5 Weak convergence of measures and the invariance principle

5.1 Maximum of the Brownian motion on interval

For $\xi_1, \xi_2, \ldots$ i.i.d. random variables with $\mathbb{E}\,\xi_j = 0, \mathrm{Var}\;\xi_j = 1$, consider the random walk $S_n = \xi_1 + \cdots + \xi_n,\; n \geq 0$, and let

$$X_n(t) = \frac{S_{\lfloor nt \rfloor}}{\sqrt{n}},\;\; t \geq 0$$

be the scaled version of the random walk. Applying Central Limit Theorem, we concluded in the previous lecture that for any $0 \leq t_1 < \cdots < t_k$ the finite-dimensional distributions converge, that is

$$(X_n(t_1), \ldots, X_n(t_k)) \stackrel{d}{\to} (B(t_1), \ldots, B(t_k)),$$

where $B = (B(t),\; t \geq 0)$ is the Brownian motion. Now, we iterate the question: is it true that also

$$\max_{t \in [0,1]} X_n(t) \stackrel{d}{\to} \max_{t \in [0,1]} B(t) \tag{1}$$

holds? Convergence of the finite-dimensional distributions alone is not sufficient to address the question, since the location of the maximum is not determined by values of the process at any fixed finite set of times. But (1) will follow from the convergence in distribution of the Brownian path 'seen as a whole', that is as a random element of some space of functions.

There is an elegant way to find the distribution of the Brownian maximum in the right-hand side of (1). To start with, recall that for constant $c > 0$

$$\widehat{B}(t) := B(t + c) - B(c), \quad t \geq 0,$$

is a BM, independent of $\mathcal{F}_c$. In turns that the identity still holds if constant c is replaced by any finite stopping time τ adapted to the natural filtration. So suppose $\tau \geq 0$ satisfies $\{\tau \leq t\} \in \mathcal{F}_t,\; t \geq 0$, and $\mathbb{P}(\tau < \infty) = 1$. Define $\mathcal{F}_\tau$ to be the σ-algebra of events A that occur before τ; which means that

$$A \in \mathcal{F}_\tau \quad \Longleftrightarrow \quad A \cap \{\tau \leq t\} \in \mathcal{F}_t, \quad \forall t \geq 0.$$

Then

$$\widehat{B}(t) := B(t + \tau) - B(\tau), \quad t \geq 0,$$

is a BM independent of $\mathcal{F}_\tau$. This fact (which is intuitively evident but requires a rigorous proof) entails the *strong Markov property* of the BM: conditionally on the value at time τ, the future of the BM after τ is independent of the pre-τ path of the process.

Taking the BM with minus sign, $(-B(t),\; t \geq 0)$, yields another BM, as is readily derived from symmetry of the normal distribution and definition of the BM. This feature has a nice generalisation known as *the reflection principle*. For $x \geq 0$, consider the hitting time

$$\tau_x := \min\{t \geq 0 : B(t) = x\},$$

then

$$\widetilde{B}(t) := B(t) 1(\tau_x \leq t) + (x - (B(t) - x)) 1(\tau_x > t), \quad t \geq 0$$

is another BM by symmetry and the strong Markov property. If in this formula '$x + \cdots$' stood in place of '$x - \cdots$' this would be a path decomposition for B. But for $\widetilde{B}$ the path coincides with the path of B before first-hitting level x at time τ_x, and after τ_x the path of $\widetilde{B}$ is obtained by reflecting the

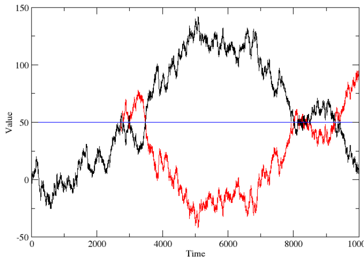

Figure 1: Brownian path reflected upon hitting level $x = 50$.

path of B about the level x, see the picture. For the maximum value $M := \max_{t\in[0,1]} B(t)$ we have by the reflection principle

$$\mathbb{P}(M > x, B(1) > x) = \mathbb{P}(M > x, B(1) \le x),$$

whence, using $M \ge B(1)$ and symmetry of the normal distribution,

$$\mathbb{P}(M > x) = \mathbb{P}(M > x, B(1) > x) + \mathbb{P}(M > x, B(1) \le x) = 2\,\mathbb{P}(M > x, B(1) > x) = \\ 2\,\mathbb{P}(B(1) > x) = \mathbb{P}(|B(1)| > x) = \frac{2}{\sqrt{2\pi}} \int_x^\infty e^{-y^2/2} dy.$$

It follows that M has the 'folded' normal distribution, with p.d.f. being the standard normal p.d.f. on $\mathbb{R}_+$ multiplied with factor 2.

