Chapter 22

Spectral Analysis and L_2 Ergodicity

Section 22.1 makes sense of the idea of white noise. This forms the bridge from the ideas of Wiener integrals, in the previous lectures, and spectral and ergodic theory, which we will pursue here.

Section 22.2 introduces the spectral representation of weakly stationary processes, and the central Wiener-Khinchin theorem connecting autocovariance to the power spectrum. Subsection 22.2.1 explains why white noise is "white".

Section 22.3 gives our first classical ergodic result, the "mean square" (L_2) ergodic theorem for weakly stationary processes. Subsection 22.3.1 gives an easy proof of a sufficient condition, just using the autocovariance. Subsection 22.3.2 gives a necessary and sufficient condition, using the spectral representation.

Any reasonable real-valued function $x(t)$ of time, $t \in \mathbb{R}$, has a Fourier transform, that is, we can write

$$\tilde{x}(\nu) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt e^{i\nu t} x(t)$$

which can usually be inverted to recover the original function,

$$x(t) = \int_{-\infty}^{\infty} d\nu e^{-i\nu t} \tilde{x}(\nu)$$

This one example of an "analysis", in the original sense of resolving into parts, of a function into a collection of orthogonal basis functions. (You can find the details in any book on Fourier analysis, as well as the varying conventions on where the 2π goes, which side gets the $e^{-i\nu t}$, the constraints on $\tilde{x}$ which arise from the fact that x is real, etc.)

There are various reasons to prefer the trigonometric basis functions $e^{i\nu t}$ over other possible choices. One is that they are invariant under translation in time, which just changes phases[1]. This suggests that the Fourier basis will be particularly useful when dealing with time-invariant systems. For stochastic processes, however, time-invariance is stationarity. This suggests that there should be some useful way of doing Fourier analysis on stationary random functions. In fact, it turns out that stationary and even weakly-stationary processes *can* be productively Fourier-transformed. This is potentially a huge topic, especially when it's expanded to include representing random functions in terms of (countable) series of orthogonal functions. The spectral theory of random functions connects Fourier analysis, disintegration of measures, Hilbert spaces and ergodicity. This lecture will do no more than scratch the surface, covering, in succession, white noise, the basics of the spectral representation of weakly-stationary random functions and the fundamental Wiener-Khinchin theorem linking covariance functions to power spectra, why white noise is called "white", and the mean-square ergodic theorem.

Good sources, if you want to go further, are the books of Bartlett (1955, ch. 6) (from whom I've stolen shamelessly), the historically important and inspiring Wiener (1949, 1961), and of course Doob (1953). Loève (1955, ch. X) is highly edifying, particular his discussion of Karhunen-Loève transforms, and the associated construction of the Wiener process as a Fourier series with random phases.

22.1 White Noise

Scientists and engineers are often uncomfortable with the SDEs in the way probabilists write them, because they want to divide through by dt and have the result mean something. The trouble, of course, is that dW/dt does not, in any ordinary sense, exist. They, however, are often happier ignoring this inconvenient fact, and talking about "white noise" as what dW/dt ought to be. This is not totally crazy. Rather, one can define $\xi \equiv dW/dt$ as a *generalized* derivative, one whose value at any given time is a random real linear functional, rather than a random real number. Consequently, it only really makes sense in integral expressions (like the solutions of SDEs!), but it can, in many ways, be formally manipulated like an ordinary function.

One way to begin to make sense of this is to start with a standard Wiener process $W(t)$, and a C^1 non-random function $u(t)$, and to use integration by

[1] If $t \mapsto t + \tau$, then $\tilde{x}(\nu) \mapsto e^{i\nu\tau}\tilde{x}(\nu)$.

parts:

$$\begin{aligned}
\frac{d}{dt}(uW) &= u\frac{dW}{dt} + \frac{du}{dt}W && (22.1)\\
&= u(t)\xi(t) + \dot{u}(t)W(t) && (22.2)\\
\int_0^t \frac{d}{dt}(uW)ds &= \int_0^t \dot{u}(s)W(s) + u(s)\xi(s)ds && (22.3)\\
u(t)W(t) - u(0)W(0) &= \int_0^t \dot{u}(s)W(s)ds + \int_0^t u(s)\xi(s)ds && (22.4)\\
\int_0^t u(s)\xi(s)ds &\equiv u(t)W(t) - \int_0^t \dot{u}(s)W(s)ds && (22.5)
\end{aligned}$$

We can take the last line to *define* ξ, and time-integrals within which it appears. Notice that the terms on the RHS are well-defined *without* the Itô calculus: one is just a product of two measurable random variables, and the other is the time-integral of a continuous random function. With this definition, we can establish some properties of ξ.

