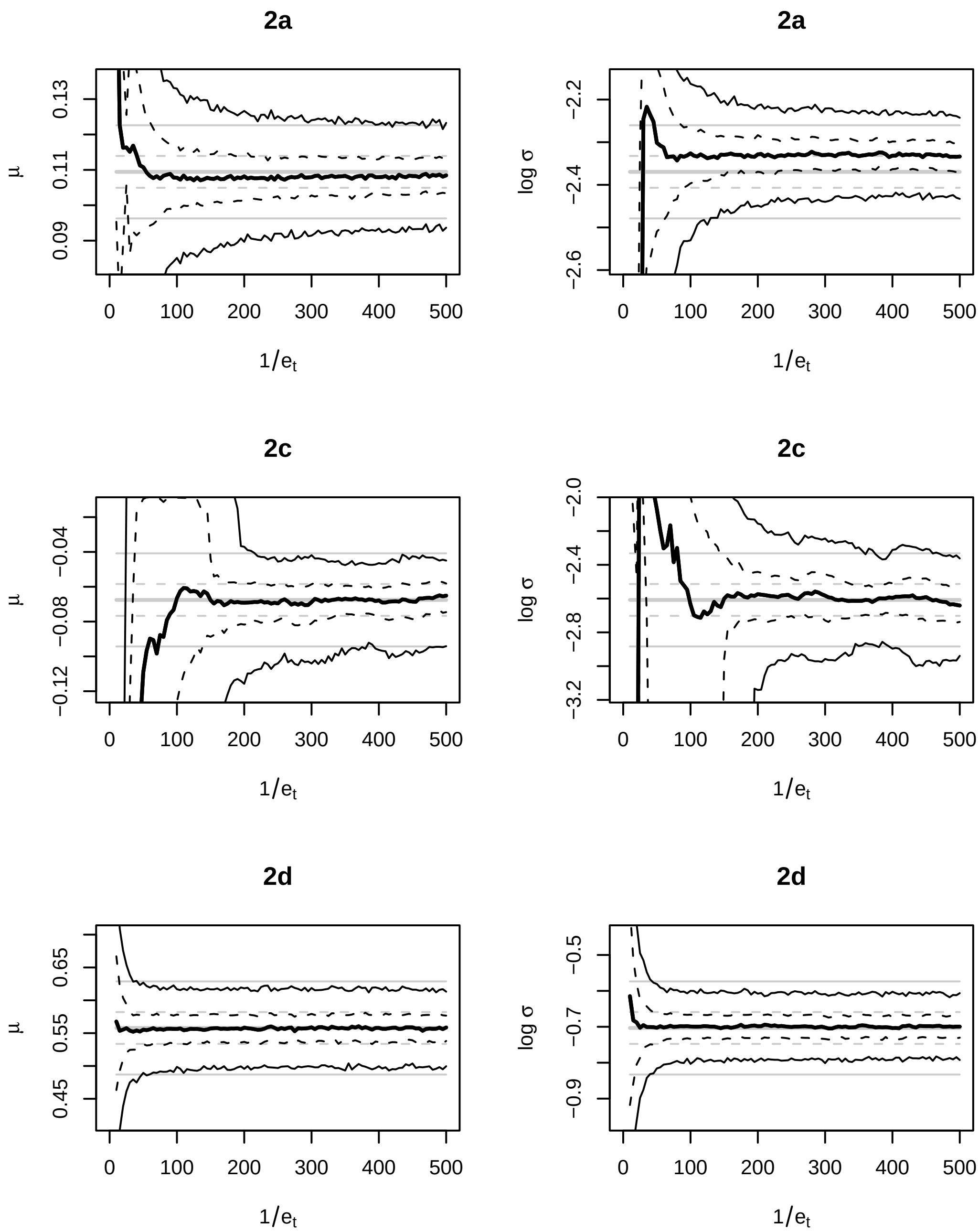


Table 1. Number of simulations per particle, $\Gamma(\epsilon)$, when applying the sequential ABC algorithm of Section 2.3 to datasets 2a-2e.

