Stochastic Calculus, Courant Institute, Fall 2018
http://www.math.nyu.edu/faculty/goodman/teaching/StochCalc2018/index.html
Jonathan Goodman, September, 2018

Lesson 3, Ito integral

1 Introduction

The main operations of ordinary calculus are differentiation and integration. One is the inverse of the other. In ordinary calculus you can define the derivative and then define the integral as the thing that undoes differentiation. We are going to go the other way. We will define the Ito integral and use it to derive the facts about differentiation. The main differentiation fact is *Ito's lemma*, which is the chain rule for differentiation. The usual chain rule may be written

$$du(X_t) = (\partial_x u(X_t))\, dX_t \ .$$

This applies when X_t is a differentiable function of t, and is often rewritten by dividing both sides by dt in the form

$$\frac{du(X_t)}{dt} = (\partial_x u(X_t))\, \frac{dX_t}{dt} \ .$$

If X_t is a diffusion process then $(dX)^2$ is on the order of dt, because

$$\mathrm{E}\big[\,(dX)^2\big] = v(X_t)dt + O(dt^2) \ .$$

(You may ΔX instead of dX, and Δt instead of dt.) Keeping terms to second order in dX gives

$$du(X_t) = (\partial_x u(X_t))\, dX_t + \frac{1}{2}\partial_x^2 u(X_t) dX^2 \ .$$

This is closer to Ito's lemma, but there still is one step left.

The integral is the "sum" of infinitely many infinitely small contributions. The expression

$$\int_0^T F_t\, dt$$

means that you divide the interval $[0,T]$ into infinitely many tiny and non-overlapping pieces of length dt and add $F_t dt$. The integral sign is a distorted S, for "sum". It is possible to give a less vague definition by defining approximate integrals of the form

$$Y_T^{(h)} = \sum_{0\le t_k<T} F_{t_k} h = \sum_{0\le t_k<T} F_{t_k}(t_{k+1}-t_k) \ .$$

Here, $h>0$ is a *time step* and $t_k = kh$ is the start of a time interval $[t_k, t_{k+1}$ of length h. It is possible to prove that the following limit exists

$$Y_T = \lim_{h\to 0} Y_T^{(h)} \ .$$

This proof depends on the function F_t – is it continuous.

You can integrate (add up) contributions of the form $F_t dX$, which gives an integral of the form

$$Y_T = \int_0^T F_t \, dX_t \; . \tag{1}$$

This may be interpreted as physic work done to a moving particle, X_t, with force F_t. The basic formula is work = Force · distance. If the force at time t is F_t and $dX_t = X_{t+dt} - X_t$ is a small displacement of the particle, the corresponding small bit of work is $dW = F\, dX_t$. The total work is $\int_0^T dW$.

There is a financial interpretation of integrals like (1). Suppose X_t is the price of an asset at time t and F_t is the amount of that asset that you own. Your "cash flow" in the next time increment dt is $F_t\, dX_t$. The "time ordering" is crucial in this interpretation. First, at time t, you "acquire" F_t "shares" of the asset. Then you "hold" the asset (don't buy or sell any) for the next time increment dt. In that time increment the price goes up or down by dX_t. The total profit (if $Y_T > 0$) or loss (if $Y_T < 0$) is given by adding up the small profit/loss amounts in the time increments dt. Optimal trading strategies are designed by studying integrals like (1) with various trading strategies F_t.

If X_t is a random process, then the "decision" F_t must be made on the basis of information available at time t. This information does not include future values X_s, for $s > t$, but it might involve predictions of future values from present information. A trading strategy F_t is *adapted*, or *non-anticipating*, or *progressively measurable*[1] if F_t is a function of $X_{[0,t]}$.

If X_t is a diffusion process and F_t is adapted, then the integral (1) is the *Ito integral*. The random processes X_t and F_t define a new random process Y_T. This Lesson explains (not in complete mathematical rigor) the proof that the following limit exits

$$Y_T = \lim_{h\to 0} Y_T^{(h)} = \lim_{h\to 0} \sum_{t_k < T} F_{t_k} \left(X_{t_{k+1}} - X_{t_k} \right) \; . \tag{2}$$

This explanation relies on facts about diffusions from Lesson 2. It also relies on F_t being continuous in a certain quantitative sense described below. The *Borel Cantelli lemma* is a mathematical trick, explained below, for proving convergence of random variables like $Y_T^{(h)}$. To make it work (experts will immediately complain), we don't take $h \to 0$ in the simple way. Instead we take a sequence $h_n = 2^{-n}$ and let $n \to \infty$.

