BOOTSTRAPPING AUTOREGRESSIVE PLUS NOISE PROCESSES

Christian Debes and Abdelhak M. Zoubir

Signal Processing Group
Darmstadt University of Technology
Merckstr. 25, D-64283 Darmstadt, Germany
e-mail: {cdebes, zoubir}@ieee.org

ABSTRACT

We adress the problem of estimating confidence intervals for the parameters of an autoregressive plus noise process, in particular when the additive noise is non-Gaussian. We demonstrate how the independent data bootstrap can be used to solve this problem. We motivate an autoregressive moving-average modeling approach and apply the recursive maximum algorithm for parameter estimation. Computer simulations are carried out to show the performance of the proposed method. Furthermore a real data example from automotive engineering has been considered for assessing our approach. Using a pressure signal from inside the combustion chamber, we show how confidence intervals for the autoregressive parameters can be calculated.

Index Terms— the bootstrap, parametric spectrum estimation

1. INTRODUCTION

Estimating the parameters of an autoregressive (AR) process is a fundamental issue in time series analysis [1] and signal processing [2]. Given estimates of the AR-parameters one can easily form a parametric estimate of the spectrum as well as predicting future samples of the process. Applications include radar [3], geophysics [4], biomedicine [5], image processing [6] and speech signal processing [7].

In all these applications the original signal is disturbed and only a mixture of signal and additive noise can be observed. This situation is depicted in Figure 1. The AR process $x(n)$ can be described by the output of a linear time-invariant filter with transfer function $1/A(\omega)$ driven by $e(n)$. $e(n)$ is stationary white Gaussian noise with $E[e(n)] = 0$ and autocovariance function $c_{ee}(m) = \sigma_e^2\delta(m)$, where $\delta(m)$ is Kronecker's delta function.

The polynomial $A(\omega)$ is given by $A(\omega) = 1+\sum_{k=1}^{p} a_k e^{-j\omega k}$ with a_k being the k-th AR parameter and p being the order of the process. The AR process is corrupted by additive noise $u(n)$ yielding the observed signal $y(n)$. In this paper we make the assumption of $u(n)$ being zero mean stationary white non-Gaussian noise, independent from $e(n)$.

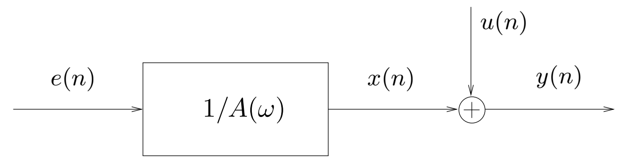

Fig. 1. AR plus noise process

1.1. Preliminaries

Given the scenario in Figure 1 parameter estimates can be obtained by means of the extended Yule-Walker equations. Assuming $u(n)$ being white noise with variance σ_u^2 we have:

$$c_{yy}(m) + \sum_{k=1}^{p} a_k c_{yy}(m-k) = \begin{cases} \sigma_e^2 + \sigma_u^2, & m = 0 \\ a_m\sigma_u^2, & m = 1, ..., p \\ 0, & m > p \end{cases}$$

with $c_{yy}(m)$ being the autocovariance function of $y(n)$.
Consistent estimates of the AR parameters can be obtained by solving the high-order Yule-Walker equations:

$$\hat{c}_{yy}(m) + \sum_{k=1}^{p} \hat{a}_k \hat{c}_{yy}(m-k) = 0, m = p+1, ..., 2p$$

Finally, consistent estimates for σ_e^2 and σ_u^2 can be obtained by solving

$$\hat{\sigma}_u^2 = \frac{1}{\hat{a}_m}(\hat{c}_{yy}(m) + \sum_{k=1}^{p} \hat{a}_k \hat{c}_{yy}(m-k))$$

for any $\hat{a}_m$ and

$$\hat{\sigma}_e^2 = \hat{c}_{yy}(0) + \sum_{k=1}^{p} \hat{a}_k \hat{c}_{yy}(k) - \sigma_u^2$$

