Statistics 212: Lecture 12 (March 12, 2025)

Convergence to Brownian Motion in Path Space

Instructor: Mark Sellke

Scribes: Kevin Liu, Haozhe (Stephen) Yang

1 Ways to Get Extra Credit

- Tell Mark about typos in HW/notes
- Future extra credit problems

2 Donsker's Theorem

Theorem 2.1 (Donsker's Theorem). *Let $X_1, X_2, \ldots$ be a simple random walk (i.e., $X_{i+1} - X_i = \pm 1$ iid) for $n \geq 1$. Consider the random function on $[0,1]$, define on $\mathbb{Z}/n$ by*

$$W^{(n)}(k/n) = X_k/\sqrt{n}$$

and interpolated linearly in between. Then $W^{(n)} \xrightarrow{d} Law(BM)$ as $n \to \infty$. In other words, for any bounded continuous $f : C([0,1]) \to \mathbb{R}$, we see $\lim_{n\to\infty} E[f(W^{(n)})] = E[f(B)]$ where B is Brownian motion.

Remark. In fact, this theorem holds for IID sums of any mean 0 and variance 1 random variable. This implies the Central Limit Theorem for IID sums under the same assumptions. Namely fix $\phi : \mathbb{R} \to \mathbb{R}$ that's bounded + continuous. Now consider $f(W) = \phi(W(1))$ which is bounded and continuous from $C([0,1]) \to \mathbb{R}$. This implies $E[\phi(W^{(n)}(1))] \to E[\phi(B(1))$ which is a CLT as ϕ is arbitrary.

Proof. This theorem can be proven using the Wald identities. The idea is to use an explicit coupling between Brownian motion and a simple random walk. We first start with a Brownian motion $B_t \sim BM$ and construct a simple random walk out of it. Consider a sequence of stopping times $\tau_1 < \tau_2 < \cdots$ where each stopping time corresponds to hitting an integer. More formally, τ_{i+1} is the first time $t \geq \tau_i$ with $|B_t - B_{\tau_i}| = 1$. (For the more general statement, one represents any mean 0 variance 1 random variable as a stopped Brownian motion, which is called the *Skorokhod embedding*.)

We claim that $\tau_1, \tau_2 - \tau_1, \tau_3 - \tau_2, \ldots$ are iid each with mean 1. By the Strong Markov property, these are iid. Recall that $B_t^2 - t$ is a martingale, so by a Wald identity, we have $1 = E[B_{\tau_j}^2] = E[\tau_j]$. Thus, by the strong law of large numbers, we can immediately deduce that $\lim_{n\to\infty} \frac{\tau_n}{n} = 1$.

We also see that $B_{\tau_1}, B_{\tau_2}, \ldots$ is a simple random walk. Let $X_j = B_{\tau_j}$. Define $W^{(n)}(k/n) = X_k/\sqrt{n}$ and $B^{(n)}(t/n) = B_t/\sqrt{n}$. We claim $d_{\sup}(W^{(n)}, B^{(n)}) \xrightarrow{p} 0$ (which also implies convergence in distribution). To prove this, we define a "re-parameterized BM" $\tilde{B}$ where

$$\tilde{B}^{(n)}(k/n) = B(\tau_k)/\sqrt{n} = B^{(n)}(\tau_k/n) = W^{(n)}(k/n)$$

where the path is linear in between defined points.

We claim $d_{\sup}(B^{(n)}, \tilde{B}^{(n)}) \xrightarrow{p} 0$. In a bit more detail, we first see that for all $\epsilon > 0$, there exists a random δ where $\sup_{|s-t|\leq\delta} |B^{(n)}(t) - B^{(n)}(s)| \leq \epsilon$. We can choose a deterministic δ_* such that

$$P(\sup_{|s-t|\leq\delta^*} |B^{(n)}(t) - B^{(n)}(s)| \leq \epsilon] \geq 1-\epsilon.$$

