A spectral estimator of ARMA parameters from thresholded data

DAVID J. ALLCROFT
Department of Mathematics and Statistics, University of Edinburgh and Biomathematics & Statistics Scotland,
King's Buildings, Edinburgh, EH9 3JZ, UK; dave@bioss.ac.uk

CHRIS A. GLASBEY
Biomathematics & Statistics Scotland,
King's Buildings, Edinburgh, EH9 3JZ, UK

Abstract

We consider computationally-fast methods for estimating parameters in ARMA processes from binary time series data, obtained by thresholding the latent ARMA process. All methods involve matching estimated and expected autocorrelations of the binary series. In particular, we focus on the spectral representation of the likelihood of an ARMA process and derive a restricted form of this likelihood, which uses correlations at only the first few lags. We contrast these methods with an efficient but computationally-intensive Markov chain Monte Carlo (MCMC) method. In a simulation study we show that, for a range of ARMA processes, the spectral method is more efficient than variants of least squares and much faster than MCMC. We illustrate by fitting an ARMA(2,1) model to a binary time series of cow feeding data.
Keywords: ARMA process; Censoring; Cow feeding data; MCMC; Simulation; Spectral likelihood; Tetrachoric correlation.

1 Introduction

Parameter estimation for fully-observed autoregressive moving average (ARMA) processes is well documented: recursive techniques based on the Kalman filter and state-space representation can be used, allowing exact Gaussian likelihoods to be computed efficiently (see, for example, Lütkepohl, 1991). However, for censored data, state-space representations do not exist and alternative methodologies have to be considered. Here we investigate the case of binary time series data, obtained by applying a threshold to a latent Gaussian process.

Our motivation comes from categorical animal behaviour data, in particular from observations on cow feeding: Figure 1 shows two days of such data. We postulate the existence of a latent,

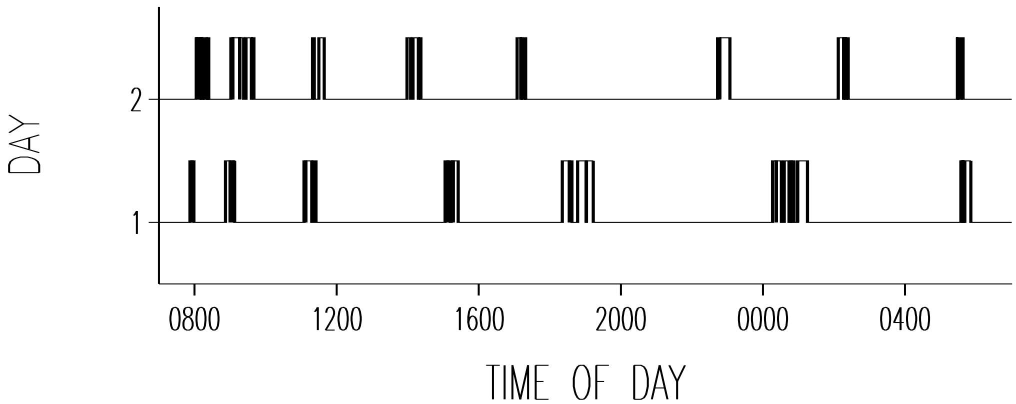

Figure 1: *Two days of feeder-visit data for one cow — raised values of the signal denote periods of feeding*

Gaussian-distributed, physiological variable, with feeding occurring when this variable exceeds a threshold. For categorical data, latent variables offer a more flexible approach than the use of either stochastic compartment models or hidden Markov models (MacDonald and Zucchini, 1997), because they simplify the inclusion of diurnal cycles, covariates and multivariate dependencies between animals. Figure 2 shows simulations of such latent and thresholded processes, in this case an ARMA(2,1) model. Note that we have chosen to use a deterministic link between the latent and categorical variables (Cox and Snell, 1989); an alternative would have been a stochastic link such as a logistic response (Keenan, 1982). We will return to this example in Section 4.

Likelihood expressions are complicated when data are censored, so our approach is to first estimate the autocorrelations of either the observed binary series or the unobserved Gaussian series, and then explore the use of a range of methods to estimate ARMA parameters by matching the autocorrelations with their expected values. In particular, we focus on the spectral representation of the likelihood of a multivariate ARMA process (Whittle, 1953), and derive a *restricted* form of this likelihood, which uses correlations at only the first few lags and so makes considerable saving in computing time. We compare this estimator with others, including least squares, use of a pairwise likelihood, and a fully-efficient but computationally-intensive method using Markov chain Monte Carlo (MCMC) techniques.

In Section 2, we consider the estimation and relationship between the autocorrelations of the observed and latent processes, derive the spectral likelihood, discuss its restricted form and outline other methods for estimating parameters in latent ARMA models. Then, in Section 3, we show simulation results to compare the efficiencies of these estimators for a range of ARMA models. In Section 4 we fit an ARMA model to the cow feeding data, and finally, in Section 5 we discuss results and ideas for further work.