Proposition 271 (Linearity of White Noise Integrals) *$\xi(t)$ is a linear functional:*

$$\int_0^t (a_1 u_1(s) + a_2 u_2(s))\xi(s)ds = a_1 \int_0^t u_1(s)\xi(s)ds + a_2 \int_0^t u_2(s)\xi(s)ds \quad (22.6)$$

PROOF:

$$\begin{aligned}
&\int_0^t (a_1 u_1(s) + a_2 u_2(s))\xi(s)ds && (22.7)\\
&= (a_1 u_1(t) + a_2 u_2(t))W(t) - \int_0^t (a_1 \dot{u}_1(s) + a_2 \dot{u}_2(s))W(s)ds\\
&= a_1 \int_0^t u_1(s)\xi(s)ds + a_2 \int_0^t u_2(s)\xi(s)ds && (22.8)
\end{aligned}$$

□

Proposition 272 (White Noise Has Mean Zero) *For all t, $\mathbf{E}[\xi(t)] = 0$.*

PROOF:

$$\begin{aligned}
\int_0^t u(s)\mathbf{E}[\xi(s)]\,ds &= \mathbf{E}\left[\int_0^t u(s)\xi(s)ds\right] && (22.9)\\
&= \mathbf{E}\left[u(t)W(t) - \int_0^t \dot{u}(s)W(s)ds\right] && (22.10)\\
&= \mathbf{E}[u(t)W(t)] - \int_0^t \dot{u}(s)\mathbf{E}[W(t)]\,ds && (22.11)\\
&= 0 - 0 = 0 && (22.12)
\end{aligned}$$

Proposition 273 (White Noise and Itô Integrals) *For all $u \in C^1$, $\int_0^t u(s)\xi(s)ds = \int_0^t u(s)dW$.*

PROOF: Apply Itô's formula to the function $f(t, W) = u(t)W(t)$:

$$d(uW) = W(t)\dot{u}(t)dt + u(t)dW \quad (22.13)$$

$$u(t)W(t) = \int_0^t \dot{u}(s)W(s)ds + \int_0^t u(t)dW \quad (22.14)$$

$$\int_0^t u(t)dW = u(t)W(t) - \int_0^t \dot{u}(s)W(s)ds \quad (22.15)$$

$$= \int_0^t u(s)\xi(s)ds \quad (22.16)$$

□

This could be used to extend the definition of white-noise integrals to any Itô-integrable process.

Proposition 274 (White Noise is Uncorrelated) *ξ has delta-function covariance:* $\mathrm{cov}\left(\xi(t_1), \xi(t_2)\right) = \delta(t_1 - t_2)$.

PROOF: Since $\mathbf{E}\left[\xi(t)\right] = 0$, we just need to show that $\mathbf{E}\left[\xi(t_1)\xi(t_2)\right] = \delta(t_1 - t_2)$. Remember (Eq. 17.14 on p. 127) that $\mathbf{E}\left[W(t_1)W(t_2)\right] = t_1 \wedge t_2$.

$$\int_0^t \int_0^t u(t_1)u(t_2)\mathbf{E}\left[\xi(t_1)\xi(t_2)\right] dt_1 dt_2 \quad (22.17)$$

$$= \mathbf{E}\left[\int_0^t u(t_1)\xi(t_1)dt_1 \int_0^t u(t_2)\xi(t_2)dt_2\right] \quad (22.18)$$

$$= \mathbf{E}\left[\left(\int_0^t u(t_1)\xi(t_1)dt_1\right)^2\right] \quad (22.19)$$

$$= \int_0^t \mathbf{E}\left[u^2(t_1)\right] dt_1 = \int_0^t u^2(t_1)dt_1 \quad (22.20)$$

using the preceding proposition, the Itô isometry, and the fact that u is non-random. But

$$\int_0^t \int_0^t u(t_1)u(t_2)\delta(t_1 - t_2)dt_1 dt_2 = \int_0^t u^2(t_1)dt_1 \quad (22.21)$$

so $\delta(t_1 - t_2) = \mathbf{E}\left[\xi(t_1)\xi(t_2)\right] = \mathrm{cov}\left(\xi(t_1), \xi(t_2)\right)$. □

Proposition 275 (White Noise is Gaussian and Stationary) *ξ is a strongly stationary Gaussian process.*

PROOF: To show that it is Gaussian, use Exercise 19.6. The mean is constant for all times, and the covariance depends only on $|t_1 - t_2|$, so it satisfies Definition 50 and is weakly stationary. But a weakly stationary Gaussian process is also strongly stationary. □

22.2 Spectral Representation of Weakly Stationary Procesess

This section will only handle spectral representations of real- and complex-valued one-parameter processes in continuous time. Generalizations to vector-valued and multi-parameter processes are straightforward; handling discrete time is actually in some ways more irritating, because of limitations on allowable frequencies of Fourier components (to the range from $-\pi$ to π).

Definition 276 (Autocovariance Function) *Suppose that, for all $t \in T$, X is real and $\mathbf{E}\left[X^2(t)\right]$ is finite. Then $\Gamma(t_1, t_2) \equiv \mathbf{E}\left[X(t_1)X(t_2)\right] - \mathbf{E}\left[X(t_1)\right]\mathbf{E}\left[X(t_2)\right]$ is the* autocovariance function *of the process. If the process is weakly stationary, so that $\Gamma(t, t+\tau) = \Gamma(0, \tau)$ for all t, τ, write $\Gamma(\tau)$. If $X(t) \in \mathbb{C}$, then $\Gamma(t_1, t_2) \equiv \mathbf{E}\left[X^\dagger(t_1)X(t_2)\right] - \mathbf{E}\left[X^\dagger(t_1)\right]\mathbf{E}\left[X(t_2)\right]$, where $\dagger$ is complex conjugation.*

Lemma 277 (Autocovariance and Time Reversal) *If X is real and weakly stationary, then $\Gamma(\tau) = \Gamma(-\tau)$; if X is complex and weakly stationary, then $\Gamma(\tau) = \Gamma^\dagger(-\tau)$.*

PROOF: Direct substitution into the definitions. □

Remarks on terminology. It is common, when only dealing with one stochastic process, to drop the qualifying "auto" and just speak of the covariance function; I probably will myself. It is also common (especially in the time series literature) to switch to the (auto)correlation function, i.e., to normalize by the standard deviations. Finally, be warned that the statistical physics literature (e.g. Forster, 1975) uses "correlation function" to mean $\mathbf{E}\left[X(t_1)X(t_2)\right]$, i.e., the *uncentered* mixed second moment. This is a matter of tradition, not (despite appearances) ignorance.

