Empirical Distribution Theory

Chun-Hei Lam, Shawn Im
Department of Mathematics, *MIT*

Submitted in part fulfillment of the requirements for the course
18.675 Theory of Probability.

Contents

1 Motivation

In 18.675 we have seen the following results:

Theorem 1.1: Strong Law of Large Numbers (SLLN)

Assume $X(\omega), X_1(\omega), X_2(\omega), ..., X_n(\omega)$ are pairwise independent, identitcally distributed (i.i.d.) L^1 *real-valued* random variables define on the space $L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{E}(X) = \mu$. Let

$$S_n(\omega) = \frac{1}{n}\sum_{i=1}^{n} X_i(\omega)$$

then $S_n \to \mu$ $\mathbb{P}$-almost surely (a.s.) with respect to ω. In other words,

$$\mathbb{P}\left(\{\omega \in \Omega \,|\, S_n(\omega) \to \mu\}\right) = 1$$

Theorem 1.2: Central Limit Theorem (CLT)

Assume $(X_i)_{i\geq 1}, X$ are i.i.d. L^2 *real-valued* random variables on space $(\Omega, \mathcal{F}, \mathbb{P})$. Write $\mathbb{E}(X) = \mu$ and Var$(X) = \sigma^2$. Let

$$S_n^*(\omega) = \sqrt{n}S_n(\omega) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i(\omega)$$

Then, in distribution,

$$S_n^*(\omega) \to \mathcal{N}(\mu, \sigma^2)$$

In other words, we have $F_{S_n^*}(x)$ converges to $\Phi_{\mu,\sigma^2}(x)$ pointwise. Here $\Phi_{\mu,\sigma^2}(x)$ is the cumulative distribution function (CDF) of normal distribution $\mathcal{N}(\mu, \sigma^2)$.

Remark. It is very important that $S_n^*(\omega)$ is just a *rescaled* version of $S_n(\omega)$. With enough regularity of $X_i(\omega)$, we may easily make connection between CLT and SLLN – if we divide the sum $\sum_{i=1}^n X_i(\omega)$ with sufficiently high power of n, then the uncertainty of X_i vanishes.

In this paper, we would like to study *empirical distributions*:

Definition 1.3: Empirical Distribution

Assume $X, (X_i(\omega))_{i\geq 1}$ are iid random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. For finite n, fix $x \in \mathbb{R}$. If $F_n := [F_n(x)](\omega)$ is a function from $(\Omega, \mathcal{F})$ to $([0,1], \mathcal{B}([0,1]))$ written as

$$[F_n(x)](\omega) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{(-\infty,x]}(X_i(\omega)) \tag{1}$$

then it is an empirical distribution (process) of X.

Example 1.4: Rolling Dices

Denote the random variable X as outcomes of rolling a six-faced dice. Assume we roll the same dice n times and record the outcomes. We may estimate the 'experimental probability' of getting values not greater than x from the dice by recording the number of times of such event and divide by n. Such 'experimental probability' is in fact an empirical distribution. Here is a result of experiment corresponding to an ω with n = 100:

x	1	2	3	4	5	6
Number of times when outcome = x	21	19	13	22	10	15
Number of times when outcome $\leq$ x	21	40	53	75	85	100
$[F_n(x)](\omega)$	0.21	0.40	0.53	0.75	0.85	1