Dataset	τ	$\epsilon = 1/125$	$\epsilon = 1/250$
2a	0.05	186	657
2b	0.05	400	1005
2c	0.05	473	1063
2d	0.05	596	5267
2e	0.05	552	9493
2a	0.01	117	285
2a	0.2	636	3582
2a	0.5	2525	18276

values (2d and 2e) than for those which only contain zero values (2a-2c). The ratio when fitting the vanilla, rather than sequential, ABC algorithm to dataset 2a with a threshold of $\epsilon = 1/125$ is equal to approximately 21000, illustrating why this approach is not viable for the problem at hand. The efficiency of the sequential algorithm depends heavily on the value of the proposal standard deviation τ: small values of τ lead to much higher acceptance rates than do larger values, but the price for this improvement in efficiency is an increase in the risk that the chain will exhibit poor mixing.

5. Application to simulated data with three components

When $D = 3$ our model contains five unknown parameters - the mean μ_1 and standard deviation σ_1 of the first component, the mean μ_2 and standard deviation σ_2 of the second component, and the correlation ρ between the first and second components. We simulate three datasets from our model, each of size $n = 200$ and each generated using the same seed for random number generation. We take $\mu_1 = 1/3$, $\mu_2 = 1/3$, $\rho = -1/2$ (all datasets), and take $\sigma_1 = \sigma_2 = 0.1$ (dataset 3a), $\sigma_1 = \sigma_2 = 0.5$ (dataset 3b) or $\sigma_1 = \sigma_2 = 1$ (dataset 3c), so that the proportion of zero values is higher for dataset 3c than for 3b and higher for 3b than for 3a. We compare results obtained using maximum likelihood against those obtained using the modified version of the sequential ABC algorithm. For the ABC algorithm the prior distributions are taken to be $\pi(\mu_i) \sim U(-5, 5)$, $\pi(\log\sigma) \sim U(-10, 2)$ and $\pi(\rho) \sim U(-1, 1)$. We take $N = 1000$, $q(\theta^*|\theta) \sim N(\theta, \tau^2 I)$ where $\tau = 0.05$, $N = 1000$, $\epsilon = 1/125$, $e_0 = 1/2$, and intermediate thresholds to be of the form $e_t = 2e_0/(t + 2)$ for $t \leq 123$.

Results are shown in Figure 2 for the parameters μ_1 and ρ; qualitatively similar results were obtained for the remaining three parameters. We see that there the median values of the ABC samples generally tend to converge towards the MLE in a reasonably smooth way as e_t becomes small, but that there are some parameters (e.g. ρ for dataset 3a) for which bias appears to persist even as ϵ becomes very small. More significantly, the ABC procedure again tends to markedly underestimate the level of parameter uncertainty, especially for those datasets (3b and 3c) in which the data exhibit a relatively high degree of variability.

We use a substantially larger value for ϵ here than in the two component case, for computational reasons - the acceptance rate of algorithm is much lower when $D = 3$ than when $D = 2$ (Table 2), and the acceptance rate drops very sharply between $\epsilon = 1/125$ and

Fig. 2. Results from fitting the latent Gaussian model to simulated datasets 2a, 2c and 2d. We plot 2.5% (solid), 25% (dotted), 50% (thick solid), 75% (dotted) and 97.5% (solid) quantiles for the parameters μ_1 and ρ, as obtained using the sequential ABC algorithm of Section 2.3. Corresponding quantiles of the maximum likelihood estimators for μ_1 and ρ are also shown (grey).