Definition 278 (Second-Order Process) *A complex-valued process X is* second order *when $\mathbf{E}\left[|X|^2(t)\right] < \infty$ for all t.*

Definition 279 (Spectral Representation, Power Spectrum) *A real-valued process X on T has a complex-valued* spectral process $\tilde{X}$, *if it has a* spectral representation*:*

$$X(t) \equiv \int_{-\infty}^{\infty} e^{-i\nu t} d\tilde{X}(\nu) \tag{22.22}$$

The power spectrum $V(\nu) \equiv \mathbf{E}\left[\left|\tilde{X}(\nu)\right|^2\right]$.

Remark 1. The name "power spectrum" arises because this is proportional to the amount of power (energy per unit time) carried by oscillations of frequency $\leq \nu$, at least in a linear system.

Remark 2. Notice that the only part of the right-hand side of Equation 22.22 which depends on t is the integrand, $e^{-i\nu t}$, which just changes the phase of each Fourier component deterministically. Roughly speaking, for a fixed ω the amplitudes of the different Fourier components in $X(t, \omega)$ are fixed, and shifting forward in time just involves changing their phases. (Making this simple is why we have to allow $\tilde{X}$ to have complex values.)

The spectral representation is another stochastic integral, like the Itô integral we saw in Section 19.1. There, the measure of the time interval $[t_1, t_2]$ was given by the increment of the Wiener process, $W(t_2) - W(t_1)$. Here, for any frequency interval $[\nu_1, \nu_2]$, the increment $\tilde{X}(\nu_2) - \tilde{X}(\nu_1)$ defines a random set function (admittedly, a complex-valued one). Since those intervals are a generating class for the Borel σ-field, the usual arguments extend this set function uniquely to a random complex measure on $\mathbb{R}, \mathcal{B}$. When we write something like $\int G(\nu) d\tilde{X}(\nu)$, we mean an integral with respect to this measure.

Rather than dealing directly with this measure, we can, as with the Itô integral, use approximation by elementary processes. That is, we should interpret

$$\int_{-\infty}^{\infty} G(t, \nu) d\tilde{X}(\nu)$$

as the L_2 limit of sums

$$\sum_{\nu_i = -\infty}^{\infty} G(t, \nu_1)(\tilde{X}(\nu_{i+1}) - \tilde{X}(\nu_i))$$

as $\sup \nu_{i+1} - \nu_i$ goes to zero. This limit can be shown to exist in pretty much exactly the same way we showed the corresponding limit for Itô integrals to exist.[2]

Lemma 280 (Regularity of the Spectral Process) *When it exists, $\tilde{X}(\nu)$ has right and left limits at every point ν, and limits as $\nu \to \pm\infty$.*

PROOF: See Loève (1955, §34.4). You can prove this yourself, however, using the material on characteristic functions in 36-752. □

Definition 281 (Jump of the Spectral Process) *The* jump of the spectral process at ν *is the difference between the right- and left- hand limits at ν, $\Delta\tilde{X}(\nu) \equiv \tilde{X}(\nu + 0) - \tilde{X}(\nu - 0)$.*

Remark 1: As usual, $\tilde{X}(\nu+0) \equiv \lim_{h \downarrow 0} \tilde{X}(\nu + h)$, and $\tilde{X}(\nu-0) \equiv \lim_{h \downarrow 0} \tilde{X}(\nu - h)$.

Remark 2: Some people call the set of points at which the jump is non-zero the "spectrum". This usage comes from functional analysis, but seems needlessly confusing in the present context.

Proposition 282 (Spectral Representations of Weakly Stationary Processes) *Every weakly-stationary process has a spectral representation.*

[2]There is a really excellent discussion of such stochastic integrals, and L_2 stochastic calculus more generally, in Loève (1955, §34).

PROOF: See Loève (1955, §34.4), or Bartlett (1955, §6.2). □

The following property will be very important for us, since when the spectral process has it, many nice consequences follow.

Definition 283 (Orthogonal Increments) *A one-parameter random function (real or complex) has* orthogonal increments *if, for $t_1 \leq t_2 \leq t_3 \leq t_4 \in T$, the covariance of the increment from t_1 to t_2 and the increment from t_3 to t_4 is always zero:*

$$\mathbf{E}\left[\left(\tilde{X}(\nu_4) - \tilde{X}(\nu_3)\right)\left(\tilde{X}(\nu_2) - \tilde{X}(\nu_1)\right)^{\dagger}\right] = 0 \quad (22.23)$$

Lemma 284 (Orthogonal Spectral Increments and Weak Stationarity) *The spectral process of a second-order process has orthogonal increments if and only if the process is weakly stationary.*

