Stationary Stochastic Processes

(Revised October 25, 2012)

by

Robert Murray, Ph.D.
Omicron Research Institute

A **stationary stochastic process** is one for which the statistical properties do not change with time. This is the easiest case to deal with, and the one for which the most mathematical results are known. The discrete stochastic time series X_t is imagined to extend to plus or minus infinity in the (integer) time index *t*. Ideally, we imagine this time series to be but one *realization* of the stochastic process, which could hypothetically be repeated an infinite number of times, so that the statistical properties could be deduced from this infinite number of realizations. In the real world, of course, the time series is usually a unique instance, and we must do our best to *estimate* the statistical properties from this one instance. In general, in a stochastic process there could be an infinite number of joint probability distributions between each set of *N* elements of the time series, for *N* different time indexes *t*, and we could deduce the joint probability distribution for each set of *N* such elements from the infinite set of realizations of the process at our disposal. So, for theoretical purposes, we will assume the existence of such an infinite set of realizations, although in reality only one realization of the process will exist.

In general, this infinite hierarchy of joint probability distributions will give rise to an infinite hierarchy of ***Nth*-order covariance functions**, which could in general be time dependent, actually a function of all *N* of the time indexes. If *N* is greater than two, then this gives rise to what are known as **higher-order statistics**. If the correlation functions are time-dependent, then this gives rise to **non-stationary statistics**. If only the lowest order correlation is considered, consisting of the two-point correlation functions, then this corresponds to **second-order statistics**. This is the usual case considered in **filter theory**, although a few excursions into higher-order statistics have been made. So the practical choice for filter theory is between **stationary** (time-independent) or **non-stationary** (time-dependent) second-order correlation. In this paper we will study the *stationary* case, and the problem of **prediction** based on stationary second-order correlation.

4.1 Stationary Processes

Given a (real) time series X_t, the most basic quantity in second-order statistics is the **covariance matrix**. This is defined as the covariance between any two elements of the time series. In the general case, this covariance will be a function of both time indexes.

For notational purposes, let us denote by *t* the *present* time, or the last *known* element of the time series, even though we suppose the stochastic process to extend from minus infinity to plus infinity in time. Then, we define a time-dependent covariance matrix as follows:

$$\Gamma_{j,n}(t) \equiv \left\langle X_{t-j}, X_{t-n} \right\rangle \equiv E\left[X_{t-j} X_{t-n} \right]$$

Note that, for a real time series, the covariance matrix is symmetric. If the covariance matrix is time-independent, corresponding to a **stationary process**, then the covariance matrix takes the form:

$$\Gamma_{j,n} \equiv \left\langle X_{t-j}, X_{t-n} \right\rangle \equiv \gamma_{|j-n|} \quad (X \text{ stationary})$$

The function $\gamma_{|j-n|}$ is called the **autocovariance** function. In this case, changing the value of *t* changes the value of both indexes of the covariance matrix by the same amount, and moves the matrix element down one row and one column, or in other words down one unit along the diagonal. Then time-independence means that all the matrix elements along each diagonal are equal. In other words, if the covariance matrix is time-independent, corresponding to a **stationary process**, then it is a ***Toeplitz*** matrix. In the general case of time-dependent covariances, therefore, the covariance matrix is not Toeplitz. However, it will still be symmetric, since the covariances are symmetric for each time *t*.

The other statistical quantity of interest in the case of a second-order stationary process is the **mean** of the process. This is defined by:

$$\mu(t) \equiv \left\langle X_t \right\rangle \equiv \mu \quad (X \text{ stationary})$$

For a second-order stationary stochastic process, the mean is assumed constant. In what follows, we will assume that the mean has been subtracted off from the time series, so that the time series is **zero-mean**.

