SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

A comparison of two proofs of Donsker's theorem

av

Sebastian Franzén

2020 - No K19

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

A comparison of two proofs of Donsker's theorem

Sebastian Franzén

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Daniel Ahlberg

2020

A comparison of two proofs of Donsker's theorem

Sebastian Franzén

May 2020

Abstract

This thesis explores Donsker's theorem: a theorem in the subject of stochastic processes that relates a Brownian motion to a limit of random walks. It states that a sequence of random walks, appropriately rescaled in time and space, and linearly interpolated between its values at integer times, converges weakly to a Brownian motion.

There are two quite different approaches to proving the theorem that involve entirely different techniques. Both of them will be described and some of the theory involved will be presented. As will be shown in this paper, one approach will prove to be a possible construction of Brownian motion. The second approach assumes its prior existence, but instead it provides, as a corollary, the Central limit theorem.

1 Introduction

This thesis explores Donsker's theorem: a theorem in the subject of stochastic processes that relates a Brownian motion to a limit of random walks. It states that a sequence $X^{(n)}$ of random walks rescaled in time and space according to n and $1/\sqrt{n}$ respectively, and with paths linearly interpolated between its values at integer times, converges weakly to a Brownian motion. This can be viewed as a strengthening of the of the central limit theorem which gives weak convergence of the random walk at a single point in time.

By observing that the rescaling continuously pushes each point of a path of a random walk to the left one may see that the convergence (almost surely) cannot be pointwise for every path. Instead, the convergence stated in the theorem is that of weak convergence of measures which means that the probabilities for the two processes of having paths being in some specified set of functions will approach each other.

Interestingly the convergence is not dependent on the distribution of the random variables that generates the random walk, but it is dependent on the fact that they are non-degenerate, independent, identically distributed with a zero mean and finite variance.

There are two quite different approaches to proving the theorem. There are two quite different approaches to proving the theorem that involve entirely different techniques. Both of them will be here described and some of the theory involved will be presented. The first approach is more analytic in nature; it relies on the fact that one sufficient condition for a sequence of measures on a metric space to have a limit point is that the measures may with arbitrary precision be guaranteed to be supported on a compact set. If we refer to Arzela-Ascoli's representation of compact sets in $C[0, 1]$ one may then supply such a condition. Due to an argument involving the central limit theorem one may then prove that the whole sequence converges to a distribution of a Brownian motion. The second approach is more probabilistic in nature and relies on concepts such as stopping times and the strong Markov property. It involves proving that any random variable with zero mean and finite variance is distributed as a Brownian motion at a random time. Using the strong Markov property one may then argue that each step in the random walk is distributed as as a Brownian motion at a random time, close to the corresponding point in time for the random walk. This binds the paths of the random walk to more and more points of corresponding paths of a Brownian motion and one may then argue that the limit is distributed as a Brownian motion.

The two proofs make different assumptions which leads to different secondary consequences. The first proof does not depend on the existence a Brownian motion. Indeed, the proof is one way to derive its existence, perhaps not the easiest, and as such is an example of a construction that is done with continuous sample paths from the outset. [1] The first proof, however, does depend on the Central limit theorem.

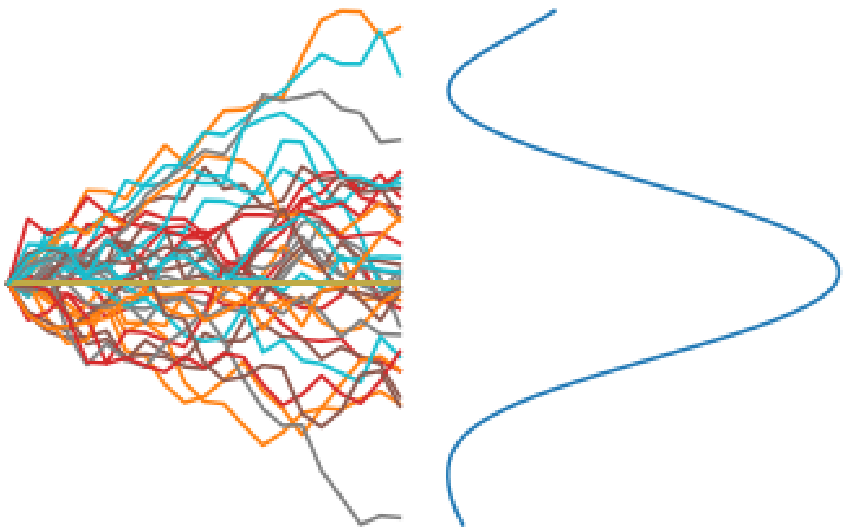
The second proof, on the other hand, does require the prior existence of Brownian motion. It does not utilize the central limit theorem but instead the central limit theorem follows as an immediate consequence. It thus provides a method to prove the central limit theorem other than via the common route of characteristic functions and Levy's continuity theorem.

2 Interlude: convergence of scaled random walks

Through the central limit theorem one may derive some elementary connections between random walks and Brownian motion (as well as between interpolated random walks and Brownian motion). We will do this in the current section before we turn to the first proof of Donsker's

[1] On the contrary, other constructions of Brownian motion may first provide a process that satisfies the axioms other than having continuous sample paths. Having established this, one then shows that the process has a continuous modification.

theorem in the next section. This will serve as an illustration on how much stronger Donsker's theorem is.



Convergence of the n:th term from the n:th scaled random walk to a normal distribution

Let $(\xi_i)_{i\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables with zero mean and second moment equal to one. Consider the random walk we get from the partial sums $S_k = \sum_{i=1}^{k} \xi_i$ from the sequence. If we, for each n, scale the random walk with $\frac{1}{\sqrt{n}}$ we get a sequence of random walks

$$\left(\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\xi_i\right)_{k\in\mathbb{N}}\right)_{n\in\mathbb{N}}.$$

From the central limit theorem (Theorem 15.37 in Klenke (2013)) we have that the the sequence

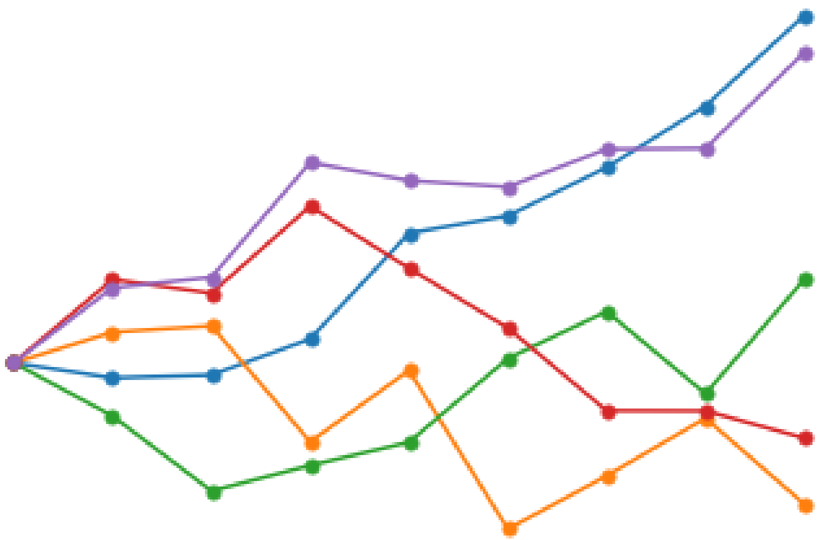
$$\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\right)_{n\in\mathbb{N}}$$

consisting of the n:th term from the n:th scaled random walk converges to a normal distribution $N(0,1)$. In the current section one may consider convergence in distribution of $S_n/\sqrt{n}$ to a normal distribution, to have the meaning that for any open interval (a, b), the probability that $S_n/\sqrt{n}$ is in (a, b) approaches the probability that a standard normal random variable is in (a, b).

Still using elementary means one may derive the more general statement that the interpolated random walk

$$X_t^{(n)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor t \cdot n \rfloor} \xi_i + (tn - \lfloor tn \rfloor) \frac{1}{\sqrt{n}} \xi_{\lfloor tn \rfloor + 1}$$

converges in distribution to $N(0, t)$ for any $t \in (0, 1)$.



Some paths of an interpolated random walk on $[0, 1]$

This is done with an application of the following theorem, Markov's inequality, Slutsky's theorem and again the central limit theorem.

Theorem 1. *Suppose that X_n, Y_n, for $n \in \mathbb{N}$, and X are random variables with values in a metric space (S, ρ). If $X_n \to X$ in distribution and $\rho(X_n, Y_n) \to 0$ in probability, then $Y_n \to X$ in distribution.*

Since

$$\left| X_t^{(n)} - \frac{1}{\sqrt{n}} S_{\lfloor tn \rfloor} \right| = (tn - \lfloor tn \rfloor)|\xi_{n+1}|$$

and we have from Markov's inequality that

$$P[(tn - \lfloor tn \rfloor)|\xi_{n+1}| \geq \varepsilon] \leq \frac{(tn - \lfloor tn \rfloor)E[\xi_{n+1}]}{\varepsilon}$$

it follows that

$$\left| X_t^{(n)} - \frac{1}{\sqrt{n}} S_{\lfloor tn \rfloor} \right| \overset{\text{in probability}}{\rightarrow} 0.$$

Referring to the theorem above, it would be sufficient to show that

$$\frac{1}{\sqrt{n}} S_{\lfloor tn \rfloor} \overset{\text{in distribution}}{\rightarrow} N(0, t),$$

and this follows from Slutsky's theorem and the central limit theorem, by noting that

$$\frac{1}{\sqrt{n}} S_{\lfloor tn \rfloor} = \frac{\sqrt{\lfloor tn \rfloor}}{\sqrt{n}} \cdot \frac{}{\sqrt{\lfloor tn \rfloor}} S_{\lfloor tn \rfloor} \overset{\text{in distribution}}{\rightarrow} \sqrt{t} \cdot N(0, 1).$$

The point of Donsker's theorem is that we get a much stronger result. Instead for getting a convergence of at a single point we get convergence over the whole interval - or differently put, convergence of the random functions instead for convergence of a random point.

3 An overview of the subject and statement of the theorem

Donsker's theorem states the convergence of a random walk to a Brownian motion. To be able to state the theorem precisely we will first discuss and define the concepts and objects involved, as well as the specific kind of convergence in the theorem.

