Estimation of Correlation Properties

George B. Rybicki

10 Jan 1995

1 Introduction

The structure function of a random function can be determined in a manner similar to that described in Press, Rybicki & Hewitt (1992) by constructing a histogram of square differences of the measured continuum as a function of time difference and fitting the resulting points with a power law or other parameterized law. However, here we shall use methods based on Bayes theorem that are better motivated and give estimates of the confidence of the fits.

   We first give a method for estimating parameters of the autocorrelation function, rather than the structure function, since the formulas are simpler. Then we present a similar method for structure functions that is independent of the mean and dispersion of the process, which are frequently poorly constrained by the data.

2 Autocorrelation Functions

Suppose we have a set of "data" f_i, at times t_i, $i = 1, 2, \ldots, N$. We model this data as the values of a Gaussian random process $f(t)$ at these times, that is, $f_i = f(t_i)$. The statistical properties of the process are completely determined by its mean $\bar{f}$ and the autocorrelation matrix $\mathbf{C}$ with components

$$C_{ij} = \langle (f_i - \bar{f})(f_j - \bar{f}) \rangle \tag{1}$$

We also assume that the data can be expressed as $f(t) = s(t) + n(t)$, the sum of a "signal" $s(t)$ and "noise" $n(t)$. We assume the noise to have zero mean, so the means of the data and signal are equal $\bar{f} = \bar{s}$. We shall assume that the signal and noise are uncorrelated (this assumption can be relaxed, if

1

necessary). Then the correlation matrix of the process can then be expressed as the sum

$$\mathbf{C} = \mathbf{C_s} + \mathbf{C}_n \tag{2}$$

where $\mathbf{C}_s$ and $\mathbf{C}_n$ are, respectively, the signal and noise autocorrelation matrices, the components of which are given by expressions analogous to Eq. (1).

For convenience of notation, let us define $y_i = f_i - \bar{f}$ to be the data with process mean subtracted, and define the corresponding vector $\mathbf{y}$. This may be written

$$\mathbf{y} = \mathbf{f} - \bar{f}\mathbf{E}, \tag{3}$$

where $\mathbf{E}$ is the vector all of whose elements are unity, $E_i = 1$.

Let us assume that the properties of the process can be defined by a number of parameters, such as the process mean $\bar{f}$ and other parameters defining the correlation matrices. We write these parameters as the single vector η. For fixed values of the parameters η, the probability of $\mathbf{f}$ is given is the multivariate Gaussian,

$$P(\mathbf{f}|\eta) = [\det(2\pi\mathbf{C})]^{-1/2} \exp\left(-\frac{1}{2}\mathbf{y}^T\mathbf{C}^{-1}\mathbf{y}\right) \tag{4}$$

For parameter estimation, we need the probability of η given the data $\mathbf{f}$. This is related to the probability (4) by Bayes theorem,

$$P(\eta|\mathbf{f}) = \frac{P(\mathbf{f}|\eta)P(\eta)}{P(\mathbf{f})} \tag{5}$$

For a given set of data the quantity $P(\mathbf{f})$ is fixed, so the only unknown on the right is the a priori probability $P(\eta)$. We now assume, as usual, that $P(\eta)$ is slowly varying on the scales of interest of the estimation problem, so that

$$P(\eta|\mathbf{f}) \propto P(\mathbf{f}|\eta) \tag{6}$$

The estimation problem is now solved by maximizing $P(\eta|\mathbf{f})$ with respect to the parameters η. From Eqs. (4) and (6), this is equivalent to maximizing the *log-likelihood function*

$$\mathcal{L} = -\frac{1}{2}\mathbf{y}^T\mathbf{C}^{-1}\mathbf{y} - \frac{1}{2}\ln\det\mathbf{C} \tag{7}$$

Equivalently, one can minimize the quantity

$$Q = \mathbf{y}^T\mathbf{C}^{-1}\mathbf{y} + \ln\det\mathbf{C} \tag{8}$$

Note that the first term of Q is just the usual χ^2 of the data, so that we can think of Q as a "corrected χ^2," where the correction is related to the normalization of the Gaussian distribution function and to the use of Bayes theorem (see, e.g., Rybicki & Kleyna (1994), Eq. [9]).

