Lecture 5: Gaussian processes & Stationary processes

Readings

Recommended:

- Pavliotis (2014), Sections 1.1, 1.2
- Grimmett and Stirzaker (2001), 8.2, 8.6
- Grimmett and Stirzaker (2001) 9.1, 9.3, 9.5, 9.6 (more advanced than what we will cover in class)

Optional:

- Grimmett and Stirzaker (2001) 4.9 (review of multivariate Gaussian random variables)
- Grimmett and Stirzaker (2001) 9.4
- Chorin and Hald (2009) Chapter 6; a gentle introduction to the spectral decomposition for continuous processes, that gives insight into the link to the spectral decomposition for discrete processes.
- Yaglom (1962), Ch. 1, 2; a nice short book with many details about stationary random functions; one of the original manuscripts on the topic.
- Lindgren (2013) is a in-depth but accessible book; with more details and a more modern style than Yaglom (1962).

We want to describe stochastic processes that are not necessarily Markov processes. In this lecture we will look at two classes of stochastic processes that are commonly used as models and that are tractable to simulate: Gaussian processes, and stationary processes.

5.1 Setup

To describe more general classes of stochastic processes we'll need two key ideas, introduced in this section: finite-dimensional distributions, and covariance functions.

5.1.1 Finite dimensional distributions

Definition. The *finite-dimensional distributions* (fdds) of a stochastic process $(X_t)_{t\in T}$ is the distribution of $(X_{t_1}, X_{t_2}, \dots X_{t_k})$ at a finite number of points $(t_1, t_2, \dots, t_k) \in T^k$. That it, it is the set of measures $\{P_{t_1,t_2,\dots,t_k}\}_{k\in\mathbb{N},(t_1,\dots,t_k)\in T^k}$ given by

$$P_{t_1,t_2,\dots,t_k}(F_1 \times F_2 \times \cdots \times F_k) = P(X_{t_1} \in F_1, X_{t_2} \in F_2, \dots X_{t_k} \in F_k),$$

where $F_i \in \mathbb{R}$ are events.

Examples 5.1 Here are some examples of processes and their fdds.

1. A sequence of i.i.d. random variables with probability measure P. The parameter set is $T = \mathbb{Z}^+$, and the fdds are $P_{t_1,\dots,t_k}(F_1 \times \cdots \times F_k) = \prod_{i=1}^k P(F_i)$.
2. Markov chains. Let X_t be a continuous-time Markov chain with generator Q and initial probability distribution μ_0, and let $0 = t_0 < t_1 < \cdots < t_k$. Then from the Chapman Kolmogorov equations, the

fdds are

$$P(X_{t_0} \in F_0, X_{t_1} \in F_1, \ldots, X_{t_k} \in F_k) = \sum_{i_0 \in F_0} \mu_0(i_0) \sum_{i_1 \in F_1} P_{i_0 i_1}(t_1 - t_0) \cdots \sum_{i_k \in F_k} P_{i_{k-1} i_k}(t_k - t_{k-1}).$$

3. Poisson process. For any $0 \le t_1 \le \ldots \le t_k$, let $\eta_1, \eta_2, \ldots, \eta_k$ be independent Poisson random variables with rates $\lambda t_1, \lambda(t_2 - t_1), \ldots, \lambda(t_k - t_{k-1})$. Then $P_{t_1,\ldots,t_k}$ is the joint distribution of $(\eta_1, \eta_1 + \eta_2, \ldots, \eta_1 + \eta_2 + \cdots + \eta_k)$.

The set of finite-dimensional distributions characterizes the pointwise statistics of a process. If one starts from a set of fdds, then one can construct a corresponding stochastic process, provided the fdds satisfy some basic consistency criteria (they are invariant under permutations, and their marginal distributions are consistent) – this result is known as the *Kolmogorov Extension Theorem* (Koralov and Sinai (2010), p.174, or Grimmett and Stirzaker (2001), Section 8.7 p.372). There is a nice description of the construction in Terry's Tao's notes: `https://terrytao.wordpress.com/2010/10/30/245a-notes-6-outer-measures-pre-measures-and-product#more-4371`.

However, knowing only the fdds of a stochastic process does not let us ask many questions that we may be interested in, such as whether the process is continuous, or differentiable, or bounded; what is the first passage time to a given set, what is the running maximum at time t, etc. To decide such questions we need the values of the process at an uncountable number of points.

Here is an example to illustrate some of the difficulties.

Example 5.2 Let $U \sim \text{Uniform}([0,1])$ be a random variable. Define two processes $X = (X_t)_{0 \le t \le 1}$ and $Y = (Y_t)_{0 \le t \le 1}$ by

$$X_t = 0 \qquad \text{for all } t, \qquad Y_t = \begin{cases} 1 & \text{if } U = t, \\ 0 & \text{otherwise} \end{cases}$$

Then X and Y have the same fdds, since $P(U = t) = 0$ for all t. But X, Y are different processes. Some differences include: X is continuous whereas Y is not; $\sup_t X = 0$ whereas $\sup_t Y = 1$; $P(X_t{=}0 \text{ for all } t) = 1$ whereas $P(Y_t{=}0 \text{ for all } t) = 0$.

This example may seem trivial, but one can often construct less trivial examples of different processes with the same fdds. Such processes are called *versions* of each other. Any theory which studies properties of sample paths must take care to specify which version of a process is being studied. Typically one considers processes which are right-continuous and have limits from the left (so-called "càdlàg" processes, an abbreviation of the French phrase "continue à droite, limites à gauche.") Conditions under which which a stochastic process has a càdlàg version have been characterized (Grimmett and Stirzaker (2001), Theorem 8.7.6, p.373; Breiman (1992), p.300). In this course we will always assume we are working with a càdlàg process.

At the same time, fdds also give us *more* information about a stochastic process than we often need – certainly more information than we are able to measure in practice. Therefore we turn to other ways to characterize stochastic processes.

5.1.2 Covariance functions

The most widely used way to characterize properties of a stochastic process is by its mean and covariance function, which give the one- and two-point fdds.

Definition. The *mean* of a stochastic process is $m(t) = \mathbb{E}X_t$.

Recall that if we have a random vector $X = (X_1, \dots, X_n)^T$, its covariance matrix

$$\Sigma = \mathbb{E}XX^T - (\mathbb{E}X)(\mathbb{E}X)^T$$

is the matrix whose elements are the covariance of X_i, X_j, i.e. $\Sigma_{ij} = \mathbb{E}X_i\mathbb{E}X_j - (\mathbb{E}X_i)(\mathbb{E}X_j)$. This concept generalizes to random functions:

Definition. The *covariance function* of a real-valued stochastic process $(X_t)_{t\in T}$ is

$$B(s,t) \equiv \mathbb{E}\left[(X_s - \mathbb{E}X_s)(X_t - \mathbb{E}X_t)\right] = \mathbb{E}X_sX_t - m(s)m(t).$$

Notice that $B(t,t) = \mathbb{E}X_t^2 - (\mathbb{E}X_t)^2$ gives the variance of the process at a single point t.

One can easily show that a covariance *matrix* is positive semidefinite, i.e. $x^T\Sigma x \geq 0$ for all $x \in \mathbb{R}^n$. (Exercise: show this!) A similar statement is true for covariance *functions.*

Definition. A function $B(s,t)$ is *positive (semi)definite* if if the matrix $(B(t_i,t_j))_{i,j=1}^k$ is positive (semi)definite for all finite time slices $\{t_i\}_{i=1}^k$.

Lemma. *The covariance function $B(s,t)$ of a stochastic process is positive semidefinite.*

Proof. Left as an exercise on the homework. □

Remark. An equivalent definition comes from considering the operator $\mathscr{K}$ on $L^2(\mathbb{R})$ defined by $\mathscr{K}f = \int B(s,t)f(t)dt$. One can verify this operator is symmetric with with respect to the L^2 inner product. Then B is positive semidefinite if the operator $\mathscr{K}$ is positive semidefinite with respect to the L^2 inner product, i.e. $\langle \mathscr{K}f, g\rangle = \int B(s,t)f(t)g(s)dtds \geq 0$ for all $f, g \in L^2(\mathbb{R})$.