3.1 Stochastic processes and Brownian motion

A stochastic process is a mathematical model of some phenomena that evolves randomly over time. As model we consider a collection of random variables indexed by numbers in some index set $I \subset [0, \infty)$ - where we let the indexing number signify the time of the occurrence of the random variable.

In what follows we will assume that a probabilty space $(\Omega, \mathcal{F}, P)$ where Ω is the set of outcomes, $\mathcal{F}$ is a σ-algebra of subsets of Ω and P is a probability measure on $\mathcal{F}$ (See e.g. Chapter 1 in Klenke (2013) for a treatment of these objects). We make the following definitions.

Definition. A random variable with values in $(S, \mathcal{S})$ is a measurable function X from $(\Omega, \mathcal{F}, P)$ to the measurable space $(S, \mathcal{S})$. If the measurable space $(S, \mathcal{S})$ is $\mathbb{R}$ with the Borel-σ algebra $\mathcal{B}(\mathbb{R})$ we will call X a real random variable - or simply a random variable.

Definition. A stochastic process is a family $(X_s)_{s\in I}$ of real random variables on $(\Omega, \mathcal{F}, P)$ indexed by some set $I \subset [0, \infty)$.

By considering a fixed outcome for the stochastic process for all indices, we get a so called *path* that evolves (non-randomly) over time. That is, for each fixed outcome ω, $(X_s(\omega))_{s\in I}$ is a map from the index set I to $\mathbb{R}$ - the map given by $s \mapsto X_s(\omega)$. One may show that this map from the probability space to the function space is measurable,[2] and thus a stochastic process may equivalently be seen as a random variable with values in the function space $\mathbb{R}^I$. Considering this we will also sometimes write $X(t)$ for this random function at the point t.

A random walk is a stochastic process $(S_n)_{n\in\mathbb{N}}$ we get by adding n independent and identically distributed random variables $\xi_1, \cdots, \xi_n$ - that is for each natural number n we let

$$S_n = \sum_{i=1}^{n} \xi_i$$

As the name suggest, a random walk is a process which at each step i moves up or down according to the value of ξ_i. The random walks we will consider in the theorem are those that are generated by random variables ξ_i:s that have expectation equal to zero. As a consequence one would believe that, on average, a path from such a process ought to evolve as much in a positive direction as in a negative direction.

A Brownian motion is a stochastic process in continuous time; originally a model for how pollen moves suspended in water (Brown 1827). As movement in space is continuous, the model has continuous sample paths. Further it has homogeneous and independent increments. The so called homogeneity of the increments means that how the process evolves between two times t_1 and t_2 only depends on the distance between t_1 and t_2 - and not on their location on the real

[2] We may motivate that this map is indeed a random variable by showing that it is measurable as follows: The product σ-algebra on R^I is the smallest σ-algebra such that every coordinate projection $\pi_s : f \mapsto f(s)$ is measurable (that is, it is genterated by the maps $\pi_t,\ t \in I$). Using the "factorization lemma" [Corollary 1.82 in Klenke] we see that X as a random function is measurable if for every t the map $\omega \to \pi_t(X(\omega)) = X_t(\omega)$ is measurable. Which is true by the assumption that X_t is a measurable map.

line; analogously the independence between the increments means that how the process evolves between two times t_2 and t_3 is independent of how the process evolved earlier between the times t_1 and t_2.

Definition. A Brownian motion $B = (B_t)_{t \geq 0}$ is a real-valued stochastic process such that

(B1) $B_0(\omega) = 0$

(B2) for any $n \in \mathbb{N}$ and any $0 = t_0 < t_1 < ... < t_n$
$B_{t_n} - B_{t_{n-1}}, ..., B_{t_1} - B_{t_0}$ are independent

(B3) $B_t - B_s \overset{\mathcal{D}}{=} N(0, t - s)$

(B4) the map $t \mapsto B_t(\omega)$ is continuous for every ω.[3]

Brownian motion has the Markov property (Lemma 2.10 in Partzsch & Schilling (2012)). That is if we consider the process evolving from some time s and onwards (and subtracts the value of the process at this time s so that it starts with value zero), that process is again a Brownian motion - and independent of the original process from time 0 to time s.

3.2 Finite dimensional distributions of a stochastic process

As written in the the previous section, a stochastic process may be seen as a random variable with values in a function space. We will discuss a property of these distributions in this section.

The distribution of a random variable is a function (namely a measure) on a collection of (measurable) subsets of the sample space of the random variable. In an analogous way as a continuous function on $\mathbb{R}$ is determined by its values on the rational numbers - meaning that any two continuous functions that are equal on the rational numbers indeed are equal on the whole real line - a distribution may be seen to be determined on particular subsets of its domain.[4] This will be an aid when we prove the convergence of the distributions of the interpolated random walks as one does not have to show that the limit distribution agrees with the distribution of a Brownian motion for any measurable set, but may instead consider a more restricted subclass of sets.

As the paths of a Brownian motion are continuous, we will consider one method to introduce a σ-algebra (the domain of a probability measure) on $C[0, 1]$. We will then show that any probability distribution on $C[0, 1]$ is determined by its finite dimensional distributions - that is determined on sets of the form

[3] Due to a Theorem by Kolmogorov and Chentsov [e.g. Theorem 21. 6 in Klenke] it may be derived that any process that satisfies (B_3) has a modification with continuous sample path - but the property is of such importance that we state it in the definition.

[4] That is on a subclass of the class of measurable set.

$$\{f \in C[0,1] : f(t_1) \in B_1, \cdots f(t_n) \in B_n\}.$$

[5]

First let $\pi_t : C[0,1] \to \mathbb{R}$ denote the projection that sends a continuous function on $C[0,1]$ to its value at the point t

$$\pi_t(f) := f(t).$$

We may then let the σ-algebra on $C[0,1]$ be the smallest such that all those projections π_t are measurable; this is written as

$$\sigma(\pi_t : t \in [0,1]).$$

We will now consider something called a $\cap$-stable generator of a σ-algebra.

Definition. A collection of subset $\mathcal{E}$ is called a $\cap$-stable generator for the σ-algebra $\sigma(\mathcal{E})$ if it is closed under intersection, that is

$$A, B \in \mathcal{E} \implies A \cap B \in \mathcal{E}.$$

One may show (e.g. Lemma 1.42 in Klenke (2013)) that any measure is determined by its values on a $\cap$-stable generator $\mathcal{E}$ of its domain, the σ-algebra $\mathcal{F} := \sigma(\mathcal{E})$. The argument goes along the following lines: assuming that two probability measures are equal on the $\cap$-stable generator, one considers the collections of all sets from the σ-algebra $\mathcal{F}$ for which the same holds. This collection, is then shown to be a σ-algebra itself that does contain the original sigma algebra $\mathcal{F}$. The two measures are thus equal on $\mathcal{F}$.

The sigma algebra generated by the $\cap$-stable generator of sets of the form

$$\begin{aligned}\{f \in C[0,1] : f(t_1) \in B_1, \cdots f(t_n) \in B_n\} \\ = \{f \in C[0,1] : \pi_{t_1}(f) \in B_1, \cdots, \pi_{t_n}(f) \in B_n\}\end{aligned}$$

[5] This property for a distribution to be determined on its finite-dimensional distribution is not particular for distributions on $C[0,1]$ though, but holds for any distribution on a function space.

certainly generates a σ-algebra that contains the σ-algebra generated by the projections π_t.[6] Thus if two measures are equal on this collection - from the $\cap$-stable generator property they are equal on a σ-algebra that does contain the σ-algebra generated by the projections π_t - thus they are equal on the latter σ-algebra.

3.3 Weak convergence

The concept of convergence we will use is that of weak convergence of measures. The current section will give a short introduction.

In mathematics we often considers if a sequence x_n of objects approaches some other object x of the same type as n increases; and if so say that the sequence x_n converges to x. This of course requires some specification of what "approaches" means. There are different ways to specify this under different generalities and depending on the structure on the collection of objects. The most straight forward one is when one has a "metric" on the space that gives a distance between any two objects. One may then say that the sequence converges if we may make the sequence become arbitrarily close to x from some n onwards. Below we define a concept of convergence for sequences of measures. This definition does not involve a metric - but it might be noted that it is possible to first introduce such, then defining the convergence via this metric and that this leads to exactly the same limits.

The concept of *weak convergence* comes from the subject of functional analysis.[7] The *weak* modifier in the name denotes that one weakens the condition for convergence[8] and thus may get limits that would not have satisfied the original stronger convergence condition. For a characterization of this weakening of the condition of convergence, consider the condition on the measure of the boundary ∂A of A in 2. in Theorem 3 below.

Definition. We say that a sequence of probability measures $(P_n)_{n\in\mathbb{N}}$ on a metric space $(S, \mathcal{S})$ converges weakly to P if for every bounded and continuous function $f : S \to \mathbb{R}$,

$$\lim_n \int f\, dP_n \to \int f\, dP$$

The motivation behind the definition is the following which tells us a measure is characterized by the collection of continuous and bounded functions, and gives us that each weak limit is unique.

[6] In fact they are equal as can be seen from that the $\cap$-stable generator of finite dimensional distributions is contained in the $\sigma(\pi_t : t \in [0,1])$. The reverse inclusion of the σ-algebras is immediate.

[7] Where to be precise the convergence under consideration in this section would be called *weak-* convergence.*

[8] namely by considering convergence in a weaker topology

Theorem 2. *Given two probability measures P and Q on* $(S, \mathcal{S})$, *if*

$$\int f \; dP = \int f \; dQ$$

for every bounded and continuous function $f : S \to \mathbb{R}$, *then* $P = Q$

The theorem above says if two measures are equal on all (measurable) sets that may be approximated by continuous and bounded functions, then in fact they are equal.

To every random variable X there corresponds the image measure $P_X := P \circ X^{-1}$. We make the following definition

Definition. We will say that a sequence X_n of random variables converges weakly to X if the corresponding measures P_{X_n} converges weakly to P_X

As the name suggest, a sequence X_n of random variables thus converges in distribution to X if their distributions P_{X_n} behaves more and more as the distribution P_X of X. One may wonder if this could have been formulated as that the distributions P_{X_n}, as functions on the collections of events A of the random variables, converges pointwise to P_X? An answer to this is given by the Portmanteau theorem.