In general, the minimization of Q with respect to the parameters has to be done numerically, but for some types of parameters it can be done analytically.

For example, suppose we consider the correlation matrix that depends on a number of parameters $a, b, c, \ldots$ such that a is a multiplicative factor, that is,

$$\mathbf{C} = a\mathbf{C}_0 \tag{9}$$

where $\mathbf{C}_0$ is independent of a. An important case where this occurs is when the signal correlation matrix to be fitted has an overall multiplicative parameter and where the noise estimates are also uncertain to within a multiplicative factor, perhaps due to uncertain observational effects; this factor can be considered as a parameter to be fitted, along with the other parameters of the correlation function of the signal. Then one can choose one parameter of the total correlation matrix to be an overall factor as in Eq. (9).

The above assumed form implies the simple scalings,

$$\mathbf{C}^{-1} = a^{-1}\mathbf{C}_0^{-1}, \qquad \det \mathbf{C} = a^n \det \mathbf{C}_0 \tag{10}$$

so that

$$Q = a^{-1}\mathbf{y}^T\mathbf{C}_0^{-1}\mathbf{y} + n \ln a + \ln \det \mathbf{C}_0 \tag{11}$$

Thus the minimization with respect to a gives $a = a_0$ where

$$a_0 = \mathbf{y}^T\mathbf{C}_0^{-1}\mathbf{y}/n \equiv \chi_0^2/n \tag{12}$$

This implies that the minimization of Q with respect to the other parameters can be done by minimization of the reduced function

$$\hat{Q} = n - n \ln n + n \ln \left(\mathbf{y}^T\mathbf{C}_0^{-1}\mathbf{y}\right) + \ln \det \mathbf{C}_0 \tag{13}$$

which no longer contains a

Note that Eq. (12) implies that $\chi^2 = a_0^{-1}\chi_0^2 = n$, which is the correct theoretical value for n degrees of freedom. This is easy to understand, since it is the result of a maximum likelihood rescaling of the errors.

3 Structure Functions

It occurs commonly that the underlying random process has strong correlations on scales longer than the length of the data set. In this case, the mean $\bar{s}$ and the variance $\mathrm{var}(s)$ of the process are not well determined by the data. It is useful therefore to derive an modified version of Q that is independent of these two parameters. The resulting quantity $\tilde{Q}$ will then depend solely on the structure function of the process, rather than the full correlation function.

We begin by defining a reduced data set $\tilde{\mathbf{f}}$ of length $(n-1)$ by subtracting the last component f_n from each of the other components of $\mathbf{f}$,

$$\tilde{f}_i = y_i - y_n = f_i - f_n, \qquad 1 \le i \le n-1 \tag{14}$$

This can be written as the matrix equation

$$\tilde{\mathbf{f}} = \mathbf{B}\mathbf{f} \tag{15}$$

where the $(n-1) \times n$ matrix $\mathbf{B}$ is defined by

$$\mathbf{B} = (\tilde{\mathbf{I}}, -\tilde{\mathbf{E}}) = \begin{pmatrix} 1 & 0 & \ldots & 0 & -1 \\ 0 & 1 & \ldots & 0 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -1 \end{pmatrix} \tag{16}$$

Here $\tilde{\mathbf{I}}$ is the $(n-1) \times (n-1)$ unit matrix, and $\tilde{\mathbf{E}}$ is the $(n-1)$-vector of ones, $\tilde{\mathbf{E}} = (1, 1, \ldots, 1)^T$ [Generally the tilde will be used to denote $(n-1)$-dimensional versions of the corresponding n-dimensional quantities.]