Example 5.3 Let's calculate the covariance function of a Poisson process $(N_t)_{t\geq 0}$ with parameter λ. When $t \geq s$, we calculate:

$$\begin{aligned} B(s,t) = \mathbb{E}N_tN_s - (\mathbb{E}N_t)(\mathbb{E}N_s) &= \mathbb{E}(N_t - N_s)N_s + \mathbb{E}N_s^2 - \lambda^2 ts \\ &= \lambda^2(t-s)s + \lambda s + \lambda^2 s^2 - \lambda^2 ts \\ &= \lambda s\,. \end{aligned}$$

We used the fact that $N_t - N_s$, N_s are independent, and the known mean and variance of a Poisson process. Repeating the calculation with $s > t$ gives that $B(s,t) = \lambda \min(s,t)$.

It will be useful later to work in complex space, so we also define these concepts for complex-valued processes.

Definition. A *complex-valued stochastic process* is one whose real and imaginary parts are real-valued stochastic processes, i.e. it has the form $X_t = Y_t + iZ_t$ where $(Y_t)_{t\in T}$, $(Z_t)_{t\in T}$ are real-valued stochastic processes.

The mean of a complex-valued stochastic process is defined in the same way as before, $m(t) = \mathbb{E}X_t$.

Definition. The *covariance function* of a complex-valued stochastic process is $B(s,t) = \mathbb{E}X_s\overline{X_t} - (\mathbb{E}X_s)\overline{(\mathbb{E}X_t)}$.

You can check that the covariance function is Hermitian: $B(s,t) = \overline{B(t,s)}$. It is also positive semidefinite, where now we use the complex definition: B is positive semidefinite if $\sum_{i,j} B(t_i,t_j)z_i\overline{z_j} \geq 0$ for all finite time slices $\{t_i\}_{i=1}^k$ and all complex-valued vectors $z = (z_1,\ldots,z_k)^T$.

5.2 Gaussian processes

An important class of processes are Gaussian processes. These arise in a number of applications because they are easy to simulate and because their fdds can be completely characterized. Also, the Central Limit Theorem suggests that they should arise from a superposition of random, uncorrelated effects.

Definition. A random vector $X = (X_1,\ldots,X_n)^T$ is a *multivariate Gaussian* if its probability density function has the form

$$f(x_1,\ldots,x_n) = \frac{1}{\sqrt{(2\pi)^n|\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}, \tag{1}$$

where $\Sigma \in \mathbb{R}^{n\times n}$ is a positive definite symmetric matrix, equal to the covariance matrix of the random vector, and $\mu \in \mathbb{R}^n$ is the mean. We write $X \sim N(\mu,\Sigma)$.

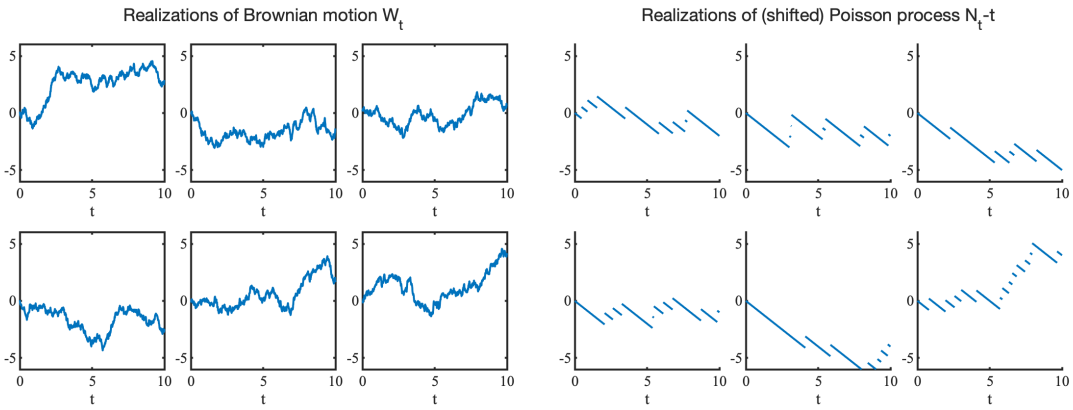
A few facts about multivariate Gaussians that may be useful are listed in the Appendix, Section 5.7.

Definition. $X = (X_t)_{t\in T}$ is a *Gaussian process* if all its fdds are Gaussian, i.e. if $(X_{t_1},\ldots,X_{t_k})$ is a multivariate Gaussian for any $(t_1,\ldots,t_k) \in T^k$.

To construct a Gaussian process we must provide two things: its mean $m(t)$, and its covariance function, $B(s,t)$. These two objects completely determine all the finite-dimensional distributions of the process. This is a powerful statement, since means and covariances are readily measurable.

Example 5.4 A *Brownian motion* or *Wiener process* is a continuous Gaussian process $W = (W_t)_{t\geq 0}$ with mean $m(t) = 0$ and covariance $B(s,t) = \min(s,t)$ for $s,t \geq 0$, and such that $W_0 = 0$. (We'll see other definitions later in the course.)

Notice that a Brownian motion $W = (W_t)_{t\geq 0}$ has the same covariance as a Poisson process with $\lambda = 1$. If we define a process $Y = (Y_t)_{t\geq 0}$ by $Y_t = N_t - t$, where N_t is a Poisson process with rate $\lambda = 1$, then Y, W both have mean $m(t) = 0$ and covariance function $B(s,t) = \min(s,t)$. However, these processes are clearly not the same; for example, the shifted Poisson process is piecewise linear, it does not have Gaussian fdds, and it is not continuous.



***Exercise* 5.1.** Show, from the definition above, that the Wiener process has stationary independent increments, i.e.

(a) the distribution of $W_t - W_s$ depends only on $t-s$;
(b) the variables $W_{t_j} - W_{s_j}$, $1 \le j \le n$ are independent whenever the intervals $(s_j, t_j]$ are disjoint.

Algorithm to simulate a Gaussian process Here is one algorithm to simulate a Gaussian process with mean $m(t)$ and positive definite covariance $B(s,t)$. Suppose we want to generate realizations evaluated at a discrete set of points $t_1, t_2, \ldots, t_n$, to obtain a random vector $X = (X_{t_1}, \ldots, X_{t_n})$.

- Form the $n \times n$ covariance matrix $\Sigma = (B(t_i, t_j))_{i,j=1}^n$.
- Find a matrix A that is a square root of Σ:

$$\Sigma = AA^T .$$

  For example, A could be the Cholesky decomposition of Σ, which exists because Σ is positive definite. Or, we could construct it from the singular value decomposition $\Sigma = UDV^{-1}$ as $A = U\sqrt{D}V^{-1}$.

- Choose an i.i.d. sequence random variables $\xi_i \sim N(0,1)$ for $i = 1, \ldots, n$, and let $\xi = (\xi_1, \xi_2, \ldots, \xi_n)^T$. Let

$$X = A\xi + \vec{m}, \qquad \vec{m} = (m(t_1), \ldots, m(t_n))^T .$$

The random vector X is a multivariate Gaussian, since it is a linear transformation of a multivariate Gaussian, ξ. To see why X is a realization of the Gaussian process at the given points, we furthermore have to show it has the correct mean and covariance matrix. We calculate

$$\mathbb{E}X = A\mathbb{E}\xi + \vec{m} = \vec{m}, \qquad \mathbb{E}(X - \vec{m})(X - \vec{m})^T = \mathbb{E}A\xi\xi^T A^T = AIA^T = \Sigma .$$

While this algorithm works in theory, it isn't always numerically stable in practice, since the covariance matrix Σ can be ill-conditioned, so algorithms to compute its square root perform poorly. In Section 5.5 we will learn another method, based on the spectral decomposition of the process, that is more numerically stable.

