Stats 150, Fall 2019
Lecture 24, Tuesday, 11/26/2019

CLASS ANNOUNCEMENTS: 1) We have 3 lecture left and then RRR week, and review sessions in RRR week will be at the normal class location and times.
2) Regarding material, we will begin a brief introduction to Brownian motion and related continuous parameters and continuous space processes.

From Durett, let $S_n := X_1 + \cdots + X_n$ as the sum of IID copies of X. Assume $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = 1$ (with $\sigma^2 = 1$ by scaling). Thanks to Wald, we know that for a stopping time T of a sequence $X_1, X_2, \ldots$, we have that

$$\mathbb{E}S_T = (\mathbb{E}T)(\mathbb{E}X),$$

if $\mathbb{E}T < \infty$ (Wald's first identity). As an aside, recall Wald's second identity that $\mathbb{E}S_T^2 = (\mathbb{E}T)(\mathbb{E}X^2)$ again if $\mathbb{E}T < \infty$.

We know that $M_n : S_n^2 - n$ is a martingale implies Wald's second identity, provided that T is **bounded**, as we have found from last class.

Taking $0 = \mathbb{E}M_0 = \mathbb{E}M_{T\wedge n}, \forall_n = \mathbb{E}M_T$, if $\mathbb{P}(T \leq b) = 1$. Then take $n \geq b$ which gives that $T \wedge n = T$. This implies $M_{T\wedge n} = M_T$, and so $\mathbb{E}M_T = 0$. Then $\mathbb{E}(S_T^2 - T) = 0$ which implies $\mathbb{E}S_T^2 = \mathbb{E}T$.

The issue arises as to how we can push this to unbounded T. Pitman says this is much harder than it looks. The issue is that we have a sequence of random variables $M_{n\wedge T}$ with $\mathbb{E}M_{n\wedge T} \equiv 0$ and $\mathbb{P}(M_{n\wedge T} \to M_T) = 1$. Then $M_{n\wedge T} = M_T$ for all large n on $(T < \infty)$.

In general, we know that

$$\mathbb{P}(Y_n \to Y) = 1,$$

and $\mathbb{E}Y_n$ has a limit does NOT imply $\mathbb{E}Y = \lim Y_n$.
Our key example of this was the "double-or-nothing" example with

$$\mathbb{E}Y_n \equiv 1, \mathbb{P}(Y_n \geq 0) = 1,$$

so that

$$\mathbb{P}(Y_n = 0 \text{ for all large } n) = 1 \implies \mathbb{P}(Y_n \to 0) = 1,$$

but $\mathbb{E}Y_n \equiv 1 \not\to 1$.
For switching limits in integrals (swapping the operations $\mathbb{E}$ and lim), we would need Math 202 or Math 105 for analysis and Math 218 and Stats 205 for the statistics portion to properly address these limits.
Let us assume the Monotone Convergence Theorem, in that if $0 \leq X_n \uparrow X$, then $0 \leq \mathbb{E}X_n \uparrow \mathbb{E}X$, allowing $+\infty$ as a value for $\mathbb{E}X$. In fact, this is the definition of $\mathbb{E}X$ for $X \geq 0$ in an advanced course. This implies

$$\mathbb{E}\sum_n Y_n = \sum_n \mathbb{E}Y_n,$$

for $Y_n \geq 0$ where provided we have nonnegative variables, we can perform these swaps.

Remark:

$$X_n \geq 0, \mathbb{P}(X_n \to X) \text{ does NOT imply } \mathbb{E}X_n \to \mathbb{E}X$$

Lemma 0.1. (Fatou's Lemma)

$$X_n \geq 0, \mathbb{P}(X_n \to X) \implies \mathbb{E}X \leq \lim_n \mathbb{E}X_n,$$

where $\mathbb{E}X < \infty$ if the limit is so.

