

## Estimation of latent Gaussian ARMA models for categorical behaviour data

David Allcroft<sup>1</sup> and Chris Glasbey<sup>2</sup>

<sup>1</sup> Department of Mathematics and Statistics, University of Edinburgh, The King's Buildings, Edinburgh, EH9 3JZ, Scotland; d.allcroft@bioss.sari.ac.uk

<sup>2</sup> Biomathematics and Statistics Scotland, The King's Buildings, Edinburgh, EH9 3JZ, Scotland

**Abstract:** We consider the fitting of latent Gaussian models to categorical time series of cow feeding data. We derive a spectral quasi-likelihood for the data, and compare it with least squares fits to autocorrelations and MCMC estimators of the parameters in thresholded ARMA processes. We show that the spectral method is more efficient than least squares and far faster than MCMC.

**Keywords:** ARMA process; Cow feeding data; MCMC; Spectral likelihood; Tetrachoric correlation.

### 1 Introduction

The analysis of categorical behaviour data (Haccou and Meelis, 1994) poses many challenges for statisticians. In particular, 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 three days of feeder-visit data for one cow, together with a derived variable, feeding data, obtained by suppressing short intervals away from feeders (Tolkamp et al., 1998). We prefer to model this variable as it is less susceptible than the feeder-visit data to herd and dominance effects. The data are recorded in continuous time, which we have discretised at one minute intervals.

We postulate a latent, Gaussian-distributed, physiological variable, with feeding occurring when this variable exceeds a threshold. Latent variables may be stochastically linked with categorical data, such as a logistic response (Keenan, 1982), or deterministically linked (Cox and Snell, 1992), as in our case. 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.

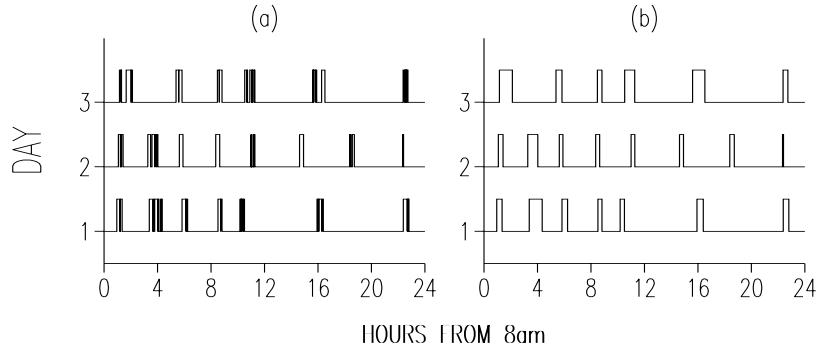


FIGURE 1. Three days of data for one cow: (a) feeder-visit data, (b) feeding data.

In Section 2, we relate the autocorrelations of the observed and latent processes, taking diurnal pattern into account. Then, in Section 3 we derive a spectral quasi-likelihood and estimate parameters in latent ARMA models using both it and a least squares fit to autocorrelations. In Section 4, we compare the efficiencies of these estimators. Finally, in Section 5 we discuss the results.

## 2 Autocorrelations

From data such as Figure 1, there is evidence for a diurnal feeding effect. We estimate the probability of feeding at a particular time of day by averaging observations at nearby times for all days. We used cross-validation to select an optimal window width of approximately one hour, by omitting each day's data in turn and then predicting it.

We estimate autocorrelations in the latent Gaussian process  $\{y_t\}$ , denoted  $\hat{\rho}_l^{(G)}$  at lag  $l$ , from the observed binary process  $\{x_t\}$ , by numerically maximising quasi-likelihoods of the form

$$\sum_t \log \left[ \int_{L_t} \int_{L_{t+l}} f(y_t, y_{t+l}; \rho_l^{(G)}) dy_t dy_{t+l} \right].$$

Here, the integration interval,  $L_t$ , is  $(-\infty, T_t)$  if  $x_t = 0$  and  $(T_t, \infty)$  if  $x_t = 1$ , threshold  $T_t$  is chosen so that the probability of not feeding,  $\Phi_{T_t}$ , matches the diurnal trend, where  $\Phi_T$  is the standard normal integral from  $-\infty$  to  $T$ , and  $f$  denotes the bivariate Gaussian density with zero mean and unit variance.

