ON THE MAXIMUM ENTROPY PROBLEM WITH AUTOCORRELATIONS SPECIFIED ON A LATTICE

by

Dimitris N. Politis
Purdue University

Technical Report #92-13

Department of Statistics
Purdue University

April 1992

On the Maximum Entropy problem with autocorrelations specified on a lattice

Dimitris N. Politis

Department of Statistics, Purdue University
West Lafayette, IN 47907-1399

Abstract

The Maximum Entropy process with autocorrelations specified on a finite lattice is identified to be a Gaussian autoregressive process with a special structure in its coefficients. The autoregressive coefficients can be obtained by means of a fast algorithm. This result extends Burg's well known Maximum Entropy theorem, where the autocorrelation is constrained for consecutive lags.

Keywords. Autoregressive processes, Gaussian time series, Maximum Entropy.

I. Introduction

Let $\{X_n, n \in \mathbf{Z}\}$ be a wide-sense stationary stochastic process with mean $EX_t = 0$, and autocovariance $\gamma(k) = EX_tX_{t+k}$, $k \in \mathbf{Z}$, and assume for simplicity that $\gamma(0) = 1$. Suppose that the first p autocorrelations of the process are known, i.e., that

$$\gamma(i) = c_i, i = 1, \ldots, p \tag{1}$$

where the given constants $c_i, i = 1, \ldots, p$ constitute the beginning of some positive definite sequence.

To obtain a full picture of the second moment structure, a model should be assumed in order to somehow extrapolate the unknown autocorrelation values $\gamma(i)$, $i > p$. A popular such model is the AR(p) autoregressive Gaussian model

$$X_t + a_1X_{t-1} + \ldots + a_pX_{t-p} = Z_t \tag{2}$$

where the sequence $\{Z_n\}$ is a sequence of independent, identically distributed, normal $N(0, \sigma^2)$ random variables, and the coefficients $a_1, \ldots, a_p$ are chosen (via the Yule-Walker equations) such that the constraints (1) are satisfied. The AR(p) model (2) has the additional feature (cf. Burg [2]) that it is the Maximum Entropy (i.e., most 'unpredictable') process that satisfies the constraints (1).

For motivation consider now a simple example that does not fall in the above framework, namely that it is known that $\gamma(3) = 1$, and that no other information is available regarding the autocorrelation sequence. By the Cauchy-Schwarz inequality, $|\gamma(3)| \leq \gamma(0) = 1$ with equality if and only if X_t and X_{t-3} are linearly dependent, for any t. Hence in our case, because of the assumption of zero mean and stationary variances, $\gamma(3) = 1$ implies that $X_t = X_{t-3}$, for any t, which can be viewed as an extreme case of an autoregressive AR(3) model, with $a_1 = a_2 = 0$, $a_3 = -1$, and noise variance $\sigma^2 = 0$. Similarly, if it is known that $\gamma(3) = -1$, it follows that $X_t = -X_{t-3}$, for any t, which is also an extreme case of an AR(3) model.

It is natural to ask what happens if the value of $\gamma(3) = c_3 \neq \pm 1$ is known, and no other information about the autocorrelation sequence is given. Intuitively, one would expect an AR(3) model to still be applicable, and hopefully that in addition $a_1 = a_2 = 0$. Indeed this guess is true, and it is shown in the next section (as a part of a more general result) that the Maximum Entropy process satisfying $\gamma(3) = c_3$, is the AR(3) model $X_t = c_3 X_{t-3} + Z_t$.

II. The Maximum Entropy process with autocorrelation specified on a lattice

The Maximum Entropy problem with non-consecutive or missing autocorrelation values was discussed in Papoulis [5] where a numerical (steepest ascent) algorithm was proposed for its solution. It was also addressed in [6] in the context of MA (moving average) processes, but again its solution involved a system of non-linear equations.

However, if there is a particular structure in the missing values, the solution to the Maximum Entropy problem with missing autocorrelations takes on a very simple form. The following result shows that, if the autocorrelation sequence is known on a finite lattice of the form $\{r, 2r, 3r, \ldots, pr\}$, then the Maximum Entropy problem has an intuitive AR (autoregressive) solution, with coefficients that vanish for lags that are not on the lattice. The remaining non-zero coefficients are easily obtainable from Yule-Walker type linear equations using a fast algorithm.

