

1 The Brownian bridge construction

The *Brownian bridge* construction is a way to build a Brownian motion path by successively adding finer scale detail. This construction leads to a relatively easy proof that Brownian motion paths are continuous. We will soon generalize this proof to show that general diffusions are continuous (cheating in places). The construction can be useful also in computations involving Brownian motion. It also is the basis of many *low discrepancy sequence* quasi-Monte Carlo methods that achieve error less than $N^{-1/2}$ from N sample paths.

Fix a final time T and consider a the *diadic* intervals

$$I_{n,k} = [kT2^{-n}, (k+1)T2^{-n}] . \quad (1)$$

To understand this formula, let $\Delta t_n = T2^{-n}$ be the time step on level n . Then $I_{n,k} = [t_{n,k}, t_{n,k+1}]$, with $t_{n,k} = k\Delta t_n$. These intervals are chosen so that going from level n to level $n+1$ divides each diadic interval exactly in half:

$$I_{n,k} = I_{n+1,2k} \cup I_{n+1,2k+1} .$$

You can get the indices on the right correct by showing that $I_{n,k}$ and $I_{n+1,2k}$ have the same starting point, which is $t_{n,k} = kT2^{-n} = (2k)T(2^{-1}2^{-n}) = (2k)2^{-(n+1)} = t_{n+1,2k}$. The level zero interval is the whole $[0, T]$. There are two level 1 intervals, 4 level 2 intervals, and so on.

The Brownian bridge construction builds the Brownian motion path W_t for $t \in [0, T]$ piece by piece and level by level. At level n we have a path $W_{n,t}$ which is linear on each level n diadic interval. Such a *piecewise linear* path is determined by its values at the diadic endpoints $t_{n,k}$, which are $W_{n,t_{n,k}}$. The construction ensures that these $W_{n,t_{n,k}}$ have the same joint distribution as the corresponding numbers $W_{t_{n,k}}$ of an exact Brownian motion path. Going from stage n to stage $n+1$ keeps the values $W_{n,t_{n,k}}$ and introduces a new value for W_{n+1} at the center of $I_{n,k}$, which is $t_{n+1,2k+1} = t_{n,k} + \Delta t_{n+1}$.

The choice of W values at level n depends on their joint probability density. We want these values to match the joint distribution of W at the same set of diadic times. There are several ways to figure out what that is, but one simple way is to write the formula for the joint density. We do this somewhat more generally. Suppose $t_0 = 0$ and $t_{k+1} > t_k$ is a sequence of increasing times. Define, for notational simplicity, $X_k = W_{t_k}$. Then conditional on \mathcal{F}_{t_k} , $X_{k+1} - X_k \sim \mathcal{N}(0, t_{k+1} - t_k)$. Write that conditional density as

$$\begin{aligned} v_k(x_{k+1} \mid x_1, \dots, x_k) &= \frac{1}{\sqrt{2\pi(t_{k+1} - t_k)}} e^{-(x_{k+1} - x_k)^2 / 2(t_{k+1} - t_k)} \\ &= C e^{-(x_{k+1} - x_k)^2 / 2(t_{k+1} - t_k)} . \end{aligned}$$

As a function of x_{k+1} , this is just the formula for a Gaussian with mean x_k and variance $t_{k+1} - t_k$. The value of the normalization constant C depends on the times t_k but not on the x values. We will treat normalization constants C like the constants in the big O stuff from last week, a C in one place need not equal

a C in a different place. With this, let $u_k(x_1, \dots, x_k)$ be the joint density of (X_1, \dots, X_k) . Then Bayes' rule gives

$$u_{k+1}(x_1, \dots, x_{k+1}) = u_k(x_1, \dots, x_k) v_k(x_{k+1} | x_1, \dots, x_k) .$$

You can write u_k in terms of u_{k-1} , and so on to the beginning. The result is

$$u_n(x_1, \dots, x_n) = C \exp \left\{ \frac{-1}{2} \left[\frac{(x_1 - x_0)^2}{t_1 - t_0} + \frac{(x_2 - x_1)^2}{t_2 - t_1} + \dots + \frac{(x_n - x_{n-1})^2}{t_n - t_{n-1}} \right] \right\} . \quad (2)$$

In this formula, $x_0 = 0$, but I put it in because it makes the formula easier to understand. It is unnecessary at this point to write the simple formula for C .