SKETCH PROOF: Assume, without loss of generality, that $\mathbf{E}[X(t)] = 0$, so $\mathbf{E}\left[\tilde{X}(\nu)\right] = 0$. "If": Pick any arbitrary t. We can write, using the fact that $X(t) = X^{\dagger}(t)$ for real-valued processes,

$$\begin{aligned}
\Gamma(\tau) &= \Gamma(t, t+\tau) && (22.24)\\
&= \mathbf{E}\left[X^{\dagger}(t)X(t+\tau)\right] && (22.25)\\
&= \mathbf{E}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i\nu_1 t}e^{-i\nu_2(t+\tau)}d\tilde{X}^{\dagger}_{\nu_1}d\tilde{X}_{\nu_2}\right] && (22.26)\\
&= \mathbf{E}\left[\lim_{\Delta\nu\rightarrow 0}\sum_{\nu_1}\sum_{\nu_2} e^{it(\nu_1-\nu_2)}e^{-i\nu_2\tau}\Delta\tilde{X}^{\dagger}(\nu_1)\Delta\tilde{X}(\nu_2)\right] && (22.27)\\
&= \lim_{\Delta\nu\rightarrow 0}\sum_{\nu_1}\sum_{\nu_2} e^{it(\nu_1-\nu_2)}e^{-i\nu_2\tau}\mathbf{E}\left[\Delta\tilde{X}^{\dagger}(\nu_1)\Delta\tilde{X}(\nu_2)\right] && (22.28)
\end{aligned}$$

where $\Delta\tilde{X}(\nu) = \tilde{X}(\nu+\Delta\nu) - \tilde{X}(\nu)$. Since t was arbitrary, every term on the right must be independent of t. When $\nu_1 = \nu_2$, $e^{it(\nu_1-\nu_2)} = 1$, so $\mathbf{E}\left[\Delta\tilde{X}^{\dagger}(\nu)\Delta\tilde{X}(\nu)\right]$ is unconstrained. If $\nu_1 \neq \nu_2$, however, we must have $\mathbf{E}\left[\Delta\tilde{X}^{\dagger}(\nu_1)\Delta\tilde{X}(\nu_2)\right] = 0$, which is to say (Definition 283) we must have orthogonal increments.

"Only if": if the increments are orthogonal, then clearly the steps of the argument can be reversed to conclude that $\Gamma(t_1, t_2)$ depends only on $t_2 - t_1$. □

Definition 285 (Spectral Function, Spectral Density) *The spectral function of a weakly stationary process is the function $S(\nu)$ appearing in the spectral representation of its autocovariance:*

$$\Gamma(\tau) = \int_{-\infty}^{\infty} e^{-i\nu\tau}dS_{\nu} \quad (22.29)$$

Remark. Some people prefer to talk about the spectral function as the Fourier transform of the autocorrelation function, rather than of the autocovariance. This has the advantage that the spectral function turns out to be a normalized cumulative distribution function (see Theorem 286 immediately below), but is otherwise inconsequential.

Theorem 286 (Weakly Stationary Processes Have Spectral Functions) *The spectral function exists for every weakly stationary process, if $\Gamma(\tau)$ is continuous. Moreover, $S(\nu) \geq 0$, S is non-decreasing, $S(-\infty) = 0$, $S(\infty) = \Gamma(0)$, and $lim_{h\downarrow 0} S(\nu + h)$ and $\lim_{h\downarrow 0} S(\nu - h)$ exist for every ν.*

Proof: Usually, by an otherwise-obscure result in Fourier analysis called Bochner's theorem. A more direct proof is due to Loève. Assume, without loss of generality, that $\mathbf{E}[X(t)] = 0$.

Start by defining

$$H_T(\nu) \equiv \frac{1}{\sqrt{T}} \int_{-T/2}^{T/2} e^{i\nu t} X(t) dt \tag{22.30}$$

and define $f_T(\nu)$ through H:

$$\begin{aligned}
2\pi f_T(\nu) &\equiv \mathbf{E}\left[H_T(\nu) H_T^{\dagger}(\nu)\right] && (22.31)\\
&= \mathbf{E}\left[\frac{1}{T}\int_{-T/2}^{T/2}\int_{-T/2}^{T/2} e^{i\nu t_1} X(t_1) e^{-i\nu t_2} X^{\dagger}(t_2) dt_1 dt_2\right] && (22.32)\\
&= \frac{1}{T}\int_{-T/2}^{T/2}\int_{-T/2}^{T/2} e^{i\nu(t_1 - t_2)} \mathbf{E}\left[X(t_1)X(t_2)\right] dt_1 dt_2 && (22.33)\\
&= \frac{1}{T}\int_{-T/2}^{T/2}\int_{-T/2}^{T/2} e^{i\nu(t_1 - t_2)} \Gamma(t_1 - t_2) dt_1 dt_2 && (22.34)\\
&= \int_{-T}^{T} \left(1 - \frac{|\tau|}{T}\right) \Gamma(\tau) e^{i\nu\tau} d\tau && (22.35)
\end{aligned}$$

Recall that $\Gamma(\tau)$ defines a non-negative quadratic form, meaning that

$$\sum_{s,t} a_s^{\dagger} a_t \Gamma(t - s) \geq 0$$

for any sets of times and any complex numbers a_t. This will in particular work if the complex numbers lie on the unit circle and can be written $e^{i\nu t}$. This means that integrals

$$\int\int e^{i\nu(t_1 - t_2)} \Gamma(t_1 - t_2) dt_1 dt_2 \geq 0 \tag{22.36}$$

so $f_T(\nu) \geq 0$.

