14
Calculation of the correlation function of a series

Let us consider a series of (real) numbers G_i, $i = 1, \dots, M$. They could be numbers coming out from measurements in a Monte Carlo simulation or any other source. We want to determine its normalized correlation function

$$\rho_G(j) = \frac{\langle G_i G_{i+j}\rangle - \langle G\rangle^2}{\langle G^2\rangle - \langle G\rangle^2}, \tag{14.1}$$

with $\langle G\rangle = \frac{1}{M}\sum_{i=1}^{M} G_i$, $\langle G^2\rangle = \frac{1}{M}\sum_{i=1}^{M} G_i^2$. Note that, by definition, $\rho_G(0) = 1$. We assume that the series is stationary, this means that the correlation function $\rho_G(j)$ defined above does not depend on i. We first make a straightforward simplification. If we define

$$z_i = \frac{G_i - \langle G\rangle}{\sqrt{\langle G^2\rangle - \langle G\rangle^2}}, \tag{14.2}$$

it can be easily proved that the correlation function $\rho_G(j)$ of the series G_i is equal to that of the series z_i,

$$\rho_z(j) = \langle z_i z_{i+j}\rangle. \tag{14.3}$$

As we have assumed that the series is stationary, we could compute this average directly including all the possible values of i. So, to compute $\rho_z(1)$ we would include the $M-1$ contributions $z_1z_2 + z_2z_3 + \cdots + z_{M-1}z_M$, to compute $\rho_z(2)$ we would include the $M-2$ contributions $z_1z_3 + z_2z_4 + \cdots + z_{M-2}z_M$, and so on. In general, we have

$$\rho_z(j) = \frac{1}{M-j}\sum_{i=1}^{M-j} z_i z_{i+j}, \quad j = 0, \dots, M-1. \tag{14.4}$$

Note that there are $M-j$ values contributing to $\rho_G(j)$, and for j close to M the statistical errors are large. For instance, for $j = M-1$, there is only one contribution to $\rho_z(j)$, namely $z_1 z_M$. On top of this, when M is large, the direct calculation of $\rho_z(j)$ using (14.4) could be slow as it takes of the order of M^2 operations. We now show that it is possible to greatly reduce the time needed to compute a correlation function by using the discrete Fourier transform.

As explained in appendix 18, the discrete Fourier transform $\hat{x} = \mathcal{F}_D[x]$ of an arbitrary set of numbers $(x_1, \dots, x_n)$ is defined as[1]:

$$\hat{x}_k = \sum_{j=0}^{n-1} e^{\frac{2\pi \mathrm{i}}{n} jk} x_{j+1}, \qquad k = 0, 1, \dots, n-1, \tag{14.5}$$

and the inverse relation, indicated as $x = \mathcal{F}_D^{-1}[\hat{x}]$,

$$x_{j+1} = \frac{1}{n} \sum_{k=0}^{n-1} e^{-\frac{2\pi \mathrm{i}}{n} jk} \hat{x}_k, \qquad j = 0, \dots, n-1. \tag{14.6}$$

With the help of (18.10) it is possible to prove the identity:

$$\sum_{i=1}^{n} x_i x_{i+j} = \frac{1}{n} \sum_{k=0}^{n-1} \mathrm{e}^{-\frac{2\pi \mathrm{i}}{n} jk} |\hat{x}_k|^2, \tag{14.7}$$

which, given (14.4), seems to imply that $\mathcal{F}_D^{-1}[|\hat{x}|^2]$, the inverse discrete Fourier transform of the series $(|\hat{x}_0|^2, \dots, |\hat{x}_{n-1}|^2)$, is related in some way to the correlation function $\rho_x(j)$ of the series x. This is true, but one has to be careful. The point is that for the previous formula (14.7) to be exact, one has to assume periodic boundary conditions, i.e. whenever x_{i+j} appears in the left hand side of this formula with $i + j > n$, it has to be understood as x_{i+j-n}. This property, namely $x_j = x_{j-n}$ for $j > n$, which derives directly from the extension of (14.6) to values $j > n$, is not present in the original series as it runs from x_1 to x_n, and x_j does not even exist for $j > n$. Note that in (14.4), for example, one never uses z_i with $i \geq M$. With a small trick, we can relate exactly the correlation function $\rho_z(j)$ to an inverse discrete Fourier transform. The trick is to introduce a new series $x = (x_1, \dots, x_n)$ of length $n = 2M$, twice the length of the original z series. The new series is defined as

$$x_i = \begin{cases} z_i, & i = 1, \dots, M, \\ 0, & i = M+1, \dots, 2M. \end{cases} \tag{14.8}$$

