An autoregressive model for irregular time series of variable stars

Susana Eyheramendy, Felipe Elorrieta and Wilfredo Palma

Department of Statistics, P. Universidad Católica de Chile,
Santiago, Chile.
email: `susana@mat.uc.cl`, `fielorrieta@mat.uc.cl`, `wilfredo@mat.uc.cl`

Abstract. This paper discusses an autoregressive model for the analysis of irregularly observed time series. The properties of this model are studied and a maximum likelihood estimation procedure is proposed. The finite sample performance of this estimator is assessed by Monte Carlo simulations, showing accurate estimators. We implement this model to the residuals after fitting an harmonic model to light-curves from periodic variable stars from the Optical Gravitational Lensing Experiment (OGLE) and Hipparcos surveys, showing that the model can identify time dependency structure that remains in the residuals when, for example, the period of the light-curves was not properly estimated.

Keywords. autoregressive, harmonic, irregular time series

1. Introduction

Time series observations are ubiquitous in astronomy, and have been used for several different purposes. For example, to distinguish between different types of supernovae, to detect and characterize extrasolar planets (e.g. Jordán *et al.* 2013) and to classify variable stars (e.g. Elorrieta *et al.* 2016). These time series are usually modeled using a parametric and/or physical model that assumes independent and homoscedastic errors, but in many cases this assumption is not accurate and there remains a time dependent structure on the errors. This commonly occurs when the proposed model cannot explain all the variability of the data. In this work we propose an autoregressive model for irregularly observed data to detect time dependencies on the data. In the next § 2, we introduce the model and its properties. Then, in § 3 we present results on simulated data and on real light-curves from variable stars observed at the OGLE and Hipparcos surveys. In § 4 we offer some discussion.

2. Methods

2.1. *Irregular Autoregressive (IAR) model*

Denote y_{t_j} an observation measured at time t_j, and consider an increasing sequence of observational times $\{t_j\}$ for $j = 1, \ldots, n$. We define the irregular autoregressive (IAR) process by

$$y_{t_1} = \sigma \, \varepsilon_{t_1}, y_{t_j} = \phi^{t_j - t_{j-1}} \, y_{t_{j-1}} + \sigma \, \sqrt{1 - \phi^{2(t_j - t_{j-1})}} \, \varepsilon_{t_j} \quad \text{for } j = 2, \ldots, n, \qquad (2.1)$$

where ε_{t_j} are independent Gaussian random variables with zero mean and unit variance. Note that

$$E(y_{t_j}) = 0 \text{ and } Var(y_{t_j}) = \sigma^2 \text{ for all } y_{t_j}, \qquad (2.2)$$

and the covariance between y_{t_k} and y_{t_j} is $E(y_{t_k} \, y_{t_j}) = \sigma^2 \, \phi^{t_k - t_j}$, for $k \geqslant j$.

259

Thus, for any two observational times t, s we can define the autocovariance function as

$$\gamma(t - s) = E(y_t \, y_s) = \sigma^2 \, \phi^{t-s}, \qquad (2.3)$$

as well as the autocorrelation function (ACF), $\rho(t - s) = \frac{\gamma(t-s)}{\gamma(0)} = \phi^{t-s}$.

Given the results above, the sequence $\{y_{t_j}\}$ corresponds to a second-order or weakly stationary process. In addition, under some conditions, the process is strictly stationary and ergodic.

2.2. Estimation

The likelihood of the data $\{y_{t_1}, \ldots, y_{t_n}\}$ can be expressed as

$$f(y_{t_1}, \ldots, y_{t_n}; \theta) = f(y_{t_1}; \theta) f(y_{t_2} | y_{t_1}; \theta) \times \ldots \times f(y_{t_n} | y_{t_{n-1}}; \theta), \qquad (2.4)$$

where $\theta = (\sigma^2, \phi)$ is the parameter vector of the model, with

$$f(y_{t_1}; \sigma^2, \phi) \sim N(0, \sigma^2) \text{ and} \qquad (2.5)$$

$$f(y_{t_j} | y_{t_{j-1}}; \sigma^2, \phi) \sim N(\phi^{t_j - t_{j-1}} \, y_{t_{j-1}}, \sigma^2 \, (1 - \phi^{2(t_j - t_{j-1})})) \text{ for } j = 2, \ldots, n. \qquad (2.6)$$

Based on equation (2.1), minus the log-likelihood of this process can be written as

$$\ell(\theta) = \frac{n}{2} \log(2\pi) + \frac{1}{2} \sum_{j=1}^{n} \log \nu_{t_j} + \frac{1}{2} \sum_{j=1}^{n} \frac{e_{t_j}^2}{\nu_{t_j}}, \qquad (2.7)$$

where we define $e_{t_1} = y_{t_1}$, $e_{t_j} = y_{t_j} - \phi^{t_j - t_{j-1}} \, y_{t_{j-1}}$ for $j > 1$ and their variances as $\nu_{t_j} = Var(e_{t_j})$.