Given estimates $\hat{a}_1, ..., \hat{a}_p, \hat{\sigma}_e^2$and $\hat{\sigma}_u^2$ the parametric spectrum can be estimated as

$$C_{yy}(\omega) = \frac{\sigma_e^2}{|A(\omega)|^2} + \sigma_u^2$$

2. ARMA MODELING

In the noisefree case it is straightforward to obtain bootstrap based confidence intervals of the AR parameters [8]. The initial estimates $\hat{a}_1, ..., \hat{a}_p$ are used to construct $\hat{A}(\omega)$ which when applied to the observations yields an estimate of the input process, $\hat{e}(n)$. This residual data can be used to obtain bootstrap resamples $e^*(n)$.
The more realistic scenario of AR plus noise cannot be handled as above. To use an independent data bootstrap scheme for this dependent data model we model the autoregressive plus noise process of order p by an autoregressive moving-average (ARMA) process of order (p, p) as follows: The spectrum of $y(n)$ is given by

$$C_{yy}(\omega) = \frac{\sigma_e^2}{|A(\omega)|^2} + \sigma_u^2 \tag{1}$$

This can be rewritten as

$$\begin{aligned} C_{yy}(\omega) &= \frac{\sigma_e^2 + \sigma_u^2 |A(\omega)|^2}{|A(\omega)|^2} \\ &= \frac{\sigma_z^2 |B(\omega)|^2}{|A(\omega)|^2} \end{aligned}$$

with $B(\omega) = 1 + \sum_{k=1}^{p} b_k e^{-j\omega k}$. This is clearly the spectrum of an ARMA process and the system is depicted in Figure 2.
We can now use ARMA spectrum estimation techniques to obtain estimates $\hat{a}_1, ..., \hat{a}_p$, $\hat{b}_1, ..., \hat{b}_p$ and $\hat{\sigma}_z^2$. By comparison of coefficients of the equation

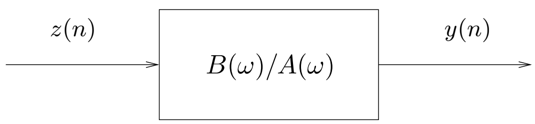

Fig. 2. ARMA process

$$\hat{\sigma}_e^2 + \hat{\sigma}_u^2 |\hat{A}(\omega)|^2 = \sigma_z^2 |\hat{B}(\omega)|^2 \tag{2}$$

estimates $\hat{\sigma}_e^2$ and $\hat{\sigma}_u^2$ can be found.

3. THE RML ALGORITHM

In this section we will shortly review the Recursive Maximum Likelihood (RML) algorithm [9] which will be used later on for ARMA parameter estimation. Given an ARMA process of order (p, q)

$$y(n) = -\sum_{i=1}^{p} a_i y(n-i) + \sum_{i=1}^{q} b_i e(n-i) + e(n) \tag{3}$$

Let θ denote the parameter vector

$$\theta = [a_1, ..., a_p, b_1, ..., b_q]^T$$

where $[\cdot]^T$ denotes matrix transpose. Further, let ϕ_n and ψ_n denote a data vector and its filtered version:

$$\begin{aligned} \phi_n &= [-y(n-1), ..., -y(n-p), r(n-1), ..., r(n-q)]^T \\ \psi_n &= [-\tilde{y}(n-1), ..., -\tilde{y}(n-p), \tilde{r}(n-1), ..., \tilde{r}(n-q)]^T \end{aligned}$$

with $r(n)$ being the residual error at time instance n. With every new data point $y(n)$ the following set of equations will be solved:

- **Prediction error:** $\epsilon_n = y_n - \phi_n^T \hat{\theta}_{n-1}$
- **Error covariance matrix:**
$$P_n = \frac{1}{\lambda} \left[\frac{P_{n-1} - P_{n-1} \psi_n \psi_n^T P_{n-1}}{\lambda + \psi_n^T P_{n-1} \psi_n^T} \right]$$
- **Parameter update:** $\hat{\theta}_n = \hat{\theta}_{n-1} + P_n \psi_n \epsilon_n$
- **Residual error:** $r_n = y_n - \psi_n^T \hat{\theta}_n$

The filtered data vectors $\tilde{y}(n)$ and $\tilde{r}(n)$ can be obtained by applying the filter $\hat{B}(\omega) = 1 + \sum_{k=1}^{q} \hat{b}_i e^{-j\omega k}$ to $y(n)$ and $r(n)$. λ is the so called forgetting factor, typically a constant close to unity.
The RML algorithm has been shown to be consistent for all ARMA models whose zeros are inside the unit circle. Further, its asymptotic efficiency has also been proven. Details can be found in [10].