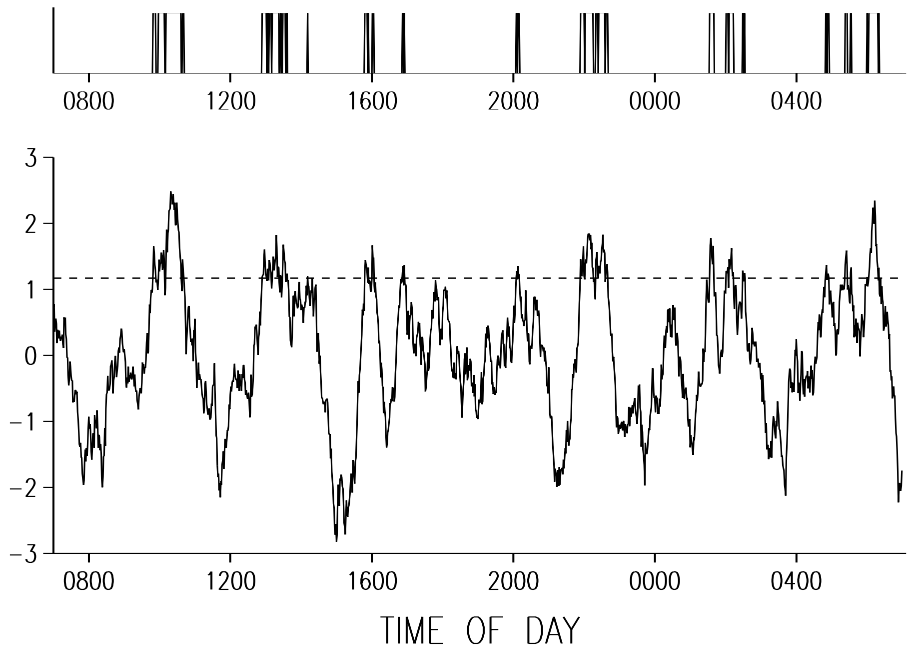

Figure 2: *A simulated ARMA(2,1) series with parameter values as estimated in Section 4. The threshold level matches that observed. Corresponding feeding events are shown above*

2 Parameter estimation

Let $Y_1, \ldots, Y_n$ denote a stationary Gaussian time series with expected autocorrelation $\rho_l^{(G)}$ at time lag l, and let $X_1, \ldots, X_n$ denote a binary time series with

$$X_t = \begin{cases} 0 & \text{if } Y_t \leq T \\ 1 & \text{otherwise} \end{cases}$$

for some threshold T. The expected autocorrelation of X_t at lag l, denoted $\rho_l^{(B)}$, is related to $\rho_l^{(G)}$ as follows:

$$\begin{aligned} \rho_l^{(B)} &= \frac{E(X_t X_{t+l}) - E(X_t)^2}{E(X_t^2) - E(X_t)^2} \\ &= \frac{\mathrm{P}(X_t = 1, X_{t+l} = 1) - \mathrm{P}(X_t = 1)^2}{\mathrm{P}(X_t = 1) - \mathrm{P}(X_t = 1)^2} \\ &= \frac{(\Phi_{T,T}(\rho_l^{(G)}) - 2\Phi_T + 1) - (1 - \Phi_T)^2}{(1 - \Phi_T) - (1 - \Phi_T)^2} \\ &= \frac{\Phi_{T,T}(\rho_l^{(G)}) - \Phi_T^2}{\Phi_T - \Phi_T^2}. \end{aligned} \tag{1}$$

Here, Φ_T is the standard normal integral from $-\infty$ to T and $\Phi_{T,T}$ is the corresponding integral for a bivariate distribution with correlation $\rho_l^{(G)}$. The threshold can be estimated as $\hat{T}$ such that

$$\Phi_{\hat{T}} = \frac{n - \sum_{i=1}^{n} x_t}{n},$$

where we follow the standard convention of denoting random variables by upper case (X_t, Y_t) and realisations by lower case (x_t, y_t). Equation (1) can also be used to transform from the sample autocorrelations of the binary time series, denoted $\hat{\rho}_l^{(B)}$ at lag l, to sample autocorrelations in the latent Gaussian process, $\hat{\rho}_l^{(G)}$, also termed *tetrachoric correlations.*

We now restrict attention to ARMA processes. The pth order autoregressive qth order moving average process is specified by

$$Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \ldots - \theta_q \epsilon_{t-q}, \quad t = -\infty, \ldots$$

where ϵ are independent normal random variables with zero mean and variance set to be such that Y has unit variance. The expected autocorrelations, $\rho_l^{(G)}$, are functionally related to the parameters $\phi_1, \ldots, \phi_p$ and $\theta_1, \ldots, \theta_q$, (see, for example, Box et al., 1994, section 3.4).

An *ad hoc* way to estimate ARMA parameters from these autocorrelations is via least squares, i.e. find the values of the parameters that minimise

$$\sum_{l=1}^{n'/2} (\hat{\rho}_l - \rho_l)^2$$

using either binary or Gaussian autocorrelations and some choice of n'. However, sample autocorrelation coefficients at different lags are highly correlated, so this is not necessarily an efficient estimation procedure. An alternative is to transform to independent statistics, for which the natural choice is by the Fourier transform, as follows.

2.1 Spectral likelihood

Whittle (1953) derived the spectral approximation for the log-likelihood, $\mathcal{L}$, of an m-dimensional stationary multivariate Gaussian process of length n. It has the form of a set of independent complex Wishart distributions (Brillinger, 1975),

$$\mathcal{L} = -\frac{1}{2} \sum_{k=-\frac{n}{2}}^{\frac{n}{2}-1} \log |S_k| - \frac{1}{2} \sum_{k=-\frac{n}{2}}^{\frac{n}{2}-1} \mathrm{trace}[{S_k}^{-1} \hat{S}_k]. \tag{2}$$

The limits of the sums here assume that n is even; if n is odd the limits become $\pm(n-1)/2$ and subsequent expressions change accordingly. S_k and $\hat{S}_k$ are, respectively, the $m \times m$ complex matrices of cross-spectral and cross-periodogram coefficients at frequency $2\pi k/n$, so

$$\hat{S}_k = \sum_{l=-\frac{n}{2}}^{\frac{n}{2}-1} \hat{\rho}_l^{(G)} e^{-2\pi ikl/n} \tag{3}$$

with S similarly defined in terms of $\rho^{(G)}$. For (2) only integer values of k need to be considered, but (3) also holds for non-integer k, required in the Appendix.