Starting here, the concept of a *martingale* will be used constantly. A stochastic process is a martingale if its increments have expected value zero at the start of each increment interval. That is, if $s > 0$, then

$$\mathrm{E}[\, X_{t+s} - X_t \mid \mathcal{F}_t] = 0 \; . \tag{3}$$

[1]These terms have slightly different meanings, but those distinctions are not relevant in this lesson.

(The full technical definition of martingale requires more technical hypotheses on X_t and a full technical definition of $\mathcal{F}_t$.) Some important facts about martingales: (1) Brownian motion is a martingale. (2) If X_t is a martingale then Y_T, defined by the Ito integral (1), is a martingale. This is sometimes called *Doob's martingale theorem*. The financial interpretation is that a trading strategy from a martingale produces a martingale – you cannot make an expected profit from a non-anticipating trading strategy. (3) A diffusion is a martingale if the infinitesimal drift is zero.

The *Ito isometry formula*, for a martingale diffusion, is

$$\mathrm{E}\left[Y_T^2\right] = \int_0^T \mathrm{E}\left[\,F_t^2\right] \mathrm{E}\left[\,v(X_t)^2\right] dt \; . \tag{4}$$

This is an integral version of a simple fact about random sums. First, suppose that U_k are random variables with $\mathrm{E}[\,U_k] = 0$, and

$$S_n = \sum_1^n U_k \; .$$

Then

$$\mathrm{E}\left[\,S_n^2\right] = \sum_1^n \mathrm{E}\left[\,U_k^2\right] \; .$$

Now suppose that $\mathcal{F}_n$ is "all the information" in the random numbers $F_1, \ldots, F_n$ and $U_1, \ldots, U_{n-1}$, and suppose that the U_n are *martingale differences* in the sense that

$$\mathrm{E}[\,U_n \mid \mathcal{F}_n] = 0 \; .$$

Define a sum that looks more like (1)

$$S_n = \sum_1^n F_k U_k \; .$$

This has

$$\mathrm{E}\left[\,S_n^2\right] = \sum_1^n \mathrm{E}\left[\,F_k^2\right] \mathrm{E}\left[\,U_k^2 \mid \mathcal{F}_k\right] \; . \tag{5}$$

This is like the Ito isometry formula, if $\mathrm{E}\left[\,F_t^2\right]$ is replaced with $\mathrm{E}\left[\,F_n^2\right]$ and $\mathrm{E}\left[\,dX^2 \mid \mathcal{F}_t\right] = v(X_t)dt$ is replaced with $\mathrm{E}\left[\,U_k^2 \mid \mathcal{F}_k\right]$. This lesson explains calculations like this.

2 Application to SDE

A *stochastic differential equation* (or *SDE*) is an expression of the form

$$dX_t = a(X_t, t)dt + b(X_t, t)dW_t \; . \tag{6}$$

In Lesson 2, we said that this does not have to be interpreted literally. It may be taken as a convenient way to express the model that $a(x,t)$ is the infinitesimal mean and $b^2(x,t)$ is the infinitesimal variance of X_t. In this lesson, we will instead interpret the SDE in the more literal *strong form*. In the strong interpretation, W_t is the Brownian motion that "drives" X_t. The position at time t, which is X_t, depends on the Brownian driver (or *forcing*) up until that time. That is X_t is a function of $W_{[0,t]}$. The SDE is to be interpreted in the integral sense

$$X_T - X_0 = \int_0^T a(X_t,t)\,dt + \int_0^T b(X_t,t)\,dW_t \ . \tag{7}$$

The first integral on the right is an ordinary Riemann integral defined as in ordinary calculus. This makes sense because $a(X_t,t)$ is a continuous function of t (assuming $a(x,t)$ is a continuous function of x and t) and dt integrals are defined for continuous integrands. Keep in mind that X_t is random, so the value of the integral is also random.

The second integral is an Ito integral with respect to Brownian motion. The theory of Ito integration has to cover this important case – the integrand $F_t = b(X_t,t)$ is a continuous but not differentiable function of t. The mathematical term *regularity* refers to the degree of smoothness (number of derivatives) or the amount by which F_t can change in a small interval of time. Regularity is *qualitative*, which means that it does not matter what the constant in the inequality is, only that there is a constant. Whatever is supposed to converge should converge no matter what (finite) value the constant has.

