Week 3
Continuous time Gaussian processes

Jonathan Goodman

September 24, 2012

1 Introduction to the material for the week

This week we take the limit $\Delta t \to 0$. The limit is a process X_t that is defined for all t in some range, such as $t \in [0, T]$. The process takes place in *continuous time*. This week, X_t is a continuous function of t. The process has *continuous sample paths*. It is natural to suppose that the limit of a Markov process is a continuous time Markov process. The limits we obtain this week will be either Brownian motion or the Ornstein Uhlenbeck process. Both of these are Gaussian. We will see how such processes arise as the limits of discrete time Gaussian processes (week 1) or discrete time random walks and urn processes (week 2).

The *scalings* of random processes are different from the scalings of differentiable paths you see in ordinary $\Delta t \to 0$ calculus. Consider a small but non-zero Δt. The net change in X over that interval is $\Delta X = X_{t+\Delta t} - X_t$. If a path has well defined velocity, $V_t = dX/dt$, then $\Delta X \approx V \Delta t$. Mathematicians say that ΔX *is on the order of* Δt, because ΔX is approximately proportional to Δt for small[1] Δt. In this *linear* scaling, reducing Δt by a factor of 2 (say) reduces ΔX approximately by the same factor of 2.

Brownian motion, and the Ornstein Uhlenbeck process, have more complicated scalings. There is one scaling for ΔX, and a different one for $\mathrm{E}[\Delta X]$. The change itself, ΔX, is on the order of $\sqrt{\Delta t}$. If Δt is small, this is larger than the Δt scaling differentiable processes have. Brownian motion moves much further in a small amount of time than differentiable processes do. The change in expected value is smaller, on the order of Δt. It is impossible for the expected value to change by order $\sqrt{\Delta t}$, because the total change in the expected value over a finite time interval would be infinite. The Brownian motion manages to have ΔX on the order of $\sqrt{\Delta t}$ through *cancellation*. The sign of ΔX goes back and forth, so that the net change is far smaller than the sum of $|\Delta X|$ over many small intervals of time. That is $|\Delta X_1 + \Delta X_2 + \cdots| << |\Delta X_1| + |\Delta X_2| + \cdots$.

[1] This terminology is different from scientists' *order of magnitude*, which means roughly a power of ten. It does not make sense to compare ΔX to Δt in the order of magnitude sense because they have different units.

Brownian motion and the Ornstein Uhlenbeck process are Markov processes. The standard filtration consists of the family of $\sigma-$algebras, $\mathcal{F}_t$, which are generated by $X_{[0,t]}$ (the path up to time t). The Markov property for X_t is that the conditional probability of $X_{[t,T]}$, conditioning on all the information in $\mathcal{F}_t$, is determined by X_t alone. The *infinitesimal mean*, or *infinitesimal drift* is $\mathrm{E}[\Delta X|\mathcal{F}_t]$, in the limit $\Delta t \to 0$. The *infinitesimal variance* is $\mathrm{var}(\Delta X \mid \mathcal{F}_t)$. We will see that both of these scale linearly with Δt as $\Delta t \to 0$. This allows us to define the infinitesimal drift coefficient,

$$\mu(X_t)\Delta t \approx \mathrm{E}[X_{t+\Delta t} - X_t \mid \mathcal{F}_t] \ , \tag{1}$$

and the infinitesimal variance, or *noise coefficient*

$$\sigma^2(X_t)\Delta t \approx \mathrm{var}(X_{t+\Delta t} - X_t \mid \mathcal{F}_t) \ . \tag{2}$$

The conditional expectation with respect to a $\sigma-$algebra requires the left side to be a function that is measurable with respect to $\mathcal{F}_t$. The X_t that appears on the left sides is consistent with this. The Markov property says that only X_t can appear on the left sides, because the right sides are statements about the future of $\mathcal{F}_t$, which depend on X_t alone.

The properties (1) and (2) are central this course. This week, they tell us how to take the *continuous time limit* $\Delta t \to 0$ of discrete time Gaussian Markov processes or random walks. More precisely, they tell us how a family of processes must be *scaled* with Δt to get a limit as $\Delta t \to 0$. You choose the scalings so that (1) and (2) work out. The rest follows, as it does in the central limit theorem. The fact that continuous time limits exist may be thought of as an extension of the CLT. The infinitesimal mean and variance of the approximating processes determine the limiting process completely.