Let us now find the probablity distribution of the reduced vector $\tilde{\mathbf{f}}$. This can clearly be written,

$$P_{n-1}(\tilde{\mathbf{f}}) = \int \delta^{(n-1)}(\tilde{\mathbf{f}} - \mathbf{B}\mathbf{f}) P_n(\mathbf{f}) \, d^n\mathbf{f}, \tag{17}$$

where $\delta^{(n-1)}$ is an $(n-1)$-dimensional delta function, and where $P_n(\mathbf{f})$ is the Gaussian distribution

$$P_n(\mathbf{f}) = [\det(2\pi\mathbf{C})]^{-1/2} \exp\left(-\frac{1}{2}\mathbf{f}^T\mathbf{C}^{-1}\mathbf{f}\right) \tag{18}$$

where $\mathbf{C} = \mathbf{C_s} + \mathbf{C_n}$. A convenient characterization of this probability distribution is through its Fourier transform, also known as its *characteristic function*, which is easily shown to be

$$\int \exp(i\mathbf{k}^T\mathbf{f}) P_n(\mathbf{f}) \, d^n\mathbf{f} = \exp\left(-\frac{1}{2}\mathbf{k}^T\mathbf{C}\mathbf{k}\right) \tag{19}$$

4

Applying the Fourier transform to Eq. (17) we find

$$\int \exp(i\tilde{\mathbf{k}}^T\tilde{\mathbf{f}})P_{n-1}(\tilde{\mathbf{f}})\,d^{n-1}\tilde{\mathbf{f}} = \int \exp(i\tilde{\mathbf{k}}^T\mathbf{B}\mathbf{f})P_n(\mathbf{f})\,d^n\mathbf{f} = \exp\left(-\frac{1}{2}\tilde{\mathbf{k}}^T\mathbf{\Gamma}\tilde{\mathbf{k}}\right) \tag{20}$$

where

$$\mathbf{\Gamma} = \mathbf{B}\mathbf{C}\mathbf{B}^T \tag{21}$$

Thus, by inverse Fourier transformation,

$$P_{n-1}(\tilde{\mathbf{f}}) = [\det(2\pi\mathbf{\Gamma})]^{-1/2}\exp\left[-\frac{1}{2}\tilde{\mathbf{f}}^T\mathbf{\Gamma}^{-1}\tilde{\mathbf{f}}\right] \tag{22}$$

Equation (22) shows that $\tilde{\mathbf{f}}$ has a Gaussian distribution with $(n-1)\times(n-1)$ correlation matrix $\mathbf{\Gamma}$. Thus, the appropriate modified expression for Q is now

$$\tilde{Q} = \tilde{\mathbf{f}}^T\mathbf{\Gamma}^{-1}\tilde{\mathbf{f}} + \ln\det\mathbf{\Gamma} \tag{23}$$

Note that the first term on the right is a modified chi-squared, given by

$$\tilde{\chi}^2 = \tilde{\mathbf{f}}^T\mathbf{\Gamma}^{-1}\tilde{\mathbf{f}} = \tilde{\mathbf{f}}^T(\mathbf{B}\mathbf{C}\mathbf{B}^T)^{-1}\tilde{\mathbf{f}} \tag{24}$$

This can be rewritten in the schematic, but suggestive, form

$$\tilde{\chi}^2 = (f_i - f_n)^T[C_{ij} - C_{in} - C_{nj} + C_{nn}]^{-1}(f_j - f_n) \tag{25}$$

where the indices i and j run from 1 to $n-1$. From this expression, one can immediately see that $\tilde{\chi}^2$ is invariant to the addition of the same arbitrary constant to each component of the data, and thus it is independent of the data mean $\bar{f}$. The matrix in brackets (a rewriting of $\mathbf{B}\mathbf{C}\mathbf{B}^T$) is likewise invariant to the addition of the same arbitrary constant to each element of the correlation matrix $\mathbf{C}$. This means that $\tilde{\chi}^2$ and $\tilde{Q}$ depend only on the structure function of the data, since the relationship between the correlation function $C(\tau)$ and the negative of the structure function $-V(\tau)$ differ only by a constant: $C(\tau) - \langle(f - \bar{f})^2\rangle = -V(\tau)$. In particular, one can replace the elements of the correlation matrix C_{ij} by $-V_{ij}$ for $1 \le i, j \le n$ without changing the values of either $\tilde{\chi}^2$ or $\tilde{Q}$. In this way $\tilde{Q}$ can be used to estimate the parameters of the structure function directly without considering the mean or variance of the data.

