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 an in-depth but accessible book; with more details and a more modern style than Yaglom (1962).

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

5.1 Setup

Here are some ideas we will need for what follows.

5.1.1 Finite dimensional distributions

Definition. The finite-dimensional distributions (fdds) of a stochastic process $X_t$ is the set of measures $P_{t_1,t_2,\ldots,t_k}$ given by

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

where $(t_1,\ldots,t_k)$ is any point in $T^k$, $k$ is any integer, and $F_i$ are measurable events in $\mathbb{R}$.

That is, we take a finite number of points $(t_1, t_2, \ldots, t_k)$ and consider the distribution of $(X_{t_1}, X_{t_2}, \ldots, X_{t_k})$.

Examples

1. A sequence of i.i.d. random variables. The parameter set is $T = \mathbb{Z}^+$, and the fdds are $P_{t_1,t_2,\ldots}(F_1 \times \cdots \times F_k) = \prod_{i=1}^k P(F_i)$. 
2. Markov chains. Let $X_t$ be a continuous-time, homogeneous Markov chain with generator $Q$ and initial probability distribution $\mu_0$, and let $0 = t_0 < t_1 < \cdots < t_k$. Then the fdd is given by

$$P(X_{t_0} \in F_0, X_{t_1} \in F_1, \ldots, X_{t_k} \in F_k) = \sum_{i_0 \in F_0} \mu_0(i_0) \sum_{i_1 \in F_1} P_{i_0 i_1}(t_1 - t_0) \cdots \sum_{i_{k-1} \in F_{k-1}} P_{i_{k-2} i_{k-1}}(t_k - t_{k-1}).$$

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

Given a set of finite-dimensional distributions, the Kolmogorov Extension Theorem says that, if the fdds satisfy some basic consistency criteria (they are invariant under permutations, and their marginal distributions are consistent), then there exists a stochastic process with the corresponding fdds. See Koralov and Sinai (2010), p. 174, or Grimmett and Stirzaker (2001), section 8.7, p. 372, or Terry’s Tao’s notes: https://terrytao.wordpress.com/2010/10/30/245a-notes-6-outer-measures-pre-measures-and-product-measures/#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, differentiable, bounded; what is the first passage time to a given set, what its maximum is at time $t$, etc. You need the values of the process at an uncountable number of points, to decide such questions.

Here is an example to illustrate some of the difficulties.

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

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

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, $Y$ is not; $\sup X = 0$, $\sup Y = 1$; $P(X_t = 0$ for all $t) = 1$, $P(Y_t = 0$ 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, from the French “continue à droite, limites à gauche.”) A theorem (see Grimmett and Stirzaker (2001), Theorem 8.7.6, p. 373, and also Breiman (1992), p. 300) gives conditions under which a stochastic process has a càdlàg version.

### 5.1.2 Covariance functions

A highly useful way to characterize properties of a stochastic process is its covariance function, which essentially characterizes the variance of the two-point fdds.

Recall that if we have a random vector $X = (X_1, \ldots, X_d)^T$, its covariance matrix is $\Sigma = \text{EX}X^T - (\text{EX})(\text{EX})^T$ is the matrix whose elements are the covariance of $X_i, X_j$, i.e. $\Sigma_{ij} = \text{EX}_i \text{EX}_j - (\text{EX}_i)(\text{EX}_j)$. This generalizes to random functions:
Definition. The covariance function of a real-valued stochastic process \((X_t)_{t \in T}\) is
\[
B(s,t) = \mathbb{E}[(X_s - \mathbb{E}X_s)(X_t - \mathbb{E}X_t)] = \mathbb{E}X_sX_t - m(s)m(t),
\]
where \(m(t) = \mathbb{E}X_t\) is the mean of the process.

Notice that \(C(t,t) = \mathbb{E}X_t^2 - (\mathbb{E}X_t)^2\) 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 the matrix \((B(t_i,t_j))_{i,j=1}^k\) is positive (semi)definite for all finite slices \(\{t_i\}_{i=1}^k\).

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

Proof. This will be an exercise on the homework.