A continuous function $b(x)$ is *Lipschitz continuous* if there is a C so that

$$|b(y) - b(x)| \leq C\,|y - x| \ .$$

A common theory of SDE like (6) applies under the hypothesis that the coefficients $a(x,t)$ and $b(x,t)$ are Lipschitz continuous functions of x. We we assume b is Lipschitz and we ignore the less important dependence of b on t (assume, for example, that b depends on x but not t). In this case, some regularity of $F_t = b(X_t)$ comes from the regularity of X_t. For a diffusion,

$$\mathrm{E}\Big[\,(\Delta X)^2 \mid \mathcal{F}_t\Big] = O(\Delta t) \ .$$

When b is Lipschitz continuous, and from the properties of big Oh, it follows that

$$\mathrm{E}\Big[\,(\Delta F)^2 \mid \mathcal{F}_t\Big] = O(\Delta t) \ , \tag{8}$$

where $F_t = b(X_t)$. We will make a definition of the Ito integral (1) under the hypotheses that X_t is a diffusion and a martingale and that F_t has the regularity property (8).

3 Borel Cantelli lemma

We will prove that the sequence $Y_T^{(h_n)}$ converges as $n \to \infty$ and $h_n = 2^{-n}$. We need a way to show that limits exist without calculating the limit explicitly. We need a way that applies to random sequences.

Suppose y_n is a sequence of numbers and we want to show the limit exists:

$$y = \lim_{n \to \infty} y_n \ .$$

One way is to study the differences $z_n = y_{n+1} - y_n$ and show that the infinite sum converges

$$\sum_1^\infty |z_n| < \infty \ . \tag{9}$$

This shows that the partial sums converge to the infinite sum, which defines the limit:

$$y_n = y_0 + \sum_{k=1}^n z_k \quad \overset{n \to \infty}{\longrightarrow} \quad y = y_0 + \sum_1^\infty z_k \ .$$

In practice, the expression for z_n may be complicated and the sum hard to calculate. We try instead to find simple bounds of the form $|z_n| \leq a_n$, where the numbers a_n are simple enough that

$$\sum_1^\infty a_n < \infty$$

is a direct explicit calculation. If this works, then the z_n sum is finite because

$$\sum_1^\infty |z_n| \leq \sum_1^\infty a_n < \infty \ .$$

Arguments like this are not quite enough for random sequences like $Y_n = Y_T^{(h_n)}$. The Y_n are random (think Gaussian, though they aren't actually Gaussian) and the differences $Z_n = Y_{n+1} - Y_n$ are random too. If Z_n is Gaussian, there is no bound of the form

$$|Z_n| \leq a_n \ .$$

At least, no bound that it true *almost surely* (i.e., with probability one). No matter how large a_n is, there is some tiny chance $|Z_n|$ is larger.

The *Borel Cantelli lemma* is the fact that convergence follows from

$$\mathrm{E}[\,|Z_n|] \leq a_n \ , \quad \sum_1^\infty a_n < \infty \ . \tag{10}$$

(Warning, the Borel Cantelli lemma is usually stated in a different but equivalent way.) To see this define the random sum (an infinite sum in the sense that there are infinitely many terms but a finite sum in the sense that the sum should be finite).

$$S = \sum_{1}^{\infty} |Z_n| \ .$$

In this sum, we set $S = \infty$ if the sum is infinite. The following fact is called the *monotone convergence theorem* (you may know it in a more general form):

$$\mathrm{E}[\,S] = \sum_{1}^{\infty} \mathrm{E}[\,|Z_n|] \ .$$

If the inequality (10) is satisfied, then $\mathrm{E}[\,S] < \infty$.

The Borel Cantelli argument is to argue that if $\mathrm{E}[\,S] < \infty$ then $S < \infty$ almost surely. This means that $\Pr(S = \infty) = 0$. If $\Pr(S = \infty) = \epsilon > 0$, then

$$\begin{aligned} \mathrm{E}[\,S] &= \mathrm{E}[\,S \mid S = \infty] \cdot \Pr(S = \infty) \ + \ \mathrm{E}[\,S \mid S < \infty] \cdot \Pr(S < \infty) \\ &\geq \infty \cdot \epsilon \\ &= \infty \ . \end{aligned}$$

If $S < \infty$ then $\sum |Z_n| < \infty$, which implies that the the limit of Y_n exists.