Define $\phi_T(\tau)$ as the integrand in Eq. 22.35, so that

$$f_T(\nu) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_T(\tau) e^{i\nu\tau} d\tau \tag{22.37}$$

which is recognizable as a proper Fourier transform. Now pick some $N > 0$ and massage the equation so it starts to look like an inverse transform.

$$f_T(\nu)e^{-i\nu t} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_T(\tau)e^{i\nu\tau}e^{-i\nu t}d\tau \quad (22.38)$$

$$\left(1-\frac{|\nu|}{N}\right)f_T(\nu)e^{-i\nu t} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_T(\tau)e^{i\nu\tau}e^{-i\nu t}\left(1-\frac{|\nu|}{N}\right)d\tau \quad (22.39)$$

Integrating over frequencies,

$$\int_{-N}^{N}\left(1-\frac{|\nu|}{N}\right)f_T(\nu)e^{-i\nu t}d\nu \quad (22.40)$$

$$= \int_{-N}^{N}\frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_T(\tau)e^{i\nu\tau}e^{-i\nu t}\left(1-\frac{|\nu|}{N}\right)d\tau d\nu$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_T(\tau)\left(\frac{\sin N(\tau-t)/2}{N(\tau-t)/2}\right)^2 N d\tau \quad (22.41)$$

For fixed N, it is easy to verify that

$$\int_{-\infty}^{\infty} N\left(\frac{\sin N(\tau-t)/2}{N(\tau-t)/2}\right)^2 dt = 1$$

and that

$$\lim_{t\rightarrow\tau} N\left(\frac{\sin N(\tau-t)/2}{N(\tau-t)/2}\right)^2 = N$$

On the other hand, if $\tau \neq t$,

$$\lim_{N\rightarrow\infty} N\left(\frac{\sin N(\tau-t)/2}{N(\tau-t)/2}\right)^2 = 0$$

uniformly over any bounded interval in t. (You might find it instructive to try plotting this function; you will need to be careful near the origin!) In other words, this is a representation of the Dirac delta function, so that

$$\lim_{N\rightarrow\infty}\int_{-\infty}^{\infty} \phi_T(\tau)\left(\frac{\sin N(\tau-t)/2}{N(\tau-t)/2}\right)^2 N d\tau = \phi_T(\tau)$$

and in fact the convergence is uniform.

Turning to the other side of Equation 22.41, $\left(1-\frac{|\nu|}{N}\right)f_T(\nu) \geq 0$, so

$$\int_{-N}^{N}\left(1-\frac{|\nu|}{N}\right)f_T(\nu)e^{-i\nu t}d\nu$$

is like a characteristic function of a distribution, up to, perhaps, an over-all normalizing factor, which will be (given the right-hand side) $\phi_T(0) = \Gamma(0) > 0$. Since $\Gamma(\tau)$ is continuous, $\phi_T(\tau)$ is too, and so, as $N \rightarrow \infty$, the right-hand side converges uniformly on $\phi_T(t)$, but a uniform limit of characteristic functions is still a characteristic function. Thus $\phi_T(t)$, too, can be obtained from a characteristic function. Finally, since $\Gamma(t)$ is the uniform limit of $\phi_T(t)$ on every bounded interval, $\Gamma(t)$ has a characteristic-function representation of the stated form. This allows us to further conclude that $S(\nu)$ is real-valued, non-decreasing, $S(-\infty) = 0$ and $S(\infty) = \Gamma(0)$, and has both right and left limits everywhere. □

There is a converse, with a cute constructive proof.

Theorem 287 (Existence of Weakly Stationary Processes with Given Spectral Functions) *Let $S(\nu)$ be any function with the properties described at the end of Theorem 286. Then there is a weakly stationary process whose autocovariance is of the form given in Eq. 22.29.*

PROOF: Define $\sigma^2 = \Gamma(0)$, $F(\nu) = S(\nu)/\sigma^2$. Now $F(\nu)$ is a properly normalized cumulative distribution function. Let N be a random variable distributed according to F, and $\Phi \sim U(0, 2\pi)$ be independent of N. Set $X(t) \equiv \sigma e^{i(\Phi - Nt)}$. Then $\mathbf{E}[X(t)] = \sigma \mathbf{E}\left[e^{i\Phi}\right]\mathbf{E}\left[e^{-iNt}\right] = 0$. Moreover,

$$
\begin{aligned}
\mathbf{E}\left[X^\dagger(t_1)X(t_2)\right] &= \sigma^2 \mathbf{E}\left[e^{-i(\Phi - Nt_1)}e^{i(\Phi - Nt_2)}\right] && (22.42)\\
&= \sigma^2 \mathbf{E}\left[e^{-iN(t_1 - t_2)}\right] && (22.43)\\
&= \sigma^2 \int_{-\infty}^{\infty} e^{-i\nu(t_1 - t_2)} dF_\nu && (22.44)\\
&= \Gamma(t_1 - t_2) && (22.45)
\end{aligned}
$$

□

Definition 288 (Jump of the Spectral Function) *The* jump of the spectral function *at ν, $\Delta S(\nu)$, is $S(\nu + 0) - S(\nu - 0)$.*

Lemma 289 (Spectral Function Has Non-Negative Jumps) $\Delta S(\nu) \geq 0$.