More generally, distribution of the maximum $M(t) := \max_{s\in[0,t]} B(s)$ follows from the identity $M(t) \stackrel{d}{=} \sqrt{t}\, M(1)$ which in turn is a consequence of the self-similarity property of the BM[(1)].

5.2 Skorokhod embedding

If τ is a stopping time with $\mathbb{E}\,\tau < \infty$, then $\mathbb{E}\,B(\tau) = 0$, because BM is a martingale and a continuous-time counterpart of Doob's Optional Sampling Theorem from Lecture 3 is applicable. The distribution of the BM value $B(\tau)$ at time τ depends, of course, on τ. Which distributions may appear that way?

Example (Gambler's ruin) Choose $-a < 0 < b$. For $\tau = \min\{t : B(t) \in \{a, b\}\}$ we have

$$\mathbb{P}(B(\tau = a) = \frac{b}{a+b}, \quad \mathbb{P}(B(\tau = b) = \frac{a}{a+b}.$$

As we vary a and b, we obtain for $B(\tau)$ any two-point distribution with mean 0.

This simple observation has a powerful generalisation.

Theorem 5.1. (Skorokhod embedding) *For any random variable ξ with $\mathbb{E}\,\xi = 0, \mathbb{E}\,\xi^2 < \infty$ there exists a stopping time τ (adapted to the natural Brownian filtration) such that $\mathbb{E}\,\tau < \infty$ and*

$$B(\tau) \stackrel{d}{=} \xi.$$

This result allows one to realise a random walk $S_n = \xi_1 + \cdots + \xi_n,\ n \ge 0$, (with i.i.d. ξ_i's, $\mathbb{E}\,\xi_j = 0$, Var $\xi_j < \infty$) in terms of the BM as

$$S_n \stackrel{d}{=} B(\tau_1 + \cdots + \tau_n).$$

[(1)]Notation $\xi \stackrel{d}{=} \eta$ stays for 'have the same probability distribution'.

Here, the identity in distribution is *joint* for $n \geq 0$, $\tau_0 = 0$, and for $n \geq 1$ the pairs

$$(\tau_n, B(\tau_1 + \cdots + \tau_n) - B(\tau_1 + \cdots + \tau_{n-1}))$$

are i.i.d. copies of (τ_1, ξ_1). The embedding can be pursued to prove functional convergence of the scaled random walk to the Brownian motion. See Section 1.11 of [1] for details of this approach.

5.3 Weak convergence of probability measures

Let us review two classic results of Probability Theory, concerning the random walk $S_n = \xi_1 + \cdots + \xi_n$.

Example Suppose $\mathbb{E}\,\xi_i = m$. By the Law of Large Numbers, as $n \to \infty$,

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} m.$$

Since convergence in probability implies convergence in distribution, we have $S_n/n \xrightarrow{d} m$, which by definition of convergence in distribution means that

$$\mathbb{E}\, f\left(\frac{S_n}{n}\right) \to \mathbb{E}\, f(m)$$

($= f(m)$ in this case) for all bounded continuous functions f. Introduce probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ as

$$P_n(A) = \mathbb{P}(S_n/n \in A), \quad P(A) = \delta_m(A)$$

(the second is the Dirac measure at point m), these are distributions of S_n/n and the constant rv m.

Re-writing the convergence of expectations in terms of the measures P_n, P we have

$$\int_{\mathbb{R}} f(x) dP_n(x) \to \int_{\mathbb{R}} f(x) dP(x).$$

This kind of convergence of measures is called *weak convergence* and denoted $P_n \Rightarrow P$.

Consider the corresponding cumulative distribution functions (c.d.f.)

$$F_n(x) = \mathbb{P}(S_n/n \leq x) = P_n((-\infty, x]), \quad F(x) = P((-\infty, x]) = 1(x \geq m).$$

The convergence $F_n(x) \to F(x)$ holds in all points $x \in \mathbb{R}$ with the sole exception of $x = m$, where F has a discontinuity.

Example Suppose further that $\mathrm{Var}\,\xi_i = \sigma^2 < \infty$. Let P_n be the probability distribution of the rv $(S_n - nm)/(\sigma\sqrt{n})$, and P be the $\mathcal{N}(0,1)$-measure on $\mathbb{R}$. Statement of the CLT can be recast as the weak convergence $P_n \Rightarrow P$. If F_n is the c.d.f. of $(S_n - nm)/(\sigma\sqrt{n})$, this amounts to

$$F_n(x) \to \Phi(x), \text{ as } n \to \infty,$$

for every $x \in \mathbb{R}$. Unlike the first example, there are no exceptional points where convergence fails, because the standard normal c.d.f. Φ is everywhere continuous.