Example. Let’s calculate the covariance function of a Poisson process \((N_t)_{t \geq 0}\) with parameter \(\lambda\). Assume WLOG that \(t \geq s\), and calculate:
\[
C(s,t) = \mathbb{E}N_sN_t - (\mathbb{E}N_s)(\mathbb{E}N_t) = \mathbb{E}(N_t - N_s)N_s + \mathbb{E}N_s^2 - \lambda^2ts
\]
\[
= \lambda^2(t-s)s + \lambda s + \lambda^2 s^2 - \lambda^2ts
\]
\[
= \lambda s.
\]

We used the fact that \(N_t - N_s, N_s\) are independent. You can repeat the calculation with \(s > t\), to find that \(C(s,t) = \lambda \min(s,t)\).

5.2 Gaussian processes

A very important class of processes are Gaussian processes. These arise in a number of applications, partly because they are tractable models that are possible to simulate and such that much is known analytically about their fdds. 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
\[
f(x_1, \ldots, x_n) = \frac{1}{\sqrt{(2\pi)^n|\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)},
\]
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 sometimes 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 iff all its fdds are Gaussian, i.e. if \((X_{t_1}, \ldots, X_{t_k})\) is a multivariate Gaussian for any \((t_1, \ldots, t_k) \in T^k\).
This means that if we are given functions \( m(t) \) and \( B(s,t) \), and \( B \) is positive definite, then we can construct a Gaussian process whose mean is \( \mathbb{E} X_t = m(t) \) and whose covariance is \( B(s,t) \), i.e. all the finite-dimensional distributions have covariance matrix \( (B(t_i, t_j))_{i,j=1}^k \).

This gives us a lot of flexibility in constructing Gaussian processes. It also means that, given only the mean and the covariance of the process, we know all the finite dimensional distributions. This is a powerful statement, since means and covariances are readily measurable. It is only true for Gaussian processes.

**Example.** 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 0 and covariance function \( \min(s,t) \). However, these are clearly not the same process; clearly the Poisson process does not have Gaussian fdds, and it is also not continuous.

**Exercise 5.1.** Show that the function \( B(s,t) = \min(s,t) \) for \( s,t \geq 0 \) is positive definite.

**Exercise 5.2.** 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 \( W_{t_j} - W_{t_j}, 1 \leq j \leq n \) are independent whenever the intervals \( (s_j,t_j) \) are disjoint.

**Simulating Gaussian processes** There is a straightforward algorithm for simulating realizations of a Gaussian process. WLOG let’s assume \( m(t) = 0 \) (otherwise we just add on the function \( m(t) \) after), and suppose we are given a positive definite function \( B(s,t) \). Suppose we want to generate realizations evaluated at a discrete set of points \( t_1,t_2,\ldots,t_n \). Proceed as follows.

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

  For example, \( A \) could be the Cholesky decomposition of \( \Sigma \), which exists, because \( \Sigma \) is positive definite. Or, we could construct it from the singular value decomposition \( \Sigma = U D V^{-1} \) as \( A = U \sqrt{D} V^{-1} \). (The Cholesky decomposition will in general be more efficient.)

- Choose a sequence of i.i.d. random variables \( \xi_i \sim N(0,1) \) for \( i = 1,\ldots,n \), and let \( \xi = (\xi_1, \xi_2, \ldots, \xi_n) \) (a column vector.) Define
  \[
  X = A \xi , \quad X = (X_{t_1}, \ldots, X_{t_n}) .
  \]

  Then \( X \) is a realization of the Gaussian process at the given points.

To see why, we just have to show it has the right covariance matrix. But

\[
\mathbb{E}XX^T = \mathbb{E}A\xi \xi^T A^T = AA^T = \Sigma .
\]

We can generate realizations of any Gaussian process using this method. However, it isn’t very stable numerically; the covariance matrix \( \Sigma \) can be very ill-conditioned, so algorithms to compute its square root
might fail. If the process we are interested in is stationary, there is a better method that we will learn about later in this lecture.

### 5.3 Stationary processes

Another important class of processes are those for which the distribution is invariant with time. For example,

- waves in the ocean
- fluctuations in annual mean temperature (prior to 1800...)
- the bond angles in a protein that is in thermodynamic equilibrium
- any 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.)