PROOF: Obvious from the fact that $S(\nu)$ is non-decreasing. □

Theorem 290 (Wiener-Khinchin Theorem) *If X is a weakly stationary process, then its power spectrum is equal to its spectral function.*

$$
V(\nu) \equiv \mathbf{E}\left[\left|\tilde{X}(\nu)\right|^2\right] = S(\nu) \qquad (22.46)
$$

PROOF: Assume, without loss of generality, that $\mathbf{E}[X(t)] = 0$. Substitute the spectral representation of X into the autocovariance, using Fubini's theorem to turn a product of integrals into a double integral.

$$\begin{aligned}
\Gamma(\tau) &= \mathbf{E}\left[X(t)X(t+\tau)\right] &(22.47)\\
&= \mathbf{E}\left[X^\dagger(t)X(t+\tau)\right] &(22.48)\\
&= \mathbf{E}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-i(t+\tau)\nu_1}e^{it\nu_2}d\tilde{X}_{\nu_1}d\tilde{X}_{\nu_2}\right] &(22.49)\\
&= \mathbf{E}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-it(\nu_1-\nu_2)}e^{-i\tau\nu_2}d\tilde{X}_{\nu_1}d\tilde{X}_{\nu_2}\right] &(22.50)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-it(\nu_1-\nu_2)}e^{-i\tau\nu_2}\mathbf{E}\left[d\tilde{X}_{\nu_1}d\tilde{X}_{\nu_2}\right] &(22.51)
\end{aligned}$$

using the fact that integration and expectation commute to (formally) bring the expectation inside the integral. Since $\tilde{X}$ has orthogonal increments, $\mathbf{E}\left[d\tilde{X}^\dagger_{\nu_1}d\tilde{X}_{\nu_2}\right] = 0$ unless $\nu_1 = \nu_2$. This turns the double integral into a single integral, and kills the $e^{-it(\nu_1-\nu_2)}$ factor, which had to go away because t was arbitrary.

$$\begin{aligned}
\Gamma(\tau) &= \int_{-\infty}^{\infty} e^{-i\tau\nu}\mathbf{E}\left[d(\tilde{X}^\dagger_\nu\tilde{X}_\nu)\right] &(22.52)\\
&= \int_{-\infty}^{\infty} e^{-i\tau\nu}dV_\nu &(22.53)
\end{aligned}$$

using the definition of the power spectrum. Since $\Gamma(\tau) = \int_{-\infty}^{\infty} e^{-i\tau\nu}dV_\nu$, it follows that S_ν and V_ν differ by a constant, namely the value of $V(-\infty)$, which can be chosen to be zero without affecting the spectral representation of X. □

22.2.1 How the White Noise Lost Its Color

Why is white noise, as defined in Section 22.1, called "white"? The answer is easy, given the Wiener-Khinchin relation in Theorem 290.

Recall from Proposition 274 that the autocovariance function of white noise is $\delta(t_1 - t_2)$. Recall from general analysis that one representation of the delta function is the following Fourier integral:

$$\delta(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\nu e^{i\nu t}$$

(This can be "derived" from inserting the definition of the Fourier transform into the inverse Fourier transform, among other, more respectable routes.) Appealing then to the theorem, $S(\nu) = \frac{1}{2\pi}$ for all ν. That is, there is equal power at all frequencies, just as white light is composed of light of all colors (frequencies), mixed with equal intensity.

Relying on this analogy, there is an elaborate taxonomy red, pink, black, brown, and other variously-colored noises, depending on the shape of their power spectra. The value of this terminology has honestly never been very clear to me, but the curious reader is referred to the (very fun) book of Schroeder (1991) and references therein.

22.3 The Mean-Square Ergodic Theorem

Ergodic theorems relate functionals calculated along individual sample paths (say, the time average, $T^{-1}\int_0^T dtX(t)$, or the maximum attained value) to functionals calculated over the whole distribution (say, the expectation, $\mathbf{E}[X(t)]$, or the expected maximum). The basic idea is that the two should be close, and they should get closer the longer the trajectory we use, because in some sense any one sample path, carried far enough, is representative of the whole distribution. Since there are many different kinds of functionals, and many different modes of stochastic convergence, there are many different kinds of ergodic theorem. The classical ergodic theorems say that time averages converge on expectations[3], either in L_p or a.s. (both implying convergence in distribution or in probability). The traditional centerpiece of ergodic theorem is Birkhoff's "individual" ergodic theorem, asserting a.s. convergence. We will see its proof, but it will need a lot of preparatory work, and it requires strict stationarity. By contrast, the L_2, or "mean square", ergodic theorem, attributed to von Neumann[4] is already in our grasp, and holds for weakly stationary processes.

We will actually prove it twice, once with a fairly transparent sufficient condition, and then again with a more complicated necessary-and-sufficient condition. The more complicated proof will wait until next lecture.

22.3.1 Mean-Square Ergodicity Based on the Autocovariance

First, the easy version, which gives an estimate of the rate of convergence. (What I say here is ripped off from the illuminating discussion in (Frisch, 1995, sec. 4.4, especially pp. 49–50).)

Definition 291 (Time Averages) *When X is a one-sided, continuous-parameter random process, we say that its* time average *between times T_1 and T_2 is $\overline{X}(T_1, T_2) \equiv (T_2 - T_1)^{-1}\int_{T_1}^{T_2} dtX(t)$. When we only mention one time argument, by default the time average is from 0 to T, $\overline{X}(T) \equiv \overline{X}(0, T)$.*

(Only considering time averages starting from zero involves no loss of generality for weakly stationary processes: why?)