Consider now the problem of finding the best linear predictor of X_{t+1}, given the N previous values of the time series. If { X_t } is a zero-mean process, we may write the best linear predictor of X_{t+1} in terms of the previous *N* values of the time series as:

$$\hat{X}_{t+1} \equiv \sum_{n=0}^{N-1} \phi_n(t) X_{t-n}$$

In the case of a **stationary process**, the Linear Prediction (LP) coefficients ϕ_n do not depend on time. Multiplying each side of the first equation by $X_{t-j}, \quad j = 0, \ldots, N-1$ and taking expectations, we arrive at:

$$\left\langle X_{t-j}, X_{t+1} \right\rangle \equiv \left\langle X_{t-j}, \hat{X}_{t+1} \right\rangle = \sum_{n=0}^{N-1} \phi_n(t) \left\langle X_{t-j}, X_{t-n} \right\rangle$$

In terms of the covariance matrix, this becomes:

$$\Gamma_{j+1,0}(t+1) = \sum_{n=0}^{N-1} \phi_n(t) \Gamma_{j,n}(t)$$

However, under the assumption of *stationarity* the covariance matrices and LP coefficients do not depend on time, so we may write this as:

$$\sum_{n=0}^{N-1}\Gamma_{j,n}\phi_n = \Gamma_{j+1,0} \quad (X \text{ stationary})$$

We now make the assumption that the covariance matrix is *nonsingular*. This will be the case for a generic **stochastic process**, provided the variance of each element of the time series is greater than zero. Otherwise, if the covariance matrix is singular, then we are dealing with a **deterministic process**, and the Linear Prediction method will not work. Thus, inverting the covariance matrix, we have the solution to the LP problem:

$$\phi_n = \sum_{j=0}^{N-1}\Gamma^{-1}_{n,j}\gamma_{|j+1|} \quad n \in \{0,\ldots,N-1\}$$

Since the covariances are stationary, they may be estimated in terms of the past data:

$$\gamma_{|k|} \approx \frac{1}{N}\sum_{n=0}^{N-1} X_{t-n-k}X_{t-n}$$

Thus the assumption of *stationarity* allows us to estimate the1-day future covariance in terms of the past covariances. The easiest way to solve these equations for the LP coefficients is to estimate the covariances by means of this estimate, form the Toeplitz covariance matrix, and then invert this matrix using a numerical routine for the inversion of Toeplitz matrix. This is the most basic version of the **Linear Prediction filter**. Other methods of solution will be described below.

The most trivial example of a **stationary process** is of course the **Random Walk** process, in which the different X_t variables are complete uncorrelated. This means that:

$$\Gamma_{j,n} \equiv \langle X_{t-j}, X_{t-n}\rangle \equiv \gamma_{|j-n|} \equiv 0 \quad (j \neq n) \quad (\text{Random Walk})$$

In this case, all the LP coefficients vanish, and we have:

$$\phi_n \equiv 0 \quad \Rightarrow \quad X_t = Z_t \qquad \{Z_t\} \sim WN\left(0,\sigma^2\right)$$

In other words, each value of the X_t variable is an independent random variable, completely un-correlated with the X_t variables at different times. This is the simplest case of a **stochastic process**, but not the generic case by any means. In this simplest case, the X_t variables are completely unpredictable, but in the generic case there will be correlation present which allows a (partial) prediction to be made.

4.2 Fourier Representation

Theoretically, the time series X_t extends from minus infinity to plus infinity, so there are an infinite number of terms. Then the Fourier transform of this process is a continuum over all possible frequencies from $-\pi$ to $+\pi$. In this case, the Fourier transform extends over the entire range of the stochastic process, for all time, and it is normally assumed that the covariance matrix is *stationary*, so that the (continuous) power spectrum does not change with time. Then the stationary time series can be specified by specifying the continuum of (independent) random variables in the Fourier basis, with frequencies ranging from $-\pi$ to $+\pi$. This then specifies the stationary time series in the time domain, for all time, in terms of the corresponding set of random variables in the

frequency domain. (It will be shown below that stationarity of the time series is equivalent to statistical independence of the continuum of Fourier components.)

For an infinite stationary stochastic process (in general complex), the Fourier components form a continuum between the (Nyquist) frequencies from $-\pi$ to $+\pi$. (More precisely, it is the interval $(-\pi, +\pi]$.) The time components of this process are then specified, in a precise mathematical formalism, by means of a *stochastic integral* [Brockwell & Davis (1991)]:

$\{X_t\}$ stationary with zero mean:

$$X_t \equiv \int_{-\pi}^{+\pi} e^{it\nu} dZ(\nu) \qquad \text{with probability one,}$$

$$\{Z(\lambda):\ -\pi \le \lambda \le +\pi\} \text{ and } Z(-\pi) = 0$$

Each infinitesimal increment $dZ(\nu)$ is regarded as an *independent* (and hence *uncorrelated*) random variable, in the Fourier basis. The specification of this entire set of random variables in the Fourier basis specifies the stationary time series X_t for all times *t*.