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

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

This definition is usually too hard to work with in practice:

- it is hard to check
- it may be too strong in many cases, when the information we are interested in depends only on the one-point and two-point fdds (mean and covariance)

A weaker definition is more commonly used in practice, based on 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: we have $\mathbb{E}X_t = \mathbb{E}X_{t+h}$, so $m(t) = m(t+h)$ for all $h$, so the mean must be constant.
- Covariance: we have $\mathbb{E}X_{t+h}X_s = \mathbb{E}X_sX_t$, 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 the weaker definition of stationarity:

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

$$m(t) \equiv m, \quad B(s,t) = C(s-t).$$

In other words, a process is weakly stationary if its mean and covariance function do not change with time. The covariance function function $C(t) = \mathbb{E}X_tX_0 = \mathbb{E}X_{t+t}X_t$ for all $s$.

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. \qed

Examples

1. Markov chains. Let $X$ be a Markov chain (discrete or continuous) and let $X_0 \sim \pi$, where $\pi$ 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 $\sigma^2$. 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 $\sigma^2$, 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 $\sigma^2$. Let $\omega \in [0, 2\pi]$ and define a process $X = (X_t)_{t \in \mathbb{R}}$ by

$$X_t = A \cos(\lambda t) + B \sin(\lambda t).$$

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

$$C(s,s+t) = \mathbb{E}X_sX_{s+t} = \mathbb{E}[(A \cos(\lambda s) + B \sin(\lambda s))(A \cos(\lambda (s+t)) + B \sin(\lambda (s+t)))]$$

$$= \mathbb{E}[A^2 \cos(\lambda s) \cos(\lambda (s+t)) + B^2 \sin(\lambda s) \sin(\lambda (s+t))]$$

$$= \sigma^2 \cos(\lambda t).$$

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 $A, B$.

From now on we will consider only weakly stationary processes.

Exercise 5.3. Show that if $(X_t)_{t \geq 0}$ is strongly stationary, then $(Y_t)_{t \geq 0}$ is too, where $Y_t = f(X_t)$, for any function $f(x)$.

5.4 Covariance functions of weakly stationary processes

The covariance function is a very useful way to analyze properties of weakly stationary processes. Let’s look at some of its properties.

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

- $C(0) \geq C(h)$ for all $h \neq 0$. The inequality is strict, except for a trivial process $X_t = Y$ for all $t$, where $Y$ is a random variable.

To see why, calculate: $C(t) = \mathbb{E}X_{t+s}X_s \leq (\mathbb{E}X_{t+s}^2)^{1/2}(\mathbb{E}X_s^2)^{1/2}$, by Cauchy-Schwartz. But the RHS is $C(0)$. Note there is equality iff $X_{t+s} = X_s$. 

6
• $C(t) = C(-t)$. This follows by definition.

• $C(t)$ 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.

From the covariance function, we can actually determine a fair amount 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

$$
\lim_{h \to 0} \mathbb{E}|X_{t+h} - X_t|^2 = 0 \tag{3}
$$

**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

$$
|C(t+h) - C(t)|^2 = |\mathbb{E}X_{t+h}X_0 - \mathbb{E}X_tX_0|^2 = \mathbb{E}|(X_{t+h} - X_t)X_0|^2 \\
\leq \mathbb{E}X_0^2(X_{t+h} - X_t)^2 \quad \text{(Cauchy-Schwartz)} \\
= C(0)(\mathbb{E}X_{t+h}^2 + \mathbb{E}X_t^2 - 2\mathbb{E}(X_{t+h}X_t)) \\
= 2C(0)(C(0) - C(h)) \to 0 \quad \text{as } h \to 0
$$

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

$$
\mathbb{E}|X_{t+h} - X_t|^2 = 2(C(0) - C(h)), \tag{4}
$$

which converges to 0 as $h \to 0$. Conversely, if $X_t$ is $L^2$-continuous, then the above implies $\lim_{h \to 0} C(h) = C(0)$. Note that (4) implies that $C(0) > C(h)$ (unless $X_{t+h} = X_t$ a.s.). Therefore the maximum of a covariance function is always at 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 more natural way to determine this is by looking at the function in Fourier space.