These properties of $\tilde{\chi}^2$ of being independent of data mean and variance are shared by the quantity χ^2_{PRH} introduced by Press, Rybicki, & Hewitt (1992; Eq. [27]), namely,

$$\chi^2_{\mathrm{PRH}} = \mathbf{y}^T\hat{\mathbf{A}}\mathbf{y} = \mathbf{y}^T\left(\mathbf{A} - \frac{\mathbf{A}\mathbf{E}\mathbf{E}^T\mathbf{A}}{\mathbf{E}^T\mathbf{A}\mathbf{E}}\right)\mathbf{y}. \tag{26}$$

where $\mathbf{A} = \mathbf{C}^{-1}$. It is possible to show that these two chi-square expressions are identical, that is,

$$\tilde{\chi}^2 \equiv \chi^2_{\mathrm{PRH}} \tag{27}$$

for any set of data. Using Eqs. (15) and (24) this relation can be written in the form

$$\mathbf{y}^T \mathbf{B}^T (\mathbf{BCB}^T)^{-1} \mathbf{B}\mathbf{y} = \mathbf{y}^T \left(\mathbf{C}^{-1} - \frac{\mathbf{C}^{-1}\mathbf{E}\mathbf{E}^T\mathbf{C}^{-1}}{\mathbf{E}^T\mathbf{C}^{-1}\mathbf{E}} \right) \mathbf{y} \tag{28}$$

which implies that Eq. (27) can be true for all $\mathbf{y}$ if and only if

$$\mathbf{B}^T (\mathbf{BCB}^T)^{-1} \mathbf{B} = \mathbf{C}^{-1} - \frac{\mathbf{C}^{-1}\mathbf{E}\mathbf{E}^T\mathbf{C}^{-1}}{\mathbf{E}^T\mathbf{C}^{-1}\mathbf{E}} \tag{29}$$

We have proved this result, but our proof is quite lengthy and will not be given here. The simplicity of this result suggests that a simple proof may exist.

Since the definition of χ^2_{PRH} treats all vector components on the same footing, the relation (27) shows that the expression for $\tilde{\chi}^2$ must be independent of the choice of which component of $\mathbf{f}$ to subtract from the others in defining $\tilde{\mathbf{f}}$; any other component could have been chosen in Eq. (14) instead of the nth one without affecting the values of $\tilde{\chi}^2$.

It is also true that the determinant of $\mathbf{\Gamma}$ is independent of the choice of which component of $\mathbf{f}$ to use in forming $\tilde{\mathbf{f}}$. We sketch the proof as follows. First, let us intoduce some notation. As a generalization of Eq. (14), we now define the vector $\tilde{\mathbf{f}}_{(k)}$ by subtracting the kth component of the data from all the others. The indices of all components with indices larger than k are then reduced by one, so that the resulting indices cover the range $1 \leq i \leq n-1$. This can be conveniently expressed by the generalization of Eq. (15),

$$\tilde{\mathbf{f}}_{(k)} = \tilde{\mathbf{B}}_{(k)}\mathbf{f} \tag{30}$$

Here the $(n-1) \times n$ matrix $\mathbf{B}_{(k)}$ is defined by

$$\mathbf{B}_{(k)} = (\mathbf{U}_{(k-1)}, -\tilde{\mathbf{E}}, \mathbf{L}_{(n-k)}) \tag{31}$$

where $\mathbf{U}_{(k-1)}$ is an $(n-1) \times (k-1)$ matrix consisting of a $(k-1) \times (k-1)$ unit matrix on top of an $(n-k) \times (k-1)$ zero matrix. Likewise $\mathbf{L}_{(n-k)}$ is an $(n-1) \times (n-k)$ matrix consisting of an $(k-1) \times (n-k)$ zero matrix on