Observe that the finite past predictor of the process at time t_j is given by

$$\widehat{y}_{t_1} = 0, \text{ and } \widehat{y}_{t_j} = \phi^{t_j - t_{j-1}} \, y_{t_{j-1}}, \text{ for } j = 2, \ldots, n. \qquad (2.8)$$

Therefore, $e_{t_j} = y_{t_j} - \widehat{y}_{t_j}$ is the prediction error with variance $\nu_{t_1} = Var(e_{t_1}) = \sigma^2$,

$$\nu_{t_j} = Var(e_{t_j}) = \sigma^2 [1 - \phi^{2(t_j - t_{j-1})}], \text{ for } j = 2, \ldots, n. \qquad (2.9)$$

By direct maximization of the log-likelihood (2.7), we can obtain the maximum likelihood estimator of σ^2,

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{n} \frac{(y_{t_j} - \widehat{y}_{t_j})^2}{\tau_{t_j}}, \text{ where } \tau_{t_j} = \nu_{t_j} / \sigma^2. \qquad (2.10)$$

It is not possible to find the maximum likelihood estimator of ϕ by direct maximization of the likelihood, but iterative methods can be used (for details see Ch. 5 of Palma *et al.* 2016). We developed a script in the statistical language/software R to estimate ϕ.

2.3. Harmonic model for fitting light curves of variable stars

In fitting light curves of periodic variable stars, the standard approach is to use an harmonic model. This model requires first to find the period of the variable star, which can be estimated using the Generalized Lomb-Scargle periodogram (Zechmeister *et al.* 2009). For more details in the procedure on periodic light-curve estimation, see for example Debosscher *et al.* (2007), or Richards *et al.* (2011).

Table 1. *Maximum likelihood estimation of simulated IAR series with mixture of Exponential distribution for the observational times, with $\lambda_1 = 130$ and $\lambda_2 = 6.5$, $w_1 = 0.15$ and $w_2 = 0.85$.*

Case	Sample size	Parameter	Estimate $\widehat{\phi}$	SD($\widehat{\phi}$)	$\widehat{\sigma}$
1	50	0.9000	0.8878	0.0456	0.8451
2	50	0.9900	0.9848	0.0085	0.5919
3	50	0.9990	0.9958	0.0033	0.5215
4	100	0.9000	0.8933	0.0284	0.8306
5	100	0.9900	0.9875	0.0050	0.5513
6	100	0.9990	0.9974	0.0018	0.3799

The residuals after subtracting a linear trend and an harmonic model with one frequency and four components denoted as $y(t)$, are $y(t) = z(t) - \hat{\alpha} - \hat{\beta}t - \sum_{j=1}^{4}(\hat{a}_j \sin(2\pi f_1 jt) + \hat{b}_j \cos(2\pi f_1 jt))$, where $\hat{a}$ is a maximum likelihood estimator and $z(t)$ is the observation at time t. Note that when fitting the IAR model to these residuals, $y(t)$, a large estimate of the parameter ϕ can reveal that the model used to fit the data is not the appropriate one due to a remaining time-dependency in the time series.

3. Results

3.1. Simulation study

This section shows the results from several Monte Carlo experiments assessing the finite sample performance of the proposed maximum likelihood estimator.

The simulated processes correspond to the model (2.1) where the observation times follow a mixture of two independent exponential distributions with means $1/\lambda_1$ and $1/\lambda_2$, and random weights w_1 and w_2, respectively. Table 1 shows a summary of the simulations based on 1000 repetitions with $\lambda_1 = 130$, $\lambda_2 = 6.5$, $w_1 = 0.15$ and $w_2 = 0.85$.

The Monte Carlo simulations suggest that the finite-sample performance of the proposed methodology is accurate. In particular, the estimation bias is small even for the reduced sample sizes used in Table 1. Note that we restrict to high values of the parameter ϕ. The reason for this is the choice of the distribution of the observational time gaps, which tend to be large.