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

$$
C(t) = \int_{-\infty}^{\infty} e^{i\lambda t} dF(\lambda) \tag{5}
$$

The integral in (5) is a **Riemann-Stieltjes integral**. The function $F(x)$ is the **spectral distribution function**.

---

1The Riemann-Stieltjes integral of a function $f(x)$ with respect to another (real-valued) function $g(x)$ over an interval $[a, b]$ is defined...
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, calculated as $\Phi(t) = \mathbb{E}e^{itX} = \int e^{itx}dG(x)$, where $G(x)$ is the cumulative distribution function of a random variable $X$. We can arbitrarily choose $G(-\infty) = 0$, so the only difference between $F$ and a cdf $G$ is that $F(\infty) = \text{const}$, while $G(\infty) = 1$. This link leads to another interpretation of (5), due to Grimmett and Stirzaker (2001) (p.382). If $\Lambda$ is a random variable with cdf $F(\lambda)$, then $g_\Lambda(t) = e^{it\lambda}$ is a pure oscillation with a random frequency, and $C(t)$ is the average value of this pure oscillation (up to a normalizing constant.)

Proof. For the “if” part, we have

$$\sum_{j,k} z_j z_k C(t_j - t_k) = \sum_{j,k} z_j z_k \int_{-\infty}^\infty e^{i\lambda t_j} e^{-i\lambda t_k} dF(\lambda)$$

$$= \int_{-\infty}^\infty \sum_{j,k} z_j e^{i\lambda t_j} z_k e^{i\lambda t_k} dF(\lambda)$$

$$= \int_{-\infty}^\infty \sum_{j} z_j e^{i\lambda t_j}^2 dF(\lambda) \geq 0.$$

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.

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

1. $C(t) \in L^1(\mathbb{R})$.

   This implies that $F(\lambda)$ is absolutely continuous with respect to the Lebesgue measure (see e.g. Lindgren (2013), Theorem 3.4 p.80), so $dF(\lambda) = f(\lambda) d\lambda$, i.e. $f(\lambda) = F’(\lambda)$ 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 $C(t)$:

$$C(t) = \int_{-\infty}^\infty f(\lambda) e^{i\lambda t} d\lambda , \quad f(\lambda) = \frac{1}{2\pi} \int_{-\infty}^\infty C(t) e^{-i\lambda t} dt . \quad (6)$$

Furthermore, since $F(\lambda)$ is non-decreasing, we must have $f(\lambda) \geq 0$.

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

by

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

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 $\int_a^b f(x) g(x) \, dx$. 

---

Miranda Holmes-Cerfon: Applied Stochastic Analysis, Spring 2019
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 we have

$$C(t) = \sum_{j=1}^{\infty} b_j e^{i\lambda_j t}, \quad F(\lambda) = \sum_{j: \lambda_j \leq \lambda} b_j.$$  

(7)

Because $F(x)$ is nondecreasing, we must have $b_i > 0 \forall i$. Clearly $C(t) \notin L^1(\mathbb{R})$, since it oscillates forever, but it is still the covariance function for a stationary process.

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

**Definition.** The *spectrum* of a process is the set of real numbers $\lambda$ 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 the spectrum is $\{\lambda : f(\lambda) > 0\}$, the support of the spectral density function $f(\lambda)$. See (6).

2. If $F(\lambda)$ is piecewise constant, then its spectrum is the set $\{\lambda_i\}_{i=1}^{\infty}$, See (7).

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.** Consider a process with covariance function

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

where $A, \alpha > 0$ are parameters. We can calculate the Fourier transform to be (see 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. If the process this models is Gaussian then it is called a stationary Ornstein-Uhlenbeck process. We will return to this example throughout the course, because it is the only example of a stationary, Gaussian, Markov process, and it is used extremely frequently in modelling.

**Example.** We can use Bochner’s theorem to make up covariance functions, say for testing code or theory. It is not easy to do this out of the blue, because it is hard to quickly tell whether a function is positive semidefinite. However, it is easy to make up functions that are non-negative and integrable, so we can take the Fourier transform to get a positive semidefinite covariance function. For example, all of these are spectral densities of some covariance function:

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

Of course, we could also incorporate jumps, or consider more complicated kinds of nondecreasing functions $F(\lambda)$. Remember that since the desired covariance function is real, the spectral density must be symmetric in $\lambda$.