To set up a fairly general framework, let $(E, \mathcal{B}, \rho)$ be a metric space with some distance function ρ. The Borel σ-algebra $\mathcal{B} = \mathcal{B}(E)$ is generated by open subsets of E.

Definition 5.2. *For $P_n, n \in \mathbb{N}$, and P probability measures on $(E, \mathcal{B}, \rho)$, we say that P_n* converge weaky *to P, denoted $P_n \Rightarrow P$, if the convergence*

$$\int_E f(x) dP_n(x) \to \int_E f(x) dP(x)$$

holds for all continuous bounded functions $f : E \to \mathbb{R}$.

For random variables ξ, ξ_n with values in E and defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let

$$P_n(A) = \mathbb{P}(\xi_n \in A), \quad P(A) = \mathbb{P}(\xi \in A), \quad A \in \mathcal{B}(\mathbb{R}).$$

Convergence in distribution $\xi_n \stackrel{d}{\to} \xi$ is equivalent to the weak convergence $P_n \Rightarrow P$. The subtle difference between the convergence concepts in that the first applies to random variables and the second to induced measures (distributions of the rv's).

Weak convergence is preserved by continuous mappings. Suppose $g : E_1 \to E_2$ is a continuous mapping of metric spaces. For measure μ on E_1 we denote $g \circ \mu$ its pushforward to E_2, defined as $g \circ \mu(A) = \mu(g^{-1}(A)),\ A \in \mathcal{B}(E_2)$. Then $P_n \Rightarrow P$ on E_1 implies $g \circ P_n \Rightarrow g \circ P$ on E_2.

Theorem 5.3. (The 'portmanteau' theorem) *The following conditions are equivalent:*

(i) $P_n \Rightarrow P$,

(ii) $\limsup_n P_n(A) \leq P(A)$ *for all closed* $A \subset E$,

(iii) $\liminf_n P_n(A) \geq P(A)$ *for all open* $A \subset E$,

(iv) $\lim_{n\to\infty} P_n(A) = P(A)$ *if* $P(\partial A) = 0$.

(The *boundary* ∂A is the intersection of the closure of A with the closure of $E \setminus A$.)

Example For $E = \mathbb{R}$, the distribution function $F(x) = P((-\infty, x])$ has a jump at z if $P(\{z\}) > 0$. On the other hand, $\partial((-\infty, z]) = \{z\}$. Thus if $P(\{z\}) > 0$, the weak convergence $P_n \Rightarrow P$ imposes no convergence condition on the sequence $F_n(z) = P_n((-\infty, z])$. But if F is continuous at z, then $F_n(z) \to F(z)$ must hold.

Let

$$\varphi_n(\theta) = \int_{\mathbb{R}} e^{i\theta x} dP_n(x), \ \ \varphi(\theta) = \int_{\mathbb{R}} e^{i\theta x} dP(x)$$

be the *characteristic functions* of probability measures P_n, P. The weak convergence $P_n \Rightarrow P$ is equivalent to the pointwise convergence of the characteristic functions: $\varphi_n(\theta) \to \varphi(\theta)$, for every $\theta \in \mathbb{R}$. This fact underlies classic proofs of the CLT and its generalisations for sums of i.i.d. rv's.

Example Consider the case of $E = \mathbb{R}^k$ endowed with the Euclidean distance. For probability measures P_n, P on $\mathbb{R}^k$ introduce the multidimensional cumulative distribution functions as

$$F_n(x_1, \dots, x_k) = P_n((-\infty, x_1] \times \cdots \times (-\infty, x_k]), \ \ F(x_1, \dots, x_k) = P((-\infty, x_1] \times \cdots \times (-\infty, x_k]).$$

If P is the distribution of random vector $(\xi_1, \dots, \xi_k)$, the function

$$F(x_1, \dots, x_k) = \mathbb{P}(\xi_1 \leq x_1, \dots, \xi_k \leq x_k)$$

is known as the joint c.d.f. Similarly to the case $k = 1$, the weak convergence $P_n \Rightarrow P$ holds if and only if

$$F_n(x_1, \dots, x_k) \to F(x_1, \dots, x_k)$$

for every point $(x_1, \dots, x_k)$, where F is continuous.

Example For $E = \mathbb{R}^\infty := \{(x_1, x_2, \dots) : x_j \in \mathbb{R}\}$ we may consider the uniform (also known as square or ℓ^∞) metric $\rho(x, y) = \sup_j |x_j - y_j|$. Let $\pi_k : (x_1, x_2, \dots) \mapsto (x_1, \dots, x_k)$ be the k-dimensional projection. Weak convergence $P_n \Rightarrow P$ holds iff $\pi_k \circ P_n \Rightarrow \pi_k \circ P$ for every $k \in \mathbb{N}$. The pushforward probability measure $\pi_k \circ P$ is the k-dimensional marginal distribution defined by $\pi_k \circ P(A) = P(\pi_k^{-1}(A)),\ A \in \mathcal{B}(\mathbb{R}^k)$.