In our application we only consider univariate series, but for full generality here we allow the series to be multivariate. For rigorous details of the proof of (2), see Coursol and Dacunha-Castelle (1983). Note that in our application the Gaussian process is latent, so (2) can only be considered as a quasi-likelihood, and we also consider the same functional expression but with $\rho^{(G)}$ replaced by $\rho^{(B)}$.

Glasbey et al. (1998) replaced n by n' in (2), to obtain a restricted likelihood, $\mathcal{L}'$:

$$\mathcal{L}' = -\frac{n}{2n'} \sum_{k=-\frac{n'}{2}}^{\frac{n'}{2}-1} \log |S'_k| - \frac{n}{2n'} \sum_{k=-\frac{n'}{2}}^{\frac{n'}{2}-1} \text{trace}[S'^{\,-1}_k \hat{S}'_k], \tag{4}$$

where S' and $\hat{S}'$ are obtained as the discrete Fourier transforms of cross-correlations up to lag $n'/2$ only, by replacing n by n' in (3). They showed empirically, using uncensored AR(1) and ARMA(1,1) models as examples, that maximising $\mathcal{L}'$ is equivalent to maximising $\mathcal{L}$, in that as long as n' is sufficiently large, the difference between the resulting parameter estimates from maximisation of $\mathcal{L}$ and $\mathcal{L}'$ is negligible. Here, we prove that for short-memory processes such as ARMA models, $\mathcal{L}' \approx \mathcal{L}$ for sufficiently large n'. In practice $n' \ll n$, which leads to considerable computational saving. The main conditions needed are that S_k is a continuous function of k and that $\rho_l^{(G)}$ and α_l are negligible for $|l| > n'/2$, where α_l is the inverse autocorrelation coefficient at lag l, defined as

$$\alpha_l = \sum_{k=-\frac{n}{2}}^{\frac{n}{2}-1} S_k^{\,-1} e^{2\pi ikl/n} \qquad \text{for } l = -\frac{n}{2}, \ldots, \frac{n}{2} - 1, \tag{5}$$

using a multivariate generalisation of inverse autocorrelations considered by Cleveland (1972) and Chatfield (1979). Full details of the proof are given in the Appendix. These conditions hold for ARMA processes, typically for small values of n', because autocorrelations and inverse autocorrelations decay exponentially (Chatfield, 1979; Box et al., 1994). The rate of convergence, illustrated in Table 1, depends on the rate of decay of ρ and α. Typically, the likelihoods in the table are approximated to within ± 0.0005 of their true value if n' is such that ρ_l and α_l are of the order 10^{-4} for $l > n'/2$.

n'	AR(1) $\phi = 0.6$ $\rho_{n'/2}$	$\alpha_{n'/2}$	$\mathcal{L}'$	MA(1) $\theta = -0.3$ $\rho_{n'/2}$	$\alpha_{n'/2}$	$\mathcal{L}'$	ARMA(1,1) $(\phi, \theta) = (0.9, 0.6)$ $\rho_{n'/2}$	$\alpha_{n'/2}$	$\mathcal{L}'$
2	0.60	-0.44	391.119	0.28	-0.30	481.004			
4	0.36	0.00	282.468	0.00	0.090	456.386	0.44	-0.11	373.575
10	0.078	0.00	281.388	0.00	-0.0024	458.353	0.32	-0.025	329.603
20	0.0060	0.00	282.237	0.00	$< 10^{-5}$	458.356	0.19	-0.0019	317.364
50	$< 10^{-5}$	0.00	282.239	0.00	$< 10^{-13}$	458.356	0.039	$< 10^{-6}$	315.662
100	$< 10^{-11}$	0.00	282.239	0.00	$< 10^{-15}$	458.356	0.0028	$< 10^{-11}$	315.770
500	$< 10^{-15}$	0.00	282.239	0.00	$< 10^{-15}$	458.356	$< 10^{-11}$	$< 10^{-15}$	315.775
1000	$< 10^{-15}$	0.00	282.239	0.00	$< 10^{-15}$	458.356	$< 10^{-15}$	$< 10^{-15}$	315.775

Table 1: *Values of $\mathcal{L}'$ for a range of n', at true parameter values, averaged over 100 simulated series of length 1000. The values of correlation (ρ) and inverse-correlation (α) coefficients are shown for lag $n'/2$. Where upper bounds are shown, this is for the magnitude only, the sign being omitted*

2.2 Other fast methods

Weighted and generalised least squares are natural methods to consider as improvements over ordinary least squares. However, analytical forms cannot be derived for the variances and covariances of the tetrachoric correlation coefficients. But, as the tetrachoric correlations are estimates of the Gaussian autocorrelations, this suggests using the variances of these sample autocorrelations instead. We therefore consider an $n'/2 \times n'/2$ weight matrix W, with elements

$$W_{kl} = \mathrm{Cov}(\hat{\rho_k}, \hat{\rho_l}) = \frac{1}{n} \sum_{i=-\infty}^{\infty} \left(\rho_i \rho_{i+l-k} + \rho_{i-k} \rho_{i+l} - 2\rho_l \rho_i \rho_{i+k} - 2\rho_k \rho_i \rho_{i+l} + 2\rho_i^2 \rho_k \rho_l \right)$$