**Example.** Turbulent fluids are frequently characterized by their spectrum, which says what are the energy-containing scales. Even though the velocity is assumed to be a stochastic process, we don’t need to know
much about its statistical properties in order to find the spectrum. The spectrum can be calculated by looking at one component of the velocity field, computing its covariance function, and finding the Fourier transform, 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.

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).

5.5 Spectral decomposition for discrete, weakly stationary processes

Stationary processes have a nice representation in spectral space. This representation also leads to an efficient, stable method to simulate realizations of them. We will work toward such a simulation method in these notes.

It will be helpful to work in complex space, so let’s first redefine some of the concepts we learned earlier.

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

\footnote{from Barbara Ryden’s notes on “Basic Turbulence”, see http://www.astronomy.ohio-state.edu/ryden/ast825/ch7.pdf}

\footnote{from http://www.physics.usyd.edu.au/cairns/teaching/lecture12/node2.html}
processes.

**Definition.** The covariance function of a complex-valued stochastic process is

\[ B(s, t) = \mathbb{E}X_s \bar{X}_t - (\mathbb{E}X_s)(\mathbb{E}X_t). \]

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: \( \sum_{s,t} B(t, t) \xi_s \bar{\xi}_t \geq 0 \). As in the real case, a complex-valued process is stationary if its mean is constant and its covariance function can be written as \( B(s, t) = C(t-s) \).

We’d like some general characterization of stationary processes. Recall the process \( X_t = A \cos(\lambda t) + B \sin(\lambda t) \) where \( A, B \) are uncorrelated, is stationary. This looks a lot like the real part of a complex exponential, so let’s try to generalize this example.

**Example.** Consider a process of the form

\[ X_t = \xi h(t), \]

where \( h(t) \) is a deterministic, complex-valued function of time, and \( \xi \) 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 if \( h(t) \) 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 \( B(s, t) = (\mathbb{E} \xi \bar{\xi}) h(s) \bar{h}(t) \). We need \( h(s) \bar{h}(t) \) to depend only on \( s-t \). Setting \( s=t \) shows we need \( |h(t)|^2 = \text{const} \). Therefore \( h(t) \) has the form

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

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

\[ B(s+t, s) = \mathbb{E}X_{s+t} \bar{X}_s = A^2 \mathbb{E} |\xi|^2 e^{i(\phi(s+t) - \phi(s))}. \]

We need

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

Since this holds for all \( t \) we must have \( \phi'(t) = \text{const} \), so \( \phi(t) = \alpha t + \beta \) for some real-valued numbers \( \alpha, \beta \).

Re-organizing constants and absorbing some into the definition of \( \xi \) shows that

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

where \( \lambda \in \mathbb{R} \) and \( \xi \) is a complex-valued random variable with \( \mathbb{E} \xi = 0 \) (if \( \lambda \neq 0 \)).

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 \( \omega \) in a probability space \( \Omega \). 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.** Now consider a sum of two exponentials, as

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

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

This is a function of \( s - t \) only if \( \mathbb{E}\xi_1\xi_2 = 0 \), since the functions \( e^{i(\lambda_1 s - \lambda_2 t)} \), \( e^{i(\lambda_2 s - \lambda_1 t)} \) are linearly independent. Therefore \( \xi_1, \xi_2 \) must be mean-zero, uncorrelated random variables. The covariance function is
\[ C(t) = b_1 e^{i\lambda_1 t} + b_2 e^{i\lambda_2 t}, \quad b_j = \mathbb{E}|\xi_j|^2. \]

**Example.** Now consider a more general superposition of frequencies, as
\[ X_t = \sum_{j=1}^{n} \xi_j e^{i\lambda_j t}. \] (8)

Similar calculations show this is stationary iff
\[ \mathbb{E}\xi_j = 0 \quad \forall j, \quad \mathbb{E}\xi_j\xi_k = 0 \quad \forall j \neq k. \] (9)

The covariance function is
\[ C(t) = \sum_{j=1}^{n} b_j e^{i\lambda_j t}, \quad b_j = \mathbb{E}|\xi_j|^2. \] (10)

We could also consider \( n = \infty \) in (8), in which case we require
\[ \sum_{j=1}^{\infty} \mathbb{E}|\xi_j|^2 = \sum_{j=1}^{\infty} b_j < \infty \]
for the covariance function to exist.