Brownian motion and Ornstein Uhlenbeck processes are characterized by their μ and σ^2. Brownian motion has constant μ and σ^2, independent of X_t. The *standard* Brownian motion has $\mu = 0$, and $\sigma^2 = 1$, and $X_0 = 0$. If $\mu \neq 0$, you have *Brownian motion with drift*. If $\sigma^2 \neq 1$, you have a general Brownian motion. Brownian motion is also called the *Wiener process*, after Norbert Wiener. We often use W_t to denote a standard Brownian motion. The Ornstein Uhlenbeck process has constant σ, but a linear drift $\mu(X_t) = -\gamma X_t$. Both Brownian motion and Ornstein are Gaussian processes. If σ^2 is a function of X_t or if μ is a nonlinear function of X_t, then X_t is unlikely to be Gaussian.

2 Path space and probability measure

The probability space Ω for continuous random processes may be taken to be the set of continuous functions of t defined for $0 \leq t \leq T$, with the extra condition that the path starts at 0 for $t = 0$. The value of the path at time t is x_t. We require $x_0 = 0$. The set of continuous functions is written $C([0,T])$. We say $x \in C([0,T])$ if x_t is a continuous function of t, so x represents the whole path and x_t is the value at a specific time.

You might be more used to writing $x(t)$ for a continuous function of t. No harm will come from expressing paths this way, but the x_t notation is more common in probability. It makes the notation for continuous time and discrete time paths similar. For discrete time, x is a vector of values and x_k is the value at index k. The set of paths with $x_0 = 0$ is called $C_0([0,T])$. (Warning: In other parts of mathematics, C_0 refers to functions that have "compact support". Not here. In still other parts of mathematics, C_0 refers to any continuous function, and C_k refers to functions that have k continuous derivatives. Not here.) For us this week, the probability space is $\Omega = C_0([0,T])$, the space of continuous functions with $x_0 = 0$.

A *probability measure* is a function that gives the probability of any measurable event A. We denote this as $P(A)$. The probability measures we have seen until now have been given in concrete ways. For discrete probability (week 2),

$$P(A) = \sum_{\omega \in A} P(\omega) \ .$$

Other probability measures are given by probability densities. If $\Omega = \mathbb{R}^d$ (the space of d component random variables), we usually call the random outcome x rather than ω. A probability density is a function $u(x)$ with the properties that $u(x) > 0$ for all x and

$$\int_{\mathbb{R}^d} u(x)\, dx = 1 \ .$$

The probability of an event $A \subseteq \Omega$ is

$$P(A) = \int_A u(x)\, dx \ .$$

The notion of abstract probability measure collects what these two examples have in common – numbers that act like probabilities associated with sets. The formal definition of a probability measure involves a $\sigma-$algebra, $\mathcal{F}$, on a probability space, Ω. For each $A \in \mathcal{F}$, there is a number $P(A)$, which is called the "probability of A. The abstract definition does not ask the numbers $P(A)$ to be defined in any specific way. They form a *probability measure* if they have the following properties:

Positivity. $P(A) \geq 0$ for every $A \in \mathcal{F}$. This really should be called "non-negativity", but often is not.

Additivity. If $A \in \mathcal{F}$ and $B \in \mathcal{F}$ are disjoint events, then $P(A \cup B) = P(A) + P(B)$.

Totality. $P(\Omega) = 1$.

Countable additivity. Let A_k be an increasing sequence of events. "Increasing" means that $A_k \subset A_{k+1}$ for all k. Let A be the union of all those events.