(see, for example, Kendall et al., 1983). Then, for weighted least squares we propose to minimise

$$\sum_{l=1}^{n'/2} \frac{1}{W_{ll}} (\hat{\rho}_l^{(G)} - \rho_l^{(G)})^2, \tag{6}$$

and for generalised least squares, we minimise

$$(\hat{\rho}^{(G)} - \rho^{(G)})^T W^{-1} (\hat{\rho}^{(G)} - \rho^{(G)}), \tag{7}$$

where $\hat{\rho}^{(G)}$ and $\rho^{(G)}$ are vectors of length $n'/2$ containing the sample and expected correlations up to lag $n'/2$. Use of (6) or (7) is equivalent to the methods used by Cressie (1985) for the fitting of variogram models. Due to concerns about the inconsistency of these estimators, alternative expressions were formed by the inclusion of the associated log-determinant terms, making the approach a maximum Gaussian likelihood one. Simulations showed the resulting estimators to be substantially more computationally intensive and no more efficient than the simpler ones and so we present results only for the estimators given directly by (6) and (7).

Applications in spatial statistics have used pairwise likelihood approaches for covariance estimation (Hjort and Omre, 1994). The quasi-likelihood is formed as the product of all bivariate probabilities for pairs of observations, (y_s, y_t), with $s < t \leq s + n'/2$,

$$Q = \prod \frac{1}{\sqrt{2\pi(1 - \rho_{(t-s)}^{(G)}{}^2)}} \exp\left[-\frac{1}{2} \frac{y_s^2 + y_t^2 - 2\rho_{(t-s)}^{(G)} y_s y_t}{1 - \rho_{(t-s)}^{(G)}{}^2}\right],$$

which simplifies to the minimisation of

$$\sum_{l=1}^{n'/2} \log(1 - \rho_l^{(G)2}) + 2\sum_{l=1}^{n'/2} \left(\frac{1 - \hat{\rho}_l^{(G)} \rho_l^{(G)}}{1 - \rho_l^{(G)2}}\right).$$

2.3 MCMC methods

All the methods above are computationally fast, especially when $n' \ll n$. Here we outline a much more computer-intensive method using Markov chain Monte Carlo (MCMC), which yields a fully efficient estimator. Gibbs sampling is used to simulate realisations of the latent Gaussian series consistent with the thresholding dictated by the binary series. For each realisation of the complete series, estimates of the parameters are obtained by sampling from the likelihood via a Metropolis-Hasting step. Finally, we average these estimates over a large number of iterations. Full details of the methodology used are given in Allcroft (2001).

The general methodology is essentially the same for the three types of process being considered: AR(1), MA(1) and ARMA(1,1). Differences occur only in the complexity of the Gibbs sampling, extra steps needing to be added in going from AR(1) to MA(1) and again in going to ARMA(1,1). For AR(1), the conditional distribution of $Y_t|y_{-t}$ (where y_{-t} is used to indicate the whole series y but omitting y_t) is simply the conditional distribution $Y_t|(y_{t-1}, y_{t+1})$, as the inverse covariance matrix is tri-diagonal. The distributional form is easily derived (Tong, 1990) and sampling is straightforward. For MA(1), the conditional distribution involves the whole series and so calculations would involve the inversion of an $n \times n$ matrix. This can be overcome using methods described by Phadke and Kedem (1978) and Ansley (1979), using the Cholesky decomposition to transform the original series into a set of independent observations for which the likelihood is easily evaluated. For ARMA(1,1), the use of a further transformation enables the same methodology to be used.

Various techniques have been developed to assess whether a Markov chain can be considered to have converged; a comprehensive review is given in Brooks and Roberts (1998). The package CODA (Best et al., 1995) was used to assess the chains using these methods. Plots were examined and diagnostics due to Geweke and to Heidelberger & Welch were used to check convergence more formally. It was decided that a burn-in of 500 iterations followed by a further 10000 realisations was adequate. For a given series, single estimates of the parameters were calculated as the mean of the 10000 realisations.

Model	n	LS		Spectral		WLS	GLS	Pair
		(B)	(G)	(B)	(G)			
AR(1)	1000	92	87	**93**	85	88	85	88
	100	**92**	86	**92**	86	87	86	88
MA(1)	100	73	59	**87**	64	59	59	59
ARMA(1,1)	100	46	45	48	**51**	41	49	47
CPU time (seconds)		2	10	6	7	19	6	7

Table 2: *Percentage efficiencies relative to MCMC. Each figure is the average over 100 simulations, over a range of parameter values and two threshold levels, for optimal n'. The most efficient estimator in each row is highlighted. CPU times are average times for a SunUltra2 to process 100 series for each estimation method, MCMC taking about 10000, 1000, 40000 and 60000 seconds respectively for the four examples shown*

3 Simulation

To compare the alternative estimators, we simulated 100 replicates of each of a range of ARMA processes. For AR(1) processes we considered parameter values $\phi = \{0.0, 0.3, 0.6, 0.9\}$, for MA(1) processes, $\theta = \{0.0, -0.3, -0.6, -0.9\}$, and for ARMA(1,1) processes, $(\phi, \theta) = \{(0.3, 0.0),\ (0.0, -0.6),\ (0.6, 0.3),\ (0.6, -0.3),\ (0.3, -0.3),\ (0.9, 0.6)\}$. Series had zero mean, unit variance, were of length $n = 100$ or 1000, and were thresholded at either 0 or 1. Parameters were estimated for a range of values of n', and using either binary (B) or Gaussian (G) correlations, using NAG optimisation routine E04JAF (Numerical Algorithms Group, 1993), This employs a quasi-Newton algorithm to minimise a function of several variables, allowing simple bounds and using function values only. All programming was done in Fortran 90.