4. THE EXPERIMENT

As described in Section 2 we model the AR plus noise process of order p by an ARMA process of order (p, p). Given initial estimates of the ARMA parameters one can now apply an inverse filter with transfer function

$$\frac{\hat{A}(\omega)}{\hat{B}(\omega)} = \frac{1 + \sum_{k=1}^{p} \hat{a}_k e^{-j\omega k}}{1 + \sum_{k=1}^{p} \hat{b}_k e^{-j\omega k}}$$

to $y(n)$ to get an estimate $\hat{z}(n)$. This residual data can be used to form bootstrap resamples $\hat{z}^*(n)$. After a large number B of bootstrap resamples a distribution function can be estimated from which we obtain confidence interval estimates. The bootstrap procedure is summarized in Table 1.

4.1. Simulated data example

For the simulations we chose an AR(2) process with parameters $a_1 = -1.4$ and $a_2 = 0.95$. Zero mean uniformly distributed noise has been added to achieve a signal-to-noise ratio (SNR) of 10 dB. $N = 1024$ data samples have been used for the simulations, the number of bootstrap resamples is $B = 1000$. We used the RML algorithm for ARMA parameter estimation.

Table 1. The Bootstrap procedure

Step 0.	*Data Collection*. Conduct the experiment and collect N observations $y(n), n = 0, ..., N-1$ from an autoregressive plus noise process of order p.
Step 1.	*Calculation of the residuals*. Model $y(n)$ by an ARMA(p,p) process and estimate its parameter vector $\hat{\theta} = [\hat{a}_1, ..., \hat{a}_p, \hat{b}_1, ..., \hat{b}_p]^T$. Apply the inverse filter with transfer function $\frac{\hat{A}(\omega)}{\hat{B}(\omega)}$ to $y(n)$ to get the residuals $\hat{z}(n)$.
Step 2.	*Resampling*. Create a bootstrap sample $y^*(0), y^*(1), ..., y^*(N-1)$ by drawing $\hat{z}^*(0), \hat{z}^*(1), ..., \hat{z}^*(N-1)$, with replacement, from the residuals $\hat{z}(0), \hat{z}(1), ..., \hat{z}(N-1)$, then passing $\hat{z}^*(0), \hat{z}^*(1), ..., \hat{z}^*(N-1)$ through the linear time-invariant filter with frequency response $\frac{\hat{B}(\omega)}{\hat{A}(\omega)}$.
Step 3.	*Calculation of the bootstrap estimate*. From $y^*(n), n = 0, ..., N-1$ obtain a bootstrap estimate of the ARMA parameter vector $\hat{\theta}^* = [\hat{a}_1^*, ..., \hat{a}_1^*, \hat{b}_p^*, ..., \hat{b}_p^*]^T$. By comparison of coefficients of Equation (2) find bootstrap estimates $(\hat{\sigma}_e^*)^2$ and $(\hat{\sigma}_u^*)^2$
Step 4.	*Repetition*. Repeat Steps 2-3 a large number of times to obtain a total of B bootstrap statistics $\hat{\theta}^{*1}, \hat{\theta}^{*2}, ..., \hat{\theta}^{*B}$, $(\hat{\sigma}_e^{*1})^2, (\hat{\sigma}_e^{*2})^2, ..., (\hat{\sigma}_e^{*B})^2$ and $(\hat{\sigma}_u^{*1})^2, (\hat{\sigma}_u^{*2})^2, ..., (\hat{\sigma}_u^{*B})^2$.
Step 5.	*Distribution Function estimation*. Sort the bootstrap estimates in increasing order to obtain $h(\hat{\theta}^*)$, $h((\hat{\sigma}_e^*)^2)$ and $h((\hat{\sigma}_u^*)^2)$. Approximate the density function of $h(\hat{\theta})$, $h(\hat{\sigma}_e^2)$ and $h(\hat{\sigma}_u^2)$ by $h(\hat{\theta}^*)$, $h((\hat{\sigma}_e^*)^2)$ and $h((\hat{\sigma}_u^*)^2)$ respectively.
Step 6.	*Confidence Interval Estimation*. Based on $h(\hat{\theta}^*)$, $h((\hat{\sigma}_e^*)^2)$ and $h((\hat{\sigma}_u^*)^2)$ estimate the desired $(1-\alpha)100\%$ bootstrap confidence intervals.

