# Topics for Today

1.) Weak Convergence / Donsker's Theorem

2.) Kolmogorov's Backward + Forward Equations

3.) "White Noise" + Stochastic Differential Equations

4.) The Ito Integral

- We showed last time that the 'right' way to define a continuous-time version of a random walk led to a process (called the 'Wiener Process'), which was continuous, but nowhere differentiable.

- This raises a number of thorny mathematical issues:

1.) What exactly does it mean for a discrete-time process to 'converge' to a continuous-time process? (They live in different 'spaces'!) We just showed their means and variances converged. What about their sample paths?

2.) How do we know a random process is continuous? (It's random!)

3.) If $W_t$ is not differentiable, in what sense (if any) can we use its innovations as the underlying i.i.d shocks driving a stochastic analog of a differential equation?

4.) In discrete time, we move interchangeably between sums and differences:

$$X_t = X_{t-1} + \varepsilon_t \quad \Big\} \text{ discrete 'differential eq.'}$$

$$X_t = \sum_{j=0}^{t} \varepsilon_j \quad \Big\} \text{ discrete 'integral eq.'}$$

This correspondence is like a discrete version of the "Fundamental Theorem of Calculus": $\boxed{F(b) - F(a) = \int_a^b F'(x)\,dx}$

Does a similar result apply to stochastic cont.-time processes? It's not obvious, since $\frac{dW}{dt}$ does not exist!

# Summary from Last Time

- A Wiener process (or standard Brownian motion) is a continuous-time process having the following 3 properties:

a.) Continuous sample paths

b.) Stationary i.i.d. 'increments'

c.) $W_t \sim N(0, t) \; \forall t \quad [W_0 = 0]$

- Therefore, over any discrete time interval, $\Delta t$, we can write

$$\Delta W_t = \varepsilon_t \sqrt{\Delta t} \qquad \varepsilon_t \sim N(0,1)$$

- Notation: As $\Delta t \to 0$, we write $dW = \varepsilon \sqrt{dt}$

$$\Rightarrow E[dW] = E[\varepsilon \cdot \sqrt{dt}] = 0$$

$$E[(dW)^2] = E[\varepsilon^2 \cdot dt] = dt$$

- More generally, we can define a Brownian motion with drift, $\mu$, and volatility, $\sigma$, as follows

$$X_t = X_0 + \mu \cdot t + \sigma \cdot W_t$$

- Theorem: If a continuous-time process, $X_t$, has continuous sample paths with stationary i.i.d. increments, then it must be a Browian motion.

That is, continuity + stationary i.i.d increments $\Rightarrow$ Normality (Levy's Theorem)

- Implication: If the data indicate a process is non-Gaussian, then it cannot be a Wiener process. (However, you might be able to convert it into one with a suitable transformation).

# A Caveat

- Before starting, we need to be aware of an important caveat concerning all our ~~results~~ results.
- Consider the following silly (but useful) example:

**Example**: Let $U$ be a r.v. which is uniformly distributed on $[0, 1]$, and consider the following 2 random processes on the continuous time interval $[0, 1]$:

1.) $X_t = 0$ for all $t \in [0, 1]$

2.) $Y_t = \begin{cases} 1 & \text{if } t = U \\ 0 & \text{otherwise} \end{cases}$

- Since $\Pr(U = t) = 0 \ \forall t$, $X_t$ and $Y_t$ have the same distributions. However, $X_t$ and $Y_t$ are different processes. In particular,