The next step is a bit of Gaussian algebra. We need to see how the density (2) depends on one of the variables x_j . Clearly, x_j appears in two terms. An abstract version of this situation is the probability density

$$u(y) = C e^{-\frac{1}{2} \left[\frac{(y-a)^2}{\sigma_a^2} + \frac{(y-b)^2}{\sigma_b^2} \right]} \quad (3)$$

Clearly this is a Gaussian. Some calculation uncovers the mean and variance. Here C refers to any number that does not depend on y .

$$\begin{aligned} \frac{(y-a)^2}{\sigma_a^2} + \frac{(y-b)^2}{\sigma_b^2} &= \frac{1}{\sigma_a^2 \sigma_b^2} [\sigma_b^2 (y-a)^2 + \sigma_a^2 (y-b)^2] \\ &= \frac{1}{\sigma_a^2 \sigma_b^2} [\sigma_b^2 y^2 - 2\sigma_a^2 b y + \sigma_a^2 y^2 - 2\sigma_b^2 a y + C] \\ &= \frac{1}{\sigma_a^2 \sigma_b^2} [(\sigma_b^2 + \sigma_a^2) y^2 - 2(\sigma_a^2 b + \sigma_b^2 a) y + C] \\ &= \frac{\sigma_b^2 + \sigma_a^2}{\sigma_a^2 \sigma_b^2} \left[y^2 - 2 \frac{a\sigma_b^2 + b\sigma_a^2}{\sigma_b^2 + \sigma_a^2} y + C \right] \\ &= \frac{(y - \bar{y})^2}{\bar{\sigma}^2} , \end{aligned}$$

where

$$\bar{y} = \frac{\sigma_b^2 a + \sigma_a^2 b}{\sigma_a^2 + \sigma_b^2} , \quad (4)$$

and

$$\bar{\sigma}^2 = \frac{\sigma_a^2 \sigma_b^2}{\sigma_a^2 + \sigma_b^2} = \frac{1}{\frac{1}{\sigma_a^2} + \frac{1}{\sigma_b^2}} . \quad (5)$$

This gives the PDF of Y as

$$u(y) = C e^{-(y-\bar{y})^2/2\bar{\sigma}^2} ,$$

which makes $Y \sim u$ a normal with mean (4) and variance (5).

You can interpret (3) as saying that Y is attracted to a with strength $\frac{1}{\sigma_a^2}$ and attracted to b with strength $\frac{1}{\sigma_b^2}$. The variance is inversely proportional to the strength of the attraction because the stronger the attraction, the closer Y stays to a or b . The result is that Y is attracted to a weighted average, weighted by the exchanged variances. The point a gets weight σ_b^2 and b gets weight σ_a^2 . If σ_b is large, then Y is weakly attracted to b and therefore is more likely to be near a . The formula (5) for the effective variance is like the formula for the net resistance of two resistors in parallel. The net resistance is less than either of the resistors separately. Here, the net variance is smaller – Y is more tightly bound to \bar{y} – than either of the individual variances.

We now apply these calculations to the distribution of the midpoint values. For that, take $t_k = t_{n+1,k}$. We want the distribution of the new midpoint value $W_{n+1,2k+1}$ conditional on the already known values $W_{n,k} = W_{n+1,2k}$ and $W_{n,k+1} = W_{n+1,2k+2}$. The variable x_{2k+1} , which corresponds to $W_{n+1,2k+1}$ appears in two terms in the sum in the exponent of (2). The variances are all equal to Δt_{n+1} . Therefore, the conditional density of $W_{n+1,2k+1}$ is

$$u(w_{n+1,2k+1}) = C e^{-\frac{1}{2} \left[\frac{(w_{n+1,2k+1} - w_{n,k})^2}{\Delta t_{n+1}} + \frac{(w_{n+1,2k+1} - w_{n,k+1})^2}{\Delta t_{n+1}} \right]} . \quad (6)$$