Definition 292 (Integral Time Scale) *The* integral time scale *of a weakly-stationary random process is*

$$\tau_{\text{int}} \equiv \frac{\int_0^\infty d\tau |\Gamma(\tau)|}{\Gamma(0)} \tag{22.54}$$

[3]Proverbially: "time averages converge on space averages", the space in question being the state space Ξ; or "converge on phase averages", since physicists call certain kinds of state space "phase space".

[4]See von Plato (1994, ch. 3) for a fascinating history of the development of ergodic theory through the 1930s, and its place in the history of mathematical probability.

Notice that τ_{int} does, indeed, have units of time.

As a particular example, suppose that $\Gamma(\tau) = \Gamma(0)e^{-\tau/A}$, where the constant A is known as the *autocorrelation time*. Then simple calculus shows that $\tau_{\text{int}} = A$.

Theorem 293 (Mean-Square Ergodic Theorem (Finite Autocovariance Time)) *Let $X(t)$ be a weakly stationary process with $\mathbf{E}[X(t)] = 0$. If $\tau_{\text{int}} < \infty$, then $\overline{X}(T) \stackrel{L_2}{\rightarrow} 0$ as $T \rightarrow \infty$.*

Proof: Use Fubini's theorem to to the square of the integral into a double integral, and then bring the expectation inside it:

$$\begin{aligned}
\mathbf{E}\left[\left(\frac{1}{T}\int_0^T dtX(t)\right)^2\right] &= \mathbf{E}\left[\frac{1}{T^2}\int_0^T\int_0^T dt_1 dt_2 X(t_1)X(t_2)\right] && (22.55)\\
&= \frac{1}{T^2}\int_0^T\int_0^T dt_1 dt_2 \mathbf{E}\left[X(t_1)X(t_2)\right] && (22.56)\\
&= \frac{1}{T^2}\int_0^T\int_0^T dt_1 dt_2 \Gamma(t_1 - t_2) && (22.57)\\
&= \frac{2}{T^2}\int_0^T dt_1 \int_0^{t_1} d\tau\Gamma(\tau) && (22.58)\\
&\leq \frac{2}{T^2}\int_0^T dt_1 \int_0^{\infty} d\tau|\Gamma(\tau)| && (22.59)\\
&= \frac{2}{T}\int_0^{\infty} d\tau|\Gamma(\tau)| && (22.60)
\end{aligned}$$

Since the integral in the final inequality is $\Gamma(0)\tau_{\text{int}}$, which is finite, everything must go to zero as $T \rightarrow \infty$. □

Remark. From the proof, we can see that the rate of convergence of the mean-square of $\left\|\overline{X}(T)\right\|_2^2$ is (at least) $O(1/T)$. This would give a root-mean-square (rms) convergence rate of $O(1/\sqrt{T})$, which is what the naive statistician who ignored inter-temporal dependence would expect from the central limit theorem. (This ergodic theorem says *nothing* about the form of the distribution of $\overline{X}(T)$ for large T. We will see that, under some circumstances, it *is* Gaussian, but that needs stronger assumptions [forms of "mixing"] than we have imposed.) The naive statistician would expect that the mean-square time average would go like $\Gamma(0)/T$, since $\Gamma(0) = \mathbf{E}\left[X^2(t)\right] = \mathbf{Var}\left[X(t)\right]$. The proportionality constant is instead $\int_0^\infty d\tau|\Gamma(\tau)|$. This is equal to the naive guess for white noise, and for other collections of IID variables, but not in the general case. This leads to the following

Corollary 294 (Convergence Rate in the Mean-Square Ergodic Theorem) *Under the conditions of Theorem 293,*

$$\mathbf{Var}\left[\overline{X}(T)\right] \leq 2\mathbf{Var}\left[X(0)\right]\frac{\tau_{\text{int}}}{T} \qquad (22.61)$$

PROOF: Since $X(t)$ is centered, $\mathbf{E}\left[\overline{X}(T)\right] = 0$, and $\left\|\overline{X}(T)\right\|_2^2 = \mathbf{Var}\left[\overline{X}(T)\right]$. Everything else follows from re-arranging the bound in the proof of Theorem 293, Definition 292, and the fact that $\Gamma(0) = \mathbf{Var}\left[X(0)\right]$. □

As a consequence of the corollary, if $T \gg \tau_{\mathrm{int}}$, then the variance of the time average is negigible compared to the variance at any one time.

22.3.2 Mean-Square Ergodicity Based on the Spectrum

Let's warm up with some lemmas of a technical nature. The first relates the jumps of the spectral process $\tilde{X}(\nu)$ to the jumps of the spectral function $S(\nu)$.

Lemma 295 (Mean-Square Jumps in the Spectral Process) *For a weakly stationary process,* $\mathbf{E}\left[\left|\Delta\tilde{X}(\nu)\right|^2\right] = \Delta S(\nu)$.