The main criterion used to compare estimators is the mean square error (MSE) which, for an estimator $\tilde{\phi}$ for ϕ can be re-expressed as

$$\mathrm{MSE}(\tilde{\phi}) = \frac{1}{100}\sum_{s=1}^{100}(\tilde{\phi}_s - \phi)^2 = \mathrm{Bias}(\tilde{\phi})^2 + \mathrm{Var}(\tilde{\phi}).$$

For models with more than one parameter, such as ARMA(1,1) with parameters (ϕ, θ), we considered the combined MSE, given by the sum of the two individual MSEs.

We present here a representative set of results for the different models and all estimation methods considered in Section 2. We assume MCMC to be a fully efficient estimator and, where results are available, express the efficiency of the other methods relative to this. Table 2 shows such percentage efficiencies, the figures shown being the ratio of the average MSE for each method to the average MSE from MCMC. The averages are over different parameter values and threshold levels, for optimal n'. CPU times are averages over the the models for all the parameters considered. The MCMC method was too slow to run for MA(1) and ARMA(1,1) processes when $n = 1000$ and so in Table 3 we simply present root mean square errors (RMSE) for an example of each model.

Model / Parameter values	T	LS (B)	LS (G)	Spectral (B)	Spectral (G)	WLS	GLS	Pair
MA(1)	0	61	61	**54**	**54**	61	61	61
$\theta = -0.3$	1	70	74	**67**	72	74	74	74
ARMA(1,1)	0	76	76	**70**	73	79	80	75
$(\phi, \theta) = (0.9, 0.6)$	1	94	95	**91**	96	103	97	93

Table 3: *1000 × RMSE of parameter estimates. Each figure is the average over 100 simulations, for the given parameter values and threshold level T, for series length 1000, for optimal value of n'. The smallest value in each row is highlighted*

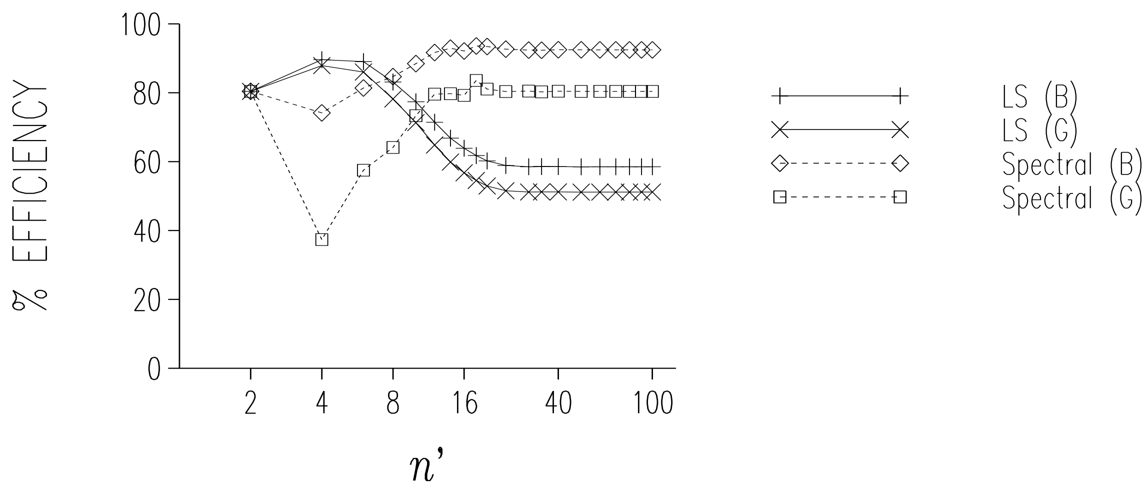

Figure 3: *Percentage efficiencies relative to MCMC for estimation in an AR(1) process of length 1000 with $\phi = 0.6$, threshold=0, for a range of values of n'*

Figure 3 shows the typical relationship between efficiency and n', for ordinary least squares and the spectral method. This illustrates one of the main advantages of the spectral method over least squares. The optimal value of n' for least squares is quite crucial, here 4, but is not known *a priori*, whereas for the spectral method the exact choice of n' is not important — it simply has to be sufficiently large. This is the general pattern, and so for example in Table 3 we see that although the spectral method using binary correlations is only marginally the most efficient, the real advantage is the lack of needing to know an optimal value for n' in advance. The value needed can be gauged from the rate of decay of the autocorrelation and inverse autocorrelation coefficients (see Section 2.1). For the majority of cases, such a value of n' resulted in the highest efficiency. It should be noted, however, that in a few cases the efficiency achieved by using this value of n' was not as high as that obtained by simply taking n' to be twice the number of parameters in the model, for which all methods agree and are equivalent to equating the first $n'/2$ sample correlations with their expected values.