Example Consider $E = C[0, 1]$, the space of continuous functions $x : [0, 1] \to \mathbb{R}$ with distance $\rho(x, y) = \sup_{t \in [0,1]} |x(t) - y(t)|$. For $0 \leq t_1 < \cdots < t_k$ let $\pi_{t_1, \dots, t_k} : x \mapsto (x(t_1), \dots, x(t_k))$. If

$P_n \Rightarrow P$ then, of course, $\pi_{t_1,\dots,t_k} \circ P_n \Rightarrow \pi_{t_1,\dots,t_k} \circ P$ by continuity of the projections, but the converse is not true.

Here is a (counter-)example. Let P_n be the Dirac measure on $C[0,1]$ supported by the function $x_n(t) = nt \cdot 1(0 \le t < 1/n) + (2 - nt) \cdot 1(1/n \le t \le 2/n)$ (where $1(\cdots) = 1$ if $\cdots$ is true and $= 0$ otherwise). Then $\pi_{t_1,\dots,t_k} \circ P_n \Rightarrow \delta_{(0,\dots,0)}$ for every projection. But P_n has no weak limit, because the sequence of functions x_n has no limit in $C[0,1]$.

Remark on the notation. If P is the probability law of a process $X = (X(t),\ t \in [0,1])$, then $\pi_{t_1,\dots,t_k} \circ P$ is the distribution of the vector $(X(t_1),\dots,X(t_k))$. In Lecture 3 we denoted this k-dimensional distribution $\mu_{t_1,\dots,t_k}$.

Definition 5.4. *A family of probability measures $\mathcal{P} = (P_j,\ j \in J)$ on E is called* relatively compact *if every sequence of measures from $\mathcal{P}$ contains a weakly convergent subsequence.*

For instance, consider $P_n(A) = \lambda(A \cap [n-1,n])$, the uniform distribution on $[n-1,n]$. The corresponding distribution functions converge pointwise to the 0 function, which does not correspond to a probability measure on $\mathbb{R}$. The sequence (P_n) has no convergent subsequence, since the measure 'escapes to infinity' as n grows.

Definition 5.5. *A family of probability measures $\mathcal{P} = (P_j,\ j \in J)$ on E is called* tight *if for every $\varepsilon > 0$ there exists a compact set $K \subset E$ such that*

$$\sup_{j \in J} P_j(E \setminus K) \le \varepsilon.$$

A metric space E is called *Polish* if it is complete (every Cauchy sequence has a limit) and separable (there exists a dense countable subset). By a result from Topology, each probability measure P on Polish space satisfies the 'compact containment condition': for every $\varepsilon > 0$, there exists a compact set $K \subset E$ such that $P(E \setminus K) \le \varepsilon$. Prokhorov's Theorem to follow generalises this property.

Theorem 5.6. (Prokhorov) *On a Polish metric space $(E, \mathcal{B}, \rho)$, a family of probability measures $\mathcal{P} = (P_j,\ j \in J)$ is relatively compact if and only if $\mathcal{P}$ is tight.*

In particular, (P_n) has a weak limit only if the sequence is tight.

Whenever P_n is the distribution of a random variable X_n with values in E, we say that (X_n) is tight if the sequence of measures (P_n) is tight.

5.4 Weak convergence in $C[0,1]$

We focus on the fundamental issue of weak convergence of probability measures on the space of continuous functions $C[0,1]$, which is a Polish space[(2)].

For two probability measures P and Q on $C[0,1]$, the equality $P = Q$ holds if and only if $\pi_{t_1,\dots,t_k} \circ P = \pi_{t_1,\dots,t_k} \circ Q$ for all $t_1 < \cdots < t_k$. That is to say, the finite-dimensional projections uniquely determine the measure. However, weak convergence of $\pi_{t_1,\dots,t_k} \circ P_n$'s does not guarantee that P_n has a weak limit at all (see example above).

Let $X = (X(t),\ t \in [0,1]), X_n = (X_n(t),\ t \in [0,1])$ be random processes with a.s. continuous paths. Recasting Prokhorov's theorem, we have

Proposition 5.7. *As $n \to \infty$, $X_n \stackrel{d}{\to} X$ if and only if the conditions hold:*

(i) $(X_n(t_1),\dots,X_n(t_k)) \stackrel{d}{\to} (X(t_1),\dots,X(t_k))$ *for every choice of times* $0 \le t_1 < \cdots < t_k$,

(2) For $C[0,T]$ the theory is completely analogous, and for $C(\mathbb{R}_+)$ the weak convergence amounts to the weak convergence of projections to each $C[0,T],\ T > 0$.