5.3 Stationary processes

Another important class of processes are those for which the distribution is invariant with time. For example, one might assume this is the case for

- waves in the ocean
- fluctuations in annual mean temperature (prior to 1800...)
- the bond angles in a polymer that is in thermodynamic equilibrium
- any ergodic Markov chain that has been run for a long time
- turbulent velocities in the atmosphere
- the Earth's background magnetic field
- etc

It would not be the case for a process that still retains memory of its initial condition, for example, (such as a protein shortly after we observe it in a particular configuration), or a process that can wander far away without bound (such as a random walker on a line or infinite lattice).

5.3.1 Strongly stationary and weakly stationary processes

Definition. A stochastic process is *strongly stationary*, or *strictly stationary*, if all the fdds are invariant with shifts in time, i.e.

$$P_{t_1+h,t_2+h,\ldots,t_k+h} = P_{t_1,\ldots,t_k} \qquad \forall\,(t_1,\ldots,t_k), \quad h \in \mathbb{R}.$$

In words, the statistics of the process don't change with shifts of time.

While this is a natural definition of stationarity, it is usually to hard to work with in practice: it is impossible to verify (how can one test an infinite collection of higher-order statistics?), and furthermore, it may be too strong in many cases, when the information we are interested in depends only on the one-point and two-point statistics. A weaker but more commonly used notion of stationarity comes from considering the mean and covariance function.

Consider what happens to the mean and covariance function of a strongly stationary process when we shift the process in time.

- Mean: $\mathbb{E}X_t = \mathbb{E}X_{t+h}$, so $m(t) = m(t+h)$ for all h, so the mean must be constant.
- Covariance: $\mathbb{E}X_{s+h}X_{t+h} = \mathbb{E}X_sX_t$, so $B(s+h,t+h) = B(s,t)$ for all h. Therefore $B(s,t) = f(s-t)$ for some function f.

This motivates a weaker definition of stationarity:

Definition. A stochastic process is *weakly stationary*, *second-order stationary*, or *wide-sense stationary* if

$$m(t) = cst, \qquad B(s,t) = C(s-t).$$

The function $C(t) = \mathbb{E}X_{s+t}X_s$ for all s is called the covariance function.

A weakly stationary process is *not* strongly stationary in general. However, if the process is Gaussian, then the two notions are equivalent.

Lemma. *If $(X_t)_{t\geq 0}$ is Gaussian and it is weakly stationary, then it is strongly stationary.*

Proof. This follows because the mean and covariance completely characterize the fdds of a Gaussian process. □

Examples 5.5

1. *Markov chains.* Let X be a Markov chain (discrete or continuous) and let $X_0 \sim \pi$, where π is the stationary distribution. Then X is strongly stationary. (Exercise: check this, by checking the fdds.)
2. *Independent sequences.* Let $X = X_0, X_1, \ldots$ be a sequence of i.i.d. random variables, with mean 0 and variance σ^2. Then X is strongly stationary, with covariance function $C(n) = \sigma^2$ if $n = 0$, $C(n) = 0$ otherwise.
3. *Identical sequences.* Let Y be a random variable with mean 0 and variance σ^2, and let $X = X_0, X_1, \ldots$ be defined by $X_n = Y$ for all n. Then X is strongly stationary, and its covariance function is $C(n) = 1$ for all n.
4. Let A, B be uncorrelated (not necessarily independent) random variables with mean 0 and variance σ^2. Let $\lambda \in [0, 2\pi]$ and define a process $X = (X_t)_{t\in\mathbb{R}}$ by

$$X_t = A\cos(\lambda t) + B\sin(\lambda t)\,. \tag{2}$$

   Then $\mathbb{E}X_t = 0$, and X has covariance function

$$\begin{aligned} B(s, s+t) &= \mathbb{E}\left[(A\cos(\lambda s) + B\sin(\lambda s))\,(A\cos(\lambda(s+t)) + B\sin(\lambda(s+t))\right] \\ &= \mathbb{E}\left[A^2\cos(\lambda s)\cos(\lambda(s+t)) + B^2\sin(\lambda s)\sin(\lambda(s+t))\right] \\ &= \sigma^2\cos(\lambda t)\,. \end{aligned}$$

   This depends only on the separation t, so X is weakly stationary. In general, X is not strongly stationary unless there are extra conditions on A, B.

   Notice this process and its covariance function can also be written as

$$X_t = \mathrm{Re}\{(A - iB)e^{i\lambda t}\} = \frac{1}{2}\left((A - iB)e^{i\lambda t} + (A + iB)e^{-i\lambda t}\right), \qquad B(s,t) = \frac{\sigma^2}{2}\left(e^{i\lambda t} + e^{-i\lambda t}\right).$$

   This representation as a sum of complex exponentials is not an accident; we'll use it in a simulation algorithm in Section 5.5.

***Exercise* 5.2.** Show that if $(X_t)_{t\geq 0}$ is strongly stationary, and $Y_t = f(X_t)$, for some function $f(x)$, then $(Y_t)_{t\geq 0}$ is also strongly stationary. Can you make up an example to show the same statement does not hold in general if strongly stationary is replaced by weakly stationary?

5.3.2 Separation of variables for weakly stationary processes*

Recall Example 5.5(4), where we showed the process $X_t = \mathrm{Re}\{(A - iB)e^{i\lambda t}\}$ is weakly stationary when A, B are uncorrelated. This suggests that stationary processes could have a representation as a sum of complex exponentials. Let's ask what such a representation must look like.

Example 5.6 Consider a process of the form

$$X_t = \xi h(t)\,,$$

where $h(t)$ is a deterministic, complex-valued function of time, and ξ is a complex-valued random variable. What conditions on $\xi, h(t)$ make X_t stationary?

The mean is $m(t) = (\mathbb{E}\xi)h(t)$. This is only constant when either $h(t)$ is constant or $\mathbb{E}\xi = 0$. Let's suppose $h(t)$ is not constant, so we must have $\mathbb{E}\xi = 0$.

The covariance function is $B(s,t) = (\mathbb{E}\xi\overline{\xi})h(s)\overline{h(t)}$. We need $h(s)\overline{h(t)} = f(s-t)$ for some function f. Setting $s = t$ shows we need $|h(t)|^2 = f(0) = cst$. Therefore $h(t)$ has the form

$$h(t) = Ae^{i\phi(t)}$$

for some real number $A \neq 0$ and some real-valued function $\phi(t)$. The covariance function is now

$$B(s+t,s) = \mathbb{E}X_{s+t}\overline{X_s} = A^2\mathbb{E}|\xi|^2 e^{i(\phi(s+t)-\phi(s))}.$$

To determine ϕ, we require $\frac{dB(s+t,s)}{ds} = 0$, which implies

$$\frac{d}{ds}\log\left(A^2\mathbb{E}|\xi|^2 e^{i(\phi(s+t)-\phi(s))}\right) = 0 \quad \Leftrightarrow \quad \phi'(s) = \phi'(s+t)\,.$$

Since this holds for all t we must have $\phi'(t) = const$, so $\phi(t) = \alpha t + \beta$ for some real-valued numbers α, β. Re-organizing constants and absorbing some into the definition of ξ shows that

$$X_t = \xi e^{i\lambda t},$$

where $\lambda \in \mathbb{R}$ and ξ is a complex-valued random variable with $\mathbb{E}\xi = 0$.

The example above can be seen as a form of separation of variables. A stochastic process depends on two variables, the time parameter t and the element ω in a probability space Ω. The representation above as $X(t,\omega) = \xi(\omega)h(t)$ is like a particular solution that one would look for using separation of variables. Therefore, if we want to generalize the example, we could consider a sum of functions that are separated in this way.

Example 5.7 Consider a sum of two exponentials, as

$$X_t = \xi_1 e^{i\lambda_1 t} + \xi_2 e^{i\lambda_2 t},$$

where ξ_1, ξ_2 are complex random variables, $\lambda_1 \neq \lambda_2$, and both $\lambda_1, \lambda_2 \neq 0$. When is X_t stationary?