For AR(1) processes, all methods can be seen to be nearly as efficient as MCMC, and so the spectral approach offers no clear benefit. For other ARMA processes the spectral method is generally seen to be more efficient than least squares, in some cases the improvement being quite substantial. Weighted and generalised least squares do not offer much advantage over ordinary least squares, and indeed are often drastically worse. The pairwise likelihood method is fairly stable with n' but generally produces estimates no more efficient than ordinary least squares. Note that, for MA(1) processes, the weighted and generalised least squares and pairwise likelihood method are all equivalent to ordinary least squares with $n' = 2$, since the expected values of autocorrelation at lags higher than 1 is zero. As demonstrated in both tables, use of binary correlations is usually to be preferred over Gaussian. It is interesting and reassuring to note from Table 2 the stability in efficiency for AR(1) processes over the two series lengths considered. CPU times for all the fast methods are small enough to be of no consequence. In contrast, times for MCMC are prohibitively high.

4 Application

We consider binary feeding data, collected for 30 days in May 1995 as part of a longer experiment at Langhill Dairy Cattle Research Centre, Edinburgh (Tolkamp et al., 1998). Thirty-six cows had continuous access to food in electronic feeders, and time spent at feeders was recorded automatically. Figure 1 shows two days of feeder-visit data for one cow. In Allcroft and Glasbey (2000) we chose to suppress short intervals away from feeders and model this derived variable instead, corresponding to bouts of feeding, i.e. meals. However our modelling approach is also capable of using the visit data directly, which is what we do here. The data are recorded in continuous time, which we have discretised at one minute intervals.

Figure 4 shows $\hat{\rho}^{(G)}$ for a typical cow. Inspection of such plots for all cows indicates that the simplest form of ARMA process that could provide an adequate model is ARMA(2,1), the autocorrelation function for which is a mixture of two exponential functions, invariant to the time scale chosen to discretise time. Parameter values of $\hat{\phi} = (1.9716, -0.9728)$ and $\hat{\theta} = -0.9927$, were obtained by numerically maximising the binary spectral likelihood. The 95% simulation-based pointwise confidence intervals shown were obtained as the 5th and 195th estimates of the ordered sample autocorrelations from 199 simulated series. The estimated autocorrelations can be seen to be consistent with the model.

Figure 2 shows data simulated from the fitted model, together with the underlying latent variable. The feeding patterns are similar to the observed data of Figure 1. To check this further, Figure 5 shows the marginal distributions of feeder visits and of intervals between visits, comparing the data values with those obtained by simulation of the fitted model. The form of these distributions played no part in the motivation for this model, so it is reassuring to observe that the marginal distributions are of the correct shape, in particular note that the model has captured the bimodality of the distribution of interval lengths.

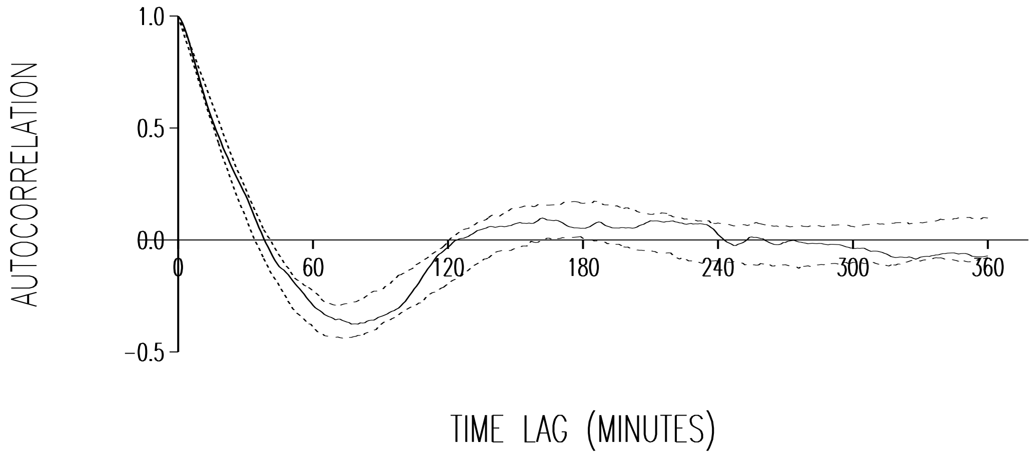

Figure 4: *Gaussian autocorrelations for one cow: (—) $\hat{\rho}^{(G)}$, (- - -) simulation-based 95% confidence envelope for $\hat{\rho}^{(G)}$ for the fitted ARMA(2,1) model*

5 Discussion

We have seen that there are fast and efficient estimation procedures available for censored ARMA processes. In particular, when the censoring is simple thresholding, the use of a spectral quasi-likelihood is computationally quick, more efficient than some other fast alternatives, and sometimes nearly as efficient as MCMC. It also has the important advantage of the number of terms needing to be retained in the optimisation being much less than n; in addition, given that n' is high enough for the autocorrelations and inverse autocorrelations to have decayed sufficiently, the exact choice is not critical. Simulations showed use of binary correlations to be preferable to Gaussian ones, both in terms of efficiency and stability. More simulation studies remain to be done to see if these results hold for more general ARMA models and other forms of censoring.