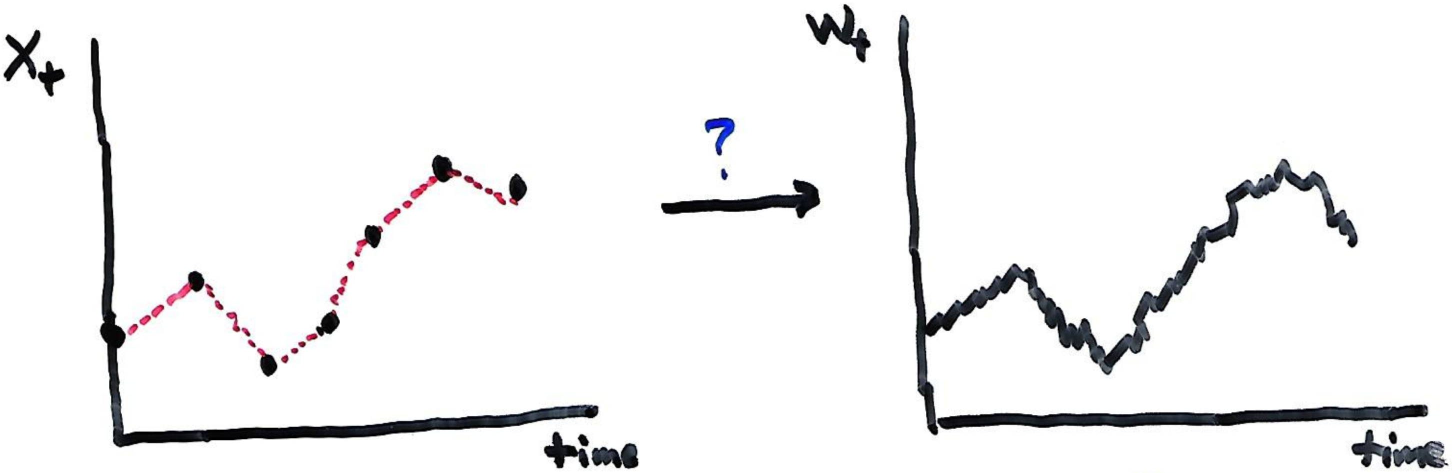
$$\Pr(X_t = 0 \ \forall t) = 1$$
$$\Pr(Y_t = 0 \ \forall t) = 0$$

## Implications

- Since we are dealing with random things, there are no guarantees
- All our statements about the properties of a random process mean they hold 'almost everywhere' or 'almost surely'
- In this way, we can say the above 2 processes are equal (They only differ on a set of measure zero).
- This qualifies our earlier claim about the continuity of a Wiener process's sample paths: more precisely, The sample paths of a Wiener process are almost surely continuous!

# Weak Convergence / Donsker's Theorem

- In what sense do the sample paths of a discrete-time random walk converge to those of a Wiener process?




continuous-time interpolation of a discrete random walk

realization of a Wiener process

- In statistics, you learned that there are many ways we can define the convergence of a random variable. One of them is:

## Convergence in Distribution

Let $X_n$ be a sequence of random variables with CDF $F_n(x_n)$. Then we say that $X_n$ converges in distribution to a r.v. $X$ with CDF $F(x)$ if $\lim_{n\to\infty} |F_n(x_n) - F(x)| = 0$ at all continuity points of $F(x)$.

[Note: This is a statement about the prob. dists. of $X_n$ & $X$, not the values of $X_n$ and $X$].

- Weak convergence can be interpreted as a function space analog of convergence in distribution (i.e., it applies to stochastic processes rather than random variables.

**Definition**: Let $(\Omega, \mathcal{F})$ be a 'measure space', where $\Omega$ is a set of continuous functions, $\mathcal{F}$ is a $\sigma$-algebra of its subsets, and $\mu_n$ be a sequence of prob. measures on $(\Omega, \mathcal{F})$. Then $\mu_n$ converges weakly to $\mu$ (denoted $\mu_n \Rightarrow \mu$) if for each continuous, bounded function $f$ we have

$$\int f(x)\mu_n(dx) \longrightarrow \int f(x)\mu(dx)$$

Equivalently, if we let $X_n$ be the random process associated with $\mu_n$ (i.e., $\mu_n(A) = \Pr(X_n \in A)$), then weak convergence implies $E f(X_n) \to E f(X)$.