Given the entire time series X_t, we can perform an inverse Fourier transform to find the random variables $dZ(\nu)$ in the Fourier basis. To accomplish this, let us use the formalism of the **Fourier transform** for ordinary functions [Brockwell & Davis (1991)]. Suppose we are given a time series $K(n)$, which is absolutely summable:

$$\sum_{n=-\infty}^{+\infty} |K(n)| < \infty$$

Then we can write this time series in terms of its Fourier transform:

$$K(h) = \int_{-\pi}^{+\pi} e^{ih\nu} g(\nu) d\nu \qquad h = 0, \pm 1, \ldots$$

The inverse Fourier transform is given by:

$$g(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{+\infty} e^{-in\lambda} K(n)$$

We may carry this over to the stochastic integral by defining the random orthogonal increment in terms of an ordinary random variable φ_ν in the Fourier basis as follows:

$$dZ(\nu) \equiv \varphi_\nu d\nu, \qquad X_t \equiv \int_{-\pi}^{+\pi} e^{it\nu} \varphi_\nu d\nu$$

Now we may express the random variable φ_ν in the Fourier basis in terms of the inverse Fourier transform of X_t:

$$\varphi_\nu \equiv \frac{1}{2\pi} \sum_{t=-\infty}^{+\infty} e^{-it\nu} X_t$$

Once again, in the Fourier basis the variables φ_ν are presumed to be *independent*, and hence *uncorrelated*, which in turn implies that the process is *stationary* (and vice-versa). This will now be shown using the formalism just introduced.

Taking the covariance between two elements of the (complex) time series, and using the assumption that the Fourier components are uncorrelated, we find:

$$E\left[X_{t+h}\bar{X}_t\right] \equiv \int_{-\pi}^{+\pi}\int_{-\pi}^{+\pi} e^{+i(t+h)\nu} e^{-it\nu'} E\left[dZ(\nu)d\bar{Z}(\nu')\right]$$

$$E\left[dZ(\nu)d\bar{Z}(\nu')\right] \equiv E\left[\varphi_\nu \bar{\varphi}_{\nu'}\right] d\nu d\nu' \equiv f(\nu)\delta(\nu-\nu')d\nu d\nu'$$

$$\therefore E\left[\varphi_\nu \bar{\varphi}_{\nu'}\right] \equiv \left\|\varphi_\nu\right\|^2 \delta(\nu-\nu') \equiv f(\nu)\delta(\nu-\nu')$$

$$\therefore E\left[X_{t+h}\bar{X}_t\right] = \int_{-\pi}^{+\pi} e^{ih\nu} f(\nu)d\nu$$

This then shows that, due to the orthogonality of the Fourier components, the delta function in the frequency variable implies that the covariance of the two time series components is independent of the time variable *t*. Conversely, if the covariance is given by an autocovariance function that is independent of t, then it may be expressed as a stochastic double integral with orthogonal increments, by reversing the argument. The autocovariance function is defined to be:

$$\gamma_X(h) \equiv E\left[X_{t+h}\bar{X}_t\right] \equiv \int_{-\pi}^{+\pi} e^{ih\nu} dF(\nu) \equiv \int_{-\pi}^{+\pi} e^{ih\nu} f(\nu)d\nu$$

From this, by taking the inverse Fourier transform, we see that the spectral density is given by:

$$f(\lambda) = \frac{1}{2\pi}\sum_{n=-\infty}^{+\infty} e^{-in\lambda}\gamma_X(n) \geq 0$$

Hence the autocovariance function of a stationary process is the Fourier transform of the spectral density [Brockwell & Davis (1991)]. We see that orthogonality of the Fourier components is a necessary and sufficient condition for stationarity.