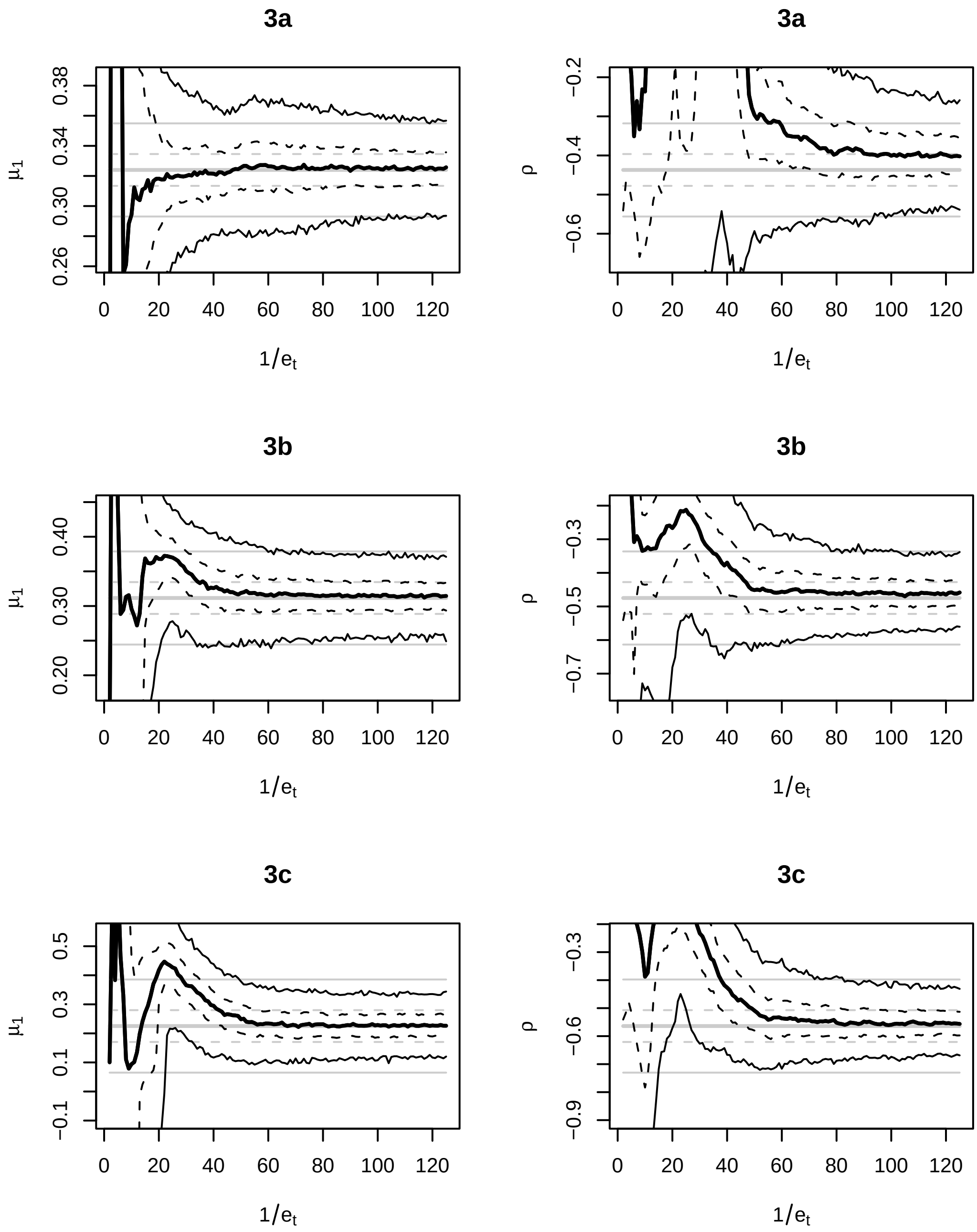


Table 2. Number of simulations per particle, $\Gamma(\epsilon)$, when applying the sequential ABC algorithm of Section 2.3 to datasets 3a-3c.

Dataset	$\epsilon = 1/50$	$\epsilon = 1/100$	$\epsilon = 1/125$
3a	319	1031	2118
3b	248	2253	8230
3c	219	2161	7253

$\epsilon = 1/150$ (not shown). The lower efficiency is related to the higher dimensionality of the parameter space (five rather than two).

6. Application to seabird diet data

We also investigate the performance of ABC methods when fitting the latent Gaussian model of Butler and Glasbey (2007) to a real dataset concerning seabird diet. The data record the relative frequencies of three different prey types - lesser sandeel *ammodytes marinus* aged less than one year (SE0), mature lesser sandeel (SE1), and other species - within samples of regurgitated material collected from four island colonies of the Black-legged Kittiwake *Rissa tridactyla* on the east coast of Britain during the period 1997-2000 (Bull et al., 2004). The data are plotted in Figure 3. The proportion of zero values is very high - only 22 of the 544 observations lie on the interior of the unit simplex, and the maximum likelihood μ lies well beyond the simplex - so these data present a challenging problem for statistical inference. For simplicity, we ignore the effects of year, colony and time of year in our analysis, although Bull et al. (2004) demonstrate that the latter two factors do have a significant effect upon dietary composition.