Theorem 3 (Portmanteau theorem)**.** *For a metric space E and probability measures* $\mu, \mu_1, \mu_2, \ldots$ *on E the following three statements are equivalent.*

1. $\mu_n \overset{weakly}{\rightarrow} \mu$
2. *For all (measurable) A with*[9] $\mu(\delta A) = 0 : \lim_n \mu_n(A) = \mu(A)$
3. *For all closed* $F \subset E$: $\limsup_n \mu_n(F) \leq \mu(F)$

We will use the equivalence between 1. and 3. in the second proof. One may also note that 2. gives rise to the equivalence between convergence in distribution of a sequence of random variables X_n to X and the pointwise convergence of their distribution functions F_n to F at all points of continuity of F - when the sequence is real valued.

3.4 Statement of the theorem

We are now almost ready to state the theorem that is the main object of the thesis.

[9] The *boundary* δA of a set A is defined as the set difference between the closure and A and the interior of A.

We first define an interpolated random walk $(X_t^{(n)})_{t\in[0,1]}$ where, for each path, we connect the values of the random walk on $\mathbb{N}$ with straight lines, and scale the index, to get a stochastic process in continuous time on $[0, 1]$.

This interpolated random walk $(X_t^{(n)})_{t\in[0,1]}$ is defined for every natural number n as follows: [10]

$$X_t^{(n)} := \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor t \cdot n \rfloor} \xi_i + (tn - \lfloor tn \rfloor) \frac{1}{\sqrt{n}} \xi_{\lfloor tn \rfloor + 1}$$

Theorem 4 (Donsker's theorem)**.** *If $\xi_1, \xi_2, \cdots$ are independent and identically distributed random variables with mean zero, variance equal to one and if $X^{(n)}$ is the interpolated random walk constructed from them as above, and $(B_t)_{t\geq 0}$ is a Brownian motion, then $(X^{(n)})_{t\geq 0}$ converges weakly to $(B_t)_{t\geq 0}$.*

4 Prokhorov's proof

We now turn to the first proof that was initially supplied by Prokhorov.[11] It relies on a theorem by Prokhorov himself, and will not just prove the convergence of the scaled random walk but also prove the existence of a Brownian motion. Prokhorov's theorem gives necessary and sufficient conditions for a sequence of probability measures to have the property that every subsequence has a further subsequence that converges to some probability measure. We will combine the sufficiency part in this theorem together with a condition on the convergence of the finite-dimensional distributions to get a sufficient condition for a sequence of probability measures on $C[0, 1]$ to converge to a limit - and where we also are able to specify the limit.

First some definitions; we assume that the the probability measures are measures on some complete and separable metric space (S, ρ).

Definition. We say that a sequence $\{P_n\}$ of probability measures is *relatively compact* if every subsequence $\{P_{n_k}\}$ contains a further subsequence $\{P_{n_{k(m)}}\}$ that converges weakly to some probability measure Q.

Definition. A sequence $\{P_n\}$ of probability measures is *tight* if for every $\varepsilon > 0$ there exists a compact set K such that for every n: $P_n[K] > 1 - \varepsilon$.

[10] We assume that every walk starts at zero, that is $X_0^{(n)} = 0$.

[11] Prokhorov, Y. V. (1956) *Convergence of Random Processes and Limit Theorems in Probability Theory* Theory of probability & its applications

Theorem 5 (Prokhorov's theorem). *A necessary and sufficient condition for the sequence $\{P_n\}$ to be relatively compact is that it is tight.*

A proof of Prokhorov's theorem is beyond the scope of this thesis but an outline is as follows: The proof that a sequence μ_n in $\mathcal{F}$, of probability measures on a metric space (E, d), has a subsequence μ_{n_k} that converges to a limit μ - and that that limit is a probability measure is done in the following steps.

1. One collects a countable number of compact subsets $\mathcal{C}$ from E (that in particular contains a sequence of compact sets K_n such that for all $\mu \in \mathcal{F}, \mu(K_n^C) < 1/n$). Via a diagonal argument one shows that there exists a subsequence n_k for which $\mu_{n_k}(C)$ converges for every C in the countable collection $\mathcal{C}$.

2. One then defines a *set function* α on the countable collection $\mathcal{C}$ that for each C takes the value of the limit of the convergent subsequence $\mu_{n_k}(C)$.

3. The goal is then to find a measure μ that on any *open* set is determined by the value of α on the compact sets in $\mathcal{C}$ - that is μ is "inner regular on the open sets with respect to the class $\mathcal{C}$"; meaning that for any open A

   $$\mu(A) = \sup \{\alpha(C) : C \in \mathcal{C} \text{ and } C \subset A\}$$

   This will make it possible to show - using The Portmanteau theorem - that the subsequence μ_{n_k} converges weakly to μ, since then for any open $A \supset C$

   $$\alpha(C) = \lim_k \mu_{n_k}(C) = \liminf_k \mu_{n_k}(C) \leq \liminf_k \mu_{n_k}(A),$$

   which implies

   $$\mu(A) \leq \liminf_k \mu_{n_k}(A).$$

4. To find such a measure μ one first defines a set function μ^* from α that is defined for every subset of E. μ^* is then shown to be an outer measure and to satisfy "inner regular on the open sets with respect to the class $\mathcal{C}$".

5. As a last step one shows that the closed sets of E are μ^*-measurable; thus in particular the Borel-sets are μ^*-measurable - and we get the desired measure μ.

A detailed proof may be found in Klenke (2013), Theorem 13.29.

It is worth to note that Prokhorov's theorem only guarantees the existence of a limit (for a subsequence of each subsequence) of a tight sequence of probability measures - but it does not specify the limit (nor does it guarantee that the limit is the same for different subsequences). As we intend to prove the convergence to a specific distribution an additional argument is required. That extra argument may be obtained from the fact that distributions on a function space are determined by their finite dimensional distributions. For the finite dimensional distributions on may namely prove convergence directly by an application of the Central limit theorem. The next section gives an outline for how the knowledge of convergence of the finite dimensional distributions may be used to prove both convergence to a specified distribution and show existence of a specified distribution.

4.1 An outline of the argument in the proof

In the argument below and further on we employ useful consequence of the definition of of weak convergence: namely that if P_n converges weakly to P on a metric space $(S, \mathcal{S})$, and if $h : S \to S'$ is a measurable mapping from S to some metric space S', then the image measures Ph^{-1} on S' converges weakly to Ph^{-1}. We will refer to this as the continuous mapping theorem (Theorem 13.25 in Klenke (2013)).

The following property is motivated in an analogous way as for a sequence of real numbers. *A sequence $\{P_n\}$ of probability measures converges weakly to some measure P if and only if for every subsequence there exists a further subsequence that converges weakly to P*

Consider the following situation. We know that $\{P_n\}$ is tight and that for any k and any $t_1, ..., t_k$ $P_n \circ \pi^{-1}_{t_1,...,t_k}$ converges weakly to $P \circ \pi^{-1}_{t_1,...,t_k}$. From Prokhorov's theorem, and the preceding paragraph, we know that for any subsequence P_{n_k} there exists a further subsequence $P_{n_{k(m)}}$ converging to some probability measure Q. By the continuous mapping theorem for any k and any $t_1, ..., t_k$ $P_{n_{k(m)}} \circ \pi^{-1}_{t_1,...,t_k}$ converges weakly to $Q \circ \pi^{-1}_{t_1,...,t_k}$. Thus since the finite dimensional distributions determines a measure, Q in fact equals P. Thus we have that for any subsequence there exists a further subsequence converging weakly to P and, again by the preceding paragraph, this means that P_n converges weakly to P.

With a similar argument is also possible to prove the *existence* of of a probability measure on a function space S with specified finite dimensional distributions. Say we want to prove the existence of a probability measure P on S with some specified finite dimensional distributions $\mu_{t_1,....,t_k}$. It would then be sufficient to exhibit a tight sequence $\{P_n\}$ whose finite dimensional distributions converge to $\mu_{t_1,....,t_k}$ for: by Prokhorov's theorem there exists a subsequence $\{P_{n_k}\}$ that converges weakly to some probability measure Q and from the continuous mapping theorem, for any given $t_1, ..., t_k$, the finite dimensional distributions of Q equals $\mu_{t_1,....,t_k}$ and since the finite dimensional distributions determine the measure, Q in fact is the desired measure P.

This is the argument we will use to prove not just the convergence of the scaled random walk to Brownian motion but also the existence of the process[12].

4.2 Convergence of the finite-dimensional distributions

As a first step of the proof as outlined above, we show in this section that the finite-dimensional distributions converge to the finite dimensional distributions of a Brownian motion -that is a joint normal distribution.

The proof of the convergence of the finite dimensional distributions utilizes the same methods as in the interlude where we proved that the scaled and interpolated random walks at a point t converges weakly to a normal distribution with mean zero and variance equal to t; and that for a vector of random variables we get weak convergence from that of the individual components. [13]

Theorem 6. *For any n, $0 \leq t_1 \leq \cdots \leq t_n$ and any Brownian motion $(B_t)_{t\geq 0}$*

$$(X^{(n)}_{t_1}, \cdots, X^{(n)}_{t_n}) \overset{weakly}{\to} (B_{t_1}, \cdots, B_{t_n})$$

To simplify notation one may prove this for $n = 2$ and write $t_1 = s$ and $t_2 = t$.

We may prove that

$$\left(X^{(n)}_s, X^{(n)}_t - X^{(n)}_s\right) \overset{\text{weakly}}{\to} (N_s, N_{t-s})$$

Where N_s and N_{t-s} are independent normally distributed random variables with mean zero and variance s and $t - s$ respectively, since then from the continuous mapping theorem it will follow that

$$\left(X^{(n)}_s, X^{(n)}_t\right) \overset{\text{weakly}}{\to} (B_s, B_t).^{14}$$

[12] One may note that as a proof of the existence of a Brownian motion, the existence of a limit will guarantee a process with distributions that satisfies B1 - B3; that it also satisfies B4 follows from the fact that the distributions are defined directly on $C[0, 1]$.

[13] This and the related converse statement are often referred to as the Cramér-Wold theorem (Theorem 15.56 in Klenke (2013)).