In the present application, will calculate an inequality

$$\mathrm{E}\Big[\, \Big| Y^{(h_{n+1})} - Y^{(h_n)} \Big| \Big] \leq a_n \ .$$

We will use the Cauchy Schwarz inequality, and first calculate

$$\mathrm{E}\Big[\, \Big(Y^{(h_{n+1})} - Y^{(h_n)} \Big)^2 \Big] \leq C_T h_n \ . \tag{11}$$

Cauchy Schwarz implies that we can take (The two numbers C_T are not the same, but they both are "constants" that depend on T and on the problem but not on n.)

$$a_n = \sqrt{C_T h_n} = C_T \left(\sqrt{2} \right)^{-n} \ .$$

This verifies the hypothesis (10) and proves that the limit exists.

4 Convergence for the Ito integral

This section explains the calculation (11). This calculation exposes the reason the Ito integral makes sense. It shows how the hypotheses (X a martingale, F_t adapted) come in. It is the most important thing in this Lesson. We compare the "Riemann sum" (2) with $\Delta t = h_n = 2^{-n}$ with the sum with $\Delta t = \frac{1}{2} h_n$. The sum with the smaller Δt has two intervals for each interval in the sum for the larger Δt. One Δt interval is $[t_k, t_{k+1}]$. The summand in the Δt sum for this interval is

$$F_{t_k} (X_{t_{k+1}} - X_{t_k}) \ .$$

For the smaller Δt, this interval is broken into two halves:

$$[t_k, t_{k+1}] = [t_k, t_{k+\frac{1}{2}}] \cup [t_{k+\frac{1}{2}}, t_{k+1}]$$

This notation is convenient. If $t_k = kh_n$, then $t_{k+\frac{1}{2}} = (k + \frac{1}{2})h_n$ makes sense. For the smaller Δt, this summand is replaced by two, one representing each half interval:

$$F_{t_k}(X_{t_{k+1}} - X_{t_k}) \;\longrightarrow\; F_{t_k}(X_{t_{k+\frac{1}{2}}} - X_{t_k}) + F_{t_{k+\frac{1}{2}}}(X_{t_{k+1}} - X_{t_{k+\frac{1}{2}}})$$

Therefore (this might be slightly wrong for the largest $t_k < T$, which we will come back to later)

$$Y_T^{(h_{n+1})} - Y_T^{(h_n)} = \sum_{t_k \leq T} V_k \;,$$

where

$$V_k = F_{t_k}(X_{t_{k+\frac{1}{2}}} - X_{t_k}) + F_{t_{k+\frac{1}{2}}}(X_{t_{k+1}} - X_{t_{k+\frac{1}{2}}}) - F_{t_k}(X_{t_{k+1}} - X_{t_k}) \;.$$

Some algebra shows that

$$V_k = \left(F_{k+\frac{1}{2}} - F_k\right)\left(X_{k+1} - X_{k+\frac{1}{2}}\right) \;.$$

The left side of (11) is

$$\mathrm{E}\left[\left(\sum_{t_k<T} V_k\right)^2\right] = \sum_{t_j<T}\sum_{t_k<T} \mathrm{E}[\,V_j V_k] \;.$$

The *diagonal* terms on the right, the ones with $j = k$, are $\mathrm{E}\big[\,V_k^2\big]$. We will look at these below. But first, we explain why the *off diagonal* terms, then ones with $j \neq k$, have

$$\mathrm{E}[\,V_j V_k] = 0 \;.$$

If $j \neq k$, then either $j > k$ or $k > j$. Without loss of generality, suppose $k > j$. The off diagonal expectation is zero because X_t is a martingale and F_t is adapted. We said $\mathcal{F}_{t_{k+\frac{1}{2}}}$ is the information you get from knowing all of history up to time $t_{k+\frac{1}{2}}$. At this time, all the quantities in $V_j V_k$ are known except X_{k+1}. The values $X_{j+\frac{1}{2}}$, X_{j+1}, and $X_{k+\frac{1}{2}}$ are known because there are part of $\mathcal{F}_{k+\frac{1}{2}}$. The values F_j, $F_{j+\frac{1}{2}}$, F_k and $F_{k+\frac{1}{2}}$ are known because F_t is adapted. This is the definition of "adapted" – knowing X_s for $0 \leq s \leq t$ determines F_s for $0 \leq s \leq t$. Since X_t is a martingale and the interval $[t_{k+\frac{1}{2}}, t_{k+1}]$ is in the future of $t_{k+\frac{1}{2}}$,

$$\mathrm{E}\Big[\,X_{k+1} - X_{t_{k+\frac{1}{2}}} \mid \mathcal{F}_{k+\frac{1}{2}}\Big] = 0 \;.$$