Theorem 0.2. (Dominated Convergence Theorem) If $\mathbb{P}(|X_n| \leq Y) = 1$ for all n and if we can check $\mathbb{E}Y < \infty$ and $\mathbb{P}(X_n \to X) = 1$, then

$$\mathbb{E}X_n \to \mathbb{E}X, \text{ and } |\mathbb{E}X| \leq \mathbb{E}Y,$$

where Y is our dominating variable.

Finally, with the dominated convergence theorem, we recall the notion of $X_n \xrightarrow{L^2} X$ means $\mathbb{E}(X_n - X)^2 \to 0$ and $\mathbb{E}X_n^2 < \infty$ for all n. Take $L^2 = \{\text{all } X : \mathbb{E}X^2 < \infty\}$ which is almost like Euclidean $\mathbb{R}^n$ just like infinite-dimensionality.

Key Fact: The key fact here is that L^2 is complete. That is, if

$$\lim_{m,n\to\infty} \mathbb{E}(X_m - X_n)^2 = 0,$$

then

$$\exists_{X \in L^2} \text{ and } X_n \xrightarrow{L^2} X.$$

If we have a Cauchy sequence in $\mathbb{R}^n$, then a Cauchy sequence converges. This fact works in a sequence of random variables. If we know that L^2 is complete (via the DCT), then the problem is easy.
L^2 has an inner product $\langle X, Y\rangle := \mathbb{E}(XY)$ and note that $\langle X, Y\rangle \leq \sqrt{\langle X, X\rangle}\sqrt{\langle Y, Y\rangle}$ via Cauchy-Schwarz, so

$$\langle X, Y\rangle := \mathbb{E}(XY) \leq \sqrt{\mathbb{E}X^2}\sqrt{\mathbb{E}Y^2}.$$

Then if M_n is a martingale, then M_n has increments:

$$\mathbb{E}M_m \underbrace{(M_n - M_m)} = 0, \forall_{m<n},$$

which follows immediately from the martingale property by conditioning.

1 Introduction to Brownian Motion

Let's continue with $S_n = X_1 + X_2 + \cdots + X_n$ as a random walk with mean 0 and variance 1 increments (simply for convenience). Then $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = 1$. We can go a long way.

Consider the simple random walk (SRW) with -1 with probability $\frac{1}{2}$ and $+1$ with probability $\frac{1}{2}$.
We have the two "martingale" facts $\mathbb{E}S_n = 0$ and $\mathbb{E}S_n^2 = n$. Now we may ask, $\mathbb{E}S_n^3 =$
Pitman jokes that if we do not notice symmetry, we will be punished on the final. Symmetry means that there is some equality in distribution. Our homework is designed to show how clever we can be with exploiting symmetry. Notice that we have:

$$X \stackrel{d}{=} -X,$$

where we can swap roles of -1 and $+1$ and get the same simple random walk. This implies that

$$(X_1, X_2, \ldots, X_n) \stackrel{d}{=} (-X_1, -X_2, \ldots, -X_n).$$

Then applying any function ψ to the coordinates gives:

$$\psi(X_1, X_2, \ldots, X_n) \stackrel{d}{=} \psi(-X_1, -X_2, \ldots, -X_n).$$

Take $\psi(X_1, \ldots, X_n) := X_1 + \cdots + X_n$, and we see that

$$\boxed{S_n \stackrel{d}{=} -S_n}.$$

Then we can get:

$$(S_n)^3 \stackrel{d}{=} (-S_n)^3 = -S_n^3.$$

We start with an n-dimensional vector, we apply a function (sum) and take the expectation. Of course, we have the obvious bound $|S_n| \leq n$, so taking the expectation is legal, and we have:

$$\mathbb{E}S_n^3 = \mathbb{E}(-S_n^3) = -\mathbb{E}S_n^3 \implies \mathbb{E}S_n^3 = 0.$$

We can continue inductively that by symmetry, we have that all odd moments of S_n are 0.
We have the parity issue that for $n \in \mathbb{Z}$,

$$\mathbb{P}(S_{2n} \text{ is even}) = 1, \text{ and } \mathbb{P}(S_{2n-1} \text{ is odd}) = 1.$$

