

Complex scalings for categorical time series

Georgi N. Boshnakov

Institute of Mathematics, Bulgarian Academy of Sciences

Abstract

The spectral envelope has been introduced recently as a means for harmonic analysis of categorical time series. The associated scalings of the categories are real. The aim of this note is to give some arguments that simpler theory can be achieved by allowing for complex scalings. The complex values seem to enrich the interpretation of the scalings.

Keywords: Spectral envelope, Scaling, Categorical time series

1 Introduction

The description of a time series through its spectrum is one of the basic tools in time series analysis. It has a highly intuitive interpretation, especially when one is interested in the study of (potential) pseudo-periodic properties of the underlying process.

The spectral theory is well developed for numerical time series. For categorical ones little has been done. The most obvious way for dealing with them is to recode the levels of the category with numerical values (scalings) and then perform standard spectral analysis on the resulting numerical time series. However, this approach needs some adaptation to the considered problem since, when applied in this way, it is not invariant with respect to the different scalings—different scalings give different spectra.

A new approach to overcome this difficulty has been recently proposed by Stoffer et al. [3]. They introduce the notion of the “spectral envelope” of a categorical time series. The spectral envelope is defined through the solution of an extremal problem. The interpretation of the spectral envelope is similar to that of the classical spectra. In particular, it represents in a sense

the distribution of the “power” of the process over different frequencies and the most interesting points in the spectrum are its “peaks”. Furthermore, the scalings associated with the spectral envelope can be used for assesment of the mutual relations between the categories. The definition of the spectral envelope given in [3] produces real scalings.

The natural mathematical statement of many problems is in the framework of the complex numbers. Nevertheless, in applications involving real data this framework is not always appropriate. The categorical time series pose no problems in this respect since their values are not numerical, anyway. Therefore, we are free to select such numerical field that best reveals the harmonic properties of the data.

The aim of this note is to give some arguments that simpler theory can be achieved by allowing for complex scalings. The complex values seem to enrich the interpretation of the scalings, without sacrificing the intuition.

2 Notations

The notations from [3] are preserved as much as possible. Unless otherwise stated the vectors are $k \times 1$ column vectors, the matrices — $k \times k$. Their elements are denoted by putting appropriate indices next to their names. A tilde over a matrix name is used for the upper-left $(k-1) \times (k-1)$ block of the corresponding matrix. For vectors the tilde selects the first $(k-1)$ elements of the vector. The eigenvalues in the diagonal representations of the Hermitian matrices are in decreasing order. The identity matrix is denoted by I , the vector of 1's — by $\mathbf{1}$. An additional index is used when the dimension is not clear. An upper index “*re*” is used for the real part of a complex matrix. A star denotes the conjugate transpose of a matrix or vector.

3 Scaling

Let $\{X_t\}$ be a stationary categorical time series which takes values in the finite set $\mathcal{C} = \{c_1, \dots, c_k\}$ with (non-zero) probabilities $p_j = \mathbf{P}(X_t = c_j) > 0$, $j = 1, \dots, k$. Any numerical vector $\beta = (\beta_1, \dots, \beta_k)^*$ can be used to transform $\{X_t\}$ into a numerical time series $X_t(\beta)$ by assigning the value β_j^* to the category c_j .

Let e_j , $j = 1, \dots, k$ be the columns of the $k \times k$ identity matrix I_k . A k -variate representation $\{Y_t\}$ of $\{X_t\}$ can be obtained by setting $Y_t = e_j$ when $X_t = c_j$. Then the relationship between Y_t and $X_t(\beta)$ is given by the formula $X_t(\beta) = \beta^* Y_t$. The variance-covariance matrix of Y_t (the lag 0 autocovariance matrix of the process $\{Y_t\}$) is denoted by V .

The spectral densities $f(\omega, \beta)$ of $X_t(\beta)$ and $f(\omega)$ of Y_t are assumed to exist and to be continuous. They are connected by the relation

$$f(\omega, \beta) = \beta^* f(\omega) \beta.$$

Although for any given β the function $f(\omega, \beta)$ expresses some harmonic properties of the underlying time series, the vectors β are not equally suited for this purpose. One extremal choice is to assign to all categories the same numerical value c , i.e. $\beta = c\mathbf{1}_k$, in which case $f(\omega, \beta)$ is identically zero.