The tower property from Lesson 2 says that the expected value of the expected value in $\mathcal{F}_{k+\frac{1}{2}}$ is the expected value:

$$\mathrm{E}[\,V_j V_k] = \mathrm{E}\Big[\,\mathrm{E}\Big[\,V_j V_k \mid \mathcal{F}_{k+\frac{1}{2}}\Big]\Big]$$

All the terms in V_j are known in $\mathcal{F}_{k+\frac{1}{2}}$, so

$$\mathrm{E}\Big[\, V_j V_k \mid \mathcal{F}_{k+\frac{1}{2}} \Big] = V_j \mathrm{E}\Big[\, V_k \mid \mathcal{F}_{k+\frac{1}{2}} \Big] \ .$$

Similarly, F_k and $F_{k+\frac{1}{2}}$ are known in $\mathcal{F}_{k+\frac{1}{2}}$, so

$$\mathrm{E}\Big[\, V_k \mid \mathcal{F}_{k+\frac{1}{2}} \Big] = \Big(F_{k+\frac{1}{2}} - F_k \Big) \mathrm{E}\Big[\Big(X_{k+1} - X_{k+\frac{1}{2}} \Big) \mid \mathcal{F}_{k+\frac{1}{2}} \Big] \ .$$

Finally, because X_t is a martingale,

$$\mathrm{E}\Big[\Big(X_{k+1} - X_{k+\frac{1}{2}} \Big) \mid \mathcal{F}_{k+\frac{1}{2}} \Big] = 0 \ .$$

This shows that the off diagonal expectations are zero.

The diagonal terms have the form (using the tower property and the fact that $F_{k+\frac{1}{2}}$ is known in $\mathcal{F}_{k+\frac{1}{2}}$)

$$\begin{aligned}
\mathrm{E}\big[V_k^2\big] &= \mathrm{E}\Big[\Big(X_{k+1} - X_{k+\frac{1}{2}} \Big)^2 \Big(F_{k+\frac{1}{2}} - X_k \Big)^2 \Big] \\
&= \mathrm{E}\Big[\mathrm{E}\Big[\Big(X_{k+1} - X_{k+\frac{1}{2}} \Big)^2 \Big(F_{k+\frac{1}{2}} - X_k \Big)^2 \mid \mathcal{F}_{k+\frac{1}{2}} \Big] \Big] \\
&= \mathrm{E}\Big[\mathrm{E}\Big[\Big(X_{k+1} - X_{k+\frac{1}{2}} \Big)^2 \mid \mathcal{F}_{k+\frac{1}{2}} \Big] \Big(F_{k+\frac{1}{2}} - F_k \Big)^2 \Big] \ .
\end{aligned}$$

The inner expectation is $O(\Delta t) = O(h_n)$ because X_t is a diffusion. Therefore

$$\mathrm{E}\big[V_k^2\big] \leq O(\Delta t) \mathrm{E}\Big[\Big(F_{k+\frac{1}{2}} - F_k \Big)^2 \Big] \ .$$

The regularity hypothesis (8) on F_t implies that $\mathrm{E}\Big[\Big(F_{k+\frac{1}{2}} - F_k \Big)^2 \Big] = O(\Delta t)$, so

$$\mathrm{E}\big[V_k^2\big] = O(\Delta t^2) \ .$$

This is the hard part. All the hypotheses have been used – regularity of F_t, X_t being a diffusion, F_t being known in $\mathcal{F}_t$, and X_t being a martingale.

Now we just put stuff together. We drop the off diagonal terms and use the inequality above for the diagonal terms. The result is

$$\mathrm{E}\Big[\left(Y_{n+1} - Y_n\right)^2 \Big] \leq C \sum_{t_k < T} \Delta t^2 \ .$$

Note that

$$\sum_{t_k < T} \Delta t \leq T \ ,$$

so

$$\mathrm{E}\Big[\left(Y_{n+1} - Y_n\right)^2 \Big] \leq C T \Delta t \ .$$

The Cauchy Schwarz inequality then gives

$$\mathrm{E}[\, |Y_{n+1} - Y_n|] \leq C_T \sqrt{\Delta t} = C_T \left(\sqrt{2} \right)^{-n} \ .$$

The Borel Cantelli lemma then shows that the limit of the Y_n exists.