Now we should ask, what is $\mathbb{E}S_n^4$? We can evaluate and expand:

$$\begin{aligned}
\mathbb{E}S_n^4 &= \mathbb{E}(X_1 + X_2 + \cdots + X_n)^4 \\
&= n(\mathbb{E}X^4) + \binom{4}{2}\binom{n}{2}\left(\mathbb{E}X^2\right)^2 + \underbrace{\binom{4}{3} \cdots + (\mathbb{E}X)(\mathbb{E}X^3) + \cdots}_{=0}.
\end{aligned}$$

Then in our example of the SRW,

$$\mathbb{E}S_n^4 = n + 3n(n-1).$$

We've all been taught the Central Limit Theorem, which says that

$$\frac{S_n}{\sqrt{n}} \xrightarrow{d} B_1 \sim N(0,1),$$

where statisticians often use Z for $N(0,1)$ and $B_1 \stackrel{d}{=} Z$, where we take the variable B for Brownian motion and B_1 to be Brownian motion at time 1.

1.1 Proof Sketch of Central Limit Theorem

We'll talk about the heart of the argument to show that

$$\frac{S_n}{\sqrt{n}} \xrightarrow{d} B_1.$$

Proof. For our simple random walk, we have

$$\begin{aligned}\mathbb{E}\left(\frac{S_n}{\sqrt{n}}\right)^4 &= \mathbb{E}\frac{S_n^4}{n^2} \\ &= \frac{n+3n(n-1)}{n^2} \to 3,\end{aligned}$$

as $n \to \infty$. □

Notice that $Y_n \xrightarrow{d} Y$ which means

$$\mathbb{P}(Y_n \le y) \to \mathbb{P}(Y \le y)$$

for all continuity points of the RHS, as we have defined in a previous lecture. We have this if and only if

$$\mathbb{E}g(Y_n) \to \mathbb{E}g(Y)$$

for all suitably 'nice' g (e.g. bounded and continuous or bounded and differentiable). Here, we see that

$$\mathbb{E}G\left(\frac{S_n}{\sqrt{n}}\right) \to 3, \text{ for } g(x) = x^4.$$

Then from our findings earlier of the moments, we have:

$$\begin{aligned}\mathbb{E}\frac{S_n}{\sqrt{n}} &= 0 \\ \mathbb{E}\left(\frac{S_n}{\sqrt{n}}\right)^2 &= 1 \\ \mathbb{E}\left(\frac{S_n}{\sqrt{n}}\right)^3 &= 0 \\ \mathbb{E}\left(\frac{S_n}{\sqrt{n}}\right)^4 &= \cdots \to 3,\end{aligned}$$

and so on. Notice that B_1 has the limit distribution with:

$$\mathbb{E}\left(\frac{S_n}{\sqrt{n}}\right)^{2m} \to (2m-1)(2m-3)\cdots 5\cdot 3\cdot 2\cdot 1 = (2m-1)!!,$$

which is the double factorial (product of the first m odd numbers).

1.2 Conclusion:

We learned then by computing moments that there is a unique distribution of B_1 with these conditions, given by:

$$\mathbb{E}(B_1^n) = \begin{cases} 0, & n \text{ is odd} \\ (2m-1)!!), & m = 2n \text{ is even}\end{cases}$$

1.3 A Better Technique:

Thanks to Euler and Laplace, we have a better technique and we can handle all moments at once using moment generating functions (MGFs).
Notice that

$$\begin{aligned} \mathbb{E}e^{\theta X} &= \frac{e^{+\theta}+e^{-\theta}}{2} \\ \implies \mathbb{E}e^{\theta S_n} &= \left(\frac{e^{\theta}+e^{-\theta}}{2}\right)^n \\ &= 1+\frac{\theta^2}{2}+ \end{aligned}$$