The algebra (4) and (5) here simply gives

$$\bar{W}_{n+1,2k+1} = \frac{1}{2} (W_{n,k} + W_{n,k+1})$$

and

$$\bar{\sigma}^2 = \frac{\Delta t_{n+1}}{2} .$$

Both of these are implemented by the formula

$$W_{n+1,2k+1} = \frac{1}{2} (W_{n,k} + W_{n,k+1}) + \sqrt{\frac{\Delta t_{n+1}}{2}} Z_{n+1,2k+1} , \quad (7)$$

where all the Z variables are independent standard normals. This gives the new value on the left as the sum of the linear interpolant (the first term on the right) and a random perturbation with variance $\Delta t_{n+1}/2$. The interpolant is the value of the level n path at the midpoint, which is $W_{n,t_{n+1,2k+1}}$. The variance of the perturbation is smaller than Δt_{n+1} because $W_{n+1,2k+1}$ is attracted to two values rather than just one, corresponding to the two terms in (6).

The Brownian bridge algorithm constructs a Brownian motion path to whatever level of detail desired. You start by choosing $W_0 = 0$ and $W_T = \sqrt{T} Z_0$. Then you apply (7) with $n = 1$ to get the midpoint value $W_{\frac{T}{2}}$. Then you apply (7) with $n = 2$ to get the two quarter points $W_{\frac{T}{4}}$ and $W_{\frac{3T}{4}}$, and so on. The application to low discrepancy sequences (high accuracy quasi-Monte Carlo) is as follows. Driving those constructions are multivariate low discrepancy sequences $U_k = (U_{k,1}, \dots, U_{k,m})$, that are uniformly distributed in the m dimensional “unit cube”. These have names such as the Sobol sequence. They have the feature that the deviation from uniformity is on the order of N^{-1} (for N samples)

rather than $N^{-1/2}$ for independent random samples. The best sequences known have quite good distributions for the first components $(U_{k,1}, U_{k,2}, \dots)$, but the quality of the $U_{k,j}$ decreases with j . The idea is to use the best sequences to get the most important W values, first the final value, then the midpoint value, then the quarter points, and so on. The later parts of the Brownian bridge construction do not change the path much, so it matters less how accurately these are sampled.

2 Hölder continuity

The Brownian bridge construction leads to a proof that Brownian motion paths are continuous functions of t . To do this, we interpret the construction as

$$W_t = W_{0,t} + \sum_{n=1}^{\infty} V_{n,t}, \quad (8)$$

where $V_{n,t}$ is the change at stage n of the construction (precise formula below). Then we show that the functions V_n have (probably have, this is random) a degree of continuity measured by the Hölder α norm (definition below). Finally, we will get an estimate of the Hölder α norm of W by adding up the norms of the terms in the sum (8). As we saw in with the definition of the Ito integral, this infinite sum converges almost surely to a finite Hölder α norm of W , which implies that W is a Hölder continuous function of t .

The Hölder α norm is a quantitative measure of continuity that is particularly suited to Brownian motion. To motivate it, note that if $f(t)$ is a differentiable function of t , then $|f(t_2) - f(t_1)| \leq C|t_2 - t_1|$, at least if $f'(t)$ is bounded in the interval $[t_1, t_2]$. A Brownian motion path is continuous, but less continuous than this. We already saw that $|W_{t_2} - W_{t_1}| \sim |t_2 - t_1|^{1/2}$, at least in the sense that $E[|W_{t_2} - W_{t_1}|^2] = |t_2 - t_1|$. You might hope that there is a C so that for all $t_1 \in [0, T]$ and $t_2 \in [0, t_1]$, we have $|W_{t_2} - W_{t_1}| \leq C|t_2 - t_1|^{1/2}$. But this is in retrospect naive. The quantity

$$R(t_1, t_2) = \frac{|W_{t_2} - W_{t_1}|}{|t_2 - t_1|^{1/2}} \quad (9)$$

is unlikely to be bounded for all choices of t_1 and t_2 as the naive inequality implies. This is because for each t_1 and t_2 , R is a random variable whose distribution is the same for each t_1 and t_2 . Moreover, R is not bounded in the sense that a Gaussian is not bounded. For any number L , no matter how large, there is some non-zero probability that $R > L$ just as there is a non-zero probability that $Z > L$ if $Z \sim \mathcal{N}(0, 1)$. If you take a large enough family of non-overlapping $[t_1, t_2]$ values, you produce that many independent chances for $R > L$. Some of them are sure to succeed (in the sense of “almost surely”).