[14] That $N_{t-s} + N_s$ is distributed as a Brownian motion at the time t follows from that the sum of two independent normal random variables is normally distributed (e.g. Example 20.6 in Billingsley (2012)).

But in the interlude we proved that

$$X_s^{(n)} \stackrel{\text{weakly}}{\to} N_s$$

and similarly one may show that

$$X_t^{(n)} - X_s^{(n)} \stackrel{\text{weakly}}{\to} N_{t-s}.$$

The independence of N_s and N_{t-s} follows from the fact that the ξ_i:s are independent, and that independence is preserved under weak limits. The convergence of the vector now follows the Cramér-Wold theorem (Theorem 15.56 in Klenke (2013)).

4.3 Compactness in $C[0,1]$

As Prokhorov's theorem states that a sequence of probability measures is tight if their masses are uniformly concentrated to compact sets we would like to know how compact sets are characterized in $C[0,1]$.

Definition. The *modulus of continuity* of a function $x \in C[0,1]$ is the function [15]

$$m_x(\delta) = \sup_{|s-t|\leq\delta} |x(s) - x(t)|$$

We have the following characterization that is a form of the Arzela-Ascoli theorem.

Theorem 7. *The set $A \subset C[0,1]$ is relatively compact[16] if and only if*

$$\sup_{x\in A} |x(0)| < \infty$$

and

$$\lim_{\delta\to 0} \sup_{x\in A} m_x(\delta) = 0$$

The first condition in the theorem above states that the functions in A are pointwise bounded at zero, and the second that the functions in A are equicontinuous over $[0,1]$ - moreover they are so in an uniform manner.

From this theorem we get the following necessary and sufficient condition for tightness of a sequence of probability measures on $C[0,1]$

[15] We will sometimes write $m(x,\delta)$ instead for $m_x(\delta)$ when that is more readable.

[16] A set A is relatively compact if the closure of A is compact

Theorem 8. *The sequence P_n of probability measures on $C[0,1]$ is tight if and only if the following two conditions hold*

For each positive η there exists an a and an n_0 such that

$$P_n[x : x(0) \geq a] \leq \eta,\ n \geq n_0$$

and for each positive ε

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P_n[x : m_x(\delta) \geq \varepsilon] = 0$$

4.4 Proof of tightness of the sequence

We now have to solve the following: we know the distribution for $X_t^{(n)}$ for each t but to characterize the sequence $P_{X^{(n)}}$ as tight we have to put an upper bound on the sequence

$$P_{X^{(n)}}[x : m_x(\delta) \geq \varepsilon], \quad n \geq k$$

for every k.

Trying to calculate the probability of some event involving an uncountable number of t:s in $[0,1]$ from that of the individual events for each t is not possible. This is so since the probability measure $P_{X^{(n)}}$ only handles countable operations - for a union of a countable number of disjoint events the probability equals the probability of the sum of their individual probabilities, and if the union is not disjoint then at least this event is bounded by the sum of the individual events.

Thus to prove tightness of the sequence we need to solve two problems: first we need to reduce the event

$$\left\{ m(X^{(n)}, \delta) \geq \varepsilon \right\} \tag{1}$$

to be contained in an event that depends on at most a countable number of coordinates; secondly we need to put a sufficiently strong bound (the bound depending on δ) on this latter event in terms of the individual distributions $P_{X_t^{(n)}}$.

The theorem that follows will let us solve the first problem: due to the piece-wise linearity of the interpolated random walk, it will let us fix a finite number of points t_i for which we check if

$$\max_{t_{i-1} \leq k \leq t_i} |x(k) - x(t_{i-1})| \geq \varepsilon, \quad k \in \mathbb{N}$$

instead for the uncountable number of $|s - t| \leq \partial$ for which

$$\sup_{|s-t| \leq \partial} |x(s) - x(t)| \geq \varepsilon.$$

Theorem 9. *Suppose that* $0 = t_0 < t_1 < \ldots < t_v = 1$ *and*

$$\min_{1<i<v}(t_i - t_{i-1}) \geq \partial$$

Then for arbitrary x and any probability measure P on $C[0, 1]$

$$P[x : m_x(\partial) \geq 3\varepsilon] \leq \sum_{i=1}^{v} P\left[x : \sup_{t_{i-1} \leq s \leq t_1} |x(s) - x(t_{i-1})| \geq \varepsilon\right]$$

The proof goes along the following lines: one considers

$$m := \max_{1<i<v} \sup_{t_{i-1} \leq s \leq t_i} |x(s) - x(t_{i-1})|.$$

For s, t in the same interval $[t_{i-1}, t_i]$, from the triangle inequality one has

$$|x(s) - x(t)| \leq |x(s) - x(t_{i-1})| + |x(t) - x(t_{i-1})| \leq 2m$$

and simarily if s, t are in adjacent intervals $[t_{i-1}, t_i]$ and $[t_i, t_{i+1}]$ respectively, then

$$|x(s) - x(t)| \leq |x(s) - x(t_{i-1})| + |x(t_{i-1}) - x(t_i)| + |x(t) - x(t_i)| \leq 3m.$$

Now given $|s - t| \leq \partial$, from the fact that $\min_{1<i<v}(t_i - t_{i-1}) \geq \partial$ it must be the case that either s and t lies in the same interval $[t_{i-1}, t_i]$ or in adjacent intervals $[t_{i-1}, t_i]$ and $[t_i, t_{i+1}]$, for some i. Thus for any $|s - t| \leq \partial$, we have that $3m$ is an upper bound of $|x(s) - x(t)|$, and thus

$$m_x(\partial) \leq 3 \max_{1<i<v} \sup_{t_{i-1} \leq s \leq t_i} |x(s) - x(t_{i-1})|.$$

Thus, given that $m_x(\delta) \geq 3\varepsilon$ we must have that at least for some $1 \leq i \leq v$ it is the case that $\sup_{t_{i-1} \leq s \leq t_i} |x(s) - x(t_{i-1})| \geq \varepsilon$ and so

$$\{x : m_x(\delta) \geq 3\varepsilon\} \subset \bigcup_{i=1}^{v} \left\{ x : \sup_{t_{i-1} \leq s \leq t_i} |x(s) - x(t_{i-1})| \right\},$$

From this the claim follows by subadditivity.

We will now turn to the second problem. We will first give an upper bound on

$$\sum_{i=1}^{v} P\left[\sup_{t_{i-1} \leq s \leq t_i} |X_s^{(n)} - X_{t_{i-1}}^{(n)}| \geq \varepsilon \right]$$

in terms of probabilities of partial sums of the random walk. The key to this will be that for every ω, $X_t^{(n)}$ is a piece-wise linear function and thus will take its supremum, on any interval $[\frac{k-1}{n}, \frac{k}{n}]$, on either of the endpoints.

We will first assume that the sequence $\{\xi_i\}$, from which we get the random walk, is normally distributed with mean and variance equal to zero and one respectively. Later we will then extend the proof. Precisely we will derive:

$$P[m(X^{(n)}, \delta) \geq 3\varepsilon] \leq \frac{4(\lambda(\delta))^2}{\varepsilon^2} P\left[\max_{k \leq m} |S_k| \geq \lambda(\delta)\sqrt{m} \right] \tag{2}$$

where $\lambda(\delta)$ is such that $\delta \to 0 \iff \lambda(\delta) \to \infty$, and $m \to \infty \iff n \to \infty$.

This is done by noting that for the t_i:s in Theorem 9 of the form m_i/n for integers $m_0 < m_1 < \cdots < m_v = n$ - the integers not necessarily consecutive - the supremum of $|X_s^{(n)} - X_{t_i}^{(n)}|$ will be taken for s at any of the nodes k/n for k between m_{i-1} and m_i. But at every such node $X_{k/n}^{(n)} = S_k/\sqrt{n}$ and thus we have

$$\begin{aligned} P\left[\sup_{t_{i-1} \leq s \leq t_1} \left| X_s^{(n)} - X_{t_{i-1}}^{(n)} \right| \geq \varepsilon \right] &= P\left[\max_{m_{i-1} \leq k \leq m_i} \left| \frac{S_k - S_{m_{i-1}}}{\sqrt{n}} \right| \geq \varepsilon \right] \\ &= P\left[\max_{k \leq m_i - m_{i-1}} |S_k| \geq \varepsilon\sqrt{n} \right], \end{aligned}$$

where the last equality follows from that S_k is the sum of identically distributed random variables, which means that each $|S_k - S_{m_{i-1}}|$ is distributed as $|S_k|$ for $0 \leq k \leq m_i - m_{i-1}$.

We may further chose $m_i, \quad i = 0, \cdots, v$ of the form $m_i = im$ so that $m_i - m_{i-1}$ is equal to m; then

$$\sum_{i=1}^{v} P\left[\max_{k \leq m_i - m_{i-1}} |S_k| \geq \varepsilon\sqrt{n}\right] = v \cdot P\left[\max_{k \leq m} |S_k| \geq \varepsilon\sqrt{n}\right]$$

What is left now to arrive at (2) is to put a bound on v in terms of n and δ. This may be done by investigating the relationships between m, n and δ. We won't do that here, but one takes $m = \lceil n\delta \rceil$ and $v = \lceil n/m \rceil$ to arrive at

$$P[m(X^{(n)}, \delta) \geq 3\varepsilon] \leq \frac{4}{2\delta} P\left[\max_{k \leq m} |S_K| \geq \frac{\varepsilon}{\sqrt{2\delta}}\sqrt{m}\right],$$

which is precisely (2) if we write $\lambda(\delta)$ for $\varepsilon/\sqrt{2\delta}$.

Now turning back to Theorem 8, we have that to get the second condition it will be sufficient to show that

$$\lim_{\lambda \to \infty} \limsup_{n \to \infty} (\lambda(\delta))^2 P\left[\max_{k \leq m} |S_k| \geq \lambda(\delta)\sqrt{m}\right] = 0. \tag{3}$$

To put a bound on

$$P[\max_{k \leq m} |S_k| \geq \lambda(\delta)\sqrt{m}]$$

we will use Etemadi's inequality.