In Figure 4 we compare the results obtained using the sequential ABC of Section 2.3 against those obtained via numerical maximum likelihood. We see that for all five of the parameters the medians of the ABC samples have converged fairly well to those of the maximum likelihood estimates by the time that we reach a threshold of $\epsilon = 1/100$, although convergence occurs more slowly than for the simulated data of Sections 4 and 5 and estimates for the correlation parameter ρ exhibit bias when $\epsilon = 1/100$. The sudden jumps in the properties of the ABC samples arise because of the discretised nature of the diet data - proportions are almost always rounded to the nearest 5%.

7. Concluding remarks

ABC methods were originally developed for use in genetics and evolutionary biology (Beaumont et al., 2002; Leman et al., 2005; Thornton and Andolfatto, 2006). Applications in these disciplines are characterised by the nature of the data (highly dependent and unreplicated), by the presence of relatively strong prior information, and by the relatively small number of parameters that are of interest. In this paper we have sought to apply the same methods within a different statistical context, in which replicated iid data arise from a model about that contains a relatively large number of parameters, about which we have little or no prior information.

Our results illustrate many of the generic issues involved in using ABC methods for statistical inference: the difficulty in choosing appropriate summary statistics S and an appropriate distance metric ρ, the difficulty in constructing an efficient algorithm if the

Fig. 3. Kittiwake diet data, plotted using a ternary diagram. The ternary diagram displays the relative frequency associated with the three different prey types, with the vertices representing those individuals which consume only a single type and the edges representing those individuals that consume two of the three types. Data points are represented by red circles whose area is proportional to the number of observations associated with that point - the area of the black circle represents the contribution of a single observation.