In the absence of trend,  $T$  is a constant and  $\hat{\rho}_l^{(G)}$  simplifies to a tetrachoric correlation, estimable from the functional relationship

$$\hat{\rho}^{(B)} = (\Phi_{T,T}(\hat{\rho}^{(G)}) - \Phi_T^2) / (\Phi_T - \Phi_T^2), \quad (1)$$

where  $\hat{\rho}^{(B)}$  denotes a sample autocorrelation of the binary process and  $\Phi_{T,T}(\hat{\rho}^{(G)})$  is the bivariate normal integral from  $-\infty$  to  $T$  with correlation  $\hat{\rho}^{(G)}$ . Expression (1) can also be used to compute expected autocorrelations of the binary process,  $\rho^{(B)}$ , from those of the Gaussian process.

### 3 Model estimation

Inspection of  $\hat{\rho}^{(G)}$  for the cows indicates that the simplest form of ARMA process that could provide an adequate model is an ARMA(2,1) model, with the additional benefit that this process has a continuous time analogue.

ARMA parameters can be estimated in an ad-hoc way via least squares, i.e. 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 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. Whittle (1953) derived the spectral approximation for the log-likelihood,  $\mathcal{L}$ , of an  $m$ -dimensional stationary multivariate Gaussian process of length  $n$ , which has the form of a set of independent complex Wishart distributions (Brillinger, 1975):

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

Here  $S_k$  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 in our case,

$$S_k = \sum_{l=-\frac{n}{2}}^{\frac{n}{2}-1} \rho_l^{(G)} e^{-2\pi i k l / n} \quad \text{for } k = -\frac{n}{2}, \dots, \frac{n}{2} - 1. \quad (3)$$

For a rigorous proof, 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  $\rho^{(B)}$ .

We have extended the proof, to show that, for short-memory processes such as ARMA models,  $\mathcal{L}$  can be approximated by a ‘restricted’ likelihood,  $\mathcal{L}'$ :

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

where  $n' \ll n$ , with considerable computational saving. Here  $S'_k$  and  $\hat{S}'_k$  are obtained as the discrete Fourier transforms of cross-correlations up to lag  $n'/2$  only, by replacing  $n$  by  $n'$  in (3).

The proof relies on re-expressing  $\mathcal{L}$  as

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

and similarly  $\mathcal{L}'$ , where  $\alpha_l$  is the inverse autocorrelation coefficient at lag  $l$ , defined as

$$\alpha_l = \sum_{k=-\frac{n}{2}}^{\frac{n}{2}-1} S_k^{-1} e^{-2\pi i k l / n} \quad \text{for } l = -\frac{n}{2}, \dots, \frac{n}{2} - 1. \quad (4)$$

Here we use a multivariate generalisation of the univariate case considered by Cleveland (1972) and Chatfield (1979). We define  $\alpha'_l$  similarly, but replacing  $n$  by  $n'$  and  $S$  by  $S'$  in (4).

The proof requires that  $\rho_l^{(\mathcal{G})}$  and  $\alpha_l$  are negligible for  $|l| > n'/2$ , and that  $S_k$  is a continuous function of  $k$ . These conditions hold for ARMA processes, typically for small values of  $n'$ , because autocorrelations and inverse autocorrelations decay exponentially (Chatfield, 1979; Box and Jenkins, 1976).

The proof then follows by showing that  $\alpha'_l \approx \frac{n'}{n} \alpha_l$ ,  $S'_k \approx S_{\frac{n'}{n}k}$  and

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

Figure 2 shows  $\rho^{(\mathcal{G})}$  and  $\hat{\rho}^{(\mathcal{G})}$  for one cow with ARMA parameter values of  $\hat{\phi} = (1.9716, -0.9728)$ ,  $\hat{\theta} = -0.9927$ , obtained by numerically maximising the Gaussian spectral likelihood. Similar values were obtained by least squares estimators for a range of values of  $n'$  and with fits directly to the binary process.