In Figure 3 a typical estimate of the residual's autocovariance function $c_{zz}(m)$ is shown. Its structure which is close to the one of a white process is an indicator for the good performance of the ARMA parameter estimator. In Figures 4 and 5 the bootstrap distribution function estimate of the AR parameters is shown versus the histograms obtained by 1000 Monte Carlo simulations. Given the estimate of the bootstrap distribution function one can obtain the 90% confidence interval estimates for the AR parameters as $(-1.411, -1.371)$ for a_1 and $(0.925, 0.963)$ for a_2.

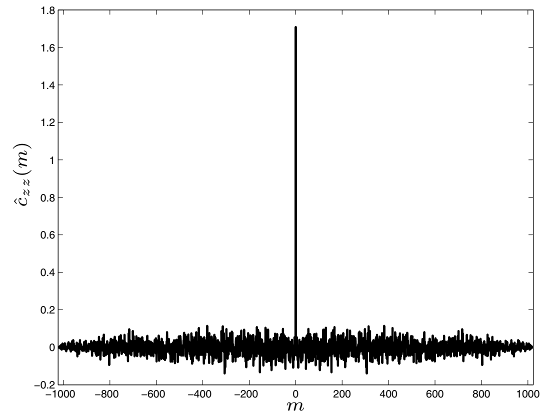

Fig. 3. Autocovariance function of the residual data

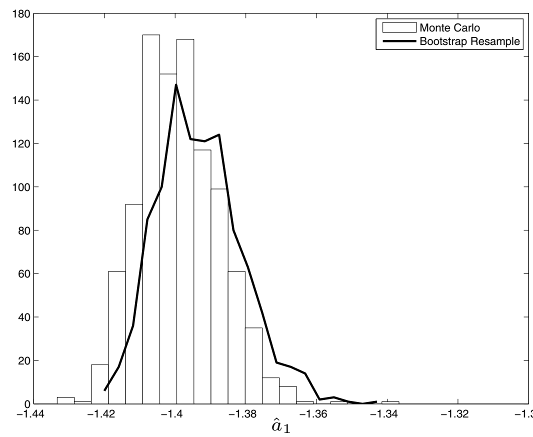

Fig. 4. Bootstrap distribution function of $\hat{a}_1$

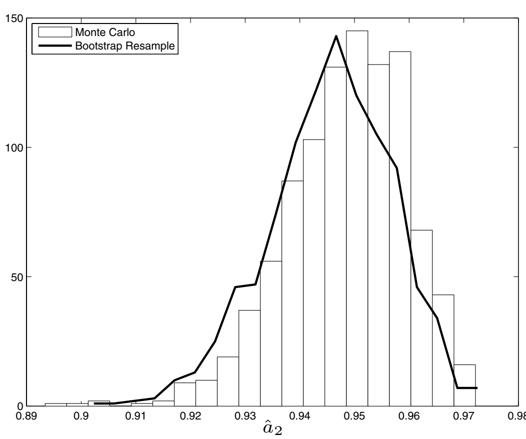

Fig. 5. Bootstrap distribution function of $\hat{a}_2$

5. REAL DATA EXAMPLE

In order to demonstrate our proposed method using a real data example we chose a problem from automotive engineering, namely estimating the spectrum of the pressure signal inside the combustion chamber. Although the signal is known to be non-stationary [11] stationarity is assumed for simplicity in this example.