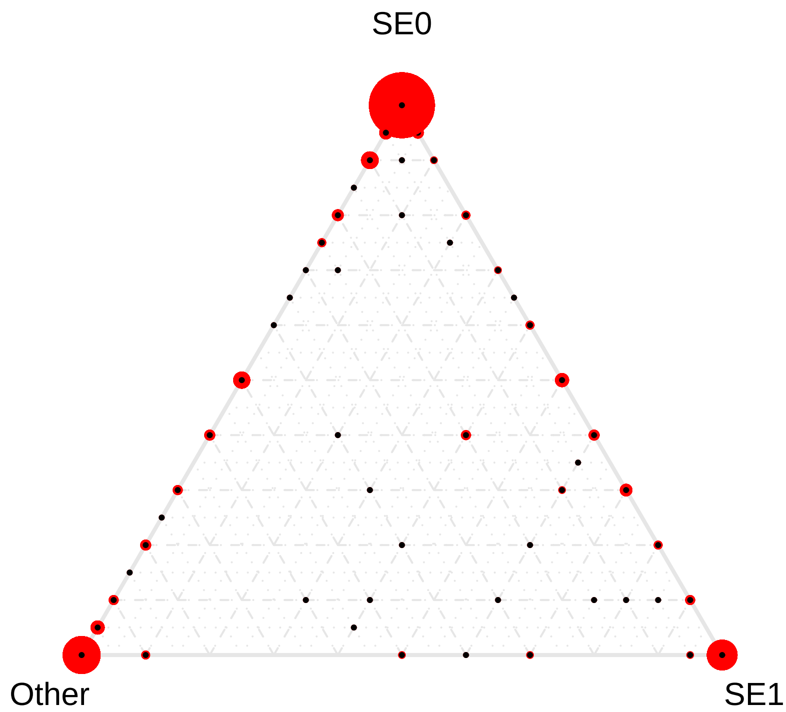
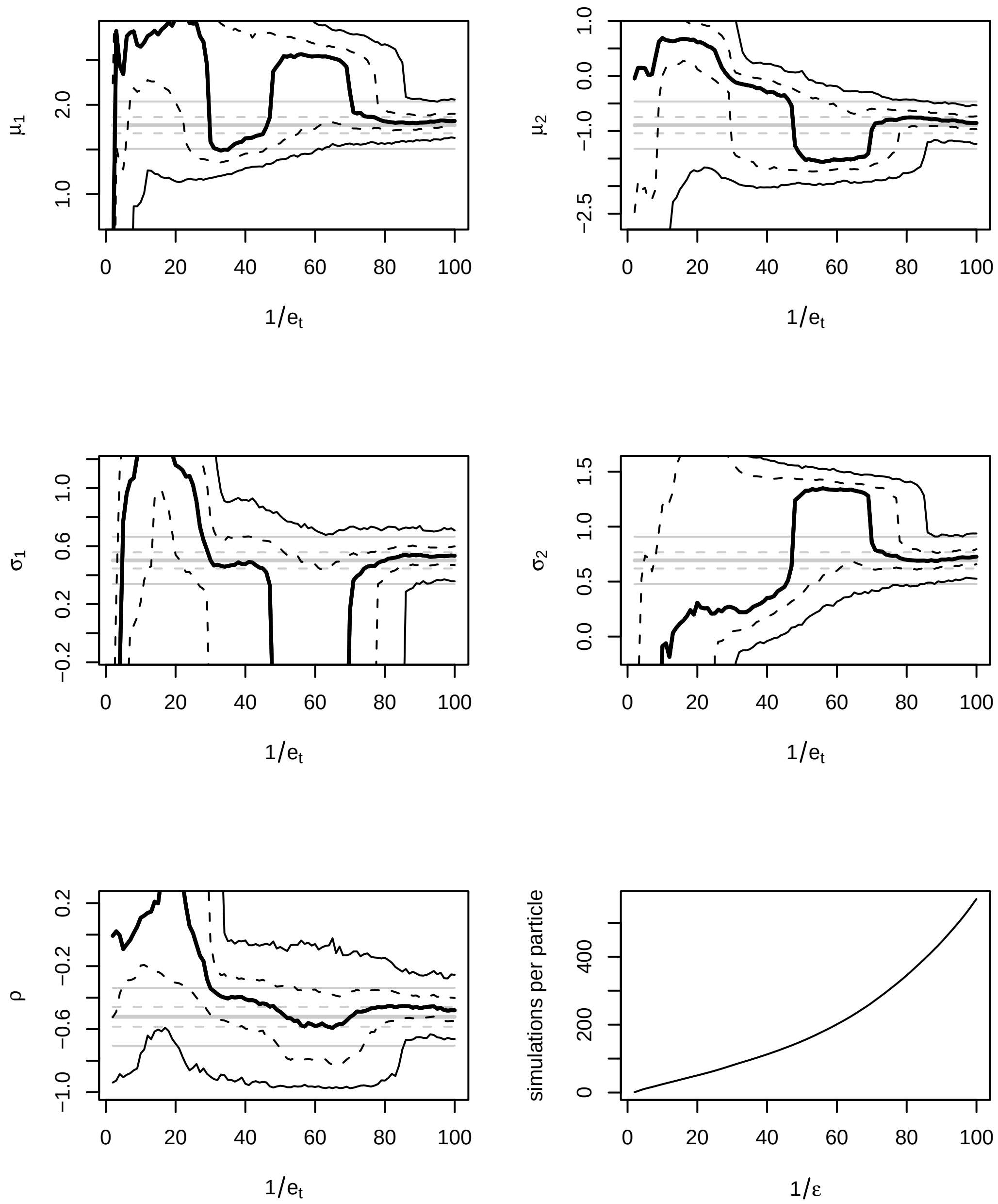


Fig. 4. Results from fitting the latent Gaussian model to kittiwake diet data. We plot 2.5% (solid), 25% (dotted), 50% (thick solid), 75% (dotted) and 97.5% (solid) quantiles for the parameters μ and $\log\sigma$, as obtained using the sequential ABC algorithm of Section 2.3. Corresponding quantiles of the maximum likelihood estimators for μ and $\log\sigma$ are also shown (grey).



prior distribution π is relatively diffuse or the size of the parameter size is moderately large, and the difficulty in evaluating the accuracy of the approximation upon which the ABC algorithm is based. ABC methods are, ultimately, able to provide a moderately good approximation to the marginal means of the posterior distribution associated with our model, even when the data contain a very high proportion of zero values, but have substantial difficulties in capturing the posterior variances and covariances. ABC methods are highly inefficient relative to the alternative methods (Gibbs sampling, or, for small D, numerical maximum likelihood) that are available for drawing inferences about the parameters of the latent Gaussian model of Butler and Glasbey (2007), but may still provide a useful tool to deal with similar, more complex, models for which alternative methods are not available.