(Slight side tangent to expound on the above: For each ϵ, δ, let $A_{\epsilon,\delta} \subseteq C([0,1])$ consist of functions with $\sup_{|s-t|\leq\delta} |B_t - B_s| \leq \epsilon$. Almost surely, we have Brownian motion is continuous, so $\forall\epsilon$, we take the largest δ so $B \in A_{\epsilon,\delta}$. This δ is random ($B_{[0,1]}$-dependent) which is inconvenient, but we can make it deterministic using countable exhaustion. Namely we see that

$$\forall\epsilon, P[\bigcup_{\delta=1/n} A_{\epsilon,\delta}] = 1$$

for Brownian Motion, which implies there exists a deterministic δ_* with

$$P[\bigcup_{\delta\geq\delta_*} A_{\epsilon,\delta}] \geq 1-\epsilon.$$

For this (ϵ, δ_*), if n is large enough, then $P[\sup_{0\leq k\leq n} \frac{|\tau_k - k|}{n} \leq \delta_*] \geq 1-\epsilon$ by the law of large numbers. (Strong LLN implies almost sure convergence of $\sup_{0\leq k\leq n} \frac{|\tau_k - k|}{n}$ to 0 as $n \to \infty$, which also implies convergence in probability.) As a result, we can conclude that

$$P[d_{\sup}(B^{(n)}, \tilde{B}^{(n)}) \leq \epsilon + 2/\sqrt{n}] \geq 1-2\epsilon.$$

We have two sources of error, which is why we have $1-2\epsilon$, and $2/\sqrt{n}$ relaxes the bound for the times in between k/n. Thus, we see $d_{\sup}(B^{(n)}, \tilde{B}^{(n)}) \xrightarrow{p} 0$.

Furthermore, we can pretty easily see that $d_{\sup}(\tilde{B}^{(n)}, W^{(n)}) \leq 2/\sqrt{n}$. Combining these two facts together, we have

$$P(d_{\sup}(B^{(n)}, W^{(n)}) \geq \epsilon] \leq \epsilon$$

so $W^{(n)} \xrightarrow{p} BM$. Whew!

It is intuitive that this convergence in probability implies convergence in distribution. Let's work through it. Fix a $f : C([0,1]) \to \mathbb{R}$ which is continuous and bounded. From the definition of continuity, there exists δ such that $|f(B^{(n)}) - f(W^{(n)})| \leq \epsilon$ if $d_{\sup}(B^{(n)}, W^{(n)}) \leq \delta$. Apriori δ is random and depends on $W^{(n)}$ in addition to ϵ. Similarly to before, given f and any $\epsilon > 0$, we can choose a deterministic δ_* so

$$P[\sup_{W;d_{\sup}(W,B^{(n)})\leq\delta_*} |f(B^{(n)}) - f(W)| \leq \epsilon] \geq 1-\epsilon.$$

So for large n with probability $1-\epsilon$, we have $d_{\sup}(B^{(n)}, W^{(n)}) \leq \delta_*$. Therefore with probability $1-2\epsilon$, we have both

$$d_{\sup}(B^{(n)}, W^{(n)}) \leq \delta_*,$$

$$\sup_{W;d_{\sup}(W,B^{(n)})\leq\delta_*} |f(B^{(n)}) - f(W)| \leq \epsilon.$$

(The first is just convergence in probability as $n \to \infty$ with (ϵ, δ_*) fixed while the latter we just obtained.) Combining shows that

$$\mathbb{P}[|f(B^{(n)}) - f(W^{(n)})| \le \epsilon] \ge 1 - 2\epsilon.$$

We thus see that $f(W^{(n)}) \to f(B^{(n)})$ in probability, which implies $E[f(B^{(n)})] \to E[f(W^{(n)})]$ since f is bounded. □

(Note: one virtue of the upcoming Portmanteau theorem is that it lets us avoid working through the avoid arguments by hand every time. It is left as an exercise to show using this theorem that convergence in probability implies convergence in distribution.)