The mean is

$$\mathbb{E}X_t = \mathbb{E}\xi_1 e^{i\lambda_1 t} + \mathbb{E}\xi_2 e^{i\lambda_2 t},$$

which is independent of t only if $\mathbb{E}\xi_1 = \mathbb{E}\xi_2 = 0$, since the functions $e^{i\lambda_1 t}$, $e^{i\lambda_2 t}$ are linearly independent.

The covariance is

$$B(s,t) = (\mathbb{E}|\xi_1|^2)e^{i\lambda_1(s-t)} + (\mathbb{E}|\xi_2|^2)e^{i\lambda_2(s-t)} + (\mathbb{E}\xi_1\overline{\xi_2})e^{i(\lambda_1 s - \lambda_2 t)} + (\mathbb{E}\overline{\xi_1}\xi_2)e^{i(\lambda_2 s - \lambda_1 t)}\,.$$

There are two ways the covariance can be a function of only $s-t$. One way is if $\mathbb{E}\xi_1\overline{\xi_2}=0$, so that each of ξ_1,ξ_2 are mean-zero, uncorrelated random variables. In this case the covariance function is

$$C(t) = b_1 e^{i\lambda_1 t} + b_2 e^{i\lambda_2 t}\,, \qquad b_j = \mathbb{E}|\xi_j|^2\,.$$

The other way is if $\lambda_2 = -\lambda_1$, and if additionally $\mathrm{Re}\{\mathbb{E}\xi_1\overline{\xi_2}\}=0$. A special case occurs when

$$\xi_1 = \overline{\xi_2} = A - iB \qquad \text{with} \quad \mathbb{E}A^2 = \mathbb{E}B^2.$$

That is, the variance of the real and imaginary parts of ξ_1 are equal, and ξ_2 is the complex conjugate of ξ_1. In this case both X_t and $C(t)$ are real-value and can be expressed as

$$X_t = 2A\cos(\lambda t) + 2B\sin(\lambda t), \qquad C(t) = 2\mathbb{E}A^2\cos(\lambda t).$$

***Exercise* 5.3.** Generalize the previous example: consider a more general superposition of frequencies, as

$$X_t = \sum_{j=1}^{\infty} \xi_j e^{i\lambda_j t}\,. \tag{3}$$

Derive conditions on the ξ_j and the λ_j for $X=(X_t)_{t\in\mathbb{R}}$ to be a weakly stationary process. Derive further conditions to ensure that X is real-valued.

In Section 5.5 we'll show how to use the representation (3) to generate realizations of a stationary Gaussian process, and in Section 5.8 we'll show how this representation generalizes to covariance functions formed from a continuous set of frequencies.

5.4 Properties and spectral decomposition of covariance functions of weakly stationary processes

The covariance function is a powerful tool to analyze properties of weakly stationary processes. We'll look at some of its properties, and then we'll introduce its spectral decomposition, which will later lead to an efficient algorithm to simulate stationary Gaussian processes.

(i) The variance of the process is $C(0) = \mathbb{E}X_t^2$.

(ii) $|C(t)| \le C(0)$ for all $t\in\mathbb{R}$.

To see why, calculate:

$$|C(t)| = \underbrace{|\mathbb{E}X_{t+s}X_s| \le (\mathbb{E}X_{t+s}^2\mathbb{E}X_s^2)^{1/2}}_{\text{Cauchy-Schwartz}} = C(0).$$

Note there is equality iff $X_{t+s}=X_s$ for some t.

(iii) $C(t)=C(-t)$. This follows by definition.

(iv) $C(t)$ is a *positive semidefinite function*: the matrix $(C(t_j-t_i))_{i,j=1}^k$ is positive semi-definite for all finite slices $\{t_i\}_{i=1}^k$. You can show this yourself as an exercise, or see Pavliotis (2014), p.7.

Remark. As for non-stationary processes, an equivalent definition of positive semidefiniteness comes from considering the operator $\mathscr{K}$ on $L^2(\mathbb{R})$ defined by $\mathscr{K}f = \int C(s-t)f(t)dt$. Then C is positive semidefinite if $\mathscr{K}$ is positive semidefinite with respect to the L^2 inner product.

From the covariance function, we can say a lot about the smoothness of the process itself.

Definition. (*) A stochastic process $X = (X_t)_{t\geq 0}$ is *continuous in the L^2 sense* if

$$\lim_{h\to 0} \mathbb{E}|X_{t+h} - X_t|^2 = 0\,. \tag{4}$$

Lemma. *(*) If the covariance function $C(t)$ of a weakly stationary stochastic process X_t is continuous at $t = 0$, then $C(t)$ is uniformly continuous for all $t \in \mathbb{R}$. Furthermore, $C(t)$ is continuous at 0 if and only if X_t is continuous in the L^2 sense.*

Proof. (Mostly from Pavliotis (2014), p.6):
To show that $C(t)$ is continuous at $t = 0$ implies it is uniformly continuous: Fix $t \in \mathbb{R}$, and suppose $\mathbb{E}X_t = 0$. Then

$$\begin{aligned}
|C(t+h) - C(t)|^2 &= |\mathbb{E}X_{t+h}X_0 - \mathbb{E}X_tX_0|^2 = \mathbb{E}|(X_{t+h} - X_t)X_0)|^2 \\
&\leq \mathbb{E}X_0^2\mathbb{E}(X_{t+h} - X_t)^2 \qquad \text{(Cauchy-Schwartz)} \\
&= C(0)\left(\mathbb{E}X_{t+h}^2 + \mathbb{E}X_t^2 - 2\mathbb{E}(X_tX_{t+h})\right) \\
&= 2C(0)(C(0) - C(h)) \;\to 0 \qquad \text{as } h \to 0
\end{aligned}$$

Now, Suppose $C(t)$ is continuous. The above showed that

$$\mathbb{E}|X_{t+h} - X_t|^2 = 2(C(0) - C(h)), \tag{5}$$

which converges to 0 as $h \to 0$. Conversely, if X_t is L^2-continuous, then the above implies $\lim_{h\to 0} C(h) = C(0)$. □

Sometimes we need to make up covariance functions, e.g. for testing code, or building models. How do we know if a given function is positive semidefinite? A natural method comes from looking at the function in Fourier space.

Bochner's Theorem. *A continuous function $C(t)$, $t \in \mathbb{R}$ is positive semidefinite if and only if there exists a non-decreasing, right continuous, bounded real function $F(\lambda)$, such that*

$$C(t) = \int_{-\infty}^{\infty} e^{i\lambda t} dF(\lambda)\;. \tag{6}$$

The integral in (6) is a *Riemann-Stieltjes integral*.[1] The function $F(x)$ is called the *spectral distribution function*.

[1] The Riemann-Stieltjes integral of a function $f(x)$ with respect to another (real-valued) function $g(x)$ over an interval $[a,b]$ is defined by

$$\int_a^b f(x)dg(x) = \lim_{\max_i(x_{i+1}-x_i)\to 0} \sum_{i=0}^{n} f(x_i^*)[g(x_{i+1}) - g(x_i)]\;,$$

where $a = x_0 < x_1 < \cdots < x_n = b$ is a partition of $[a,b]$, and $x_i^* \in [x_i, x_{i+1}]$. A sufficient, though not necessary, condition for this limit to exist is that f be continuous and g be of bounded variation. If $g(x)$ is differentiable and bounded, then the Riemann-Stieltjes integral equals the Riemann integral $\int_a^b f(x)g'(x)dx$. See Lecture 1 for more details.