The Hölder norm of a function $f(t)$ is

$$\|f\|_{\alpha} = \max_{t_1 \neq t_2} \frac{|f(t_2) - f(t_1)|}{|t_2 - t_1|^{\alpha}}. \quad (10)$$

The space of *Hölder α continuous* functions is the set of functions with $\|f\| < \infty$. We argued using (9) that the natural in Brownian motion to try $\alpha = \frac{1}{2}$, but that W is not Hölder $\frac{1}{2}$ because $\|W\|_{\frac{1}{2}} = \infty$. However, it is easier for a function to be Hölder α if α is smaller. To see this, suppose $T = 1$, and $\beta > \alpha$. Then $|t_2 - t_1|^\beta \leq |t_2 - t_1|^\alpha$ (because $|t_2 - t_1| \leq 1$). Then, looking at (10) we see that $\|f\|_\beta \geq \|f\|_\alpha$ because

$$\frac{|f(t_2) - f(t_1)|}{|t_2 - t_1|^\beta} \geq \frac{|f(t_2) - f(t_1)|}{|t_2 - t_1|^\alpha} \quad \text{for all } t_1 \in [0, 1] \text{ and } t_2 \in [0, 1] \text{ with } t_2 \neq t_1.$$

If $\beta > \alpha$ and $\|f\|_\beta < \infty$, then $\|f\|_\alpha < \infty$, so Hölder β implies Hölder α . But it is possible to be Hölder α without being Hölder β . We will indeed show that Brownian motion is (almost surely) Hölder continuous for some $\alpha < \frac{1}{2}$.

In the previous paragraphs we talked about the Hölder *norm* without saying what a norm is. A norm is a measure of the size of a vector in a vector space. If V is a vector space and $f \in V$, we write $\|f\|$ for the norm of f . The abstract definition is that a norm is a real valued function of a vector (for any vector, f , the norm $\|f\|$ is a real number). A norm is thought of as some measure of the size of f . To be a norm, this function must have the following properties:

1. $\|f\| \geq 0$ for all $f \in V$ and $\|f\| = 0$ only if $f = 0$.
2. If $a \in \mathbb{R}$, then $\|af\| = |a| \cdot \|f\|$.
3. If $f \in V$ and $g \in V$, then $\|f + g\| \leq \|f\| + \|g\|$.

The first property is pretty clear. A measure of size is never negative, and the only vector with size zero is the zero vector. The second property is *homogeneity* (of degree 1). If you double a vector than you double its size. This makes you think of $\|f\|$ as something like length rather than something like area, say. If you multiply by -1 , the size does not change.

The third property is the *triangle inequality*. The triangle refers to two ways (paths) to go from the origin to $f + g$. One way goes first to f and then to $f + g$. The total length of that path is the sum of the lengths of its pieces, which is $\|f\| + \|g\|$. The other way to go directly from the origin to $f + g$. That path has length $\|f + g\|$. The triangle inequality says that the direct path ($0 \rightsquigarrow f + g$) is shorter than the indirect path ($0 \rightsquigarrow f \rightsquigarrow f + g$), when “lengths” are taken to be the norms of f and g .

One familiar example of norm is the euclidean norm of a vector $x \in \mathbb{R}^n$, which is $\|x\| = (\sum_k x_k^2)^{1/2}$. For another example, let V be the set of continuous functions of t defined for $t \in [0, T]$. The *sup norm* is¹

$$\|f\| = \max_{t \in [0, T]} |f(t)| .$$

¹There is a distinction between *supremum* and *maximum*. If A is a set of numbers, the maximum is the largest element of A . The problem with this definition is that A may not have a largest element. For example, if A is the set of all numbers with $a < 2$, the “maximum” would be 2, but $2 \notin A$. The *supremum* is the smallest number, b so that $a \leq b$ for all $a \in A$. I write max when I should write sup because the distinction is an irrelevant mathematical technicality here. It might be that $\sup A = \infty$, for example if A is the set of integers.