The countable additivity of $\mathcal{F}$ implies that $A \in \mathcal{F}$. Countable additivity of P is

$$P(\cup_{k=1}^{\infty} A_k) = \lim_{k\to\infty} P(A_k) \; . \tag{3}$$

We can derive many other properties of probability measures in a straightforward way from these. For example, if $A \in \mathcal{F}$, then $P(A^c) = 1 - P(A)$. This is because A and A^c are disjoint, so $P(A) + P(A^c) = P(A \cup A^c) = P(\Omega) = 1$. For another example, suppose A_k are a decreasing family of events (meaning that $A_{k+1} \subseteq A_k$). Then countable additivity implies that

$$\lim_{k\to\infty} P(A_k) = P(\cap_{k=1}^{\infty} A_k) \; . \tag{4}$$

This just restates the countable additivity property (3). You can see that (3) implies (4) as follows. If A_k is a decreasing family of events, then $B_k = A_k^c$ is an increasing family of events (think this through). It is easy to show that if $A = \cup A_k$ and $B = \cup B_k$, then $A^c = B$. (If $\omega \in A$ then $\omega \in A_k$ for all k – the definition of intersection – so $\omega \notin B_k$ for all k, thus $\omega \notin \cup B_k$. Conversely, if $\omega \notin A$, then there is a k with $\omega \notin A_k$. But this implies that $\omega \in A_k^c = B_k$, so $\omega \in B$. This is what mathematicians call a "routine verification".)

A probability measure on our path space $\Omega = C_0([0,T])$ must be of the abstract kind. It cannot be given by probabilities because the space is not discrete. It cannot be given by a probability density because the space is not finite dimensional. In most of the examples in this course, our probability measures can be defined in a $\Delta t \to 0$ process. Define $t_k^{\Delta t} = k\Delta t$. Let $\mathcal{F}_{\Delta t}$ be the $\sigma-$algebra of information you get by observing a path at the set of times $t_k^{\Delta t}$. For example, the event

$$A = \{X_{t_k} \leq 1, \text{ for all } t_k < T\} \tag{5}$$

is measurable in $\mathcal{F}_{\Delta t}$ (We often neglect to write Δt everywhere it should go, as in t_k rather than $t_k^{\Delta t}$.). If Δt is not a rational number, then the event $\{X_2 > 0\}$ is not measurable in $\mathcal{F}_{\Delta t}$, because $t = 2$ is not one of the observation times t_k. A more important event not in $\mathcal{F}_{\Delta t}$ is

$$A = \{X_t \leq 1, \text{ for all } t \in [0,T]\} \; . \tag{6}$$

3 Kinds of convergence

Suppose we have a family of processes $X_t^{\Delta t}$, and we want to take $\Delta t \to 0$ and find a limit process X_t. There are two kinds of convergence, *distributional* convergence, and *pathwise* convergence. Distributional convergence refers to the probability distribution of $X_t^{\Delta t}$ rather than the numbers. It is written with a half arrow, $X_t^{\Delta t} \rightharpoonup X_t$ as $\Delta t \to 0$, or possibly $X_t^{\Delta t} \overset{\mathcal{D}}{\rightharpoonup} X_t$. The CLT is an example of distributional convergence. If $Z \sim \mathcal{N}(0,1)$ and Y_k are i.i.d., mean zero, variance 1, then $X_n = \frac{1}{\sqrt{n}}$ converges to Z is distribution, which means that the distribution of X_n converges to $\mathcal{N}(0,1)$. But the numbers X_n have nothing to do with the numbers Z, so we do not expect that $X_n \to Z$ as $n \to \infty$. We write $X_n \stackrel{\mathcal{D}}{\rightharpoonup} \mathcal{N}(0,1)$ of $X_n \stackrel{\mathcal{D}}{\rightharpoonup} Z$ as $n \to \infty$, which is convergence in distribution.

Later in the course, starting in week 5, there will be examples of sequences that converge pathwise.

4 Discrete time Gaussian process

Consider a linear Gaussian recurrence relation of the form (27) from week 1, but in the one dimensional case. We write this as

$$X_{n+1} = aX_n + bZ_n \ . \tag{7}$$

We want X_n to represent, perhaps approximately, the value of a continuous time process at time $t_n = n\Delta t$. We guess that we can substitute (7) into (1) and (2) to find the scalings of a and b with Δt. We naturally use $\mathcal{F}_n$ instead of $\mathcal{F}_t$ on the right. The result for (1) is

$$\mu(X_n)\Delta t = aX_n \ .$$

To get $\mu = 0$ for Brownian motion, we would take $a = 0$. To get $\mu(x) = -\gamma x$ for the Ornstein Uhlenbeck process, we should take $a = -\gamma\Delta t$. The a coefficient in the recurrence relation should scale linearly with Δt to get finite, non-zero, drift coefficient in the continuous time limiting process.