Remark. This theorem was discovered independently by Khinchin slightly after its publication by Bochner, so it is sometimes called the *Bochner-Khinchin theorem*. It is usually stated for characteristic functions of a real-valued random variable X, defined by $\phi(t) = \mathbb{E}e^{itX} = \int e^{itx} dG(x)$, where $G(x)$ is the cumulative distribution function of X. We can arbitrarily choose $G(-\infty) = 0$, so the only difference between F above and a cdf G is that $F(\infty) = const$, while $G(\infty) = 1$. This link leads to another interpretation of (6), due to Grimmett and Stirzaker (2001) (p.382). If Λ is a random variable with cdf $F(\lambda)$, then $g_\Lambda(t) = e^{it\Lambda}$ is a pure oscillation with a random frequency, and $C(t) = \mathbb{E}g_\Lambda(t)$ is the average value of this pure oscillation (up to a normalizing constant).

Proof. For the "if" part, assume $C(t)$ has representation (6). We show it is positive semidefinite. Let $\{t_i\}_{i=1}^k$ be a finite time slice, and let $\{z_i\}_{i=1}^k$ be a collection of complex numbers. We have

$$\begin{aligned}\sum_{j,k} z_j \overline{z_k} C(t_j - t_k) &= \sum_{j,k} z_j \overline{z_k} \int_{-\infty}^{\infty} e^{i\lambda t_j} e^{-i\lambda t_k} dF(\lambda) \\ &= \int_{-\infty}^{\infty} \sum_{j,k} z_j e^{i\lambda t_j} \overline{z_k e^{i\lambda t_k}} dF(\lambda) \\ &= \int_{-\infty}^{\infty} \left| \sum_j z_j e^{i\lambda t_j} \right|^2 dF(\lambda) \geq 0.\end{aligned}$$

The "only if" part is more work. See Lindgren (2013), p. 75-78. See also Grimmett and Stirzaker (2001) (p.381 and p.182), for a proof based on similar results for characteristic functions. □

Remark. The partial proof above shows why the Fourier transform of a positive semidefinite function must be nonnegative – if there are frequencies λ where $F(\lambda)$ decreases, then by considering a function whose Fourier transform has support only near these frequencies, we obtain an inner product above that is negative.

In applications we usually only need two special cases of Bochner's theorem.

1. $F(\lambda)$ is absolutely continuous with respect to the Lebesgue measure. Then

$$dF(\lambda) = f(\lambda) d\lambda \qquad \Leftrightarrow \qquad f(\lambda) = F'(\lambda) \quad \text{almost everywhere.}$$

The function $f(\lambda)$ is called the *spectral density function*. Then, Bochner's theorem says that $f(\lambda)$ is the Fourier transform of $C(t)$:

$$C(t) = \int_{-\infty}^{\infty} f(\lambda) e^{i\lambda t} d\lambda \ , \qquad f(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} C(t) e^{-i\lambda t} dt \ . \tag{7}$$

Furthermore, since $F(\lambda)$ is non-decreasing, we must have $f(\lambda) \geq 0$. The above representation will hold when $C(t) \in L^1(\mathbb{R})$ Lindgren (2013), Theorem 3.4 p.80).

This is an important result – *the Fourier transform of a covariance function is always nonnegative* (provided the Fourier transform exists). Conversely, *given a nonnegative function, its inverse Fourier transform is a positive semidefinite function*, hence it can be the covariance function for a weakly stationary process.

2. $F(x)$ is piecewise constant, with jumps of size $\{b_i\}_{i=1}^{\infty}$ at a countable set of points $\{\lambda_i\}_{i=1}^{\infty}$. Then

$$F(\lambda) = \sum_{j:\lambda_j \le \lambda} b_j \quad \Leftrightarrow \quad F'(\lambda) = \sum_j b_j \delta(\lambda - \lambda_j) \ .$$

Because $F(x)$ is nondecreasing, we must have $b_i > 0 \ \forall i$. The covariance function has the form

$$C(t) = \sum_{j=1}^{\infty} b_j e^{i\lambda_j t} \, . \tag{8}$$

We obtain a representation of $C(t)$ that looks like a discrete inverse Fourier transform, although the frequencies λ_i are not restricted to be equally spaced. Thus, although $C(t) \notin L^1(\mathbb{R})$ since $C(t)$ oscillates forever, and it may not be periodic over any interval, we still obtain a representation of the function in spectral space.

The spectral distribution function is an important quantity in applications. It tells us which frequencies are represented in the covariance function, and hence also in the process X_t.

Definition. The *spectrum* of a process is the set of real numbers λ for which $F(\lambda+\varepsilon) - F(\lambda-\varepsilon) > 0$ for all $\varepsilon > 0$.

For the two cases above we have

1. If $F(\lambda)$ is absolutely continuous, then from (7) the spectrum is $\{\lambda : f(\lambda) > 0\}$, the support of the spectral density function $f(\lambda)$.
2. If $F(\lambda)$ is piecewise constant, then from (8) its spectrum is the set $\{\lambda_i\}_{i=1}^{\infty}$.

A process's spectrum could also have a mixture of discrete and continuous components (or "singular" components, which are of interest theoretically but typically don't arise in applications).

Example 5.8 Consider a process with covariance function

$$C(t) = Ae^{-\alpha|t|},$$

where $A, \alpha > 0$ are parameters. Its Fourier transform is (Pavliotis (2014), p.8)

$$f(\lambda) = \frac{A}{\pi} \frac{\alpha}{\lambda^2 + \alpha^2}.$$

Since $f(\lambda) > 0$, we know that $C(t)$ is positive definite. A Gaussian with this covariance function is called a *stationary Ornstein-Uhlenbeck process*. This is the only example of a stationary, Gaussian, Markov process, and it is used extremely frequently in modelling.

Example 5.9 We can use Bochner's theorem to make up covariance functions, e.g. to test code or theory. It is not easy to make up a covariance function out of the blue, because it is hard to quickly tell whether a function is positive semidefinite. However, it is easy to make up integrable functions such that $f(\lambda) \ge 0$, and since their Fourier transforms are positive semidefinite they can be used as covariance functions. For example, all of these are spectral densities of some covariance function:

$$f(\lambda) = 1_{1\le\lambda\le 2}(\lambda), \quad f(\lambda) = (1-\lambda^2)1_{|\lambda|\le 1}(\lambda), \quad f(\lambda) = Ae^{-B\lambda^2/2}, \quad f(\lambda) = Ae^{-B\lambda^2/2}1_{|\lambda|\le 5} + 1_{10\le\lambda\le 12}.$$

When the desired covariance function is real, the spectral density must be symmetric in λ.