The last example looks a lot like the second special case of Bochner’s theorem, see (7). Indeed, from Bochner’s theorem we know that this is the most general form of covariance function for a process with a discrete spectrum. Therefore, let us suppose we have a covariance function constructed as in (10). We can generate realizations of a process with this covariance function, by applying the argument above in reverse:

- Generate uncorrelated random variables \( \xi_i \) with mean zero, variance \( b_i \);
- Set \( X_t = \sum_{j=1}^{\infty} \xi_j e^{i\lambda_j t} \).

The process \( X_t \) will have the desired mean and covariance.

Of course, not all processes with the same means and variances are the same. The choice of distribution for the \( b_i \)’s matters a lot. A type of process for which we can say something concretely are Gaussian processes. Clearly, \( X \) will be Gaussian, if and only if the \( b_i \)’s are Gaussian. In this case, if they are uncorrelated, they must also be independent. We obtain a method for simulating a stationary Gaussian random process.

**Simulating stationary Gaussian processes** Suppose you wish to simulate a real-valued, stationary, Gaussian process with mean zero and covariance \( C(t) \), on a discrete set of equally-spaced points \( \{t_j\}_{j=0}^{N-1} \) which discretize the interval \([-L/2, L/2]\). That is, set \( t_j = j \Delta t - L/2 = j \frac{L}{N} - \frac{L}{N}, j = 0, 1, \ldots, (N-1) \) [4] The grid spacing is \( \Delta t = \frac{L}{N} \).

---

[4] You don’t have to shift by \( L/2 \). We do, to avoid numerical problems when \( C(t) \) is not periodic on the chosen interval.
Compute the discrete Fourier transform \( \hat{C}(k) \) of \( C(t) \) and its inverse as
\[
\hat{C}(k_n) = \Delta t \sum_{j=0}^{N-1} C(t_j) e^{-it_j k_n}, \quad C(t_j) = \frac{\Delta k}{2\pi} \sum_{n=0}^{N-1} \hat{C}(k_n) e^{it_j k_n},
\]
(11)
where \( \Delta k = \frac{2\pi}{L} \), and \( k_n = n\frac{2\pi}{L} - \frac{nN}{L} = n\Delta k - \Delta k \frac{N}{2} \).

Construct \( X \) as
\[
X(t_j) = \text{Re} \left\{ \sum_{n=0}^{N-1} \xi_n e^{it_j k_n} \right\}, \quad \text{where} \quad \xi_n = \sqrt{\hat{C}(k_n) \frac{\Delta k}{2\pi} (A_n + iB_n)}.
\]
(12)
and \( A_n, B_n \sim N(0,1) \) are independent standard normals. Here \( \text{Re}\{\cdot\} \) means take the real part. (You could also take the imaginary part, and get another independent realization of \( X \).)

**Exercise 5.4.** Verify directly that the process \( X \) constructed in (12) above has the correct covariance.

### 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 \( X_t \overset{m.s}{\to} Y \), 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 \( \mu \) and covariance \( C(t) \), and suppose that \( C(t) \in L^1(\mathbb{R}) \). Then
\[
\lim_{T \to \infty} \mathbb{E} \left[ \frac{1}{T} \int_0^T X_s \, ds - \mu \right]^2 = 0.
\]
(13)

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.

---

\(^5\)You don’t have to shift by \( \Delta k \frac{N}{2} \). We do, because then it is easier to make up functions \( \hat{C}(k) \) which are symmetric in \( k \), as is required if \( C(t) \in \mathbb{R} \).
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:

\[
\mathbb{E} \left| \frac{1}{T} \int_0^T X_s \, ds - \mu \right|^2 = \frac{1}{T^2} \int_0^T \mathbb{E} \left| X_s - \mu \right|^2 \, ds
\]

\[
= \frac{1}{T^2} \int_0^T \mathbb{E} (X_s - \mu) (X_s - \mu) \, ds = \frac{1}{T^2} \int_0^T C(t-s) \, ds.
\]