The latent Gaussian model appears to be a flexible one for categorical behaviour data, the ARMA(2,1) process fitting well to the cow feeding data. In Allcroft (2001), this type of model was compared with others, one of them being the hidden Markov model (MacDonald and Zucchini, 1997), which has a discrete latent structure and is again considered in a discrete time framework. The data occur naturally in continuous time, and one of the advantages of the latent Gaussian variable model over other potential models is that it has a continuous time analogue. However it can be shown (Cox and Miller, 1965, pages 293–296) that if the autocorrelation function is not twice-differentiable at lag zero, then realisations of the continuous-time process have short term 'jitter' and so if the process crosses the threshold at a time t_0, then for any small $\eta > 0$, the process will cross the threshold infinitely often in the interval $(t_0, t_0 + \eta)$. Therefore although the process can be defined in continuous time, it would be a poor model for cow behaviour. A solution would be to consider a higher-order ARMA process such as ARMA(3,2), with enough parameters to be able to constrain the autocorrelation function to

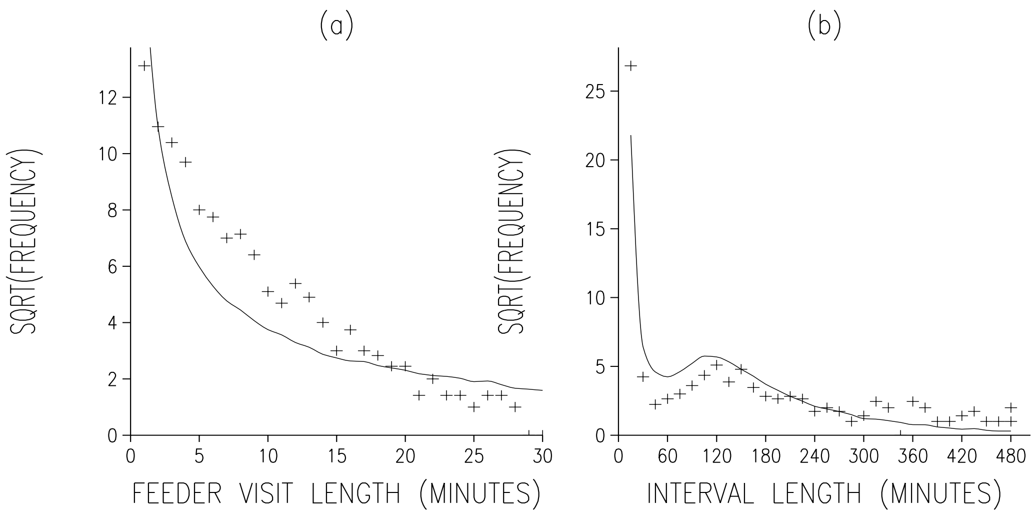

Figure 5: *Marginal distributions of (a) feeder-visit lengths, (b) interval lengths: (+) sample frequencies, (—) frequencies based on 100 realisations of the fitted ARMA(2,1) model*

be twice-differentiable at lag zero. The details of this remain to be investigated.

Further work could also investigate the use of multivariate processes to model a herd of cows simultaneously, with the inclusion of covariates. More general categorical data could also be modelled, e.g. three ordered categories and hence two thresholds for the latent variable. A possibility for unordered behaviours would be to have a latent variable associated with each category, the current behaviour corresponding to the one with the highest value, as described by Bartholomew and Knott (1999) for econometric data. Other forms of censoring could be also considered, for example that encountered in rainfall data (Glasbey et al., 1998), for which zero rainfall corresponds to a censored value below the threshold, and above the threshold, the variable is directly observed. This case requires the variance of the Gaussian variable to be estimated, whereas we were able to work with standardised variables.

Acknowledgements

We thank Ilias Kyriazakis, Bert Tolkamp and Langhill Dairy Cattle Research Centre, Edinburgh, for the data. We also acknowledge Colin Aitken and Elizabeth Austin for their input to this work and the Scottish Executive Rural Affairs Department for financial support.

Appendix

We outline the mains steps in the proof that $\mathcal{L}$ is approximated by $\mathcal{L}'$ under the conditions outlined in Section 2.1.

First, we re-express the full likelihood, $\mathcal{L}$, given by (2), as

$$\mathcal{L} = -\frac{1}{2} \sum_{k=-\frac{n}{2}}^{\frac{n}{2}-1} \log |S_k| - \frac{1}{2} \sum_{l=-\frac{n}{2}}^{\frac{n}{2}-1} \text{trace } [\alpha_l \; \hat{\rho}_l^{(G)}],$$

using (3) and (5). Similarly, the restricted likelihood, $\mathcal{L}'$ given by (4), can be re-expressed as

$$\mathcal{L}' = -\frac{n}{2n'} \sum_{k=-\frac{n'}{2}}^{\frac{n'}{2}-1} \log |S'_k| - \frac{n}{2n'} \sum_{l=-\frac{n'}{2}}^{\frac{n'}{2}-1} \text{trace } [\alpha'_l \; \hat{\rho}_l^{(G)}],$$

where α'_l is defined similarly to α_l in (5), replacing n by n' and S by S'.

We will show that

$$\frac{n}{n'} \sum_{k=-\frac{n'}{2}}^{\frac{n'}{2}-1} \log |S'_k| \approx \sum_{k=-\frac{n}{2}}^{\frac{n}{2}-1} \log |S_k| \tag{8}$$

and

$$\frac{n}{n'} \sum_{l=-\frac{n'}{2}}^{\frac{n'}{2}-1} \text{trace } [\alpha'_l \; \hat{\rho}_l^{(G)}] \approx \sum_{l=-\frac{n}{2}}^{\frac{n}{2}-1} \text{trace } [\alpha_l \; \hat{\rho}_l^{(G)}]. \tag{9}$$

For (8), consider

$$\begin{aligned} S_k - S'_{\frac{n'}{n}k} &= \sum_{l=-\frac{n}{2}}^{\frac{n}{2}-1} \rho_l^{(G)} e^{-2\pi i k l/n} - \sum_{l=-\frac{n'}{2}}^{\frac{n'}{2}-1} \rho_l^{(G)} e^{-2\pi i k l/n} \\ &= \sum_{l=-\frac{n}{2}}^{\frac{n'}{2}-1} \rho_l^{(G)} e^{-2\pi i k l/n} + \sum_{l=\frac{n'}{2}}^{\frac{n}{2}-1} \rho_l^{(G)} e^{-2\pi i k l/n} \quad \approx \quad 0. \end{aligned} \tag{10}$$