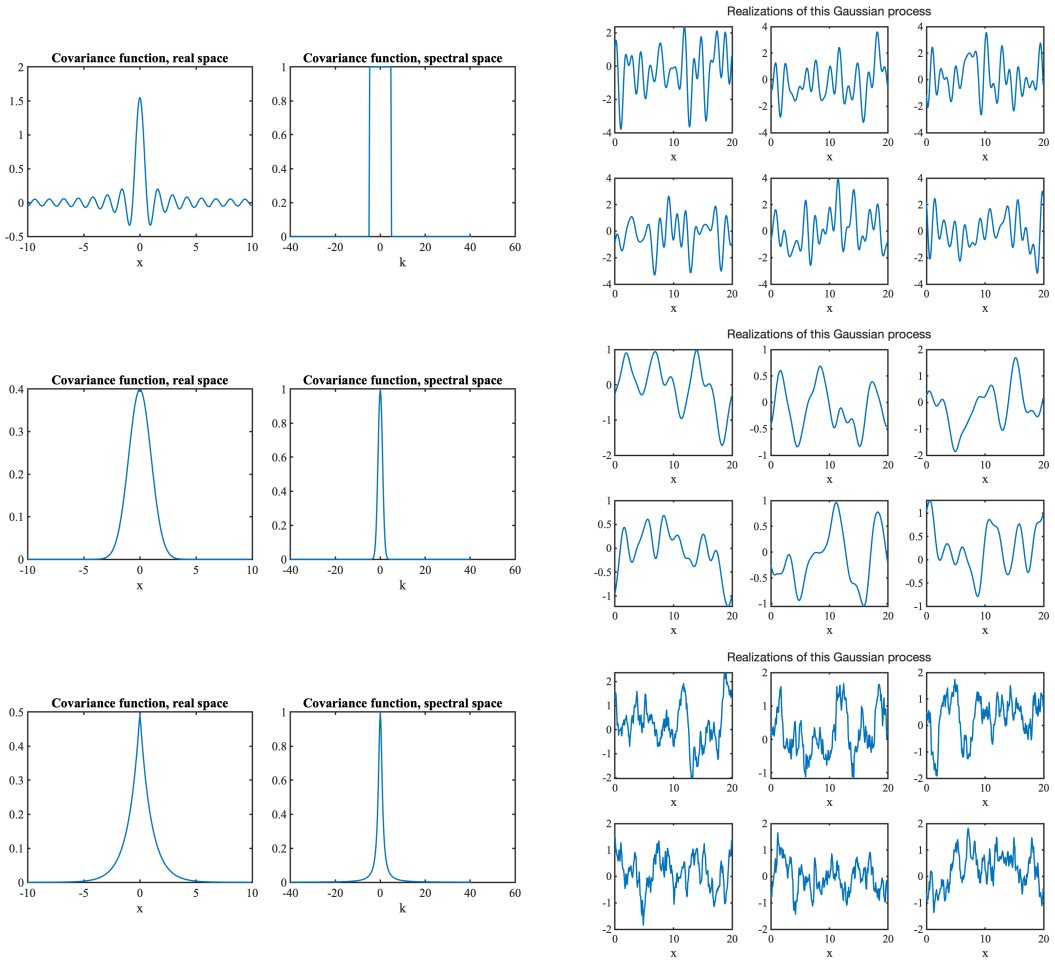


Figure 1: Some example covariance functions and realizations of stationary Gaussian processes with these covariance functions. Notice the similarity between the smoothness of $C(t)$ near 0 and the smoothness of the realizations. Notice also that frequencies that are present in the spectral decomposition of $C(t)$, also appear strong in the realizations of each process.

Example 5.10 Turbulent fluids are frequently characterized by their energy spectrum. The spectrum can be calculated by looking at one component of the velocity field, computing its covariance function, and finding the Fourier transform of the covariance function, which gives the spectral density.

Here is an example that shows a typical velocity field, and schematic of a typical spectrum, for turbulence in the atmosphere:[2]

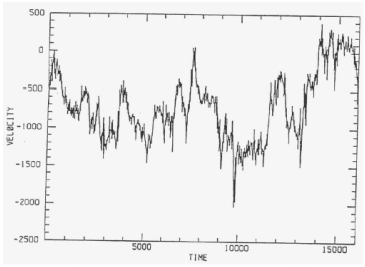


Figure 7.2: Turbulent velocities within the Earth's atmosphere, measured 18 m above the ground. Time and velocity are in arbitrary units.

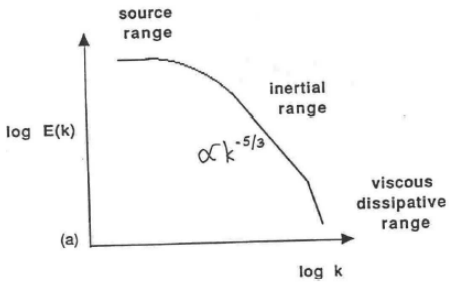
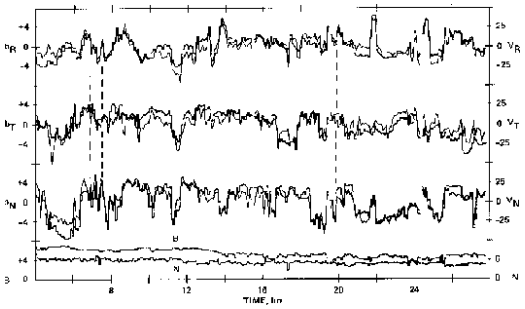
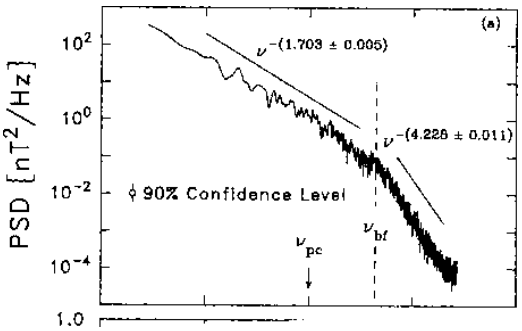


Figure 7.3: The Kolmogorov spectrum for incompressible turbulence.

Here is another example, that shows measurements of the three components of magnetic field and corresponding velocity field in the sun (left), and a typical computed spectrum for magnetic field turbulence (right):[3]





5.5 Simulating stationary Gaussian processes

Stationary processes have a representation in spectral space, which leads to an efficient, stable method to simulate realizations of them.

Example 5.7 and its generalization in Exercise 5.3 showed that given a collection of complex-valued random variables $\{\xi_i\}_{i=1}^{\infty}$ satisfying $\mathbb{E}\xi_i = 0$ and $\mathbb{E}\xi_i\overline{\xi_j} = 0$ for all $j \neq k$, we can construct a real or complex-valued stationary process $X = (X_t)_{t\in\mathbb{R}}$ and its covariance function $C(t)$ as

$$X_t = \sum_{j=1}^{\infty} \xi_j e^{i\lambda_j t}\,, \qquad C(t) = \sum_{j=1}^{\infty} b(\lambda_j) e^{i\lambda_j t} \quad \text{where} \quad b(\lambda_j) = \mathbb{E}|\xi_j|^2. \tag{9}$$

[2] from Barbara Ryden's notes on "Basic Turbulence", see http://www.astronomy.ohio-state.edu/ ryden/ast825/ch7.pdf

[3] from http://www.physics.usyd.edu.au/ cairns/teaching/lecture12/node2.html

In fact, for a process with a discrete spectrum, there is no other way to represent the covariance function: Bochner's theorem (see (8)) tells us that all such covariance functions have this form. Interestingly, the representation of X shows that a stationary stochastic process can be thought of as one whose Fourier coefficients are random variables.

We can generate realizations of a stationary process with a discrete spectrum, by following the recipe in (9):

- Generate uncorrelated complex-valued random variables ξ_i with mean zero, variance b_i;
- Set $X_t = \sum_{j=1}^n \xi_j e^{i\lambda_j t}$.

Making this algorithm useful in practice requires dealing with a few additional considerations: how to choose the distribution of the ξ_i, how to generate a real-valued process, and how to approximate a process with a continuous spectrum, as one with a discrete spectrum.

Clearly, the choice of distribution for the ξ_i lead to different types of processes. X will be Gaussian, if and only if the ξ_i are Gaussian. In this case, if the ξ_i are uncorrelated, they must also be independent.

A real-valued process can be ensured in one of two ways. One way is to impose a condition on the spectral coefficients corresponding to symmetric frequencies. That is, if λ_j is in the spectrum of X, with spectral coefficient ξ_j, then $-\lambda_j$ is also in the spectrum, with spectral coefficient $\overline{\xi_j}$. A perhaps simpler way is to first ensure the covariance function is real, by ensuring that $b(\lambda) = b(-\lambda)$. Then, one can generate the process as in (9) but with energies twice the desired ones, , and finally take its real part.

Finally, a process with a continuous spectrum, can be turned into one with a discrete spectrum by considering the process on a long but finite interval, at a set of equally-spaced grid points, and with periodic boundary conditions.

We obtain a method for simulating a real-valued, stationary Gaussian random process.