Theorem. *Suppose that $\{X_n, n \in \mathbf{Z}\}$ is the Gaussian autoregressive process satisfying the difference equation*

$$X_t + \sum_{k=1}^{rp} a_k X_{t-k} = Z_t \tag{3}$$

where r, p are two positive integers, and the sequence $\{Z_n\}$ is a sequence of independent, identically distributed, normal $N(0, \sigma^2)$ random variables. Also suppose that $\gamma(0) = 1$, and that

$$\gamma(ri) = c_{ri}, i = 1, \ldots, p \tag{4}$$

where $c_r, c_{2r}, \ldots, c_{pr}$ are some given constants.

If $a_i = 0$ for all $i \notin \{r, 2r, 3r, \ldots, pr\}$, then the process $\{X_n\}$ has Maximum Entropy rate among all wide-sense stationary processes whose autocovariances satisfy the constraints (4).

Proof. The fact that the maximum entropy process turns out to be Gaussian with mean zero is a consequence of the fact that equation (4) represents a constraint on the second order moments only. As is well known [3] among processes with identical autocovariance function, the mean zero Gaussian such process has maximum entropy rate.

Now the conditions that a Gaussian process should satisfy in order to have Maximum

Entropy rate subject to the constraints (4) are (cf. [4], [6], [7])

$$\tilde{\gamma}(i) = 0, \forall i \notin \{r, 2r, 3r, \ldots, pr\} \tag{5}$$

where $\tilde{\gamma}(k)$, $k \in \mathbf{Z}$, is the inverse autocorrelation sequence of the process $\{X_t\}$, which is defined as the sequence satisfying $\sum_{i=-\infty}^{\infty} \tilde{\gamma}(i)\gamma(i+k) = 0$ for all $k \neq 0$, $\sum_{i=-\infty}^{\infty} \tilde{\gamma}(i)\gamma(i) > 0$, and $\tilde{\gamma}(0) = 1$.

Because $\tilde{\gamma}(i) = 0$ for all $i > pr$, it follows (cf. [1], [6], [7]) that the Maximum Entropy process satisfying (4) is an AR process of order pr satisfying the difference equation (3), and hence its inverse autocorrelation can be easily calculated as

$$\tilde{\gamma}(k) = \frac{\sum_{i=1}^{pr-k} a_i a_{i+k}}{\sum_{i=1}^{pr} a_i^2} \tag{6}$$

for $k = 1, \ldots, pr$.

A direct calculation now shows that, if $a_i = 0$ for all $i \notin \{r, 2r, 3r, \ldots, pr\}$, then the maximum entropy conditions (5) are indeed satisfied, and the theorem is proven.□

The above theorem includes as a special case (with $r = 1$) the original Maximum Entropy result of Burg [2]. As another special case (with $p = 1$ and $r = 3$), the AR(3) example $X_t = c_3 X_{t-3} + Z_t$ of the Introduction is shown to have Maximum Entropy among all processes satisfying $\gamma(3) = c_3$.

The theorem does not immediately extend to the case where there is no such lattice pattern in the missing autocorrelation values. To see this, consider the following simple example. Suppose it is given that

$$\gamma(i) = c_i \neq 0, i = 3, 4 \tag{7}$$

Then the Maximum Entropy process subject to constraint (7) is certainly an AR(4) Gaussian process, but its coefficients can *not* satisfy $a_1 = a_2 = 0$; for if $a_1 = a_2 = 0$ *were* satisfied, then the maximal condition $\tilde{\gamma}(1) = 0$ would imply that either a_3 or a_4 is zero, and consequently that either c_3 or c_4 should be zero, which is a contradiction.

To turn to the practical problem of solving for the AR coefficients in the Maximum Entropy model of the theorem, that is,

$$X_t + a_r X_{t-r} + a_{2r} X_{t-2r} + \cdots + a_{pr} X_{t-pr} = Z_t \tag{8}$$

using the given information $\gamma(0) = 1$, and $\gamma(ri) = c_{ri}, i = 1, \ldots, p$, one would proceed the usual way, i.e., multiply both sides of equation (8) with $X_{t-ri}, i = 0, 1, \ldots, p$, and take expectations to obtain the following Yule-Walker type system of equations

$$\begin{aligned}
\gamma(0) + a_r\gamma(r) + a_{2r}\gamma(2r) + \cdots + a_{pr}\gamma(pr) &= \sigma^2 \\
\gamma(r) + a_r\gamma(0) + a_{2r}\gamma(r) + \cdots + a_{pr}\gamma(pr - r) &= 0 \\
\gamma(2r) + a_r\gamma(r) + a_{2r}\gamma(0) + \cdots + a_{pr}\gamma(pr - 2r) &= 0 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\gamma(pr) + a_r\gamma(pr - r) + a_{2r}\gamma(pr - 2r) + \cdots + a_{pr}\gamma(0) &= 0
\end{aligned} \tag{9}$$