This follows from the definitions of S and S' and, because $\rho^{(G)}$ decays exponentially, for sufficiently large n', $\rho_l^{(G)} \approx 0$ for $|l| > \frac{n'}{2}$. Hence

$$\frac{n}{n'} \sum_{k=-\frac{n'}{2}}^{\frac{n'}{2}-1} \log |S'_k| \approx \frac{n}{n'} \sum_{k=-\frac{n'}{2}}^{\frac{n'}{2}-1} \log |S_{\frac{n}{n'}k}|. \tag{11}$$

In general $\frac{n}{n'}k$ is non-integer, but note that definition (3) still holds. Then if $|S_k|$ is continuous in k, using the compound trapezoidal rule for integration, we can approximate the right hand side of (8) by the right hand side of (11), to within order $1/n'$, and thus obtain (8).

For (9), we assume that α decays exponentially, and therefore for sufficiently large n', $\alpha_l \approx 0$ for $|l| > \frac{n'}{2}$. So, we simply need to show that

$$\frac{n}{n'} \alpha'_l \approx \alpha_l.$$

From the definitions of α and α', and (10), we have

$$\frac{n}{n'}\alpha'_l = \frac{n}{n'}\sum_{k=-\frac{n'}{2}}^{\frac{n'}{2}-1} {S'_k}^{-1} e^{2\pi i l k/n'} \approx \frac{n}{n'}\sum_{k=-\frac{n'}{2}}^{\frac{n'}{2}-1} {S_{\frac{n}{n'}k}}^{-1} e^{2\pi i l k/n'} \approx \sum_{k=-\frac{n}{2}}^{\frac{n}{2}-1} {S_k}^{-1} e^{2\pi i k l/n} = \alpha_l$$

where the second approximation again depends on the continuity of the function being summed.

References

Allcroft, D. J. (2001). *Statistical models for short-term animal behaviour*. PhD thesis, University of Edinburgh.

Allcroft, D. J. and Glasbey, C. A. (2000). Estimation of latent Gaussian ARMA models for categorical behaviour data. In V. Núñez-Antón and E. Ferreira, editors, *Proceedings of the 15th International Workshop on Statistical Modelling*, pages 294–299. The University of the Basque Country, Bilbao.

Ansley, C. F. (1979). An algorithm for the exact likelihood of a mixed autoregressive-moving average process. *Biometrika*, 66:59–65.

Bartholomew, D. J. and Knott, M. (1999). *Latent Variable Models and Factor Analysis*. Arnold, London, second edition.

Best, N. G., Cowles, M. K., and Vines, S. K. (1995). *CODA: Convergence Diagnostics and Output Analysis Software for Gibbs sampling output, Version 0.30*. MRC Biostatistics Unit, Cambridge, UK.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*. Prentice Hall, Englewood Cliffs, New Jersey, third edition.

Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory*. Holt, Rinehart and Winston, New York.

Brooks, S. P. and Roberts, G. O. (1998). Convergence assessment techniques for Markov chain Monte Carlo. *Statistics and Computing*, 8:319–335.

Chatfield, C. (1979). Inverse autocorrelations. *Journal of the Royal Statistical Society, Series A*, 142:363–377.

Cleveland, W. S. (1972). The inverse autocorrelations of a time series and their applications. *Technometrics*, 14:277–293.

Coursol, J. and Dacunha-Castelle, D. (1983). Remarks on the approximation of the likelihood function of a stationary Gaussian process. *Theory of Probability and its Applications*, 27:162–167.

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

Cox, D. R. and Snell, E. J. (1989). *Analysis of Binary Data*. Chapman & Hall, London, second edition.

Cressie, N. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology*, 17:563–586.

Glasbey, C. A., Nevison, I. M., and Hunter, A. G. M. (1998). Parameter estimators for Gaussian models with censored time series and spatio-temporal data. In Payne, R. and Green, P., editors, *COMPSTAT98 Proceedings in Computational Statistics*, pages 323–328, Heidelberg. Physica-Verlag.

Hjort, N. L. and Omre, H. (1994). Topics in spatial statistics. *Scandinavian Journal of Statistics*, 21:289–357.

Keenan, D. (1982). A time series analysis of binary data. *Journal of the American Statistical Association*, 77:816–821.

Kendall, M. G., Stuart, A., and Ord, J. K. (1983). *The Advanced Theory of Statistics, Volume 3: Design and Analysis, and Time-Series*. Griffin, High Wycombe, fourth edition.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.

MacDonald, I. L. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-valued Time Series*. Chapman & Hall, London.

Numerical Algorithms Group (1993). *Library Manual Mark 16*. Oxford.

Phadke, M. S. and Kedem, G. (1978). Computation of the exact likelihood function of multivariate moving average models. *Biometrika*, 65:511–519.

Tolkamp, B. J., Allcroft, D. J., Austin, E. J., Nielsen, B. L., and Kyriazakis, I. (1998). Satiety splits feeding behaviour into bouts. *Journal of Theoretical Biology*, 194:235–250.

Tong, Y. L. (1990). *The Multivariate Normal Distribution*. Springer, New York.

Whittle, P. (1953). The analysis of multiple stationary time series. *Journal of the Royal Statistical Society, Series B*, 15:125–139.