Lemma 1 (Etemadi's inequality). *If $S_1 \cdots S_n$ are partial sums of independent random variables, then*

$$P\left[\max_{k \leq n} S_k \geq 3\alpha\right] \leq 3 \max_{k \leq n} P\left[S_k \geq \alpha\right]$$

Now applying Etemadi's inequality it will be sufficient to show that

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} (\lambda(\delta))^2 \max_{k\le m} P[|S_k| \ge \lambda(\delta)\sqrt{m}] = 0$$

For the particular case when ξ_i are independent standard normal random variables then from the inequality[17] $P[|N| \ge \lambda] \le E[N^4]\lambda^{-4} \le 3\lambda^{-4}$ we get that

$$P[|S_k| \ge \lambda\sqrt{m}] = P[\sqrt{k}|N| \ge \lambda\sqrt{m}] \le 3\lambda^{-4}, \text{for } k \le n$$

and thus of course

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} (\lambda(\delta))^2 \max_{k\le m} P[|S_k| \ge \lambda\sqrt{m}] \le \lim_{\lambda\to\infty} \limsup_{n\to\infty} (\lambda(\delta))^2 3\lambda^{-4} = 0,$$

which proves (3).

We have thus shown that X_t^n is tight - under the assumption that ξ_i are independent and distributed according to $N(0,1)$.

To extend the result to the case when ξ_i is not normally distributed one may use the central limit theorem. First one breaks the maximums of $|S_k|$ into two cases: one for k is sufficiently large so that $S_k/\sqrt{k}$ may be approximated sufficiently close with a $N(0,1)$ distribution, and two for k less than this one may get a bound from Chebyshev's inequality.

As we now have shown that the finite dimensional distributions of X_t^n converges to the finite dimensional distributions of a Brownian motion. Per Prokhorov's theorem and the argument outlined in section 4.1 this then proves the weak convergence to - and the existence of - a Brownian motion.

5 A second proof due to Skorokhod

For the second proof we introduce an additional concept in the theory of stochastic process - this is the concept of a stopping time. A stopping time associated with a stochastic process is a random variable taking values in the index set of the stochastic process.

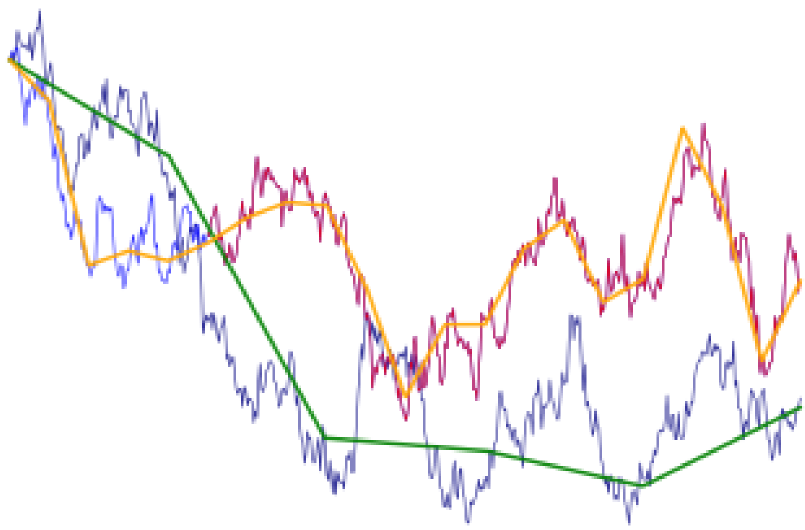
[17] The first inequality here follows from Markov's inequality and the second from that the forth moment of a standard normal random variable equals 3. This latter fact may be attained by integrating x^2 with respect to the standard normal density.

The only requirement for a random variable, with values in the index set, to be a stopping time is the following: if at any time t, one asks which outcomes has led the stopping time to take a value before t - we should be able to do so from the outcomes of the process up to that time t.

With a stopping time and a stochastic process one may define a random variable that for each outcome takes the value of the path of the stochastic process, for this same outcome, at the time of the stopping time.

For example one could define a stopping time τ such that it takes the value t for which the Brownian motion for the first time equals some number a. With the knowledge that almost surely Brownian motion hits every point, should one considers the Brownian motion at this stopping time one would get a random variable that almost surely takes the constant value a.

This random variable defined through a stopping time and a Brownian motion we will say is embedded into the Brownian motion. In the proof that follows we will "embed" the random walk into the Brownian motion. This means that for every k we construct a stopping time $\tau^{(k)}$ such that $B(\tau^{(k)})$ is distributed as the $k:th$ step S_k of the random walk and such that $\tau^{(k)}$ is close to k.



A path of a random walk on $[0, 1]$ is read from a path of a Brownian motion and follows the path of the Brownian motion more closely as the number of steps increases.

Changing the perspective to that of paths we will then show that this means that, with a probability that may be controlled, a path of the random walk up to time n is such that there corresponds a path of the Brownian motion to it, and the value at each step of the path of the random walk is close to the value of Brownian path at the random point $\tau^{(k)}$ (where the random point is close to k). Rescaling the time of both the Random walk and the Brownian motion so that they are indexed in $[0, 1]$ and letting the number of steps n in the random walk increase, we have that the paths of the processes gets bounded at more and more points in time. Thus also the paths of the linear interpolations $X^{(n)}$ of the random walks must be closer and closer to the Brownian paths over the whole interval. [See the next figure.]

As the proof hinges on having a stopping time $\tau^{(k)}$ such that the stopped Brownian motion $B(\tau^{(k)})$ is distributed as the k:th step in a random walk and such that τ_k is close to k, we will first give one solution to the problem of representing a centered random variable (with finite expectation) as a stopped Brownian motion. Then we will use this to prove that the interpolated random walk converges to the Brownian motion. But first we define what a stopping time is and state the Strong Markov property.

Definition. Given a filtration[18] $(\mathcal{F}_t)_{t\geq 0}$, a map $\tau : \Omega \to [0, \infty]$ is called a stopping time if

$$\text{for all } t \geq 0\text{: } \{\tau \leq t\} \in \mathcal{F}_t$$

Definition. Given a $\sigma(\{B_s \ : \ s \leq t\})_{t\geq 0}$- stopping time τ we define the stopped Brownian motion B_τ as

$$B_\tau(\omega) = \begin{cases} B_{\tau(\omega)}(\omega) & \text{if } \tau(\omega) < \infty \\ 0 & \text{if } \tau(\omega) = \infty \end{cases}$$

and the Brownian motion $B_{\tau+t} - B_\tau$ starting at the random time τ as

$$(B_{\tau+t}(\omega) - B_\tau(\omega))_{t\geq 0}$$

A key ingredient in the current approach to Donsker's theorem will be a strengthening of the Markov property (as introduced in the introduction) to what is called the strong Markov property. The time s from which the new process evolves from may be picked randomly. That is one considers what is called a stopping time - a random variable τ with values in the the index set of Brownian motion. Then the process starting at the random time τ is again a Brownian motion and independent of the original process up to the random time τ.

[18] Recall that a filtration is an increasing sequence of σ-algebras.

Theorem 10 (Strong Markov property of Brownian motion[19]). *For a Brownian motion* $(B_t)_{t\geq 0}$ *with filtration* $(\sigma(\{B_s : s \leq t\}))_{t\geq 0}$, *and any almost surely finite stopping time* τ,

$$(B_{\tau+t} - B_\tau)_{t\geq 0}$$

is again a Brownian motion - and independent of the stopped sigma algebra

$$\mathcal{F}_{\tau+} := \left\{ A \in \sigma(\bigcup_{s\geq 0} \sigma(B_s) : \forall t,\ A \cap \{\tau \leq t\} \in \sigma(\{B_s : s \leq t\})_+ \right\}^{20}$$

5.1 Skorokhod stopping problem

The problem of finding a stopping time τ such that the stopped Brownian motion at the time τ follows some specified distribution μ is called the Skorokhod stopping problem. The problem has a trivial solution if one defines $\tau = \inf\{t \geq 0 : B_t = X\}$, where X follows the distribution μ.[21] This solution won't be sufficient for us though since we later will apply the law of large numbers to the sequence of stopping times $\tau_1, \tau_2, \ldots$ - which requires them to have finite expectation, which is not the case for τ as defined here. To show that τ has infinite expectation, one may fill out the details in the sketch below.

The idea is to use the fact that for any real number $a \neq 0$ the expectation of

$$\inf\{t \geq 0 : B_t = a\}$$

is infinite, [22] and to do so one proceeds as follows: Since $W_t := B_{t+1} - B_1 \overset{\mathcal{D}}{=} B_t$ it follows that

$$\inf\{t \geq 0 : B_t = Z\} - 1 \overset{\mathcal{D}}{=} \inf\{t \geq 0 : W_t = Z - B_1\}.$$

As $\inf\{t \geq 0 : W_t = Z - B_1\}$ is a nonnegative function, its expectation is greater than its expectation restricted to the two regions

$$\{Z - B_1 < -1\} \text{ and } \{Z - B_1 > 1\}.$$

[19] Theorem 6.5 in Partzsch & Schilling (2012)

[20] Here $\sigma(\{B_s : s \leq t\})_+$ denotes $\cap_{v>t}\sigma(B_v)$.

[21] Here one uses the fact that almost surely a one-dimensional Brownian motion is recurrent - meaning it hits any value.

[22] This follows from an application of Wald's identities (Theorem 5.10 in Partzsch & Schilling (2012)).

Further

$$\inf\{t \geq 0 : W_t > 1\} \geq \inf\{t \geq 0 : W_t = 1\}.$$

Combining the above one arrives at

$$E[\tau] \geq \mathbb{E}[\inf\{t \geq 0 : W_t = 1\}]\mathbb{P}[|Z - B_1| > 1],$$

and as one may verify that $\mathbb{P}[|Z - B_1| > 1] > 0$,[23] it follows that the expectation of τ is infinite.

As the name suggest, the stopping problem was first solved by A. V. Skorokhod in 1961.[24] One may also note that from Wald's identities [25], for a stopping time with finite expectation, we get the equation

$$E[B_\tau] = 0$$

which implies that a necessary condition for a random variable to be embedded is that it is centered about zero.