Acknowledgements

Funding for this work was provided by the Scottish Government, and helpful comments on the ABC methodology were provided by Tony Pettitt and Scott Sisson.

The seabird monitoring work was co-commissioned by JNCC and NERC as part of its programmes of monitoring and research into nature conservation. We are grateful to Bernadette Clarke, David Hammond and Robin Cox for help with collecting diet samples on the Isle of May, to John Davies, Alan Leitch, Martin Moss and other members of the Lothian Ringing Group for samples from Inchkeith, to Brian and Stephanie Little and other members of the Fife Ringing Group for samples from Inchcolm, and to John Walton and the National Trust wardens for samples from the Farne Islands. Scottish Natural Heritage allowed us to work on the Isle of May. We would like to thank Iain Gibb for help with otolith identification.

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data.* London: Chapman and Hall.

Allcroft, D. J. and C. A. Glasbey (2003). A latent Gaussian Markov random field model for spatio-temporal rainfall disaggregation. *Applied Statistics 52*, 487–498.

Beaumont, M. A., W. Zhang, and D. J. Balding (2002). Approximate Bayesian computation in population genetics. *Genetics 162*, 2025–2035.

Bortot, P., S. G. Coles, and S. A. Sisson (2007). Inference for stereological extremes. *JASA 102*, 84–92.

Bull, K., S. Wanless, D. A. Elston, F. Daunt, S. Lewis, and M. P. Harris (2004). Local-scale variability in the diet of Black-legged Kittiwakes *rissa tridactyla*. *Ardea 92*(1), 43–52.

Butler, A. and C. A. Glasbey (2007). A latent gaussian model for compositional data with structural zeroes. In preparation.

Del Moral, P., D. A., and J. A. (2006). Sequential monte carlo samplers. *J. Roy. Statist. Soc. Ser. B 68*(3), 411–436.

Fry, J., T. R. L. Fry, and K. R. McLaren (2000). Compositional data analysis and zeros in micro data. *Applied Economics 32*(8), 953–959.

Fu, Y. X. and W. H. Li (1997). Estimating the age of the common ancestor of a sample of DNA sequences. *Mol. Biol. Evol. 14*, 195–199.

Leman, S. C., Y. Chen, J. E. Stajich, M. A. F. Noor, and M. K. Uyenoyama (2005). Likelihoods from summary statistics: recent divergence between species. *Genetics 171*(3), 1419–1436.

Marjoram, P., J. Molitor, V. Plagnol, and S. Tavare (2003). Markov chain Monte Carlo without likelihoods. *PNAS 100*(26), 15324–15328.

Martín-Fernández, J. A., C. B. Barceló-Vidal, and V. Pawlowsky-Glahn (2003). Dealing with zeros and missing values in compositional data sets using nonparametric imputation. *Mathematical Geology 35*(3), 253–278.

Oigard, T. A., H. Rue, and F. Godtliebsen (2006). Bayesian multiscale analysis for time series data. *Computational Statistics & Data Analysis 31*, 1719–1730.

Pritchard, J. K., M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol. 16*, 1791–1798.

Robert, C. P. (2004). *Monte Carlo Statistical Methods.* Springer Texts in Statistics. New York: Springer-Verlag.

Rue, H., M. S., and N. Chopin (2007). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. submitted.

Sisson, S. A., Y. Fan, and M. M. Tanaka (2007). Sequential Monte Carlo without likelihoods. *PNAS 104*, 1760–1765.

Thornton, K. and P. Andolfatto (2006). Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *drosophila melanogaster*. *Genetics 172*(3), 1607–1619.

Tobin, J. (1958). Estimation for relationships with limited dependent variables. *Econometrica 26*(1), 24–36.

Weiss, G. and A. von Haeseler (1998). Inference of population history using a likelihood approach. *Genetics 149*, 1539–1546.