Table 1.1: A result of experiment with n = 100

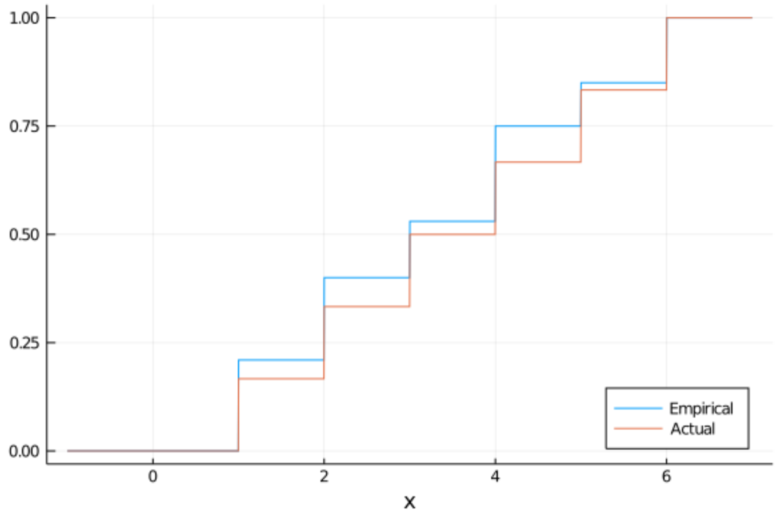

Figure 1.1: Plot of $F_n(x)$ and $F_X(x)$

Remark. Note that $\mathbb{I}_{(-\infty,x]}(X_i(\omega)) = 1 \iff X_i(\omega) \leq x \iff \mathbb{I}_{[X_i(\omega),\infty)}(x) = 1$. Therefore (1) can also be written as

$$[F_n(x)](\omega) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{[X_i(\omega),\infty)}(x) \tag{2}$$

Often the first step of understanding unknown distributions is to obtain samples from the distribution by experiments and histogram the sample. This is equivalent to obtain an empirical distribution from the unknown distribution. In this paper we would like to understand how well can we understand the actual distribution from empirical distributions with large n. In other words, *in what way* does an empirical distribution converges to the actual distribution itself? Without further clarification we assume that $X, (X_i(\omega))$ are iid L^2 random variables on space $(\Omega, \mathcal{F}, \mathbb{P})$ with distribution F_X, and $[F_n(x)](\omega)$ as defined in Definition 1.3.

It is important to note that an empirical distribution is in fact a *real-valued* (continuous time) stochastic process, that is a collection of random variables from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with index (in our case the index is $x \in \mathbb{R}$). The Ionescu-

Tulcea theorem suggests that we can view a stochastic process as a *single* random variables from $(\Omega, \mathcal{F}, \mathbb{P})$ to a large space of sample paths. In our case, the random variable takes 'value' from the *Skorokhod Space* $D([0,\infty], \mathbb{R})$, the collection of all right continuous functions from $[-\infty, \infty]$ to $\mathbb{R}$ with left limits, known as *catalag*. We may discuss the convergence of empirical distributions as convergence of random variables, except the random variables are not real-valued in our case.

Remark. The following is the metric for the Skorokhod Topology

$$d_{D([0,\infty])}(f,g) = \sum_{k=1}^{\infty} \frac{1}{2^k} \min\left\{1, \inf_{\lambda} \sup_{0\le s\le k} (|\lambda(s) - s| + d_{\mathbb{R}}(f(\lambda(s)), g(s))\right\} \tag{3}$$

where λ is any strictly increasing continuous function from $[0,k] \to [0,\infty)$ and $d_{\mathbb{R}}$ is the metric on $\mathbb{R}$.

1.1 A First Try: Glivenko Cantelli Theorem

Perhaps it is easier to establish the a.s. uniform convergence of empirical processes. We start by noting a trivial observation:

Lemma 1.5: Pointwise Convergence of Empirical Processes

For all fixed $x \in \mathbb{R}$, we have $[F_n(x)](\omega) \to F_X(x)$ almost surely w.r.t. ω.

Proof. Fix $x \in \mathbb{R}$. We look at the collection $(\mathbb{I}_{(-\infty,x]}(X_i))_{i\ge1}$. This is in L^1 since

$$\mathbb{E}(\mathbb{I}_{(-\infty,x]}(X_i)) = \mathbb{P}(X_i \le x) = \mathbb{P}(X \le x) = F_X(x) \le 1$$

In addition, the collection is independent. This is clear from the fact that $(X_i)_{i\ge1}$ (hence $(\sigma(X_i))_{i\ge1}$) are independent and $\sigma(\mathbb{I}_{(-\infty,x]}(X_i))$ is a sub-σ-algebra of $\sigma(X_i)$.