(ii) *the sequence* (X_n) *is tight.*

To apply the proposition we need explicit conditions ensuring tightness in $C[0,1]$. These rely on the following quantity that measures fluctuations of a continuous function.

Definition 5.8. *For* $x \in C[0,1]$, *the* modulus of continuity *is*

$$w(x;h) = \sup_{|t-s|\le h} |x(t) - x(s)|, \quad h > 0.$$

Each particular $x \in C[0,1]$ is *uniformly* continuous, or, stated in terms of the modulus of continuity, $\lim_{h\to 0} w(x;h) = 0$. Since convergence 'almost surely' implies convergence in probability, for every probability measure P on $C[0,1]$, it holds that $\lim_{h\to 0} P\{x : w(x;h) > \varepsilon\} = 0$, for $\varepsilon > 0$.

Recall the Arzela-Ascoli theorem from Analysis: a family of functions $(x_j, j \in J) \subset C[0,1]$ is relatively compact if the functions are uniformly bounded and equicontinuous. These conditions hold precisely when (a) $\sup_{j\in J} |x_j(0)| \le r$ for some $r > 0$, and (b) $\lim_{h\to 0} \sup_{j\in J} w(x_j;h) = 0$. The next tightness criterion is a probabilistic counterpart of conditions (a), (b).

Theorem 5.9. *A sequence of probability measures* (P_n) *on* $C[0,1]$ *is tight if and only if the following two conditions hold.*

(i)

$$\lim_{r\to\infty} \limsup_n P_n\{x : |x_n(0)| > r\} = 0$$

(i.e. $\pi_0 \circ P_n$ *is tight).*

(ii) *For every* $\varepsilon > 0$

$$\lim_{h\to 0} \limsup_n P_n\{x : w(x;h) > \varepsilon\} = 0.$$

For practical use the condition (ii) is inconvenient, because to assess $w(x;h)$ we need to examine all s, t at distance at most h. To get a *sufficient* condition for tightness, one can replace (ii) by better accessible

(ii′) *For every* $t \in [0,1]$ *and* $\varepsilon > 0$

$$\lim_{h\to 0} \limsup_n \frac{1}{h} \max_{1\le j\le h^{-1}} P_n\{x : \sup_{(j-1)h\le s\le jh} |x(t) - x(s)| > \varepsilon\} = 0.$$

That (ii′) implies (ii) follows from the estimate (valid for any P on $C[0,1]$)

$$P\{x : w(x;h) > 3\varepsilon\} \le \sum_{1\le j\le h^{-1}} P\{x : \sup_{(j-1)h\le s\le jh} |x(t) - x(s)| > \varepsilon\}$$

(the reason for factor 3 is the same as in Lecture 4, p. 4).

5.5 Donsker's Invariance Principle

We have now tools to address the convergence in distribution of a scaled random walk

$$\frac{S_{\lfloor nt\rfloor}}{\sqrt{n}}, \quad t \ge 0,$$

(where $S_n = \xi_1 + \cdots + \xi_n$ with i.i.d. $\xi_1, \xi_2, \ldots$ having $\mathbb{E}\,\xi_j = 0, \mathrm{Var}\ \xi_j = 1$). A small nuisance is that the process is a piecewise constant function of t, increasing by jumps. To circumvent this, hence stay within the $C[0,1]$ setting, we resort to a linear interpolation

$$X_n(t) = \frac{1}{\sqrt{n}} \left(\sum_{j=1}^{\lfloor nt \rfloor} \xi_j + (nt - \lfloor nt \rfloor) \xi_{\lfloor nt \rfloor + 1} \right) \tag{2}$$

(so, we update the notation and now X_n stays for a continuous process).

Since $X_n(0) = 0$, condition (i) of tightness is trivially satisfied. For (ii′) we apply a maximum inequality (due to Ottaviani, see [2])[(3)]

$$\mathbb{P}(\max_{1 \le i \le n} S_i > 2r\sqrt{n}) \le \frac{\mathbb{P}(|S_n| \ge r\sqrt{n})}{1 - 1/r^2}.$$