top of an $(n-k) \times (n-k)$ unit matrix. Note that $\mathbf{B} = \mathbf{B}_{(n)}$. The reduced correlation matrix for the choice k is then

$$\mathbf{\Gamma}_{(k)} = \mathbf{B}_{(k)} \mathbf{C} \mathbf{B}_{(k)}^T \tag{32}$$

which is the appropriate generalization of the reduced correlation matrix appearing in Eq. (22), for which $\mathbf{\Gamma} = \mathbf{\Gamma}_{(n)}$.

We now define $\mathbf{T}_{(k,l)}$ to be an $(n-1) \times (n-1)$ matrix equal to $\mathbf{B}_{(k)}$ with its lth column deleted. Then we state without proof the following two properties of $\mathbf{T}_{(k,l)}$ for all k and l, $1 \le k, l \le n-1$,

$$\mathbf{B}_{(k)} = \mathbf{T}_{(k,l)} \mathbf{B}_{(l)}; \qquad |\det \mathbf{T}_{(k,l)}| = 1 \tag{33}$$

Therefore

$$\mathbf{\Gamma}_{(k)} = \mathbf{T}_{(k,l)} \mathbf{B}_{(l)} \mathbf{C} \mathbf{B}_{(l)}^T \mathbf{T}_{(k,l)}^T = \mathbf{T}_{(k,l)} \mathbf{\Gamma}_{(l)} \mathbf{T}_{(k,l)}^T \tag{34}$$

and finally,

$$\det \mathbf{\Gamma}_{(k)} = (\det \mathbf{T}_{(k.l)})^2 \det \mathbf{\Gamma}_{(l)} = \det \mathbf{\Gamma}_{(l)} \tag{35}$$

Thus $\tilde{Q}$ is independent of which component f_k is subtracted from the others in defining $\tilde{\mathbf{f}}$, and we may write simply

$$\tilde{Q} = \tilde{\mathbf{f}}_{(k)}^T \mathbf{\Gamma}_{(k)}^{-1} \tilde{\mathbf{f}}_{(k)} + \ln \det \mathbf{\Gamma}_{(k)} \tag{36}$$

In view of Eq. (26), it is interesting to ask if there is a simple expression for $\tilde{Q}$ based on $n \times n$ matrices that treats all components alike. This would involve finding an appropriate $n \times n$ expression for the determinant term in Eq. (23) or (36). It should be pointed out $\det \hat{\mathbf{A}}$ itself vanishes and is not the desired quantity.

We conjecture that the desired relation is

$$\det \mathbf{\Gamma}_{(k)} = (\mathbf{E}^T \mathbf{C}^{-1} \mathbf{E}) \det \mathbf{C}, \tag{37}$$

which has been checked numerically. In that case we may write $\tilde{Q}$ in the form

$$\tilde{Q} = \mathbf{y}^T \hat{\mathbf{A}} \mathbf{y} + \ln \det \mathbf{C} + \ln(\mathbf{E}^T \mathbf{C}^{-1} \mathbf{E}) \tag{38}$$

which is invariant to $\mathbf{y} \to \mathbf{y} + \lambda \mathbf{E}$ and $\mathbf{C} \to \mathbf{C} + \mu \mathbf{E} \mathbf{E}^T$.

We note that the method for finding the minimum with respect to an overall multiplicative factor will also work for the case of a structure function with overall multiplicative factor, providing that we use the reduced matrices given in this section.

References

[1] Press, W. H., Rybicki, G. B., & Hewitt, J. N. 1992, *Astrophysical Journal*,**385**, 404–415.

[2] Rybicki, G. B., & Kleyna, J. T. 1994, in *Reverberation Mapping of the Broad-Line Region of Active Galactic Nuclei*, ed. P.M. Gondhalekar, K. Horne, and B.M. Peterson (San Francisco: Astronomical Society of the Pacific), 85–93.