The linear system of equations (9) can then be solved for $a_r, a_{2r}, \ldots, a_{pr}$, and σ^2, without a matrix inversion by using one of the well known fast algorithms (Burg's or Levinson's [1]). To see this, observe that if we define the new stationary process $\{Y_n\}$ by $Y_n = X_{nr}$, for all $n \in \mathbf{Z}$, then the autocovariance $\gamma_Y(k) = EY_tY_{t+k}$ of the $\{Y_n\}$ sequence coincides with the autocovariance of the $\{X_n\}$ sequence on the lattice $\{0, \pm r, \pm 2r, \ldots\}$, that is, $\gamma_Y(k) = \gamma(kr)$, for all $k \in \mathbf{Z}$. Hence, solving the system (9) is tantamount to fitting the AR(p) model

$$Y_t + a_rY_{t-1} + a_{2r}Y_{t-2} + \ldots + a_{pr}Y_{t-p} = Z_t \tag{10}$$

to the $\{Y_n\}$ sequence, using the information that $\gamma_Y(0) = 1$, and $\gamma_Y(i) = c_{ri}, i = 1, \ldots, p$, which can be done by one of the fast algorithms.

Looking at the sequence $\{Y_n\}$, which is actually the original sequence $\{X_n\}$ sampled at a reduced rate, points to a practical situation where our theorem might be useful. For an example, consider a daily time series $\{X_n\}$, where X_n is a certain random quantity at day n, and suppose that the time series was observed weekly (instead of daily) over a year (=52 weeks), i.e., the data were $X_0, X_7, X_{14}, \ldots, X_{357}$. In other words, the *weekly* time series $Y_n = X_{7n}$, was observed for $n = 0, 1, 2, \ldots, 51$.

Since the $\{Y_n\}$ observations are consecutive and complete, the $\{Y_n\}$ autocovariances $\gamma_Y(k)$ can be estimated from the data for $k = 0, 1, \ldots, p$, (with $p << 52$), and an AR(p) model (10) can be fit by solving the Yule-Walker system of equations (9). If however the objective was to model the *daily* time series $\{X_n\}$ using the incomplete $X_0, X_7, X_{14}, \ldots, X_{357}$ data, our theorem could be invoked to infer that the Maximum Entropy model for the $\{X_n\}$ sequence is the AR(pr) model (8), with the *same* AR coefficients as in the AR(p) model for $\{Y_n\}$. Note that model (8) specifies uniquely a probability model for the $\{X_n\}$ sequence, and can be used for all the practical purposes of spectral estimation, prediction, interpolation, and so forth.

III. Conclusions

The Maximum Entropy problem with autocorrelation specified on a lattice of the form $\{r, 2r, 3r, \ldots, pr\}$ was discussed, and its solution was found to be a particular Gaussian AR(pr) model (cf. equation (8)), characterized by coefficients that vanish for lags not on the lattice. The remaining non-zero coefficients were shown to be obtainable from Yule-Walker type equations using a fast algorithm. This finding (which extends the well known Maximum Entropy result of Burg [2]) might be applied in situations where the stationary process of interest is sampled at a rate smaller than the desirable one.

References

[1] Brockwell, P.J. and Davis, R.A., *Time series: Theory and Methods*, Springer-Verlag, 1991.

[2] Burg, J.P., Maximum Entropy Spectral Analysis, *Proc. 37th Ann. Int. Mtg. Soc. Explor. Geophys.*, Oklahoma City, 1967.

[3] Cover, T.M. and Thomas, J., *Elements of Information Theory*, John Wiley, New York, 1991.

[4] Lev-Ari, H., Parker, S.R., and Kailath, T., Multidimensional Maximum-Entropy Covariance Extension, *IEEE Trans. Inform. Theory*, vol. 35, 3, pp. 497-508, 1989.

[5] Papoulis, A., On entropy rate, in *Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems*, C.R.Smith and G.J.Erickson, eds., D.Reidel Publication Company, Norwell, Mass., 1987.

[6] Politis, D. N., Moving average processes and maximum entropy, *IEEE Trans. Information Theory*, IT-38, No. 3, May 1992.

[7] Robinson, E.A., A historical perspective of spectral estimation, *Proc. IEEE*, vol. 70, No. 9, pp. 885-907, 1982.