Calibrating the noise coefficient gives scalings characteristic of continuous time stochastic processes. Inserting (7) into (2), with $\mathcal{F}_n$ for $\mathcal{F}_t$, we find

$$\sigma^2(X_n)\Delta t = b^2 \ .$$

Brownian motion and the Ornstein Uhlenbeck process have constant σ^2, which suggests the scaling $b = \sigma\sqrt{\Delta t}$. These results, put together, suggest that the way to approximate the Ornstein Uhlenbeck process by a discrete Gaussian recurrence relation is

$$X^{\Delta t}_{n+1} = -\gamma \Delta t X^{\Delta t}_n + \sigma\sqrt{\Delta t} Z_n \ . \tag{8}$$

Let X_t be the continuous time Ornstein Uhlenbeck process. We define a discrete time approximation to it using (8) and

$$X^{\Delta t}_{t_n} = X^{\Delta t}_n \ .$$

Note the inconsistent notation. The subscript on the left side of the equation refers to time, but the subscript on the right refers to the number of time steps. These are related by $t_n = n\Delta t$.

Let us assume for now that the approximation converges as $\Delta t \to 0$ in the sense of distributions. There is much discussion of convergence later in the course. But assuming it converges, (8) gives a Monte Carlo method for estimating things about the Ornstein Uhlenbeck process. You can approximate the values of X_t for t between time steps using linear interpolation if necessary. If $t_n < t < t_{n+1}$, you can use the definition

$$X_t^{\Delta t} = X_{t_n}^{\Delta t} + \frac{t - t_n}{t_{n+1} - t_n} \left(X_{t_{n+1}}^{\Delta t} - X_{t_n}^{\Delta t} \right) .$$

The values $X_{t_n}^{\Delta t}$ are defined by, say, (8). Now you can take the limit $\Delta t \to 0$ and ask about the limiting distribution of $X_{[0,T]}^{\Delta t}$ in path space. The limiting probability distribution is the distribution of Brownian motion or the Ornstein Uhlenbeck process.

5 Brownian motion

Many of the most important properties of Brownian motion follow from the limiting process described in Section 4.

5.1 Independent increments property

The *increment* of Brownian motion is $W_t - W_s$. We often suppose $t > s$, but in many formulas it is not strictly necessary. We could consider the increment of a more general process, which would be $X_t - X_s$. The increment is the net change in W over the time interval $[s, t]$.

Consider two time intervals that do not overlap: $[s_1, t_1]$, and $[s_2, t_2]$, with $s_1 \leq t_1 \leq s_2 \leq t_2$. The *independent increments property* of Brownian motion is that increments over non-overlapping intervals are independent. The random variables $X_1 = W_{t_1} - W_{s_1}$ and $X_2 = W_{t_2} - W_{s_2}$ are independent. The intervals are allowed to touch endpoints, which would be $t_1 = s_2$, but they are not allowed to have any interior in common. The cases $s_1 = t_1$ and $s_2 = t_2$ are allowed but trivial.

The independent increments property applies to any number of non-overlapping intervals. If $s_1 \leq t_1 \leq s_2 \leq t_2 \leq \cdots$, then the corresponding increments, X_1, X_2, $X_3, \ldots$, are an independent family of random variables. Their joint PDF is a product.

The approximate sample paths for standard Brownian motion are given by (8) with $\gamma = 0$ and $\sigma = 1$. The exact Brownian motion distribution in path space is the limit of the distributions of the approximating paths, as discussed in Section 4. Suppose that the interval endpoints are time step times, such as $s_1 = t_{k_1}$, $t_1 = t_{l_1}$, and so on. The increment of W in the interval $[s_1, t_1]$ is determined by the random variables Z_j for $s_1 \leq t_j < t_1$. These Z_j are independent for non-overlapping intervals of time. In the discrete approximation there may be small dependences because one Δt time step variable Z_j is a member of, say, $[s_1, t_1]$ and $[s_2, t_2]$. This overlap disappears in the limit $\Delta t \to 0$. The possible dependence between the random variables disappears too.