4.3 LP in the Fourier Basis

We now wish to express the Linear Prediction equation in the Fourier basis. We write the LP equation for the *h*-step-ahead predictor in the form (letting *N* go to infinity):

$$\sum_{n=0}^{\infty}\gamma_{|j-n|}\phi_n = \gamma_{|j+h|} \quad \left(X \text{ stationary}\right)$$

In order to represent the autocovariance functions in the Fourier basis, we note that when the time series is real, the autocovariance function is given by:

$$\gamma_X(h) \equiv E\left[X_{t+h}X_t\right] \equiv \int_{-\pi}^{+\pi} \cos\left[h\nu\right] f(\nu)d\nu \quad \left(X \text{ real}\right)$$

The spectral density is in turn given by:

$$f(\nu) \equiv \left\|\varphi_\nu\right\|^2, \qquad \varphi_\nu \equiv \frac{1}{2\pi}\sum_{t=-\infty}^{+\infty} e^{-it\nu} X_t$$

By analogy with the expansion for the stochastic variable φ_ν, let us define the spectral density as the absolute square of an ordinary complex function. However, it can be

shown that, as for an ARMA (Auto-Regressive Moving Average) process, if the process is *causal* and *invertible*, then the sum is restricted to positive frequencies only:

$$f(\lambda) \equiv \frac{\sigma^2}{2\pi}\left|\psi\left(e^{-i\lambda}\right)\right|^2, \qquad \psi\left(e^{-i\lambda}\right) \equiv \sum_{k=0}^{+\infty}\psi_k e^{-ik\lambda} \qquad \left(\sigma^2 \equiv \mathrm{var}\left[X_t\right]\right)$$

This corresponds to the idea, comparing this expression with the corresponding one for the time series, that $\left(\psi^*\right)$ corresponds to a Fourier sum over past values of the time series only. Thus the autocovariance function in this new notation is given by:

$$\gamma_{|h|} \equiv \int_{-\pi}^{+\pi}\cos\left[h\lambda\right] f(\lambda) d\lambda = \frac{\sigma^2}{2\pi}\int_{-\pi}^{+\pi}\cos\left[h\lambda\right]\left|\psi\left(e^{-i\lambda}\right)\right|^2 d\lambda$$

Now if we use this in the LP equation, we get:

$$\int_{-\pi}^{+\pi}\sum_{n=0}^{+\infty}\phi_n \cos\left[(j-n)\lambda\right]\left|\psi\left(e^{-i\lambda}\right)\right|^2 d\lambda = \int_{-\pi}^{+\pi}\cos\left[(j+h)\lambda\right]\left|\psi\left(e^{-i\lambda}\right)\right|^2 d\lambda$$

We note from the derivation of the LP equations that the index *j* can range over all non-negative values, so we have a separate equation for each such value of the index *j*. Expressing the cosines in terms of exponentials, we may write:

$$\int_{-\pi}^{+\pi}\sum_{n=0}^{+\infty}\phi_n\left(e^{+i(j-n)\lambda}+e^{-i(j-n)\lambda}\right)\left|\psi\left(e^{-i\lambda}\right)\right|^2 d\lambda = \int_{-\pi}^{+\pi}\left(e^{+i(j+h)\lambda}+e^{-i(j+h)\lambda}\right)\left|\psi\left(e^{-i\lambda}\right)\right|^2 d\lambda$$

Now, in view of the fact that the LP coefficients are real, let us make the following definitions:

$$\phi\left(e^{-i\lambda}\right) \equiv \sum_{n=0}^{+\infty}\phi_n e^{-in\lambda}, \quad \phi^*\left(e^{-i\lambda}\right) \equiv \sum_{n=0}^{+\infty}\phi_n e^{+in\lambda}, \quad \left(\phi_n \text{ real}\right)$$

Now we may write the equation in the form:

$$\int_{-\pi}^{+\pi}\left(e^{+ij\lambda}\phi\left(e^{-i\lambda}\right)+e^{-ij\lambda}\phi^*\left(e^{-i\lambda}\right)\right)\left|\psi\left(e^{-i\lambda}\right)\right|^2 d\lambda = \int_{-\pi}^{+\pi}\left(e^{+ij\lambda}e^{+ih\lambda}+e^{-ij\lambda}e^{-ih\lambda}\right)\left|\psi\left(e^{-i\lambda}\right)\right|^2 d\lambda$$