It is a straightforward consequence of this definition that the sum needed in (14.4) for the calculation of $\rho_z(j)$ can be written as

$$\sum_{i=1}^{M-j} z_i z_{i+j} = \sum_{i=1}^{n} x_i x_{i+j}, \quad j = 0, 1, \dots, M-1 \tag{14.9}$$

with periodic boundary conditions now assumed in the sum of the right hand side[2]. Finally, using (14.4), (14.8) and the identity (14.7) with $n = 2M$, we get

$$\rho_z(j) = \frac{1}{M-j} \left[\frac{1}{n} \sum_{k=0}^{n-1} \mathrm{e}^{-\frac{2\pi \mathrm{i}}{n} jk} |\hat{x}_k|^2 \right], \quad j = 0, 1, \dots, M-1, \tag{14.10}$$

1) For the difference in notation with respect to appendix 18 we refer to the foonote 13.2.

2) The reader might find it useful to check by himself (14.7) and (14.9).

showing that it is possible to obtain, after multiplication by the factor $\frac{1}{M-j}$, the correlation function $\rho_z(j)$ from the inverse discrete Fourier series of $|\hat{x}_k|^2$, being $\hat{x}_k$ the direct discrete Fourier series of x_i. The big advantage of this procedure is that the use of fast Fourier routines allows one to implement this algorithm in a time that scales as $M \log M$, much smaller for large M than the time of order M^2 needed by the direct algorithm.

Once we have computed $\rho_G(j)$ we can compute the correlation time τ_G, using (4.23). Just a final word of warning. If you plot the resulting correlation function, it will typically behave smoothly only for not too large values of j and it will have wild oscillations for large j.The reason being that the larger j, the fewer values contribute to its estimator and the error greatly increases. As a consequence, when computing τ_G, using (4.23), do not extend the upper sum all the way to $M-1$. Restrict the sum to those values of j for which the correlation function, having decayed to a value near zero, is still well behaved. In fact, if all you want is to have an estimate of the correlation time τ_G in order to determine faithfully the error of your estimator, it is enough to know the characteristic decay time of the correlation function[3)].

Here comes a program listing that computes the correlation function of a given series. Remember to read the instructions of the fast Fourier transform routines to use to find out if you need to correct for any factors of M. In this subroutine, the correlation function $\rho_z(i)$ overwrites the input value x_i. It returns also the mean `xm` and the root-mean-square `x2`.

```
subroutine correla(x,m,xm,x2)
implicit double precision (a-h,o-z)
dimension x(0:m-1)
double complex z(0:2*m-1)

  xm=sum(x)/m
  x2=sqrt(sum(x**2)/m-xm*xm)
  x=(x-xm)/x2
  z(0:m-1)=dcmplx(x,0.0d0)
  z(m:2*m-1)=dcmplx(0.0d0,0.0d0)
  call fft1d(z,2*m,1)
  z=dcmplx(abs(z)**2,0.0d0)
  call fft1d(z,2*m,-1)
  do j=0,m-1
    x(j)=real(z(j))/(m-j)
  enddo

end subroutine correla
```

3) There is another, deeper, reason not to sum all the way up to $j = M-1$ to obtain the correlation time as it can be shown that this procedure leads to $\tau_G = -1/2$ and hence a zero error in the estimator. The reason is simple to understand, we are computing the error of an estimator using the own estimator as the true value. In other words, we are taking formula (4.15) with $I = \dfrac{1}{N}\sum_{k=1}^{M} G_k$ and hence the error is 0.