Algorithm to simulate a stationary Gaussian process Suppose you wish to simulate a real-valued, stationary, Gaussian process with mean zero and covariance function $C(t)$. We will generate realizations of the process at a set of N equally-spaced points $\{t_j\}_{j=0}^{N-1}$ on an interval of length L. The grid spacing is $\Delta t = \frac{L}{N}$ and the corresponding grid spacing in spectral space is $\Delta k = \frac{2\pi}{L}$. It is convenient to choose the time and spectral grid points as[4]

$$t_j = j\Delta t - L/2, \quad j = 0,1,\ldots,(N-1), \qquad k_n = n\Delta k - \Delta k\frac{N}{2}, \quad n = 0,1,\ldots,(N-1).$$

Compute the discrete Fourier transform $\hat{C}(k)$ of $C(t)$ and its inverse as

$$\hat{C}(k_n) = \Delta t \sum_{j=0}^{N-1} C(t_j) e^{-it_j k_n}, \qquad C(t_j) = \frac{\Delta k}{2\pi} \sum_{n=0}^{N-1} \hat{C}(k_n) e^{it_j k_n}, \tag{10}$$

Construct X as

$$X(t_j) = \mathrm{Re}\left\{ \sum_{n=0}^{N-1} \xi_n e^{it_j k_n} \right\}, \qquad \text{where} \quad \xi_n = \sqrt{\hat{C}(k_n)\frac{\Delta k}{2\pi}}\,(A_n + iB_n), \tag{11}$$

[4] We shift space by $L/2$, to avoid numerical problems when $C(t)$ is not periodic on the chosen interval. Similarly we shift frequencies by $\Delta k\frac{N}{2}$, because then it is easier to make up functions $\hat{C}(k)$ which are symmetric in k, as is required if $C(t) \in \mathbb{R}$.

and $A_n, B_n \sim N(0,1)$ are independent standard normals. Here $\mathrm{Re}\{\cdot\}$ means take the real part. (You could also take the imaginary part, and obtain another independent realization of X.)

***Exercise* 5.4.** Verify directly that the process X constructed in (11) above has the correct covariance.

This algorithm is numerically stable, and it is efficient, particularly if you use the fast Fourier transform to evaluate (11). It should be preferred to the Cholesky decomposition method from Section sec:gaussian, when it applies.

Remark. We have introduced this method for stationary Gaussian processes, but it can be extended to non-stationary Gaussian processes, by replacing the complex exponential $e^{i\lambda_j t}$ by functions $e_j(t)$ which are eigenfunctions of the operator defined by $\mathscr{K} f := \int B(s,t) f(t) dt$ (see remark toward the end of Section 5.1.2). In fact, you can check that complex exponentials are the eigenfunctions of the operator $\mathscr{C} f := \int C(s-t) f(t) dt$, which is why they appear in our representations.

Remark. Notice that $\xi_i \sim \sqrt{\Delta k}$. One might expect that $\xi_i \sim \Delta k$, so the sum in (11) approaches an integral as the grid spacing $\Delta x \to 0$. In fact, because the ξ_i are uncorrelated mean-zero random variables, they must be scaled differently in order to obtain a finite, non-zero function in the limit – we'll return to this point later in the course when we construct a stochastic integral. See Section 5.8 for a brief discussion of the spectral representation of a stationary process with a continuous spectrum.

5.6 Ergodic properties of weakly stationary processes*

If a process is stationary, then we may hope to be able to extract all the statistical information we need from a single trajectory of the process. This would be useful, for example, if we only have access to one trajectory: perhaps it is the temperature fluctuations in New York, the velocity field at a single location in the ocean, etc; or a single run of a molecular dynamics or Monte Carlo simulation, where is can be hard to obtain many independent realizations since it can have a long burn-in time before one can start collecting statistics. However, it is not obvious that the time average of a function of of a single trajectory should equal the average over an ensemble of trajectories, and in fact this is only true when the covariance function decays quickly enough.

Ergodic theorems in general relate time averages to ensemble averages. We will look at one version of the Ergodic theorem for stationary processes, which gives conditions under which the time average of a stationary process converges in *mean-square* to its mean value. Recall that a sequence of random variables X_t converges in mean-square to another random variable Y, written $X_t \xrightarrow{m.s.} Y$, if $\lim_{t\to\infty} \mathbb{E}|X_t - Y|^2 = 0$. This is not the strongest possible Ergodic theorem, but it is easy to prove.

Theorem (Ergodic Theorem for stationary processes)**.** *(See Pavliotis (2014)) Let $(X_t)_{t\geq 0}$ be a weakly stationary process with mean μ and covariance $C(t)$, and suppose that $C(t) \in L^1(\mathbb{R})$. Then*

$$\lim_{T\to\infty} \mathbb{E}\left| \frac{1}{T}\int_0^T X_s\, ds - \mu \right|^2 = 0. \tag{12}$$

This theorem generalizes trivially to weakly stationary sequences $X_0, X_1, X_2, \ldots$.

The ergodic theorem shows that if the correlation function decays quickly enough, then we may compute an expectation of a random variable, simply by time-averaging a *single* trajectory. So, the following are equivalent: either we can simulate or observe a system for a very long time and compute the average, or we can average over many short independent simulations or observations.

Remark. Some stronger and weaker statements are also possible. If X_t is strongly stationary then the convergence in the Ergodic Theorem is almost surely – see Grimmett and Stirzaker (2001), section 9.5. For a weaker version of the theorem, see Grimmett and Stirzaker (2001), section 9.5, which shows that, with no restrictions on the covariance function, there exists a random variable Y such that $\frac{1}{T}\int_0^T X_t\,dt \xrightarrow{m.s.} Y$.

Proof of Ergodic theorem. Calculate:

$$\begin{aligned}
\mathbb{E}\left|\frac{1}{T}\int_0^T X_s\,ds-\mu\right|^2 &= \frac{1}{T^2}\mathbb{E}\left|\int_0^T (X_s-\mu)\,ds\right|^2 \\
&= \frac{1}{T^2}\int_0^T\int_0^T \mathbb{E}(X_t-\mu)(X_s-\mu)dtds \qquad \text{Fubini's Theorem to interchange } \mathbb{E},\int \\
&= \frac{1}{T^2}\int_0^T\int_0^T C(t-s)dtds\,.
\end{aligned}$$

Now we change variables to $u=t-s$, $v=t+s$. The domain of integration in (u,v) is $[-T,T]\times[|u|,2T-|u|]$, and the Jacobian is $\frac{\partial(t,s)}{\partial(u,v)}=\frac{1}{2}$. The integral becomes

$$\int_0^T\int_0^T C(t-s)dtds = \int_{-T}^T\int_{|u|}^{2T-|u|}\frac{1}{2}C(u)dvdu = \int_{-T}^T (T-|u|)C(u)du = 2\int_0^T (T-u)C(u)du,$$

where the last step follows because $C(u)$ is symmetric. Substituting into the above calculations, we have

$$\begin{aligned}
\mathbb{E}\left|\frac{1}{T}\int_0^T X_s\,ds-\mu\right|^2 &= \frac{2}{T^2}\int_0^T (T-u)C(u)du \\
&\le \frac{2}{T}\int_0^T \left|\left(1-\frac{u}{T}\right)C(u)\right|du \\
&\le \frac{2}{T}\int_0^\infty C(u)du \qquad\qquad \text{Dominated Convergence Theorem}
\end{aligned}$$

which $\to 0$, since $C\in L^1((0,\infty))$. □

Remark. We see from the proof that we can actually weaken the condition that $C\in L^1$ – it is enough to know that $\lim_{T\to\infty}\frac{1}{T}\int_0^T C(u)(1-\frac{u}{T})du=0$.

***Exercise* 5.5.** Show that the Ergodic Theorem implies the Law of Large Numbers for i.i.d. random variables.

***Exercise* 5.6.** Let Y be a random variable and let $X_t=Y$ for all t. Show that the statement of the Ergodic theorem fails to hold, and identify the condition in the statement of the Ergodic theorem that is not satisfied by this process.

From the Ergodic theorem above, we may obtain an ergodic theorem for Markov chains.