5.2 Mean and variance

The mean of a Brownian motion increment is zero.

$$\mathrm{E}[W_t - W_s] = 0 \ . \tag{9}$$

The variance of a Brownian motion increment is equal to the size of the time interval:

$$\mathrm{var}(W_t - W_s) = \mathrm{E}\Big[(W_t - W_s)^2\Big] = t - s \ . \tag{10}$$

This is an easy consequence of (8) with $\gamma = 0$ and $\sigma = 1$. If $s = t_k$ and $t = t_n$, then the increment is

$$W_{t_n}^{\Delta t} - W_{t_k}^{\Delta t} = \sqrt{\Delta t}\,(Z_k + \cdots + Z_{n-1}) \ .$$

Since the Z_j are independent, the variance is

$$\Delta t\,(1 + 1 + \cdots + 1) = \Delta t\,(n-k) = t_n - t_k = t - s \ .$$

5.3 The martingale property

The independent increments property upgrades the simple statements to conditional expectations. Suppose $\mathcal{F}_s$ is the $\sigma-$algebra that knows about $W_{[0,s]}$. This is the $\sigma-$algebra generated by $W_{[0,s]}$. The is part of the Brownian motion path is determined by the increments of Brownian motion in the interval $[0,s]$. But all of these are independent of $W_t - W_s$. The increment $W_t - W_s$ is independent of any information in $\mathcal{F}_s$. In particular, we have the conditional expectations

$$\mathrm{E}[W_t - W_s | \mathcal{F}_s] = 0 \ . \tag{11}$$

The variance of a Brownian motion increment is equal to the size of the time interval:

$$\mathrm{var}(W_t - W_s | \mathcal{F}_s) = \mathrm{E}\Big[(W_t - W_s)^2 \,|\mathcal{F}_s\Big] = t - s \ . \tag{12}$$

The formula (11) is called the *martingale property.* It can be expressed as

$$\mathrm{E}[W_t | \mathcal{F}_s] = W_s \ . \tag{13}$$

To understand this, recall the the left side is a function of the path that is known in $\mathcal{F}_s$. The value W_s qualifies; it is determined (trivially) by the the path $W_{[0,s]}$. The variance formula (12) may be re-expressed in a similar way:

$$\begin{aligned}
\mathrm{E}\big[W_t^2 | \mathcal{F}_s\big] &= \mathrm{E}\Big[(W_s + [W_t - W_s])^2 \,|\mathcal{F}_s\Big] \\
&= \mathrm{E}\Big[W_s^2 + 2W_s\,[W_t - W_s] + [W_t - W_s]^2 \,|\mathcal{F}_s\Big] \\
&= W_s^2 + 2W_s \mathrm{E}[W_t - W_s | \mathcal{F}_s] + \mathrm{E}\Big[(W_t - W_s)^2 \,|\mathcal{F}_s\Big] \\
\mathrm{E}\big[W_t^2 | \mathcal{F}_s\big] &= W_s^2 + (t - s) \ .
\end{aligned}$$

We used the martingale property, and the fact that a number can be pulled out of the expectation if it is known in $\mathcal{F}_s$.

6 Ornstein Uhlenbeck process

The Ornstein Uhlenbeck process is the continuous time analogue of a scalar Gaussian discrete time recurrence relation. Let X_t be a process that satisfies (1) and (2) with $\mu(X_t) = -\gamma X_t$ and constant σ^2. Suppose $u(x,t)$ is the probability density for X_t. Since u is the limit of Gaussians, as we saw in Section 4, u itself should be Gaussian. Therefore, $u(x,t)$ is completely determined by its mean and variance. We give arguments that are derived from those in Week 1 to find the mean and variance.

The mean is simple

$$\begin{aligned}
\mu_{t+\Delta t} &= \mathrm{E}[X_{t+\Delta t}] \\
&= \mathrm{E}[X_t + \Delta X] \\
&= \mathrm{E}[X_t] + \mathrm{E}[\mathrm{E}[\Delta X|\mathcal{F}_t]] \\
&= \mu_t + \mathrm{E}[-\gamma X_t \Delta t + (\text{smaller})] \\
&= \mu_t - \gamma \mathrm{E}[X_t]\,\Delta t + (\text{smaller}) \\
\mu_{t+\Delta t} &= \mu_t - \gamma \mu_t \Delta t + (\text{smaller})\ .
\end{aligned}$$

The last line shows that

$$\partial_t \mu_t = -\gamma \mu_t\ . \tag{14}$$

This is the analogue of (28) from Week 1.