which them implies

$$\begin{aligned} \mathbb{E}e^{\theta\frac{S_n}{\sqrt{n}}} &= \left(\frac{e^{\theta/\sqrt{n}}+e^{-\theta/\sqrt{n}}}{2}\right)^n \\ &= \left(1+\frac{\theta^2}{2n}+\frac{\theta^4}{mn^2}+\cdots\right)^n, \end{aligned}$$

which converges to Exponential $\left(\theta^2/2\right)$, which we can see by recalling that if $x_n \to x$ then $\left(1+\frac{x_n}{x}\right)^n \to e^x$. Pitman tells us to check and expand

$$\begin{aligned} e^{\theta^2/2} &= 1+\frac{\theta^2}{2}+\frac{\left(\frac{\theta^2}{2}\right)^2}{2!}+\frac{\left(\frac{\theta^2}{2}\right)^3}{3!}+\cdots \\ &= 1+\underbrace{(\mathbb{E}B_1)}_{=0}\theta+\underbrace{\left(\mathbb{E}B_1^2\right)}_{=1}\frac{\theta^2}{2}+\underbrace{\mathbb{E}(B_1^3)}_{=0}\frac{\theta^3}{3!}+\frac{\mathbb{E}(B_1^4)\theta^4}{4!}+\cdots \end{aligned}$$

Take $n=10,000$ steps and consider the plot of $t \to \frac{S_{nt}}{\sqrt{n}}$.
At time 1, we have

$$\frac{S_{10,000}}{100} \stackrel{d}{\equiv} B_1.$$

Then at time 2, we have the sum of two independent copies of something that is $\mathcal{N}(0,1)$, which is $\mathcal{N}(0,2)$ via the basic additive property of normal distribution. Lok at the process:

$$\left(\frac{S_{nt}}{\sqrt{n}}, t \geq 0\right) \xrightarrow{d} \left(B_t, t \geq 0\right),$$

where $\xrightarrow{d}$ means convergence in distribution of dinite discrete distributions, and where we take S_{nt} for nt which we find by linear interpolation. In particular,

$$\left(\frac{S_{nt_i}}{\sqrt{n}}, 1 \leq i \leq n\right) \xrightarrow{d} \left(B_{t_i}, 1 \leq i \leq n\right).$$

Then the limit process $(B_t, t \geq 0)$ must have certain properties.
Notice that $\mathbb{E}B_t^2 = t \cdot \mathbb{E}B_1^2$ for positive integers t, so that

$$\boxed{B_t \stackrel{d}{=} t^{1/2}B_1},$$

which we call **Brownian Scaling**. This gives us a 1-dimensional distribution of B_t. To get finite discrete distributions (FDD's), we can take $0 < t_1 < t_2 < \cdots < t_n$ and look at

$$B_t, B_{t_2}-B_{t_1}, \ldots, B_{t_n}-B_{t_{n-1}},$$

which are independent copies of $B_t, B_{t_2-t_1}, B_{t_3-t_2}, \ldots$. We scale it so that the variance increment is the interval length in time (we've hidden the σ at the start).

1.4 Conclusion:

Then for all bounded continuous g and some unbounded g (i.e. product of powers), we have convergence in distribution ($\xrightarrow{d}$) defined as:

$$\mathbb{E}g\left(\frac{S_{nt_1}}{\sqrt{n}}, \frac{S_{nt_2}}{\sqrt{n}}, \ldots, \frac{S_{nt_m}}{\sqrt{n}}\right) \to \mathbb{E}g(B_{t_1}, \ldots, B_{t_m}).$$

So in terms of Brownian motion, we have one final fact. It is possible to construct the limit process B (Brownian motion) with **continuous paths**. This is a theorem due to Weiner in the 1920s, and Pitman warns this is not easy.

Recall that our poisson process holds and jumps; on the other hand, Brownian motion 'wiggles' and are incredibly variable (with probability 1, the paths are **everywhere** continuous but **nowhere** differentiable).

In the next lectures, we'll see how various properties of random walks can be understood as Brownian limits.