$$\sigma(\mathbb{I}_{(-\infty,x]}(X_i)) = \left\{\phi,\ X_i^{-1}((-\infty,x]),\ X_i^{-1}((-\infty,x])^C, \Omega\right\} \subseteq \sigma(X_i)$$

Finally, the distribution of $\mathbb{I}_{(-\infty,x]}(X_i)$ is identical for all i. Therefore by SLLN:

$$\forall x \in \mathbb{R},\ \mathbb{P}(\{\omega \in \Omega \,|\, [F_n(x)](\omega) \to F_X(x)\}) = 1$$

$\square$

Remark. Using same reasoning we may show that for all x, the left limits converges almost surely w.r.t. ω as well. In other words, $[F_n(x-)](\omega) \to F_X(x)$.

It says that for almost sure ω, if we take a slice from the sequence of sample paths $[F_n(x)](\omega)$, then we know that the slice will converge to slice of $F_X(x)$. However, we cannot conclude the (uniform) convergence of the sample path itself. We will show that the fact that $[F_n(x)](\omega)$ and $F_X(x)$ is a non-decreasing catalag guarantees that pointwise convergence of $[F_n(x)](\omega)$ implies uniform convergence.

Theorem 1.6: Glivenko-Cantelli Theorem

$$\lim_{n\to\infty} \|[F_n(x)](\omega) - F_X(x)\|_{D([-\infty,\infty])} = 0$$

almost surely w.r.t. ω.

Proof. The idea of proof is to partition $\mathbb{R}$ into finitely many intervals (given $\epsilon > 0$ by finitely many "grid points". We know that at those grid points, $[F_n(x)](\omega)$ is closed to $F_X(x)$. The monotonicity of $F_n(x)$ and $F_X(x)$ guarantees that even if x is not those grid points, $F_n(x)(\omega)$ is not too far from $F_X(x)$.

Fix $\epsilon > 0$ and let k be a positive integer such that $1/k < \epsilon/2$. We define the "grid points" $x_{j,k} = \inf A_{j/k}$ for all $j = 1, ..., k-1$. Such an infinmum always exists and is unique, otherwise either $F_n(-\infty)(\omega) > 0$ or $F_n(\infty)(\omega) < 0$, in both ways contradicting to the fact that $F_n(x)$ is a CDF itself. For our convenience define $x_{0,k} = -\infty$ and $x_{k,k} = \infty$. Note that the sequence of grid points $(x_{j,k})_{j=0}^k$ is non-decreasing. With this we ensure that any x in $\mathbb{R} \cup \{\pm\infty\}$ must fall into exactly one of the interval $[x_{j,k}, x_{j+1,k})$ (note that the intervals might be empty sets.)

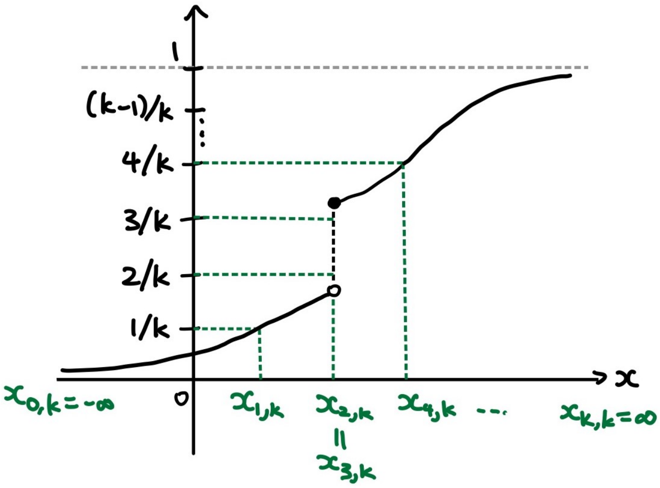

Figure 1.2: Grid Points

By pointwise convergence, for all $j = 1, ..., k-1$ there are positive integers $N_{j,k,\omega}$ such that for all $n \geq N_{j,k,\omega}$ we have

$$|[F_n(x_{j,k})](\omega) - F_X(x_{j,k})| < 1/2k$$

There are also positive integers $N^-_{j,k,\omega}$ such that for all $n \geq N_{j,k,\omega}$ we have