By the i.i.d. assumption, the probabilities in (ii′) are the same for all j, and choosing $j = 1$, we estimate using Ottaviani's inequality

$$\frac{1}{h}\,\mathbb{P}(\sup_{s \le h} |X_n(s)| \ge \varepsilon) = \frac{1}{h}\,\mathbb{P}(\max_{1 \le i \le nh} |S_i| \ge \varepsilon\sqrt{n}) \le$$

$$\frac{1}{h}\,\frac{\mathbb{P}(|S_{\lfloor nh \rfloor}| \ge \varepsilon\sqrt{n}/2)}{1 - 4h/\varepsilon} = \frac{1}{h}\frac{\mathbb{P}\left(\frac{|S_{\lfloor nh \rfloor}|}{\sqrt{nh}} \ge \frac{\varepsilon}{2\sqrt{h}}\right)}{1 - 4h/\varepsilon} \to \frac{1}{h}\frac{2\left(1 - \Phi\left(\frac{\varepsilon}{2\sqrt{h}}\right)\right)}{1 - 4h/\varepsilon}$$

as $n \to \infty$, where the limit holds by the CLT for $S_n/\sqrt{n}$. Letting $h \to 0$, the right-hand side converges to 0, in consequence of the familiar bound, for $x > 0$

$$1 - \Phi(x) < \frac{e^{-x^2/2}}{x\sqrt{2\pi}}$$

on the tail of the normal distribution and by elementary properties of the exponential function.

Therefore, (X_n) is tight. Taken together with the convergence of finite-dimensional distributions, we have completed the proof of

Theorem 5.10. (Donsker's Invariance Principle) *The scaled random walk* (2) *satisfies*

$$X_n \xrightarrow{d} B,$$

where B is the Brownian motion.

This important result is also called the 'functional CLT'. The word 'invariance' appears in this context to emphasize that $\psi(X_n) \xrightarrow{d} \psi(B)$ *for every continuous functional* $\psi : C[0,1] \to \mathbb{R}$. The example we started with, was $\psi(f) = \max_{t \in [0,1]} f(t)$, hence

$$\max_{1 \le j \le n} \frac{S_j}{\sqrt{n}} \xrightarrow{d} \max_{t \in [0,1]} B(t),$$

the latter distributed like $|B(1)|$, i.e. with folded normal distribution.

We could have been dealing with the piecewise constant version of the scaled random walk. But this would require working in a larger *Skorokhod space* of functions with possible jumps. This space extension does not capture any new features because jumps of the scaled random walk become negligible in the limit.

[(3)]In the simplest case of the random walk generated by fair coin-tossing, the reflection principle allows one to express the distribution of $\max_{0 \le i \le n} S_i$ in terms of the binomial distribution.

5.6 The invariance principle via Skorokhod embedding

For r.v. ξ with $\mathbb{E}\xi = 0, \mathrm{Var}\xi = 1$ the stopping time τ from by Theorem 5.1 should satisfy $\mathbb{E}\tau = 1$, by the virtue of Wald's identity

$$\mathrm{Var}\, B(\tau) = \mathrm{Var}\, \xi\, \mathbb{E}\, \tau.$$

Iterating Skorokhod's theorem and appealing to the strong Markov property of the BM readily yields embedding of a random walk.

Corollary 5.10.1. *Let $S_n = \xi_1 + \cdots + \xi_n$ be a random walk with i.i.d. increments having $\mathbb{E}\,\xi_1 = 0, \mathrm{Var}\,\xi_1 = 1$. There exist nonnegative r.v.'s $\tau_1, \tau_2, \ldots$ such that $\tau_1 + \cdots + \tau_n$ is a stopping time adapted to the natuaral filtration of the BM, and*

$$(B(\tau_1 + \cdots + \tau_n),\ n > 0) \stackrel{d}{=} (S_n,\ n > 0).$$

In particular,

$$\frac{S_n}{\sqrt{n}} \stackrel{d}{=} \frac{B(\tau_1 + \cdots + \tau_n)}{\sqrt{n}}.$$

If we could introduce n inside B to have instead

$$\frac{B(\tau_1 + \cdots + \tau_n)}{n}$$

(which is straighforward if τ_n's are constant), then the Law of Large Numbers for the arithmetic averages of τ_n's could be applied to conclude that $S_n/\sqrt{n}$ converges in distribution to $\mathcal{N}(0, 1)$, thus concluding the CLT from the LLN[(4)].

By self-similarity

$$B_n(t) := \sqrt{n} B(t/n),\ t \geq 0$$

is a BM. Therefore applying Corollary 5.10.1 to B_n we see that there exist $\tau_{1,n}, \ldots, \tau_{n,n}$ such that

$$\left(\frac{S_n}{\sqrt{n}},\ 1 \leq k \leq n\right) \stackrel{d}{=} \left(\frac{B(\tau_{1,n} + \cdots + \tau_{k,n})}{\sqrt{n}}, 1 \leq k \leq n\right).$$

As the notation shows the $\tau_{k,n}$'s depend on n, because B_n depends on n, although $(\tau_{1,n}, \ldots, \tau_{n,n}) \stackrel{d}{=} (\tau_1, \ldots, \tau_n)$ for a single i.i.d. sequence $\tau_1, \tau_2, \ldots$.