5.2 Dubins embedding

The intuition for Dubin's embedding is as follows: consider the problem of trying to represent a random variable X with uniform distribution on $\{-4, -2, 2, 4\}$ as a stopped Brownian motion $B(\tau)$ for some stopping time τ. We may define a sequence of stopped Brownian motions such that the last is distributed as X. This is done as follows:

Let $\tau_{-a,a}$ denote the stopping time $\inf_t\{B_t \in \{-a, a\}\}$. From Wald's identities it follows that $B(\tau_{-a,a})$ has distribution $\delta_{-a}/2 + \delta_a/2$.[26] Then applying the strong Markov property, which says that

$$B\big(\tau_{-3,3} + t\big) - B(\tau_{-3,3})\big)_{t\geq 0}$$

[23] This may be seen from that $P[|Z - B_1| > a] = E[P(|Z - B_1| > a|Z)] = E[g(Z)]$ by the law of iterated expectations and where $g(z) = P[|z - B_1| > a]$. As g is minimized at zero the claim follows.

[24] An English translation may be found in A. V. Skorokhod. (1965). Studies in the theory of random processes.

[25] Theorem 5.10 in Partzsch & Schilling (2012)

[26] Especially see Corollary 5.11 in Partzsch & Schilling (2012)

is a Brownian motion and independent of $B(\tau_{-3,3})$, and writing $\tau_{-1,1}$ for the stopping time of this Brownian motion[27], we get that

$$B(\tau_{-3,3} + \tau_{-1,1})$$

Is uniformly distributed on $\{-4, -2, 2, 4\}$, since e.g.

$$P[(B(\tau_{-3,3} + \tau_{-1,1}) = 2] = P[B(\tau_{-3,3} = 3, B(\tau_{-3,3} + \tau_{-1,1}) - B(\tau_{-3,3}) = -1] = \frac{1}{2} \times \frac{1}{2}$$

The structure of the preceding example is that the finite sequence $B(\tau_{-3,3})$, $B(\tau_{-3,3} + \tau_{-1,1})$ is a martingale[28] that converges to X; $B(\tau_{-3,3} + \tau_{-1,1})$ restricted to one of the two atoms of the distribution of $B(\tau_{-3,3})$ is supported on two values and may be written as

$$f_2(B(\tau_{-3,3}), D_2)$$

Where D_2 is measurable w.r.t. $B(\tau_{-3,3} + \tau_{-1,1})$ and takes values in $\{-1, 1\}$. [29]

The next lemma show that this type of structure is more generally possible: given any centered random variable X with a finite second moment, there exists a martingale - with the properties described above - that converges almost surely and in L^2 to X. Knowing of the existence of that type of martingale, we may then use it to construct a sequence of stopping times $\tau_n \to \tau$ such that $B(\tau_n) \stackrel{\mathcal{D}}{=} X_n$ and $B(\tau) \stackrel{\mathcal{D}}{=} X$. But first we recall a convergence theorem from the theory of martingales.[30]

Theorem 11. *Let $(X_n)_{n\in\mathbb{N}}$ be a L^2-bounded[31] martingale. Then there exists an $\sigma(\cup_n \mathcal{F}_n)$-measurable random variable X_∞, with $E[X_\infty^2] < \infty$, such that X_n converges to X_∞ almost surely and in L^2.*

Definition. A stochastic process $(X_n)_{n\geq 0}$ is called a binary splitting if $X_0 = x_0$, and for every n there exists a random variable $D_n : \Omega \to \{-1, 1\}$ and and a function $f_n : \mathbb{R}^{n-1} \times \{-1, 1\} \to \mathbb{R}$ such that

[27] Here $\tau_{-1,1}$ is defined as the first time $B(\tau_{-3,3} + t) - B(\tau_{-3,3}))_{t\geq 0}$ Brownian motion hits -1 or 1.

[28] We recall that a martingale is stochastic process of integrable random variables $(X_n)_{n\geq 0}$ and a filtration $(\mathcal{F}_n)_{n\geq 0}$ such that for every n: we have that $E[X_{n+1}|\mathcal{F}_n] = X_n$.

[29] Explicitly we may here take $f_2(x, y) = x + y$ and D_2 the random variable that equals one on $\{B(\tau_{-3,3} + \tau_{-1,1}) > B(\tau_{-3,3})\}$ and minus one otherwise.

[30] Theorem 11.10 in Klenke (2013).

[31] A sequence $(X_n)_{n\geq 0}$ is said to be L^2-bounded if $\sup_n E[X_n^2] < \infty$.

$$X_n = f_n(X_1, \dots, X_{n-1}, D_n)$$

Lemma 2. *For a square integrable centered random variable X there exists a binary splitting martingale $(X_n)_{n\geq 0}$ such that $X_0 = 0$ and $X_n \to X$ almost surely and in L^2.*

We will only sketch the proof of the lemma. We follow the proof of Theorem 22.10 in Klenke (2013).

The idea is to inductively define the the sequences $(D_n)_{n\geq 1}$, $(\mathcal{F}_n)_{n\geq 1}$ and $(X_n)_{n\geq 0}$ as: $X_0 = E[X] = 0$, then

$$D_n := \begin{cases} 1, & \text{if } X \geq X_{n-1} \\ -1, & \text{if } X < X_{n-1} \end{cases},$$

$$\mathcal{F}_n := \sigma(D_1, \dots, D_n)$$

and

$$X_n := E[X|\mathcal{F}_n].$$

Through measurability and the structure of D_n one may show that for each n there exists a function f_n from $\mathbb{R}^{n-1} \times \{-1, 1\}$ to $\mathbb{R}$ such that

$$f_n(X_1, \dots, X_{n-1}, D_n) = X_n.$$

From the so called tower property of conditional expectation and the definition of X_{n+1} it follows that

$$E[X_{n+1}|\mathcal{F}_n] = X_n$$

so that $(X_n)_{n\geq 0}$ is a martingale.[32] From Jensen's inequality for conditional expectation[33] we have that

[32] To see the integrability of each X_n we note that $E[X_n] = E[E[X|\mathcal{F}_n|] = E[X] = 0$.

[33] Theorem 8.20 in Klenke (2013).

$$E\left[X_n^2\right] = E\left[(E[X|\mathcal{F}_n])^2\right] \leq E\left[E[X^2|\mathcal{F}_n]\right] = E\left[X^2\right] < \infty,$$

so that from the martingale convergence theorem above, X_n converges to some square integrable random variable X_∞ almost surely and in L^2. What remains then is to show that this random variable equals X almost surely.

This is done by showing that

$$\lim_n D_n(X - X_n) \overset{\text{a.s.}}{=} |X - X_\infty|,$$

and then noting that since D_n is $\mathcal{F}_n$-measurable

$$E[D_n(X - X_n)] = E[E[D_n(X - X_n)|\mathcal{F}_n]] = E[D_n E[X - X_n|\mathcal{F}_n]] = 0.$$

Since from this we may apply the bounded convergence theorem to the sequence $D_n(X - X_n)$ to conclude that

$$E[|X - X_\infty|] = \lim_n E[D_n(X - X_\infty)] = 0,$$

and thus $X = X_\infty$ almost surely.

Theorem 12. *For a binary splitting martingale $(X_n)_{n\geq 0}$ with $X_0 = 0$, bounded in L^2, and a Brownian motion B there exists stopping times $\tau_1 \leq \tau_2 \leq \cdots$ such that*

$$(X_n)_{n\geq 0} \overset{D}{=} \left(B(\tau_n)_{n\geq 0}\right)$$

and such that for every $n :\ E[\tau_n] = E[X_n^2]$. Further if

$$\tau := \lim_n \tau_n$$

then $E[\tau] = Var[\lim X_n]$ and

$$\lim_n X_n \overset{D}{=} B(\tau)$$

Again we follow Klenke (2013), Theorem 22.9.

First one lets $(f_n)_{n\geq 1}$ and $(D_n)_{n\geq 0}$ denote the functions associated with the binary splitting martingale. Without loss of generality one may assume that each f_n is such that for any numbers $a_1, \dots, a_{n-1}: \quad f(a_1, \dots, a_{n-1}, -1) < f_n(a_1, \dots, a_{n-1}, 1)$. Inductively one defines the sequences $(\tau_n)_{n\geq 0}$, $(B_{\tau_n})_{n\geq 1}$ and $(\widetilde{D}_n)_{n\geq 1}$ as

$$\tau_n := \inf\left\{t \geq \tau_{n-1} : B_t \in \{f_n(B_{\tau_1}, \dots, B_{\tau_{n-1}}, -1), f_n(B_{\tau_1}, \dots, B_{\tau_{n-1}}, 1)\}\right\},$$

$$B_{\tau_n}$$

and

$$\widetilde{D}_n := \begin{cases} 1, & \text{if } \widetilde{X}_n \geq \widetilde{X}_{n-1} \\ -1, & \text{if } \widetilde{X}_n < \widetilde{X}_{n-1} \end{cases}$$

The first step is to show that $(B_{\tau_n})_{n\geq 0} \overset{\mathcal{D}}{=} (X_n)_{n\geq 0}$. To do so we first we claim that

$$P[\widetilde{D}_n = 1 | B_{\tau_1}, \dots, B_{\tau_{n-1}}] = \frac{B_{\tau_{n-1}} - f_n(B_{\tau_1}, \dots, B_{\tau_{n-1}}, -1)}{f_n(B_{\tau_1}, \dots, B_{\tau_{n-1}}, 1) - f_n(B_{\tau_1}, \dots, B_{\tau_{n-1}}, -1)} \tag{4}$$

To see this, consider that for $n = 2$ we may consider the set

$$A(y) = \{x \in C([0, \infty)) \mid \exists t > 0 : x(t) = f_2(y, 1) - y, \forall s \in [0, t], x(s) > f_2(y, -1) - y\}$$

Then using the strong Markov property and the formula, that for X, Y independent and ϕ bounded and measurable: $E[\phi(X, Y)|Y] = E[\phi(X, y)]|_{y=Y}$[34], we have that

$$P[B_{\tau_2} - B_{\tau_1} > 0 \mid B_{\tau_1}] = P\left[(B_{t+\tau_1} - B_t)_{t\geq 0} \in A(y)\right]|_{y=B_{\tau_1}} = P[B \in A(y)]|_{y=B_{\tau_1}}.$$

From the fact that $P[B_{\tau_{-a,b}} = b] = a/(a+b)$[35] we conclude that

[34] Lemma A.3 in Partzsch & Schilling (2012)