**Theorem (Donsker)**: Let $\varepsilon_t \sim$ i.i.d with $E\varepsilon_t = 0$ and $E\varepsilon_t^2 = 1$. Let $X_T = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_T$ be the $T^{th}$ partial sum. Form the following linear continuous-time interpolation of $X_T$ on the interval $[0, 1]$

$$W_t^n = \begin{cases} \frac{1}{\sqrt{n}} X_m & \text{if } t = \frac{m}{n} \quad (0 \le m \le n) \\ \text{linear} & \text{if } t \in [m/n, \frac{(m+1)}{n} \end{cases}$$

Then, as $n \to \infty$ $W_t^n \Rightarrow W_t$, where $W_t$ is a Wiener process.

[Note: No assumption was made about the distribution of $\varepsilon_t$ ]

- Believe it or not, this is actually a really useful + practical result. Why? Because the function $f$ can be used to define many useful properties of a path (e.g., the maximum value reached before time $T$).
- Donsker's Theorem then tells us that the sample path properties of a random walk can be approximated by those of a Wiener process (and vice versa). This means computer programs can be used to approx. the path properties of $W_t$ (or, going the way) analytically derived results for $W_t$ approx. those for $X_t$!

# Kolmogorov's Backward + Forward Equations

- Last time we saw that Markov processes are a convenient type of stochastic process.

- Markov processes are fully characterized by their <u>transition probabilities</u> (and an initial condition).

<u>Example</u>: A discrete-state/discrete-time process (called a <u>Markov Chain</u>) is characterized by its transition probability matrix, $P_{ij}$.

- A standard Wiener process is Markov, and from our results so far, its transition probabilities are given by:

$$f(y,t \mid x,s) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left\{ \frac{-(y-x)^2}{2(t-s)} \right\}$$

= prob. that $W_t$ will be at $y$ at time $t$ given it starts at $x$ at time $s < t$.

- By taking derivatives, one can readily verify that $f$ is the solution of the following 2 (partial) differential equations:

1.) $\frac{\partial f}{\partial s} = -\frac{1}{2} \frac{\partial^2 f}{\partial x^2}$ } Backward Eq.

2.) $\frac{\partial f}{\partial t} = \frac{1}{2} \frac{\partial^2 f}{\partial y^2}$ } Forward Eq.

- The backward eq. conditions on a future time+value, and describes how the distribution of this value changes with different initial conditions. It is solved 'backwards'.

- The forward eq. conditions on an initial distribution, and describes how this distribution evolves over time. It is solved 'forwards'.

- Later we will derive more general versions of these equations using a different method, based on <u>Ito's Lemma</u>.
- Physicists refer to the forward eq. as the "Fokker-Planck" eq.

# "White Noise" + Stochastic Differential Equations

- Discrete time processes are built from discrete-time i.i.d. random variables, $\varepsilon_t$

$$X_t = \underbrace{\alpha_1 X_{t-1} + \alpha_2 X_{t-2}}_{\text{propagation mech.}} + \underbrace{\varepsilon_t}_{\text{impulse}}$$

- It would be nice if the same were true for cont.-time processes
- The "sum" of a cont.-time i.i.d. process makes sense. That's what a Wiener process is: $W_t = \int^t \varepsilon_s \, ds$
- **Problem**: We can't differentiate this to get: $\frac{dW}{dt} = \varepsilon_t$

  $W_t$ is too erratic. The stochastic analog of the Fundamental Theorem of Calculus breaks down!

- Next time I will show how we can develop new calculus rules that make this work (in a certain sense).
- For now, I just want to alert you to the fact that there is a mathematically sophisticated way of defining the above derivative by appealing to the notion of "generalized functions", which are motivated by the following

  Dirac $\delta$-function

$$\delta(x) = \begin{cases} +\infty & x = 0 \\ 0 & x \neq 0 \end{cases} \qquad \int_{-\infty}^{\infty} \delta(x)\,dx = 1$$

$$\Rightarrow \int_{-\infty}^{\infty} f(x)\,\delta(x)\,dx = f(0)$$

- Engineers + physicists use these all the time. Using $\delta$-functions we can say that $\frac{dW}{dt}$ "exists", and engineers call it "white noise".