What this equation says is that, after integrating over all frequencies, only the "zero frequency" modes of the integrands survive, since the integral of a periodic sine wave over a range of 2π in frequency is zero. The integers *j* range from zero to infinity, and there is a separate equation for each integer *j*. So corresponding to the integer *j*, and regarding $\phi\left(e^{-i\lambda}\right)$ as being "positive frequency", and the *j*'th mode the term proportional to $e^{-ij\lambda}$, only the *j*'th mode of the terms proportional to $e^{+ij\lambda}$ survive, and only the –*j*'th mode of the terms proportional to $e^{-ij\lambda}$ survive, on both sides of the equation. Since each side of the equation involves the sum of a quantity and its complex conjugate, the conditions on the quantity and its complex conjugate are equivalent and redundant. Hence, for a given $j \geq 0$ the mode that survives on the left hand side is:

$$j\text{'th mode of:} \quad \phi\left(e^{-i\lambda}\right)\psi\left(e^{-i\lambda}\right)\psi^*\left(e^{-i\lambda}\right) \propto e^{-ij\lambda} \qquad (\text{lhs})$$

On the right hand side for a given $j \geq 0$ the mode that survives is:

$$j\text{'th mode of:} \quad e^{+ih\lambda}\psi\left(e^{-i\lambda}\right)\psi^*\left(e^{-i\lambda}\right) \propto e^{-ij\lambda} \qquad (\text{rhs})$$

The equation says that for each of these corresponding modes on each side of the equation, for a given $j \geq 0$, the sum of the modes and their complex conjugates are

equal. Now, the quantity $\psi^*\left(e^{-i\lambda}\right)$ consists of negative frequency modes, by our definition. Each term of this quantity shifts the mode in the same direction as some positive *j*, so this quantity is redundant with the exponential $e^{+ij\lambda}$ for some non-negative *j*. Thus no information is lost if we simply drop this term from both sides of the equation. Hence, for a given $j \geq 0$ the mode that survives on the left hand side is:

$$j\text{'th mode of:} \quad \phi\left(e^{-i\lambda}\right)\psi\left(e^{-i\lambda}\right) \propto e^{-ij\lambda} \qquad (\text{lhs})$$

On the right hand side for a given $j \geq 0$ the mode that survives is:

$$j\text{'th mode of:} \quad e^{+ih\lambda}\psi\left(e^{-i\lambda}\right) \propto e^{-ij\lambda} \qquad (\text{rhs})$$

On the right-hand side, only the modes of non-negative frequency survive. Writing this term out, we have:

$$j\text{'th mode of:} \quad e^{+ih\lambda}\psi\left(e^{-i\lambda}\right) = \sum_{k=0}^{+\infty}\psi_k e^{-i(k-h)\lambda} = \sum_{k=-h}^{+\infty}\psi_{k+h}e^{-ik\lambda}$$

We see that the first *h* modes, corresponding to negative frequency, do not survive the integral, so the surviving terms on the right-hand side are given by:

$$j \geq 0 \quad \Rightarrow \quad j\text{'th modes of} \quad \sum_{k=-h}^{+\infty}\psi_{k+h}e^{-ik\lambda} \quad \text{are} \quad \sum_{k=0}^{+\infty}\psi_{k+h}e^{-ik\lambda}$$

Thus, we find that the general solution is given by:

$$\text{Re}\left\{\left[\sum_{n=0}^{+\infty}\phi_n e^{-in\lambda}\right]\left[\sum_{k=0}^{+\infty}\psi_k e^{-ik\lambda}\right]\right\} = \text{Re}\left\{\left[\sum_{k=0}^{+\infty}\psi_{k+h}e^{-ik\lambda}\right]\right\} \quad (\phi_n \text{ real})$$

If we make the assumption that the ψ_k are also real, then the complex conjugate quantities are obtained just by reversing the sign of the frequency λ. Hence we may write for the general solution:

$$\left[\sum_{n=0}^{+\infty}\phi_n e^{-in\lambda}\right] = \left[\sum_{k=0}^{+\infty}\psi_{k+h}e^{-ik\lambda}\right] \Big/ \left[\sum_{k=0}^{+\infty}\psi_k e^{-ik\lambda}\right] \quad (\phi_n \text{ and } \psi_k \text{ real})$$

This, then, is the general solution of the Linear Prediction problem for the *h*-step-ahead predictor, in the Fourier domain, for a causal, invertible time series.

References

Peter J. Brockwell & Richard A. Davis, Time Series: Theory and Methods, 2nd ed. Springer-Verlag, New York (1991)

Donald B. Percival & Andrew T. Walden, Wavelet Methods for Time Series Analysis, Cambridge University Press (2000)