Proof of Theorem 5.10 via embedding Consider

$$T_n(t) := \frac{\tau_{1,n} + \cdots + \tau_{\lfloor nt \rfloor, n}}{n}.$$

The processes have the same distribution:

$$\left(\frac{S_{\lfloor nt \rfloor}}{\sqrt{n}},\ t \in [0, 1]\right) \stackrel{d}{=} (B(T_n(t)),\ t \in [0, 1]).$$

For $\delta > 0$ we have that

$$\mathbb{P}\left(\sup_{0 \leq t \leq 1} |B(T_n(t) - B(t)| \geq \varepsilon\right) \leq \mathbb{P}\left(\sup_{0 \leq t \leq 1} |T_n(t) - t| \geq \delta\right) + \mathbb{P}\left(\sup_{0 \leq t \leq 1,\ 0 \leq s, t \leq 1 + \delta} |B(s) - B(t)| \geq \varepsilon\right).$$

[(4)]These two standard results are presented separately in most Probability courses.

For $\delta \to 0$ the 2nd term approaches 0 by the continuity of the BM.

The 1st term satisfies

$$\sup_{0\le t\le 1} |T_n(t) - t| \xrightarrow{\mathbb{P}} 0 \tag{3}$$

as $n \to \infty$. Indeed, for $\varepsilon > 0$

$$\sup_{0\le t\le 1} |T_n(t) - t| \le$$
$$\frac{1}{n} + \sup_{1\le k\le n} \left| \frac{\tau_{1,n} + \cdots + \tau_{k,n} - k}{n} \right| \stackrel{d}{=} \frac{1}{n} + \sup_{1\le k\le n} \frac{k}{n} \left| \frac{\tau_1 + \cdots + \tau_k}{k} - 1 \right| \le$$
$$\frac{1}{n} + \sup_{k\le n\varepsilon} \frac{k}{n} \left| \frac{\tau_1 + \cdots + \tau_k}{k} - 1 \right| + \sup_{k> n\varepsilon} \frac{k}{n} \left| \frac{\tau_1 + \cdots + \tau_k}{k} - 1 \right| \le$$
$$\frac{1}{n} + \varepsilon \sup_{k>0} \left| \frac{\tau_1 + \cdots + \tau_k}{k} - 1 \right| + \sup_{k> n\varepsilon} \frac{k}{n} \left| \frac{\tau_1 + \cdots + \tau_k}{k} - 1 \right|.$$

By the strong LLN

$$\sup_{k>0} \left| \frac{\tau_1 + \cdots + \tau_k}{k} - 1 \right|$$

is bounded, and as $n \to \infty$

$$\sup_{k> n\varepsilon} \frac{k}{n} \left| \frac{\tau_1 + \cdots + \tau_k}{k} - 1 \right| \to 0 \text{ a.s..}$$

Since ε can be chosen arbitrarily small

$$\sup_{1\le k\le n} \frac{k}{n} \left| \frac{\tau_1 + \cdots + \tau_k}{k} - 1 \right| \to 0 \quad \text{a.s.}$$

This gives (3) since convergence almost surely for $\tau_1, \tau_2, \ldots$ implies convergence in probability for the triangular array $\tau_{k,n}, 1 \le k \le n$.

5.7 Functional limit for the empirical distribution function

Fitting distribution to the empirical data is a central theme in Statistics, which motivated study of maxima of random processes.

Suppose $\xi_1, \xi_2, \ldots$ are i.i.d. rv's, uniformly distributed on $[0, 1]$. Think of the realisation of $\xi_1, \ldots, \xi_n$ as sample data collected in a statistical experiment. We call

$$F_n(t) := \frac{\#\{j \le n : \xi_j \le t\}}{n}$$

the *empirical distribution function*. For instance, $F_n(1/2)$ is the proportion of the sample that falls in the left half of $[0, 1]$.

In fact, F_n is a *random* c.d.f. corresponding to a *random* discrete probability measure $n^{-1} \sum_{j=1}^n \delta_{\xi_j}$. For each fixed $t \in [0, 1]$, $nF_n(t)$ is a Binomial(n, t) random variable.

As $n \to \infty$, the Law of Large Numbers tells us that for $t \in [0, 1]$

$$F_n(t) \xrightarrow{\text{a.s.}} t.$$

A stronger version of the LLN in this context is known as the *Glivenko-Cantelli theorem*, which states that

$$\sup_{t\in[0,1]} |F_n(t) - t| \xrightarrow{\text{a.s.}} 0.$$

The identity function on $[0, 1]$ is the c.d.f. of the uniform distribution, thus the theorem asserts convergence of the empirical distribution function to the 'true c.d.f.' in the metric of $C[0, 1]$.