This satisfies the axioms. It is zero only if $f = 0$. It is homogenous of degree 1. It satisfies the triangle inequality because $\max |f + g| \leq \max |f| + |g|$. The Hölder norm (10) also is homogeneous and satisfies the triangle inequality. But $\|f\|_\alpha = 0$ implies that f is a constant, not that the constant is zero. That is OK for Brownian motion because a Brownian motion path starts with $W_0 = 0$. Therefore a constant path is the zero path. The vector space for Brownian motion is continuous functions defined for $t \in [0, t]$ with $W_0 = 0$. The Hölder norm (10) is indeed a norm on that space. To be slightly more precise, the Hölder space H_α is the space of continuous functions so that $\|f\|_\alpha < \infty$. The triangle inequality implies that if $f \in H_\alpha$ and $g \in H_\alpha$, then $f + g \in H_\alpha$, which is necessary for H_α to be a vector space.

A *Banach space* is a vector space with a norm so that the triangle inequality works for infinite sums. That is, if $g_n \in V$ is a sequence of vectors and

$$\sum_{n=1}^{\infty} \|g_n\| < \infty ,$$

then

$$f = \sum_{n=1}^{\infty} g_n$$

is in V and

$$\|f\| \leq \sum_{n=1}^{\infty} \|g_n\| .$$

The Hölder space H_α is a Banach space, as long as you require $f(0) = 0$. The proof is not hard if you have a good undergraduate analysis class, but I don't to spend class time on it.

3 Hölder norms of Brownian bridge components

We now estimate the sizes of the terms $V_n = W_{n+1} - W_n$ in the Brownian bridge construction. For that purpose we describe $W_{n+1} = W_n$ in a slightly different way. The approximate path $W_{n,t}$ is linear on level n diadic intervals $I_{n,k}$. The midpoint of that interval is $\frac{1}{2}(t_{n,k} + t_{n,k+1}) = t_{n+1,2k+1}$. The value of W_n at that midpoint is the average of the values at the endpoints: $W_{n,t_{n+1,2k+1}} = \frac{1}{2}(W_{n,t_{n,k}} + W_{n,t_{n,k+1}})$. To go to W_{n+1} , (7) says you add the value $C_n Z_{n+1,2k+1}$, with $C_n = \sqrt{\frac{\Delta t_{n+1}}{2}}$. We give a more formal description of this using *hat functions*. The hat function $h_{n,k}(t)$ describes $W_{n+1,t} - W_{n,t}$ inside the diadic interval $I_{n,k}$. It is the function $h_{n,k}(t) = 0$ if $t \notin I_{n,k}$. It is piecewise linear inside $I_{n,k}$ and has $h_{n,k}(t_{n+1,2k+1}) = 1$. The graph of $h_{n,k}$ has the shape of a triangular witch hat. Within $I_{n,k}$ we have $W_{n+1,t} = W_{n,t} + C_n Z_{n+1,2k+1} h_{n,k}(t)$.

The hat function description of $W_{n+1} - W_n$ relies on the fact that the differences $W_{n+1,t} - W_{n,t}$ are independent for disjoint diadic intervals on level n .

That, plus the fact that $h_{n,j}(t) = 0$ if $j \neq k$, implies that for $t \in I_{n,k}$ only the $j = k$ term of the following sum is not zero:

$$W_{n+1,t} - W_{n,t} = V_{n,t} = C_n \sum_{j=0}^{2^n-1} Z_{n+1,2j+1} h_{n,j}(t). \quad (11)$$

To repeat, only one term of this sum is different from zero at any t . We already understood why the one term formula is correct for $t \in I_{n,k}$.

Now we need to estimate the Hölder α norm of the sum (11). First, it is “clear” (after some minutes of checking) that

$$\|h_{n,k}\|_\alpha = \frac{1}{|\Delta t_{n+1}|^\alpha}.$$

This corresponds to taking $t_1 = t_{n,k}$ and $t_2 = t_{n,k} + \Delta t_{n+1} = t_{n+1,2k+1}$. In this time interval, $|h_{n,k}(t_2) - h_{n,k}(t_1)| = 1$. Given that $\Delta t_n = T2^{-n}$, some algebra turns this into

$$\|h_{n,k}\|_\alpha = (2^\alpha T^{-\alpha}) 2^{\alpha n}. \quad (12)$$

The large n hat functions have large Hölder α norm because they go from 0 to 1 quickly. The parameter α determines how rapidly $\|h_{n,k}\|_\alpha$ grows with n . The growth is less rapid when α is closer to zero.