One natural question is: What is the maximal power that could be attributed to a given frequency ω ? The spectrum $f(\omega, \beta)$ itself can not be used directly, so the following criterion function is introduced,

$$s(\omega, \beta) = \frac{\{\beta^* f(\omega) \beta\}}{\beta^* V \beta},$$

where β obeys the following condition: $\beta \neq \mathbf{0}$ and $\beta \neq c\mathbf{1}$. Note that both the nominator and the denominator are different from zero if and only if β obeys this condition. The sets of real and complex vectors β obeying this condition will be denoted by \mathcal{R}_1^k and \mathcal{C}_1^k , respectively.

Consider the functionals $\lambda_S(\omega)$ and $\lambda(\omega)$, defined as

$$\lambda_S(\omega) = \sup_{\beta \in \mathcal{R}_1^k} \frac{\{\beta^* f(\omega) \beta\}}{\beta^* V \beta} = \sup_{\beta \in \mathcal{R}_1^k} \frac{\{\beta' f^{re}(\omega) \beta\}}{\beta' V \beta} \quad (1)$$

$$\lambda(\omega) = \sup_{\beta \in \mathcal{C}_1^k} \frac{\{\beta^* f(\omega) \beta\}}{\beta^* V \beta}. \quad (2)$$

Clearly $\lambda_S(\omega) \leq \lambda(\omega)$ for all ω .

The following definition of the spectral envelope is given in [3].

Definition 1 *The function $\lambda_S(\omega)$, $\omega \in [-\pi, \pi]$ is said to be the spectral envelope of $\{X_t\}$.*

So defined, the spectral envelope $\lambda_S(\omega)$ does not use the full information about the imaginary part of the spectrum f .

On the other hand, the spectral density matrix is Hermitian and non-negative definite. Hence the values of $s(\omega, \beta)$ are real and nonnegative even for complex vectors β . This property allows for the definition of $\lambda(\omega)$ given above (see equation (2)). The following definition of the spectral envelope is used in the next sections.

Definition 2 *The function $\lambda(\omega)$, $\omega \in [-\pi, \pi]$ is said to be the spectral envelope of $\{X_t\}$.*

With this definition the values of the spectral envelope are again real, while the scalings are (possibly) complex.

4 The nature of the extremal problem

The spectral envelope has led us to the solution of a generalized eigenvalue problem for the couple of matrices (f, V) or to the study of the extremal values of a Rayleigh quotient, see [1, Chapter 8]. Care must be taken, since both f and V are of non-full rank ($k - 1$). However, since the upper-left blocks of f and V are non-singular, Hermitian and positive definite we can contemporarily diagonalize them with non-singular matrices ([2, § 15.3]). Namely, let G be a $(k - 1) \times (k - 1)$ matrix such that

$$\tilde{f} = G\tilde{M}G^*, \quad \tilde{V} = G\tilde{D}G^*,$$

where $\tilde{M} = \text{diag}(m_1, \dots, m_{k-1})$ and $\tilde{D} = \text{diag}(d_1, \dots, d_{k-1})$ are diagonal matrices with positive real main diagonals. Then

$$f = FMF^*, \quad V = FDF^*,$$

where

$$F = \begin{pmatrix} G & 0 \\ 0' & 1 \end{pmatrix}$$

$$M = \begin{pmatrix} \tilde{M} & G^{-1}f_{\cdot k} \\ f_{k\cdot}(G^*)^{-1} & f_{kk} \end{pmatrix}$$

$$D = \begin{pmatrix} \tilde{D} & G^{-1}V_{\cdot k} \\ V_{k\cdot}(G^*)^{-1} & V_{kk} \end{pmatrix}$$

(f_{k-} is the k -th row of f , f_{-k} — its k -th column.)

Let $\gamma = F^*\beta$. We can suppose without loss of generality that $\beta_k = 0$. It is clear that then $\gamma_k = 0$ too. Therefore

$$\frac{\beta^* f \beta}{\beta^* V \beta} = \frac{\beta^* F M F^* \beta}{\beta^* F D F^* \beta} = \frac{\gamma^* M \gamma}{\gamma^* D \gamma} = \frac{\tilde{\gamma}^* \tilde{M} \tilde{\gamma}}{\tilde{\gamma}^* \tilde{D} \tilde{\gamma}} = \frac{\sum_{i=1}^{k-1} \mu_i \gamma_i^2}{\sum_{i=1}^{k-1} d_i \gamma_i^2}.$$