An orthogonality property of conditional expectation makes the variance calculation easy. Suppose $\mathcal{F}$ is a $\sigma-$algebra, X is a random variable, and $Y = \mathrm{E}[X|\mathcal{F}]$. Then

$$\mathrm{var}(X) = \mathrm{E}\Big[(X-Y)^2\Big] + \mathrm{var}(Y)\ . \tag{15}$$

The main step in the proof is to establish the simpler formula

$$\mathrm{E}\big[X^2\big] = \mathrm{E}\Big[(X-Y)^2\Big] + \mathrm{E}\big[Y^2\big]\ . \tag{16}$$

The formula (16) implies (15). If $\mu = \mathrm{E}[X]$, then $\mathrm{var}(X) = \mathrm{E}\Big[(X-\mu)^2\Big]$, and $\mathrm{E}[X-\mu|\mathcal{F}] = Y - \mu$. So we get (15) by applying (16) with $X-\mu$ instead of X. Note that

$$\begin{aligned}
\mathrm{E}\big[X^2\big] &= \mathrm{E}\Big[([X-Y]+Y)^2\Big] \\
&= \mathrm{E}\Big[(X-Y)^2\Big] + 2\mathrm{E}[(X-Y)\,Y] + \mathrm{E}\big[Y^2\big]\ .
\end{aligned}$$

The formula (16) follows from the orthogonality relation that, in turn, depends on the tower property and the fact that $\mathrm{E}[Y|\mathcal{F}] = Y$:

$$\begin{aligned}
\mathrm{E}[(X-Y)\,Y] &= \mathrm{E}[\ \mathrm{E}[(X-Y)\,Y \mid \mathcal{F}]\ ] \\
&= \mathrm{E}[\ \mathrm{E}[\ X-Y \mid \mathcal{F}]\ \ Y] \\
&= \mathrm{E}[\ (Y-Y)\ Y] \\
&= 0\ .
\end{aligned}$$

To summarize, we just showed that $X - \mathrm{E}[X]$ is *orthogonal* to $\mathrm{E}[X]$ in the sense that $\mathrm{E}[(X-Y)\,Y] = 0$. The formula (16) is the Pythagorean relation that follows from this orthogonality. The variance formula (15) is just the mean zero case of this Pythagorean relation.

The variance calculation for the Ornstein Uhlenbeck process uses the Pythagorean relation (15) in $\mathcal{F}_t$. The basic mean value relation (1) may be re-written for Ornstein Uhlenbeck as

$$\mathrm{E}[X_{t+\Delta t}|\mathcal{F}_t] = X_t + \mu(X_t)\Delta t + (\text{smaller}) = X_t - \gamma X_t \Delta t + (\text{smaller}) \ .$$

Let $\sigma_t^2 = \mathrm{var}(X_t)$. Then (third line justified below)

$$\begin{aligned}
\sigma^2_{t+\Delta t} &= \mathrm{var}(X_{t+\Delta t}) \\
&= \mathrm{E}\Big[(X_{t+1} - X_t + \gamma X_t \Delta t + (\text{smaller}))^2\Big] + \mathrm{var}\Big((X_t - \gamma X_t \Delta t + (\text{smaller}))^2\Big) \\
&= \sigma^2 \Delta t + (1-\gamma\Delta t)^2 \,\mathrm{var}(X_t) + (\text{smaller}) \\
&= \sigma_t^2 + \left(\sigma^2 - 2\gamma\sigma_t^2\right)\Delta t + (\text{smaller}) \ .
\end{aligned}$$

This implies that σ_2^2 satisfies the scalar continuous time version of (29) from Week 1, which is

$$\partial_t \sigma_t^2 = \sigma^2 - 2\gamma\sigma_t^2 \ . \tag{17}$$

Both (14) and (17) have simple exponential solutions. We see that μ_t goes to zero at the exponential rate γ, while σ_t^2 goes to $\sigma^2/(2\gamma)$ at the exponential rate 2γ. The probability density of X_t is

$$u(x,t) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-(x-\mu_t)^2/(2\sigma_t^2)} \ . \tag{18}$$

This distribution has a limit as $t \to \infty$ that is the statistical steady state for the Ornstein Uhlenbeck process.