[35] Where as previously $\tau_{-a,b} := \inf\{t \geq 0 : B_t \in \{-a, b\}\}$ and the distribution of $B_{\tau_{-a,b}}$ may be derived from Wald's identities (See Corollary 5.11 in Partzsch & Schilling (2012)).

$$P[B_{\tau_2} - B_{\tau_1} > 0 \mid B_{\tau_1}] = \frac{B_{\tau_1} - f_2(B_{\tau_1}, -1)}{f_2(B_{\tau_1}, 1) - f_2(B_{\tau_1}, -1)},$$

Continuing, since X_n is a martingale we have that

$$\begin{aligned} X_{n-1} &= E[X_n|X_1,\dots,X_{n-1}] = E[f_n(X_1,\dots,X_{n-2},D_n)|X_1,\dots,X_{n-1}] \\ &= E[f_n(X_1,\dots,X_{n-1},-1)1_{D_n=-1} + f_n(X_1,\dots,X_{n-1},1)1_{D_n=1}|X_1,\dots,X_{n-1}] \\ &= f_n(X_1,\dots,X_{n-1},-1)E[1_{D_n=-1}|X_1,\dots,X_{n-1}] + f_2(X_1,\dots,X_{n-1},1)E[1_{D_n=1}|X_1,\dots,X_{n-1}] \end{aligned}$$

From which we get[36]

$$\frac{X_1 - f_n(X_1,\dots,X_{n-1},-1)}{f_n(X_1,\dots,X_{n-1},1) - f_n(X_1,\dots,X_{n-1},-1)} = P[D_n = 1|X_1,\dots,X_{n-1}] \tag{5}$$

From (4) and (5) one derives that $(B_{\tau_n})_{n\geq 0} \overset{\mathcal{D}}{=} (X_n)_{n\geq 0}$ using induction as follows:

For $n = 1$, (4) and (5) says that

$$P[D_1 = 1] = P[D_1 = 1|X_0] = \frac{-f_1(-1)}{f_1(1) - f_1(-1)}$$

and

$$P[\widetilde{D}_1 = 1] = P[\widetilde{D}_1 = 1|X_0] = \frac{-f_1(-1)}{f_1(1) - f_1(-1)},$$

which says that $D_1 \overset{\mathcal{D}}{=} \widetilde{D}_1$.

Now [37]

$$B_{\tau_1} = f_1(1) \iff \widetilde{D}_1 = 1$$

and

[36] Noting that $P[D_n = -1|X_1,\dots,X_{n-1}] = 1 - P[D_n = 1|X_1,\dots,X_{n-1}]$

[37] Remember that $B_\tau = f(1) \iff B_\tau \geq B_{\tau_0} = 0$

$$X_1 = f_1(1) \iff D_1 = 1$$

implying that

$$P[B_{\tau_1} = f_1(1)] = P[\widetilde{D}_1 = 1] = P[D_1 = 1] = P[X_1 = 1]$$

so that $B_{\tau_1} \overset{\mathcal{D}}{=} X_1$.

In the inductive step we assume that $(B_{\tau_i})_{0\leq i\leq n-1} \overset{\mathcal{D}}{=} (X_i)_{0\leq i\leq n-1}$. Then equations (4) and (5) implies that

$$P[\widetilde{D}_n|B_{\tau_1},\ldots,B_{\tau_{n-1}}] \overset{\mathcal{D}}{=} P[D_n|X_1,\ldots,X_{n-1}].$$

In particular this means that for any $a_1,\ldots,a_{n-1}$ such that $P[B_{\tau_1} = a_1,\ldots,B_{\tau_{n-1}} = a_{n-1}] = P[X_1 = a_1,\ldots,X_{n-1} = a_{n-1}] > 0$

$$\begin{aligned} P[\widetilde{D}_n = 1, B_{\tau_1} = a_1,\ldots,B_{\tau_{n-1}} = a_{n-1}]& \\ = P[\widetilde{D}_n = 1|B_{\tau_1} = a_1,\ldots,B_{\tau_{n-1}} = a_{n-1}]P[B_{\tau_1} = a_1,\ldots,B_{\tau_{n-1}} = a_{n-1}]& \\ = P[D_n = 1|X_1 = a_1,\ldots,X_{n-1} = a_{n-1}]P[X_1 = a_1,\ldots,X_{n-1} = a_{n-1}]& \\ = P[D_n = 1, X_1 = a_1,\ldots,X_{n-1} = a_{n-1}]& \end{aligned}$$

so that

$$(\widetilde{D}_n, B_{\tau_1},\ldots,B_{\tau_{n-1}}) \overset{\mathcal{D}}{=} (D_n, X_1,\ldots,X_{n-1}). \tag{6}$$

Now

$$B_{\tau_n} = f(a_1,\ldots,a_{n-1},1) \iff \widetilde{D}_n = 1 \wedge B_{\tau_1} = a_1 \wedge \cdots \wedge B_{\tau_{n-1}} = a_{n-1}$$

and

$$X_n = f(a_1,\ldots,a_{n-1},1) \iff D_n = 1 \wedge X_1 = a_1 \wedge \cdots \wedge X_{n-1} = a_{n-1}.$$

Thus (6) implies $(B_{\tau_n})_{n\geq 0} \stackrel{\mathcal{D}}{=} (X_n)_{n\geq 0}$.

We now proceed to show that $B_\tau = \lim_n X_n$. From the strong Markov property applied to the Brownian motion $B_{\tau_{n-1}+t} - B_{\tau_{n-1}}$ we note that

$$\tau_n - \tau_{n-1} = \inf\left\{t \geq 0 : B_{t+\tau_{n-1}} - B_{\tau_{n-1}} \in \{f_1(-1), f_1(1)\}\right\}$$
$$\stackrel{\mathcal{D}}{=} \inf\left\{t \geq 0 : B_t \in \{f_1(-1), f_1(1)\}\right\},$$

and conclude that

$$E[\tau_n - \tau_{n-1}] = E[(B_{\tau_n} - B_{\tau_{n-1}})^2].$$

And since a martingale has uncorrelated increments implying that $E[(X_n - X_{n-1})^2] = E[X_n^2] - E[X_{n-1}^2]$, one may with induction conclude that

$$E[X_n^2] = E[\tau_n].$$

We may now apply Theorem 11 to conclude that there exists some square integrable random variable X_∞ such that

$$X_n \to X_\infty \text{ almost surely and in } L^2.$$

Since $|E[X_n^2] - E[X_\infty^2]| \leq E[|X_n^2 - X_\infty^2|] \to 0$ we also have $E[X_n^2] \to E[X_\infty^2]$, and applying the monotone convergence theorem[38] to the nonnegative increasing sequence $(\tau_n)_{n\geq 0}$ we conclude

$$E[\tau] = \lim_n E[\tau_n] = \lim_n E[X_n^2] = E[X_\infty^2].$$

As a last step we conclude that since Brownian motion has continuous sample paths,

$$B_\tau = \lim B_{\tau_n} = \lim_n X_n \stackrel{\mathcal{D}}{=} X_\infty.$$

The last theorem says that from a binary splitting martingale we may construct a sequence of stopping times; for the stopping time τ that is the limit of these stopping times, the Brownian

[38] Theorem 4.20 in Klenke (2013)

motion stopped at τ is distributed as the limit of the binary splitting martingale. Further the expectation of the stopping time τ equals the variance of the limit of the binary splitting martingale. Thus given a centered random variable X with finite second moment, Lemma 2 guarantees a binary splitting martingale converging to it; and Theorem 12 let us construct stopping time τ such that $B(\tau)$ is distributed as X. This concludes Dubins solution to the Skorokhod stopping problem.

5.3 From an embedding to Donsker's invariance principle

Now that we may embed a centered random variable into a Brownian motion the idea is to use the strong Markov property of Brownian motion:[39] first we embed ξ_1 into B_t, then we embed ξ_2 into the new Brownian motion $B_2(t) := B(\tau_1 + t) - B(\tau_1)$ starting at τ_1, then we embed ξ_3 into $B_3(t) := B_2(\tau_2 + t) - B(\tau_2)$ starting at τ_2 etc. The important thing here is of course the strong Markov property that says that $B(\tau_1 + t) - B(\tau_1)$ is a Brownian motion and independent of $\mathcal{F}_{\tau_1+}$, which allows us to define a new stopping time τ_2 that is measurable w.r.t.

$$\mathcal{F}^{(2)}_{\tau_2+} := \{A \in \mathcal{F}^{(2)}_{\infty} : \forall t,\ A \cap \{\tau_2 \le t\} \in \mathcal{F}^{(2)}_{t+}\},$$

where $\mathcal{F}^{(2)}$ denotes the filtration generated by B_2. Since $\mathcal{F}^{(2)}_{\tau_2+}$ per definition is a sub-sigma algebra of $\mathcal{F}^{(2)}$ which is independent of $\mathcal{F}_{\tau_1+}$, and $B_2(\tau_2)$ is measurable w.r.t. $\mathcal{F}^{(2)}_{\tau_2+}$, $B_2(\tau_2)$ is independent of $\mathcal{F}_{\tau_1+}$ - and thus of $B(\tau_1)$. Thus the embedded random variables are independent.

Then we claim that for every n there are stopping times $\tau_1 \le \tau_2 \le \cdots$ such that

$$B(\tau_1 + \cdots \tau_n) \stackrel{\mathcal{D}}{=} \xi_1 + \cdots + \xi_n \text{[40]}$$

This may be derived as follows: Find τ_1 such that

$$\xi_1 \stackrel{\mathcal{D}}{=} B(\tau_1);$$

then (to see the pattern) find τ_2 such that

$$\xi_2 \stackrel{\mathcal{D}}{=} B_2(\tau_2) := B(\tau_1 + \tau_2) - B(\tau_1) \implies B(\tau_1 + \tau_2) = B(\tau_1) + B_2(\tau_2) \stackrel{\mathcal{D}}{=} \xi_1 + \xi_2$$

[39] This approach to Donsker's theorem is due to D. Freedman. (1971). *Brownian motion and diffusion* Springer New York.

[40] Note that a sum of stopping times is a stopping time.