AR Parameter	Confidence Interval
a_1	$(-0.94, -0.60)$
a_2	$(1.25, 1.66)$
a_3	$(-0.42, 0.20)$
a_4	$(0.68, 1.26)$
a_5	$(-0.07, 0.59)$
a_6	$(0.82, 1.45)$
a_7	$(-0.07, 0.65)$
a_8	$(0.41, 1.07)$
a_9	$(0.36, 1.03)$
a_{10}	$(0.12, 0.71)$
a_{11}	$(0.29, 0.86)$
a_{12}	$(0.03, 0.54)$
a_{13}	$(0.12, 0.63)$
a_{14}	$(-0.08, 0.23)$
a_{15}	$(0.15, 0.39)$

Table 2. 90% confidence intervals for the AR parameters

We model the signal by an AR plus noise process and apply the ARMA modeling described in Section 2 using the RML algorithm for parameter estimation. The order has been fixed to $p = 15$. $B = 1000$ bootstrap resamples have been considered for estimating the confidence intervals. The 90% confidence intervals for the AR parameters can be calculated as shown in Table 2. Finally Figure 6 shows the spectrum using the AR plus noise method versus the spectrum obtained using Welch's periodogram. It can clearly be seen that the parametric spectrum estimate is capable of detecting the different modes in the signal.

Finding confidence intervals for the spectrum of the pressure signal is straightforward using the bootstrap approach depicted in Table 1 and presents a real alternative to asymptotic method [8]. Results on confidence interval estimation for the spectrum of the pressure signal will be presented elsewhere.

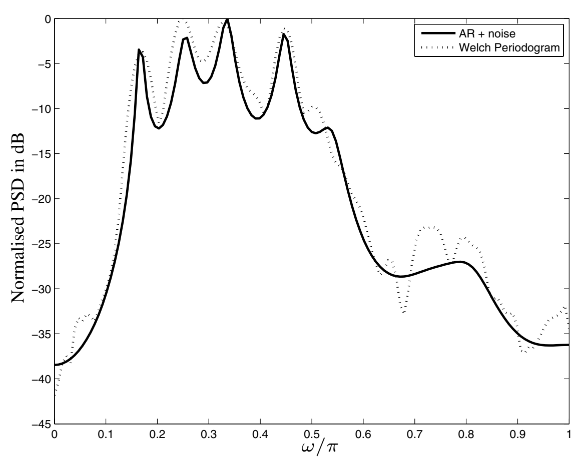

Fig. 6. Spectra of the pressure signal by AR + noise method and Welch's periodogram

6. CONCLUSION

We presented a bootstrap based approach for confidence interval estimation for autoregressive plus noise processes. The need for ARMA modeling of the process was justified. We used the independent data bootstrap to successfully estimate confidence intervals in a computer simulated as well as in a real world example. Future work will be done in the more sophisticated task of estimating confidence bands for parametric spectra.

7. REFERENCES

[1] D. R. Brillinger, *Time Series: Data Analysis and Theory*, Holt, R & W, 1975.

[2] S.M. Kay and S.L. Marple, "Spectrum analysis - a modern perspective," *Proceedings of the IEEE*, vol. 69, no. 11, pp. 1380–1419, 1981.

[3] S. Kesler and S. Haykin, "The maximum entropy method applied to the spectral analysis of radar clutter," *IEEE Transactions on Information Theory*, vol. 24, no. 2, pp. 269–272, 1978.

[4] L.J. Griffiths and R. Prieto-Diaz, "Spectral analysis of natural seismic events using autoregressive techniques," *IEEE Transactions on Geoscience Electronics*, vol. 15, no. 1, pp. 13–25, 1977.

[5] W. Gersch, "Spectral analysis of EEG's by autoregressive decomposition of time series," *Math. Biosc.*, vol. 7, pp. 205–222, 1970.

[6] J. Cadzow and K. Ogina, "Two-dimensional spectral estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, pp. 396–401, 1981.

[7] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[8] A. M. Zoubir and D. R. Iskander, *Bootstrap Techniques for Signal Processing*, Cambridge University Press, 2004.

[9] B. Friedlander, "A recursive maximum likelihood algorithm for ARMA spectral estimation," *IEEE Transactions on Information Theory*, vol. 28, no. 4, pp. 639–646, 1982.

[10] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*, MIT Press, 1983.

[11] S. Carstens-Behrens, M. Wagner, and J.F. Böhme, "Improved knock detection by time variant filtered structure-borne sound," in *Proceedings of the 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1999, vol. 4, pp. 2255–2258 vol.4.