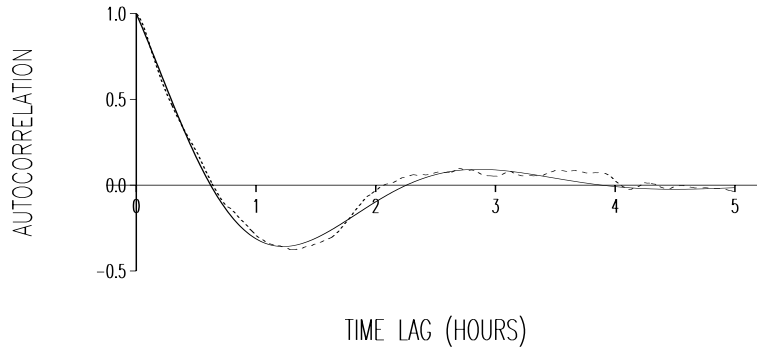


FIGURE 2. Gaussian autocorrelations for one cow: (—)  $\rho^{(\mathcal{G})}$ , (---)  $\hat{\rho}^{(\mathcal{G})}$ .

## 4 Efficiencies of estimators

We compare efficiencies of least squares and spectral estimators of ARMA parameters by simulation, for a range of models, of values of  $T$ ,  $n$  and  $n'$  and use of either binary or Gaussian correlations. MCMC, combining Gibbs sampling with a Metropolis-Hastings algorithm, was also used for smaller values of  $n$ , to obtain a fully efficient estimator by maximum likelihood. Note that, in general, the MCMC method uses too much CPU time to be a practical alternative. A representative set of results is given in Table 1. For the AR(1) processes, all the methods are seen to be nearly as efficient as MCMC and so the spectral approach offers no clear benefit. For the other processes the spectral method is generally seen to be more efficient than the least squares, in some cases the improvement being quite substantial. The optimal value of  $n'$  for least squares is quite crucial, typically 2 or 4, whereas for the the spectral method, the exact choice of  $n'$  is not important, it simply has to be sufficiently large, typically greater than about 10.

Model / Parameter values	$n$	$T$	LS		Spectral		MCMC
			(B)	(G)	(B)	(G)	
AR(1) $\phi = 0.6$	100	0	145	146	<b>142</b>	145	140
		1	183	190	<b>181</b>	191	172
	1000	0	<b>36</b>	37	<b>36</b>	38	34
		1	<b>43</b>	48	<b>43</b>	50	39
MA(1) $\theta = -0.6$	100	0	291	291	256	<b>247</b>	143
		1	327	339	<b>321</b>	333	202
	1000	0	148	148	<b>95</b>	104	*
		1	135	155	<b>110</b>	155	*
ARMA(1,1) $(\phi, \theta) = (0.6, -0.3)$	100	0	487	489	451	<b>311</b>	275
		1	621	612	586	<b>431</b>	300
	1000	0	107	107	104	<b>82</b>	*
		1	243	212	234	<b>122</b>	*
CPU time (seconds)			0.9	2.3	3.7	7.8	50000

TABLE 1:  $1000 \times \text{RMSE}$  of parameter estimates, averaged over 100 simulations, at optimal value of  $n'$ . The smallest value in each row, excluding MCMC, is highlighted. (MCMC was too slow to apply for MA(1) and ARMA(1,1) processes when  $n = 1000$ .) CPU times are for a SunUltra2 to process 100 series of length 100 for each estimation method.

## 5 Discussion

Latent Gaussian processes are flexible models for categorical behaviour data. In particular, we have seen that a latent ARMA(2,1) process shows

promise as a model for cow feeding data. We have explored alternative estimators, both analytically and by simulation, and found that the use of a spectral quasi-likelihood is both computationally quick and more efficient than least squares alternatives.

Further work will involve a more comprehensive simulation study for ARMA(2,1) processes. Also, we will explore ways of modelling a group of cows simultaneously in a multivariate framework, and the relationship between the latent Gaussian model and stochastic compartment and hidden Markov models.

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

### References

- Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis Forecasting and Control*. Holden-Day, San Francisco.
- Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory*. Holt, Rinehart and Winston, New York.
- 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 Snell, E.J. (1992). *Analysis of Binary Data*. Second Edition. Chapman & Hall, London.
- Haccou, P. and Meelis, E. (1994). *Statistical Analysis of Behavioural Data: An Approach Based on Time-structured Models*. Oxford University Press, Oxford.
- Keenan, D.M. (1982). A time series analysis of binary data. *Journal of the American Statistical Association* **77**, 816-821.
- MacDonald, I.L. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-valued Time Series*. Chapman & Hall, London.
- Tolcamp, 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.
- Whittle, P. (1953). The analysis of multiple stationary time series. *Journal of the Royal Statistical Society, Series B* **15**, 125-139.