PROOF: This follows directly from the Wiener-Khinchin relation (Theorem 290). □

Lemma 296 (The Jump in the Spectral Function) *The jump of the spectral function at ν is given by*

$$\Delta S(\nu) = \lim_{T\to\infty} \frac{1}{T}\int_0^T \Gamma(\tau) e^{i\nu\tau} d\tau \tag{22.62}$$

PROOF: This is a basic inversion result for characteristic functions. It should become plausible by thinking of this as getting the Fourier transform of Γ as T grows. □

Lemma 297 (Existence of L_2 Limits for Time Averages) *If X is weakly stationary, then for any real f, $\overline{e^{ift}X}(T)$ converges in L_2 to $\Delta\tilde{X}(f)$.*

PROOF: Start by looking at the squared modulus of the time average for finite time.

$$\left|\frac{1}{T}\int_0^T e^{ift} X(t) dt\right|^2 \tag{22.63}$$

$$= \frac{1}{T^2}\int_0^T \int_0^T e^{-if(t_1-t_2)} X^{\dagger}(t_1) X(t_2) dt_1 dt_2$$

$$= \frac{1}{T^2}\int_0^T \int_0^T e^{-if(t_1-t_2)} \int_{-\infty}^{\infty} e^{i\nu_1 t_1} d\tilde{X}_{\nu_1} \int_{-\infty}^{\infty} e^{-i\nu_2 t_2} d\tilde{X}_{\nu_2} \tag{22.64}$$

$$= \frac{1}{T^2}\int_0^T \int_{-\infty}^{\infty} dt_1 d\tilde{X}_{\nu_1} e^{it_1(f-\nu_1)} \int_0^T \int_{-\infty}^{\infty} dt_2 d\tilde{X}_{\nu_2} e^{-it_2(f-\nu_2)} \tag{22.65}$$

As $T \to \infty$, these integrals pick out $\Delta\tilde{X}(f)$ and $\Delta\tilde{X}^{\dagger}(f)$. So, $\overline{e^{ift}X}(T) \stackrel{L_2}{\rightarrow} \Delta\tilde{X}(f)$. □

Notice that the limit provided by the lemma is a random quantity. What's really desired, in most applications, is convergence to a *deterministic* limit, which here would mean convergence (in L_2) to zero.

Lemma 298 (The Mean-Square Ergodic Theorem) *If X is weakly stationary, and $\mathbf{E}[X(t)] = 0$, then $\overline{X}(t)$ converges in L_2 to 0 iff*

$$\lim T^{-1} \int_0^T d\tau \Gamma(\tau) = 0 \tag{22.66}$$

PROOF: Taking $f = 0$ in Lemma 297, $\overline{X}(T) \stackrel{L_2}{\rightarrow} \Delta\tilde{X}(0)$, the jump in the spectral function at zero. Let's show that the (i) expectation of this jump is zero, and that (ii) its variance is given by the integral expression on the LHS of Eq. 22.66. For (i), because $\overline{X}(T) \stackrel{L_2}{\rightarrow} Y$, we know that $\mathbf{E}\left[\overline{X}(T)\right] \rightarrow \mathbf{E}[Y]$. But $\mathbf{E}\left[\overline{X}(T)\right] = \overline{\mathbf{E}[X]}(T) = 0$. So $\mathbf{E}\left[\Delta\tilde{X}(0)\right] = 0$. For (ii), Lemma 295, plus the fact that $\mathbf{E}\left[\Delta\tilde{X}(0)\right] = 0$, shows that the variance is equal to the jump in the spectrum at 0. But, by Lemma 296 with $\nu = 0$, that jump is exactly the LHS of Eq. 22.66. □

Remark 1: Notice that if the integral time is finite, then the integral condition on the autocovariance is automatically satisfied, but not vice versa, so the hypotheses here are strictly weaker than in Theorem 293.

Remark 2: One interpretation of the theorem is that the time-average is converging on the zero-frequency component of the spectral process. Intuitively, all the other components are being killed off by the time-averaging, because the because time-integral of a sinusoidal function is bounded, but the denominator in a time average is unbounded. The only part left is the zero-frequency component, whose time integral can also grow linearly with time. If there is a jump at 0, then this has finite variance; if not, not.

Remark 3: Lemma 297 establishes the L_2 convergence of time-averages of the form

$$\frac{1}{T}\int_0^T e^{ift} X(t) dt$$

for any real f. Specifically, from Lemma 295, the mean-square of this variable is converging on the jump in the spectrum at f. Multiplying $X(t)$ by e^{ift} makes the old frequency f component the new frequency zero component, so *it* is the surviving term. While the ergodic theorem itself only needs the $f = 0$ case, this result is useful in connection with estimating spectra from time series (Doob, 1953, ch. X, §7).

22.4 Exercises

Exercise 22.1 (Mean-Square Ergodicity in Discrete Time) *It is often convenient to have a mean-square ergodic theorem for discrete-time sequences rather than continuous-time processes. If the dt in the definition of $\overline{X}$ is re-interpreted as counting measure on $\mathbb{N}$, rather than Lebesgue measure on $\mathbb{R}^+$, does the proof of Theorem 293 remain valid? (If yes, say why; if no, explain where the argument fails.)*

Exercise 22.2 (Mean-Square Ergodicity with Non-Zero Mean) *State and prove a version of Theorem 293 which does not assume that* $\mathbf{E}[X(t)] = 0$.

Exercise 22.3 (Functions of Weakly Stationary Processes) *Suppose* X *is a weakly stationary process, and* f *is a measurable function such that* $\|f(X_0)\|_2 < \infty$. *Is* $f(X)$ *a weakly stationary process? (If yes, prove it; if not, give a counter-example.)*

Exercise 22.4 (Ergodicity of the Ornstein-Uhlenbeck Process?) *Suppose the Ornstein-Uhlenbeck process is has its invariant distribution as its initial distribution, and is therefore weakly stationary. Does Theorem 293 apply?*