Putting $\delta_i = \sqrt{d_i} \gamma_i$ we obtain

$$\frac{\beta^* f \beta}{\beta^* V \beta} = \frac{\sum_{i=1}^{k-1} (\mu_i/d_i) \delta_i^2}{\sum_{i=1}^{k-1} \delta_i^2} = \sum_{i=1}^{k-1} \frac{\mu_i}{d_i} \left(\frac{\delta_i^2}{\sum_{i=1}^{k-1} \delta_i^2} \right).$$

Therefore

$$\min_{1 \leq i \leq k-1} \frac{\mu_i}{d_i} \leq \frac{\beta^* f \beta}{\beta^* V \beta} \leq \max_{1 \leq i \leq k-1} \frac{\mu_i}{d_i}, \quad (3)$$

and in particular

$$\max \frac{\beta^* f \beta}{\beta^* V \beta} = \max_{1 \leq i \leq k-1} \frac{\mu_i}{d_i}. \quad (4)$$

Denoting by s the index at which this maximum is achieved we have for the δ 's for which $s(\omega, \beta)$ achieves its maximum that $\delta_s = 1$, and $\delta_i = 0$ when $i \neq s$. Hence $\gamma_s = 1/\sqrt{d_s}$. Since $\beta = (F^*)^{-1} \gamma$, we can see that the optimal β is obtained from the s -th column of $(F^*)^{-1}$ (multiplied by $1/\sqrt{d_s}$).

It is known also that $\mu_i/d_i, i = 1, \dots, k-1$ do not depend on the particular choice of the matrix F . For the maximal among them this is obvious from equation 4.

Another look at the maximum of $s(\omega, \beta)$ can be taken as follows. Let $f = U \Lambda U^*, V = Q D Q^*$, with unitary matrices U and Q , diagonal matrices Λ and D . As usual the elements on their diagonals are in non-increasing order, so that $\lambda_k = d_k = 0$. Then we have

$$\frac{\beta^* f \beta}{\beta^* V \beta} = \frac{\beta^* U \Lambda U^* \beta}{\beta^* Q D Q^* \beta} = \frac{\gamma^* Q^* U \Lambda U^* Q \gamma}{\gamma^* D \gamma} = \frac{\tilde{\gamma}^* \tilde{M} \tilde{\gamma}}{\tilde{\gamma}^* \tilde{D} \tilde{\gamma}} + \frac{g M_k \gamma}{\gamma^* D \gamma}, \quad (5)$$

where $\gamma = Q^* \beta$, $M = Q^* U \Lambda U^* Q$, $g = (0, \dots, 0, \gamma_k)$, the matrices \tilde{M} and \tilde{D} are the upper $(k-1) \times (k-1)$ blocks of M and D respectively. The last term in equation 5 is zero, because $\gamma_k = 0$. Hence the problem of the

maximization of $s(\omega, \beta)$ is equivalent to a generalized eigenvalue problem with positive definite \tilde{D} . Proceeding further, let $\delta_i = \sqrt{d_i}\gamma_i$. Then

$$\frac{\beta^* f \beta}{\beta^* V \beta} = \frac{\tilde{\delta}^* \tilde{D}^{-1/2} \tilde{M} \tilde{D}^{-1/2} \tilde{\delta}}{\tilde{\delta}^* \tilde{\delta}}.$$

This is the Rayleigh quotient whose maximum is at the maximal eigenvalue of the matrix $\tilde{D}^{-1/2} \tilde{M} \tilde{D}^{-1/2}$ in the nominator. It is attained at the corresponding eigenvector. In the time domain these transformations are equivalent to the transformation $Z_t = \tilde{V}^{-1/2} \tilde{X}_t$.

For this problem the setting $\beta_k = 0$ is natural and non-restrictive. Nevertheless, it is useful to see what happens if the condition $\beta_k = 0$ is replaced by the seemingly more general one $a^* \beta = 0$. The problem is to maximize $s(\omega, \beta)$ under this condition.

Let a be a fixed vector, whose k -th element is nonzero. Let us consider the quadratic form $x^* f x$, for vectors x , such that $a^* x = 0$. Let M be the matrix

$$M = \begin{pmatrix} I_{k-1} & 0 \\ \tilde{a}^* & a_k^* \end{pmatrix}.$$

The inverse matrix of M is the following

$$M^{-1} = \begin{pmatrix} I_{k-1} & 0 \\ \tilde{b}^* & b_k^* \end{pmatrix},$$

where $b_k^* = 1/a_k^*$, $\tilde{b}^* = -\tilde{a}^*/a_k^*$.