5 Homework 3, Problem 7

Here is background for problem 7 of homework 3. In probability, an *event* is something that may or may not happen at random. If A is an event, then $\Pr(A)$ is the probability that A happens. If B is another event, then $A \cap B$ is the event that both A and B happen. The event $A \cup B$ is the event that at least one of the events happened (or both). We will see that events may be thought of as sets, so $A \cap B$ is the intersection of A and B (the people in both A and B) and $A \cup B$ is the union of A and B (the people in at least one of the events A, B).

The *indicator function*, written $\mathbf{1}_A$ or χ_A,[2] is a random variable associated with A. The indicator function is equal to 1 if A happens and 0 if A does not happen. Therefore

$$\mathrm{E}[\,\mathbf{1}_A] = \Pr(A)\ .$$

The indicator function is useful because you can add them and do other algebra. For example, for any two events,

$$\mathbf{1}_A + \mathbf{1}_B = \mathbf{1}_{A\cup B} + \mathbf{1}_{A\cap B}\ .$$

(Check: if neither A nor B happens, then both sides are zero. If A happens but not B, then the left side is $1 + 0 = 1$. On the right, $A \cup B$ happens but $A \cap B$ does not, so the right side is $1+0 = 1$. If both A and B happen then both sides are 2.)

Suppose A is some event that only concerns a Brownian motion path W. We say $W \in A$ if A happens and $W \notin A$ if A does not happen. For example, if A is the event $W_t \leq 1$ for $0 \leq t \leq 2$, then $W \in A$ if $W_t \leq 1$ for $0 \leq t \leq 2$. The indicator function is a function of W. In this case

$$\mathbf{1}_A(W) = \begin{cases} 1 & \text{if } W_t \leq 1 \text{ for all } 0 \leq t \leq 2 \\ 0 & \text{if } W_t > 1 \text{ for some } t \text{ between 0 and 2.} \end{cases}$$

Suppose A_n is an infinite sequence of events. Let N be the number of events A_n that happen (example below). This is a random variable. Then

$$\mathrm{E}[\,N] = \sum_1^\infty \Pr(A_n)\ .$$

If $N = \infty$, we say the events A_n happen *infinitely often* and abbreviate it as *i.o.* For example, if A_n is an event involving Brownian motion, we write "$W \in A_n$ i.o" if there are infinitely many n with $W \in A_n$. If $N < \infty$ we say "$W \in A_n$ f.o" (for "finitely often"). The *Borel Cantelli lemma* – the form given in most probability books, is

$$\text{if } \sum_1^\infty \Pr(A_n) < \infty \quad \text{then} \quad A_n \text{ happens finitely often, almost surely.}$$

[2] The letter χ is the Greek letter "chi". Here it stands for *characteristic function.* In probability, the characteristic function is something else, so we call this only the indicator function. Nevertheless, we sometimes use χ to represent the indicator function.

Said in a slightly different way

$$\sum_1^\infty \Pr(A_n) < \infty \quad \Longrightarrow \quad \Pr(\, A_n \text{ happens i.o. }) = 0 \;.$$

You can prove this by using the "obvious" formula (think about it for at most an hour)

$$\mathrm{E}[\,N] = \mathrm{E}\left[\sum_1^\infty \mathbf{1}_{A_n}\right] \le \sum_1^\infty \Pr(A_n) \;.$$

The last sum on the right over-counts because more than one event happen might happen.

Let $s > 0$ be some "speed". Problem 7 concerns the size of W_t for large t. The goal is to show that for any s then almost surely there it a T_s so that

$$W_t < st \text{ for } t > T \;.$$

This may be done by finding a family of events related to hitting times A_n so that only finitely many of the A_n happens. If A_n happens finitely often, then there is an M so that A_n does not happen if $n > M$. The "upper bound" M may be random. In our example, M will be related to T, the last time $W_t \ge st$.

A final hint: you can show that an integral is small without working the integral. This is important, since most integrals cannot be "solved" explicitly. For example, consider the integral

$$I = \int_0^r e^{-\frac{a^2}{2t}}\, dt \;.$$

There is no formula for I, but clearly ($r \cdot \max$ is the area of a box that contains the curve)

$$I < r \cdot \max_{t \le r} e^{-\frac{a^2}{2t}} = re^{-\frac{a^2}{2r}} \;.$$

This is "exponentially small" as $r \to 0$ because exponentials beat powers.