3.2. Implementation of the IAR model to variable stars from OGLE and Hipparcos

In order to show an application of our model to the fitting of light curves from variable stars, we select forty variable stars from the OGLE and Hipparcos surveys for which the harmonic model gives a precise fit of the light-curve. In this way, we are confident that the estimate of the period is correct. Figure 1(a) shows an example of one of the variable stars selected. We then apply our IAR model to the residuals of the best harmonic model (2.3). For the forty chosen light-curves we obtain small values close to zero for the parameter ϕ, as shown in the boxplot on the right in Figure 1(c). We then vary the frequency f_1 in the interval $[f_1 - 0.45 f_1, f_1 + 0.45 f_1]$, and after doing so we fit the harmonic model with the wrong frequency. The residuals of the harmonic model have now temporal structure that can be captured with the IAR model, and in particular by the inferred value of ϕ. Figure 1(b) shows for each value of the frequency (with the right frequency f_1 at the center of the plot at zero), the estimated ϕ parameter of the IAR model applied to the residuals of the harmonic fit. Note that as we move away from the correct frequency, the value of ϕ increases. Figure 1(c) shows in the boxplot on the left the distribution of ϕ for
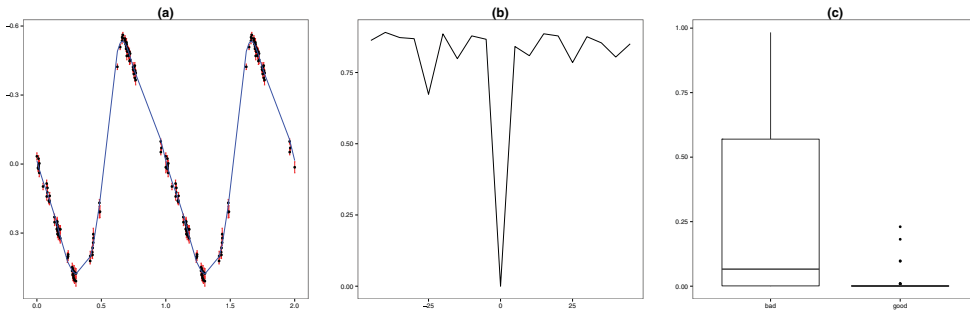
Figure 1. a) Example of the light curve of a Classical Cepheid. The continuous blue line is the harmonic fit. b) On the x-axis is the % of variation from the correct frequency, and on the y-axis is the estimate of the parameter ϕ of the IAR model. c) Boxplot of the distribution of ϕ, for the light-curves using the correct frequency (on the right) and for the light-curves using the incorrect frequency (on the left).

the light-curves with the incorrect frequency. This distribution is now more spread-out, taking larger values away from zero.

4. Discussion

In this work we have introduced an autoregressive model for irregularly observed time series (IAR), and show that it is weakly stationary, and under some conditions, it is strictly stationary and ergodic. We show on simulated data, that the maximum likelihood estimators for the parameters of the model are accurate. Also, we show on real data from variable stars from the OGLE and Hipparcos surveys, that the IAR model can detect when temporal dependency remains on the residuals of a model. In particular, we use the harmonic model with wrong periods of the light-curve to show how the parameter $\widehat{\phi}$ of the IAR model has a distribution that takes larger values away from zero, in contrast with the distribution of the $\widehat{\phi}$ when the correct periods where used to estimate the harmonic model. In the latter case, the distribution concentrates around zero with few exceptions.

Acknowledgements

References

Debosscher, J., Sarro, L. M., Aerts, C., Cuypers, J., Vandenbussche, B., Garrido, R., & Solano, E. 2007, *Astronomy & Astrophysics*, 475, 1159.
Elorrieta, F., Eyheramendy, S., Jordán, A., Dékány, I., Catelan, M., Angeloni, R., Alonso-García, J., Contreras-Ramos, R., Gran, F., Hajdu, G., Espinoza, N., Saito, R., & Minniti, D. 2016, *Astronomy & Astrophysics*, 595, A82.
Jordán, A., Espinoza, N., Rabus, M., Eyheramendy, S., Sing, D. K., Désert, J. M., Bakos, G. A., Fortney, J. J., López-Morales, M., Maxted, P. F. L., Triaud, A., & Szentgyorgyi, A. 2013, *The Astrophysical Journal*, 778, 184.
Palma W. 2016, *Wiley Series in Probability and Statistics*.
Richards, J. W., Starr, D. L., Butler, N. R., Bloom, J. S., Brewer, J. M., Crellin-Quick, A., Higgins, J., Kennedy, R., & Rischard, M. 2011, *The Astrophysical Journal*, 733, 10.
Zechmeister, M. & Kürster, M. 2009, *Astronomy & Astrophysics*, 496, 577.