This theorem introduces more questions:

- How do we think about convergence in distribution? Might not always be able to do this coupling
- Can we prove Donsker's directly from a (multidimensional) CLT, without needing this clever stopping time argument?

We'll say something about these today and next class (after spring break).

Definition 2.2. Let (S, d) be a complete separable metric space. $\mu_n \to \mu$ if $\int f\,d\mu_n \to \int f\,d\mu$ for all bounded and continuous functions $f : S \to \mathbb{R}$.

(**Added after lecture**:) given a probability measure μ on S and a measurable function $g : S \to S'$, the pushforward $g_\#\mu$ is the probability measure on S' given by

$$(g_\#\mu)(A) = \mu(g^{-1}(A))$$

for all measurable subsets $A \subseteq S'$. In other words, if $X \sim \mu$, then $g(X) \sim g_\#\mu$.

Theorem 2.3 (Continuous Mapping Theorem)**.** *If $\mu_n \to \mu$ in distribution and $g : S \to S'$ continuous, then also $g_\#\mu_n \to g_\#\mu$ in distribution. This directly follows from the fact that if $\phi : S' \to \mathbb{R}$ is bounded continuous, then $\phi \circ g : S \to \mathbb{R}$ is as well.*

Theorem 2.4 (Portmanteau Theorem)**.** *The following are equivalent:*

(a) $\mu_n \to \mu$

(b) $\int f\,d\mu_n \to \int f\,d\mu$ *for bounded Lipschitz f.*

(c) $\forall C \subseteq S$ *closed,* $\limsup_{n\to\infty} \mu_n(C) \le \mu(C)$.

(d) $\forall U \subseteq S$ *open,* $\limsup_{n\to\infty} \mu_n(U) \ge \mu(U)$.

(e) $\mu_n(A) \to \mu(A)$ *if A is a measurable set with $\mu(\partial A) = 0$.*

Proof. Some easy implicatures are $(a) \to (b)$ and $(c) \leftrightarrow (d)$. Another one is $(c), (d) \to (e)$. Let the closure be $\bar{A}$ and interior be A°. We see $\mu(\bar{A}) = \mu(A^\circ)$ and $\mu_n(\bar{A}) \ge \mu_n(A^\circ)$ as $\bar{A} \supseteq A \supseteq A^\circ$. Then we see $\limsup \mu_n(\bar{A}) \le \mu(\bar{A}) = \mu(A^\circ) \le \liminf \mu_n(A^\circ)$, so we easily concluce that everything is equal.

For $(b) \to (c)$, the idea is to approximate the indicator I_C from above by continuous functions (think about a hump with round corners). An explicit construction in a general metric space is

$$g_\epsilon(x) = \frac{d(x, (C^\epsilon)^c}{d(x, (C^\epsilon)^c) + d(x, C)}$$

where $C^\epsilon = \{x, d(x, C) \le \epsilon\}$ and $(\cdot)^c$ denotes complement. This is Lipschitz because both distances are Lipschitz in x and the denominator is always $\ge \epsilon$. By definition, we have $\int g_\epsilon\,d\mu_n \to \int g_\epsilon\,d\mu$ for all ϵ with

$1 \ll 1/\epsilon \ll n$. For ϵ small, we have $\int g_\epsilon d\mu \to \int g d\mu$ by dominated convergence with $\int g_\epsilon d\mu_n \geq \int I_C d\mu_n$.

For $(e) \to (a)$, assume $f : S \to [0,1]$ is continuous. Then

$$\int f d\mu = \int_0^1 \mu(\{x \in S : f(x) \geq y\}) dy$$

and

$$\int f d\mu_n = \int_0^1 \mu_n(\{x \in S : f(x) \geq y\}) dy.$$

Letting $A_y = \{x \in S : f(x) \geq y\}$, it is not hard to show that $\partial A_y \subseteq f^{-1}(y)$. This means $\mu(A_y) = 0$ except for countably many y. So by dominated convergence, the bottom integral converges to the top one (since the integrands are $[0,1]$-valued and countable sets have Lebesgue measure 0). $\square$