For each fixed t, by the normal approximation to the binomial distribution,

$$\sqrt{n}(F_n(t) - t) \xrightarrow{d} \mathcal{N}(0, t(1 - t)).$$

Comparing with the variance function of the Brownian bridge B° (see Lecture 4), one might guess that the left-hand side viewed as a random process converges in distribution to B°.

Let $Y_n = (Y_n(t),\ t \in [0, 1])$ be the random continuous function whose graph is obtained by spanning a broken line on the points $(0, 0), (1, 1)$ and $(\xi_j, \sqrt{n}(F_n(\xi_j) - \xi_j)), 1 \le j \le n$. This is just a linear interpolation, similar to that employed for the scaled random walk.

Theorem 5.11. *As $n \to \infty$*

$$Y_n \xrightarrow{d} B^\circ.$$

Let $K = \max_{t\in[0,1]} |B^\circ(t)|$. The rv K has the Kolmogorov-Smirnov distribution

$$\mathbb{P}(K \le y) = 1 - 2\sum_{j=1}^{\infty}(-1)^{j-1}e^{-2j^2y^2}.$$

The theorem implies that

$$\max_{t\in[0,1]} |Y_n(t)| \xrightarrow{d} K.$$

For big samples, to test the statistical null-hypothesis that 'the data come from the uniform distribution' one can calculate the sample value of $\max_{t\in[0,1]} |Y_n(t)|$ and check if it falls in a critical region depending on the desired confidence level.

The procedure is readily adapted to test if 'the data $\eta_1, \dots, \eta_n$ come from the probability distribution with given continuous c.d.f. F'. To that end, just recall that $\xi_j := F(\eta_j)$ follow the Uniform$[0, 1]$ distribution, by the virtue of the probability integral transform you have seen in Statistics courses.

Exercises

1. Let $S_n = \xi_1 + \cdots + \xi_n,\ n \ge 0$, be a random walk. (a) Find a condition on the distribution of ξ_1 to ensure that $(S_n, n \ge 0)$ and $(-S_n, n \ge 0)$ have identical distributions. (b) Assuming the RW is simple, i.e. $\mathbb{P}(\xi_i = 1) = \mathbb{P}(\xi_i = -1) = 1/2$, use the Reflection Principle to relate the distribution of $\max_{0\le i\le n} S_i$ to the distribution of S_n.

2. For $x \ge 0$ let τ_x be the first time the BM hits level x. (a) Show that $\tau_{cx} \stackrel{d}{=} c^2\tau_x$. (b) Derive the p.d.f. of τ_x. (c) Show that the process $(\tau_x,\ x \ge 0)$ has stationary, independent increments. (d) We have $\mathbb{E}\, B(\tau_x) = x$ and not 0. Explain why Doob's Optional Sampling Theorem is not applicable.

3. For $a > 0$ let $\hat{\tau}_a = \min\{t : B(t) = t + a)\}$. (a) Using the strong Markov property, show that

   $$P(\hat{\tau}_{a+b} < \infty | \hat{\tau}_a < \infty) = \mathbb{P}(\hat{\tau}_b < \infty).$$

   (b) Using (a), prove that $\sup_{t\ge0}(B(t) - t)$ has exponential distribution.

4. Show that $\max_{t\in[0,1]} x(t)$ is a continuous functional on $C[0, 1]$.

5. Formulate the invariance principle for $(S_{\lfloor nt\rfloor},\ t \in [0, T])$.

6. Let $(N(s),\ s \geq 0)$ be a Poisson process with rate 1. (a) State a CLT for $N(t) - t$ as $t \to \infty$. (b) Let for $T > 0$

$$X_T(t) := \frac{N(tT) - tT}{\sqrt{T}},\ t \in [0, 1].$$

Find the *functional* limit for the process X_T as $T \to \infty$. You are not expected to show tightness, but justify the convergence of finite-dimensional distributions. (You may assume that T is an integer parameter, although this is not necessary).

7. Calculate and compare the covariance functions of the empirical process $(F_n(t), t \in [0, 1])$ (derived from the Uniform$[0, 1]$ sample) and of the Brownian bridge $(B^\circ(t),\ t \in [0, 1])$.

Literature

1. T. Liggett, Continuous time Markov processes, AMS, 2010.

2. O. Kallenberg, Foundations of modern Probability, Springer, 1997.

3. P. Billingsley, Convergence of probability measures, Wiley, 1968.

4. P. Mörters, Y. Peres, Brownian motion, Cambridge, 2010.