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

\[
\int_0^T \int_0^T C(t-s) \, ds \, dv = \int_{-T}^T \int_{|u|}^{2T-|u|} \frac{1}{2} C(u) \, dv \, du = \int_{-T}^T (T-|u|) C(u) \, du = 2 \int_0^T (T-u) C(u) \, du,
\]

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

\[
\mathbb{E} \left| \frac{1}{T} \int_0^T X_s \, ds - \mu \right|^2 = \frac{2}{T^2} \int_0^T \mathbb{E} (T-u) C(u) \, du
\]

\[
\leq \frac{2}{T} \int_0^T \left( 1 - \frac{u}{T} \right) C(u) \, du
\]

\[
\leq \frac{2}{T} \int_0^\infty C(u) \, du
\]

Dominated Convergence Theorem

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

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

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

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

Example (Ergodic theorem for Markov chains). Let $X = \{X_n\}_{n \in \mathbb{N}}$ be a regular Markov chain with finite state space $\mathcal{S}$, and suppose $X_0 \sim \pi$, where $\pi$ is the unique stationary distribution of the chain. Suppose you run the chain and want to empirically determine $\pi_k$ 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 $n$th jump.
You can show that $X$ is strongly stationary. From this, you can show that the process $(I_n)_{n \in \mathbb{N}}$ is strongly stationary (see exercise 5.3). The mean of $(I_n)_{n \in \mathbb{N}}$ is $E[I_0] = \pi_k$, and the covariance is $C(n) = E[I_0I_n - \pi_k^2] = \pi_k P_{kk}^n - \pi_k^2$.

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 $\beta > 0$ is the absolute value of the eigenvalue with the second-largest norm. Therefore

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

Therefore by the Ergodic theorem,

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

**Example** (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

$$D \equiv \lim_{t \to \infty} \frac{1}{2} \frac{d}{dt} \mathbb{E}(x(t))^2 = \lim_{t \to \infty} \frac{1}{2} \frac{d}{dt} \mathbb{E} \left( \int_0^t u_L(t) dt \right)^2.$$  \hspace{1cm} (14)

Let’s assume the Lagrangian velocity is a stationary stochastic process with mean zero and covariance function $C(t) \in L^1$. Then, following the calculations in the proof of the ergodic theorem, we find

$$\mathbb{E} \left( \int_0^t u_L(t) dt \right)^2 \approx 2t \int_0^t C(u) du$$

for $t$ large.

Therefore we find that

$$D = \int_0^\infty C(u) du.$$ \hspace{1cm} (15)

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.

---

\[6\] Let’s check:

$$P(X_{i+1} = x_1, \ldots, X_{i+h} = x_0) = P(X_{i+1} = x_1) P(X_{i+2} = x_2 | X_{i+1} = x_1) \cdots P(X_{i+h} = x_0 | X_{i+h-1} = x_{h-1})$$

$$= P(X_{i+1} = x_1) P(X_{i+2} = x_2 | X_{i+1} = x_1) \cdots P(X_{i+h} = x_0 | X_{i+h-1} = x_{h-1})$$

$$= P(X_{i+1} = x_1, \ldots, X_{i+h-1} = x_{h-1}).$$
5.7 Some facts about multivariate Gaussian random variables

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

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

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

- 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.

- If $(\mathbf{X}, \mathbf{Y})$ is a multivariate Gaussian, and if $\mathbf{X}, \mathbf{Y}$ are uncorrelated, then $\mathbf{X}, \mathbf{Y}$ are independent.

  **Proof.** This follows directly from the form of the probability density function. If $\mathbf{X}, \mathbf{Y}$ 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

  \[
  f(x, y) = f_x(x)f_y(y) = \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) \cdot \frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right).
  \]

  Therefore $\mathbf{X}, \mathbf{Y}$ are independent.\footnote{Here is a general expression for the pdf of a bivariate Gaussian. If $(\mathbf{X}, \mathbf{Y})$ is a multivariate Gaussian and the correlation coefficient between $\mathbf{X}, \mathbf{Y}$ is $\rho$, then the pdf has the form

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

References


Note that uncorrelated ⇒ 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.

- 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 than 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 5.6.** 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 $X, Y$.)