$$|[F_n(x_{j,k}-)](\omega) - F_X(x_{j,k}-)| < 1/2k$$

Take $N_{k,\omega} = \max\left(\max_{j=1,...,k-1} N_{j,k,\omega}, \max_{j=1,...,k-1} N^-_{j,k,\omega}\right)$, we have for all $n \geq N_{k,\omega}$,

$$\forall j = 1, ..., k-1, \quad |[F_n(x_{j,k})](\omega) - F_X(x_{j,k})| < 1/2k$$

Now let $x \in \mathbb{R}$ be falling into one of the intervals $[x_{j,k}, x_{j+1,k})$. We immediately have

$$[F_n(x_{j,k})](\omega) \leq [F_n(x)](\omega) \leq [F_n(x_{j,k+1}-)](\omega) \leq [F_n(x_{j,k+1})](\omega)$$

In particular, when $n \geq N_{k,\omega}$, we have

$$F_X(x_{j,k}) - 1/2k < [F_n(x_{j,k})](\omega) F_X(x_{j,k}) + 1/2k$$
$$F_X(x_{j+1,k}-) - 1/2k \leq [F_n(x_{j+1,k}-)](\omega) < F_X(x_{j+1,k}-) + 1/2k$$

This lead to

$$[F_n(x)](\omega) \in \left(F_X(x_{j,k}) - \frac{1}{2k},\ F_X(x_{j+1,k}) + \frac{1}{2k} \right)$$

But we also have

$$F_X(x) \in [F_X(x_{j,k}),\ F_X(x_{j+1,k}-)) \subseteq \left(F_X(x_{j,k}) - \frac{1}{2k},\ F_X(x_{j+1,k}) + \frac{1}{2k} \right)$$

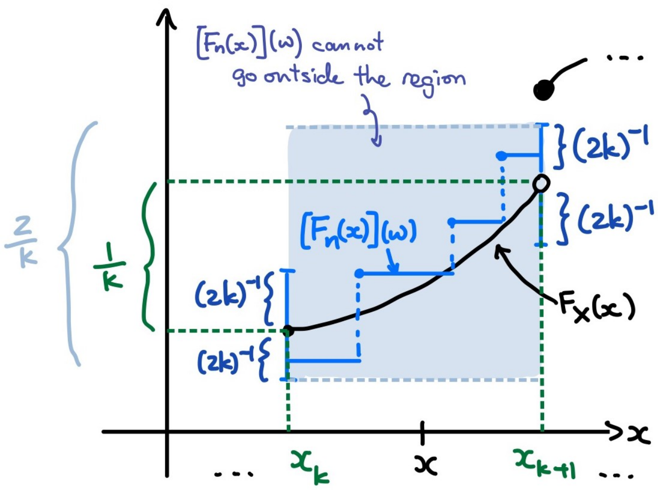

Figure 1.3: The bound of $[F_n(x)](\omega)$ for $x \in [x_k, x_{k+1})$

The interval at RHS has length $2/k$, so $|[F_n(x)](\omega) - F_X(x)| < 2/k < \epsilon$. This leads to uniform convergence. □

1.2 Central Limit Theorem?

Notice that

$$[F_n(x)](\omega) - F_X(x) = \frac{1}{n} \sum_{i=1}^{n} (\mathbb{I}_{(-\infty,x]}(X_i(\omega)) - F_X(x)) \tag{4}$$

We may see Glivenko-Cantelli Theorem an anology of SLLN on the stochastic process $(\mathbb{I}_{(-\infty,x]}(X_i(\omega)) - F_X(x))$. Is there an analogy of CLT for the stochastic process $(\mathbb{I}_{(-\infty,x]}(X_i(\omega)) - F_X(x))$ – in particular what is the limiting behavior of

$$\sqrt{n}\, \|[F_n(x)](\omega) - F_X(x)\|_{D([-\infty,\infty])} = \sup_x \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathbb{I}_{(-\infty,x]}(X_i(\omega)) - F_X(x)) \right)$$