Put $y = Mx$. For the x 's under consideration the k -th element y_k of y is zero. Therefore

$$x^* f x = y^* (M^{-1})^* f M^{-1} y = y^* G y = \tilde{y}^* \tilde{G} \tilde{y},$$

where $G = (M^{-1})^* f M^{-1}$.

In this way the more general case $a^* x = 0$ is reduced to the case $y_k = 0$. If \tilde{y} is known then for every a the corresponding x can be found easily, $x = M^{-1}y$, with appropriate matrix M . The requirement $a_k \neq 0$ is not a restriction since it corresponds to reordering of the variables in x (or equivalently to multiplication by a permutation matrix).

The connection between \tilde{G} and \tilde{f} can be seen more explicitly as follows

$$\begin{aligned}
G &= (M^{-1})^* f M^{-1} = \begin{pmatrix} I & \tilde{b} \\ 0' & b_k \end{pmatrix} \begin{pmatrix} \tilde{f} & f_{\cdot k} \\ f_{k\cdot} & f_{kk} \end{pmatrix} \begin{pmatrix} I & 0 \\ \tilde{b}^* & 0' \end{pmatrix} b_k^* \\
&= \begin{pmatrix} \tilde{f} + \tilde{b}f_{\cdot k} & f_{\cdot k} + \tilde{b}f_{kk} \\ b_k f_{k\cdot} & b_k f_{kk} \end{pmatrix} \begin{pmatrix} I & 0 \\ \tilde{b}^* & 0' \end{pmatrix} b_k^* \\
&= \begin{pmatrix} \tilde{f} + \tilde{b}f_{k\cdot} + f_{\cdot k}\tilde{b}^* + \tilde{b}f_{kk}\tilde{b}^* & f_{\cdot k}b_k^* + \tilde{b}f_{kk}b_k^* \\ b_k f_{k\cdot} + b_k f_{kk}\tilde{b}^* & b_k f_{kk}b_k^* \end{pmatrix}
\end{aligned}$$

Therefore $\tilde{G} = \tilde{f} + \tilde{b}f_{k\cdot} + f_{\cdot k}\tilde{b}^* + \tilde{b}f_{kk}\tilde{b}^*$.

5 Estimation of the spectral envelope

Periodogram- and smoothed periodogram- based estimators of the spectral envelope from real scalings have been considered in [3]. The asymptotic distribution of the periodogram is complex Wishart. The distribution of the largest root of a Wishart matrix is not easily tractable, but the distribution of the largest root of its real part is even more difficult. In the former case there are some known results, see the discussion in [3].

The theorems from [3] are easily adapted to the case of complex scalings, the main difference being that there is no need to take real parts of the complex Wishart distributions. For example, in [3, Theorem 3.1] the only change is to remove words like “real part” and the upper index re of W . We are not going to reproduce these results here, because they require a lot of additional notation to be introduced.

6 Complex scalings

When the scalings β are real, the following interpretation can be given to them:

1. If two categories obtain the same (or close) values then these are merged into one greater category.
2. Categories that obtain values with opposite sign, but with close absolute values — of equal importance.

The complex scalings have an additional “degree of freedom” — while in the case of real numbers equal absolute values occur only if the numbers are equal or with opposite signs, in the complex case equal absolute values have numbers which lie on the same circle. In other words, the complex sign (or “phase”) takes on values $\exp(i\phi)$, where $\phi \in (-\pi, \pi]$. The + and – signs correspond to $\phi = 0$ and $\phi = \pi$, respectively. Categories whose scalings have almost equal absolute values appear in the time series with similar “dynamic”. The phase characterizes the time shift between them.

7 Acknowledgements

The paper has been completed during the author’s visit to the Division of Quality Technology and Statistics, Luleå University, Sweden. This research was partly supported also by contract No. I 19/91 with the Bulgarian Ministry of Science and Education.

References

- [1] P. Lancaster and M. Tismenetsky. *The Theory of Matrices*. Academic Press, Second edition, 1985.
- [2] B. N. Parlett. *The Symmetric Eigenvalue Problem*. Prentice-Hall, 1980.
- [3] D. S. Stoffer, D. E. Tyler, and A. J. McDougall. Spectral analysis for categorical time series: Scaling and the spectral envelope. *Biometrika*, 80(3), 611–622, 1993.