Example 5.11 (Ergodic theorem for Markov chains) Let $X = (X_n)_{n\in\mathbb{N}}$ be a regular Markov chain with finite state space S, and suppose $X_0 \sim \pi$, where π is the unique stationary distribution of the chain. Suppose you run the chain and want to empirically determine π_k for some state k. You can do this as follows: define a collection of indicator functions $I = \{I_n : n \geq 0\}$ by

$$I_n = \begin{cases} 1 & \text{if } X_n = k \\ 0 & \text{otherwise.} \end{cases}$$

Then the partial sum $S_n = \sum_{j=0}^{n-1} I_j$ is the number of visits to the state k before the nth jump.

You can show that X is strongly stationary.[5] From this, you can show that the process $(I_n)_{n\in\mathbb{N}}$ is strongly stationary (see exercise 5.2.) The mean of $(I_n)_{n\in\mathbb{N}}$ is $\mathbb{E}I_0 = \pi_k$, and the covariance is $C(n) = \mathbb{E}I_0 I_n - \pi_k^2 = \pi_k P_{kk}^n - \pi_k^2$.

To use the Ergodic theorem above, we must check that $C(n) \in L^1$. We use the fact that $|P_{ij}^n - \pi_j| \leq C_0 e^{-\beta n}$ for some $C_0 > 0$ and n large enough, where $\beta = -\log|\lambda_2| > 0$ and λ_2 is the eigenvalue of P with the second-largest norm. Therefore

$$\sum_{m=0}^{\infty} |C(m)| = \pi_k \sum_{m=0}^{\infty} |\|P_{kk}^n - \pi_k^2\| \leq \pi_k \sum_{m=0}^{\infty} C_0 e^{-\beta m} = \frac{C_0 \pi_k}{1 - e^{-\beta}}.$$

Therefore by the Ergodic theorem,

$$\frac{1}{n} S_n = \frac{1}{n} \sum_{j=0}^{n-1} I_j \xrightarrow{m.s.} \pi_k.$$

Example 5.12 (Taylor dispersion) Consider a particle in a turbulent velocity field with position $x(t)$ and whose Lagrangian velocity is $u_L(t)$. For simplicity let's just focus on one coordinate, so $x(t), u_L(t) \in \mathbb{R}$. If the velocity field is "random" enough, then we expect the particle to be pushed and pulled in many different directions so over time, it tends to move away from its initial condition, or diffuse, in the same way as a random walk: it should behave like a Brownian motion over long enough time scales. Recall that for a Brownian motion the mean-square displacement is proportional to t. Inspired by this, we can define a particle's "effective" diffusion coefficient by considering its mean-square displacement, as

$$D \equiv \lim_{t\to\infty} \frac{1}{2} \frac{d}{dt} \mathbb{E}(x(t))^2 \quad = \lim_{t\to\infty} \frac{1}{2} \frac{d}{dt} \mathbb{E} \left(\int_0^t u_L(t) dt \right)^2. \tag{13}$$

Let's assume the Lagrangian velocity is a stationary stochastic process with mean zero and covariance func-

[5] Let's check:

$$\begin{aligned} P(X_{t_1+h} = x_1, \cdots, X_{t_n+h} = x_n) &= P(X_{t_1+h} = x_1) P(X_{t_2+h} = x_2 | X_{t_1+h} = x_1) \cdots P(X_{t_n+h} = x_n | X_{t_{n-1}+h} = x_{n-1}) \\ &= P(X_{t_1} = x_1) P(X_{t_2} = x_2 | X_{t_1} = x_1) \cdots P(X_{t_n} = x_n | X_{t_{n-1}} = x_{n-1}) \\ &= P(X_{t_1} = x_1, \cdots, X_{t_n} = x_n). \end{aligned}$$

tion $C(t) \in L^1$. Then, following the calculations in the proof of the ergodic theorem, we find

$$\begin{aligned}
\mathbb{E}\left(\int_0^t u_L(t)dt\right)^2 = \ldots = 2t\int_0^t \Big(1-\frac{u}{t}\Big)C(u)du && \text{see calcs in proof}\\
\approx 2t\int_0^t C(u)du && \text{for } t \text{ large}\\
\approx 2t\int_0^\infty C(u)du\,. &&
\end{aligned}$$

Therefore we find that

$$D = \int_0^\infty C(u)du\,. \tag{14}$$

Therefore the diffusion coefficient is simply the integral of the covariance function of the Lagrangian velocity. This is a useful result that is used in a wide variety of contexts; for example one can use it to calculate the effective diffusivity due to interacting waves in the ocean, or to calculate diffusion coefficients of a large particle in a sea of smaller particles, such as a grain of dust in a water bath. This formula was derived by Taylor in an early paper in 1930 (?)

References

Breiman, L. (1992). *Probability*. SIAM.

Chorin, A. and Hald, O. (2009). *Stochastic Tools in Mathematics and Science*. Springer, 2nd edition.

Grimmett, G. and Stirzaker, D. (2001). *Probability and Random Processes*. Oxford University Press.

Koralov, L. B. and Sinai, Y. G. (2010). *Theory of Probability and Random Processes*. Springer.

Lindgren, G. (2013). *Stationary Stochastic Processes: Theory and Applications*. CRC Press.

Pavliotis, G. A. (2014). *Stochastic Processes and Applications*. Springer.

Yaglom, A. M. (1962). *An Introduction to the Theory of Stationary Random Functions*. Dover.

5.7 Some facts about multivariate Gaussian random variables

Here are some facts about multivariate Gaussian random variables that may be useful.

(1) Let X be a multivariate Gaussian random vector, let $A \in \mathbb{R}^{m\times n}$ with $m \le n$ be a constant matrix, and let $Y = AX$. Then Y is also a multivariate Gaussian random vector, with $Y \sim N(\mu, A\Sigma A^T)$.

Proof. This is an exercise is multivariable calculus; Transform the density using the correct Jacobian. See e.g. Grimmett and Stirzaker (2001), Theorem 4.9.6 p.117. □

(2) The marginal distributions are Gaussian, with mean and covariance given by dropping the relevant rows and columns in the mean vector and covariance matrix.

Proof. Another exercise in calculus; integrate the joint density over a subset of variables, and show it has the correct form. □

(3) If (X,Y) is a multivariate Gaussian, and if X,Y are uncorrelated, then X,Y are independent.

Proof. This follows directly from the form of the probability density function. If X,Y are uncorrelated then the covariance matrix has the form $\Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$. We can substitute this form into (1) and find the density factors in a product of marginals, as

$$f(x,y) = f_x(x)f_y(y) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1^2)}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2^2)}{2\sigma_2^2}}.$$

Therefore X, Y are independent.[6] □

Note that uncorrelated $\Rightarrow$ independent is *only* true for jointly Gaussian random variables; for other distributions this is not true in general. This is one reason why calculations with Gaussians are significantly more tractable than with other kinds of distributions.

(4) The conditional distributions are Gaussian. For example, if $X = (X_1,\ldots,X_n)$ and $Y = (X_m,\ldots,X_n)$, then the random vector $Z = (X|Y)$, whose probabilities are calculated as $P(Z \in A) = P(X \in A|Y)$, is a multivariate Gaussian. The mean and covariance may also be calculated using calculus and algebra; the formulas are somewhat more complicated that for the marginals, involving Schur complements.

Note it is not true that if random variables X, Y are separately Gaussian, then the random vector (X,Y) is a multivariate Gaussian. For a counterexample, do this exercise.

***Exercise* 5.7.** Let $X \sim N(0,1)$, let $a > 0$, and let $Y = X$ if $|X| < a$, and $Y = -X$ if $|X| \geq a$. Show that $Y \sim N(0,1)$, but the random variable (X,Y) does not have a bivariate normal distribution. (Hint: calculate the covariance of X,Y.)

[6] Here is a general expression for the pdf of a bivariate Gaussian. If (X,Y) is a multivariate Gaussian and the correlation coefficient between X,Y is ρ, then the pdf has the form

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ -\frac{\frac{(x-\mu_1^2)}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}}{2(1-\rho^2)} \right\}.$$