And by induction, if

$$
\begin{aligned}
&B(\tau_1 + \cdots + \tau_{n-1}) \overset{\mathcal{D}}{=} \xi_1 + \cdots + \xi_{n-1} \\
\text{and } &B_{n-1}(\tau_{n-1} + \tau_n) = B(\tau_1 + \cdots + \tau_{n-1} + \tau_n) - B(\tau_1 \cdots \tau_{n-2}) \\
\text{and } &B_{n-1}(\tau_{n-1}) = B(\tau_1 + \tau_2 + \cdots \tau_{n-1}) - B(\tau_1 + \cdots + \tau_{n-2})
\end{aligned}
$$

then

$$
\begin{aligned}
\xi_n \overset{\mathcal{D}}{=} B_n(\tau_n) &:= B_{n-1}(\tau_{n-1} + \tau_n) - B_{n-1}(\tau_{n-1}) \\
&= B(\tau_1 + \cdots + \tau_n) - B(\tau_1 + \cdots + \tau_{n-1})
\end{aligned}
$$

which implies

$$
\begin{aligned}
B(\tau_1 + \cdots + \tau_n) &= B(\tau_1 + \cdots + \tau_{n-1}) + B_n(\tau_n) \\
&\overset{\mathcal{D}}{=} (\xi_1 + \cdots + \xi_{n-1}) + \xi_n.
\end{aligned}
$$

Now, as in the first proof, we use the fact that the interpolated random walks are monotone between the points k/n, $k = 0, \cdots n$ and that S_n may be written as a stopped Brownian motion, to derive that the random functions $\omega \mapsto B_\omega(nt)/\sqrt{n}$ and $\omega \mapsto X^{(n)}(\omega)$ converge in probability.

Lemma 3. *For a Brownian motion $(B_t)_{t\geq 0}$ and a random variable ξ with mean zero and variance one, the interpolated random walk $X_t^{(n)}$ with increments from the law of ξ satisfies*

$$
\lim_{n\to\infty} P\left(\sup_{0\leq t\leq 1}\left|\frac{B(nt)}{\sqrt{n}} - X_t^{(n)}\right| > \varepsilon\right) = 0
$$

for any $\varepsilon > 0$.

The proof of the Lemma proceeds as follows: From the preceding construction we may write the random walk S_n - from which $X_t^{(n)}$ is defined - as

$$
S_n = B(\tau_1 + \cdots + \tau_n)
$$

One then considers the set

$$A_n := \left\{ \sup_{0\le t\le 1} \left| \frac{B(nt)}{\sqrt{n}} - X_t^{(n)} \right| > \varepsilon \right\}$$

which per the definition of the supremum equals the set

$$\left\{ \exists t \in [0,1) : \left| \frac{B(nt)}{\sqrt{n}} - X_t^{(n)} \right| > \varepsilon \right\}$$ [41]

Define $k(t)$ to be the unique integer k such that

$$\frac{k-1}{n} \le t < \frac{k}{n}$$

then using the fact that $X_t^{(n)}$ is monotone between the "nodes" $(k(t)-1)/n$ and $k(t)/n$ we get

$$A_n \subset \left\{ \exists t \in [0,1) : \left| \frac{X_{k(t)-1}^{(n)}}{\sqrt{n}} - \frac{B(nt)}{\sqrt{n}} \right| > \varepsilon \right\} \bigcup \left\{ \exists t \in [0,1) : \left| \frac{X_{k(t)}^{(n)}}{\sqrt{n}} - \frac{B(nt)}{\sqrt{n}} \right| > \varepsilon \right\} =: A_n^*$$

Now we rewrite $X_{k(t)}^{(n)}$ as a stopped Brownian motion $B(\tau_{k(t)})$,

$$A_n^* = \left\{ \exists t \in [0,1) : \left| \frac{B(\tau_{k(t)-1})}{\sqrt{n}} - \frac{B(nt)}{\sqrt{n}} \right| > \varepsilon \right\} \bigcup \left\{ \exists t \in [0,1) : \left| \frac{B(\tau_{k(t)})}{\sqrt{n}} - \frac{B(nt)}{\sqrt{n}} \right| > \varepsilon \right\}$$

For any $0 < \delta < 1$ we may write this as

$$\left\{ \exists s, t \in [0,1) : |t-s| \le \delta \text{ and } \left| \frac{B(s)}{\sqrt{n}} - \frac{B(nt)}{\sqrt{n}} \right| > \varepsilon \right\}$$
$$\bigcup \left\{ \exists t \in [0,1) : \max\left\{ \left| \frac{\tau_{k(t)}}{n} - t \right|, \left| \frac{\tau_{k(t)-1}}{n} - t \right| \right\} \ge \delta \right\}$$

[41]Note that by continuity of the paths, $|\frac{B(n\cdot 1)}{\sqrt{n}} - X_1^{(n)}| > \varepsilon$ iff there exists $t < 1$ such that $|\frac{B(n\cdot t)}{\sqrt{n}} - X_t^{(n)}| > \varepsilon$. Thus we need not include the right endpoint of $[0,1]$.

As a consequence of the Kolmogorov-Chentsov[42] theorem one may derive that Brownian motion is Hölder-continuous with probability one and thus the first set in the union above has probability one for δ sufficiently small.

From our construction of stopping times above one see that for every n

$$\tau_{n+1} - \tau_n \stackrel{\mathcal{D}}{=} \tau_1$$

and the increments are independent. Thus since

$$E[\tau_{n+1} - \tau_n] = E[\tau_1] = E[\xi_1] = 1$$

we may apply the law of large numbers to conclude that

$$\lim_{n\to\infty} \frac{\tau_n}{n} = \lim \frac{1}{n} \sum_{k=1}^{n} (\tau_{n+1} - \tau_n) = 1. \text{ a.s.}$$

Now as $k(t)$ is such that $(k(t) - 1)/n \leq t < k(t)$,

$$\begin{aligned} &\left\{ \exists t \in [0,1) : \max\{ \left| \frac{\tau_{k(t)}}{n} - t \right|, \left| \frac{\tau_{k(t)-1}}{n} - t \right| \} \geq \delta \right\} \\ &\quad \subset \left\{ \exists t \in [0,1) : \max\{ \left| \frac{\tau_{k(t)} - (k(t) - 1)}{n} \right|, \left| \frac{\tau_{k(t)-1} - k(t)}{n} \right| \} \geq \delta \right\} \\ &\qquad \subset \left\{ \max_{0\leq k\leq n} \left| \frac{\tau_k - (k-1)}{n} \right| \geq \delta \right\} \bigcup \left\{ \max_{0\leq k\leq n} \left| \frac{\tau_{k-1} - k}{n} \right| \geq \delta \right\} \end{aligned}$$

For any sequnce of real numbers $\{a_n\}$ one may derive

$$\lim_{n\to\infty} \frac{a_n}{n} = 1 \implies \lim_{n\to\infty} \max_{0\leq k\leq n} \frac{|a_k - (k-1)|}{n} = 0 \text{ and } \lim_{n\to\infty} \max_{1\leq k\leq n} \frac{|a_{k-1} - k|}{n} = 0$$

From which it follows that the probability of the last union goes to zero. We have thus shown what we set out to prove.

[42] Theorem 21.6 in Klenke (2013)

5.4 Proof of Donsker's theorem

We may now apply the part of the Portmanteau theorem that states the equivalence of 1. and 3. to prove weak convergence by proving that

$$\limsup_n P_{X^{(n)}}(F) \le P_B(F)$$

for every closed set F of $C[0,1]$.

Given any closed set F of $C[0,1]$ and $\rho(f,F) := \inf_{g\in F}\sup_{0\le t\le 1}|f(t)-g(t)|$, we define

$$F[k] := \left\{ f \in C[0,1] : \rho(f,F) \le \frac{1}{k} \right\}, \qquad k \in \mathbb{N}$$

it is known that $f \mapsto \rho(x,F)$ is a continuous map and thus $F[k]$ is a closed set. Then

$$P[X^{(n)} \in F] \le P\left[\frac{B(nt)}{\sqrt{n}} \in F[k]\right] + P\left[\sup_{0\le t\le 1}\left|\frac{B(nt)}{\sqrt{n}} - X_t^{(n)}\right| > \varepsilon\right]$$

where we proved in the preceding lemma that the second term on the right hand side goes to zero as $n \to \infty$. Since $F[k]$ is closed we have that

$$F = \bigcap_{k=1}^{\infty} F[k]$$

and from the continuity of measure we get

$$P[F] = \lim_{k\to\infty} P[F[k]]$$

We may thus conclude

$$\limsup_n P[X_n^{(n)} \in F] \le P\left[\frac{B(nt)}{\sqrt{n}} \in F\right].$$

Since $\frac{B(nt)}{\sqrt{n}}$ is again a Brownian motion on $[0,1]$, the proof is complete.

6 Concluding remark

It is hard to directly compare the two proofs apart from stating, as in the introduction, that they make different assumptions and that different consequences thus may be derived from them. As mentioned in the introduction, and as the reader might now be aware of, the first theorem is a possible construction of a Brownian motion. From the second proof, however, one may derive the central limit theorem without having to deal with characteristic functions.

One may want to note that Prokhorov's proof may be extended to where $X_t^{(n)}$ is an element of the space $D[0,1]$ of right-continuous functions with left-sided limits. A similar proof will do except for the characterization we have from Arzela-Ascoli for compact sets in $C[0,1]$ which is different, and not as easy, in $D[0,1]$.

7 Acknowledgements

I would like to thank my advisor Daniel Ahlberg for guiding me through the subject of this thesis. I would also like to thank Alan Sola, who originally suggested the subject to me, and helped me spot errors in the draft.

This study constitutes 15 credits and is done for a bachelor's degree in mathematics at Stockholm University.

8 References

Billingsley P. (1999) *Convergence of probability measures* Wiley New York
Billingsley P. (2012) *Probability and measure* Wiley New York
Brown R. (1827) *A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants: and on the general existence of active molecules in organic and inorganic bodies* Philosophical magazine
Karatzas I. & Shreve E.S. (1991) *Brownian motion and stochastic calculus* Springer New York
Klenke A. (2013) *Probability theory* Springer Berlin
Mörters P. & Peres Y. (2010) *Brownian motion* Cambridge university press United Kingdom
Partzsch L. & Schilling R. (2012) *Brownian motion* Walter de Gruyter Berlin/Boston
Revuz D. & Yor M. (1999) *Continuous Martingales and Brownian Motion* Springer Berlin