One should note that this random variable is much harder to study since it does not converge to zero almost surely. We will break down into two steps:

1. Study the convergence of $\sqrt{n}([F_n(x)](\omega) - F_X(x))$. This leads us to the *Empirical CLT*. In particular $\sqrt{n}\,\|[F_n(x)](\omega) - F_X(x)\|_{D([-\infty,\infty])}$ converges by an appropriate version of Continuous Mapping Theorem. (CMT)
2. Study the distribution of the limiting distribution of $\sqrt{n}\,\|[F_n(x)](\omega) - F_X(x)\|_{[-\infty,\infty]}$.

We first study the protypical example when $X, (X_i)_{i\geq 1}$ are uniformly distributed over $[0,1]$, or U$[0,1]$ distributed. In this case, both $F_n(x)$ and $F_X(x)$ take trivial values outside $[0,1]$ (0 on $[-\infty, 0)$ and 1 on $(1,\infty]$). We may therefore restrict $[F_n(x)](\omega)$ to $[0,1]$ and treat them as *(random) elements* in $D([0,1])$, i.e. catalag defined over $[0,1]$. We will see that results of convergence of $[F_n(x)](\omega)$ in this case can be extended to more general settings via CMT.

Again we may define the uniform norm on $D([0,1])$

$$\|\,.\,\|_{D([0,1])} : f \mapsto \sup_{x\in[-\infty,\infty]} |f(x)| \tag{5}$$

which induces a metric $d_{D([0,1])}(f,g)$. Furthermore, we write

We further define a coordinate map

Definition 1.7: Coordinate Map from Skorokhod Space

Let $S = \{s_1, ..., s_n\}$ be a finite subset of $[0,1]$. The coordinate map $\pi_S : D([0,1]) \to \mathbb{R}^n$ is defined as

$$\pi_S(F) = (F(s_1), F(s_2), ..., F(s_n))^T \tag{6}$$

We abuse notation and write $\pi_t(F) = \pi_{\{t\}}(F) = F(t)$ for all $t \in [0,1]$.

It is trivial to see that when S is fixed, π_S is continuous (the components is continuous and the codomain is finite dimensional). Finally we write (in the case when X and $(X_i)_{i\geq 1}$ are U$[0,1]$ distributed)

$$U_n := [U_n(x)](\omega) = \sqrt{n}([F_n(x)](\omega) - F_X(x)) \tag{7}$$

Fix $s,t \in [0,1]$ with $s \leq t$, we have $\mathbb{E}(U_n(t)) = 0$ for all n. Moreover,

$$\begin{aligned}
\text{cov}(\mathbb{I}_{(-\infty,s]}(X), \mathbb{I}_{(-\infty,t]}(X)) &= \mathbb{E}(\mathbb{I}_{(-\infty,s]}(X)\mathbb{I}_{(-\infty,t]}(X)) - \mathbb{E}(\mathbb{I}_{(-\infty,s]}(X))\mathbb{E}(\mathbb{I}_{(-\infty,t]}(X)) \\
&= \mathbb{E}(\mathbb{I}_{(-\infty,s]}(X)) - \mathbb{E}(\mathbb{I}_{(-\infty,s]}(X))\mathbb{E}(\mathbb{I}_{(-\infty,t]}(X)) \\
&= s - st \qquad (8) \\
\text{Var}(\mathbb{I}_{(-\infty,t]}(X)) &= t - t^2 \qquad (9)
\end{aligned}$$

This lead to

$$\begin{aligned}
\text{cov}(U_n(s), U_n(t)) &= s - st \qquad (10) \\
\text{Var}(U_n(t)) &= t - t^2 \qquad (11)
\end{aligned}$$

Using our notation, Lemma 1.5 can now be rewritten as $\pi_t([F_n(x)](\omega)) \to F_X(t)$ a.s. (w.r.t. ω) for all fixed t. In fact CLT asserts further that for fixed $t \in [0,1]$, one has $\pi_t([U_n(x)](\omega)) \to \mathcal{N}(0, t-t^2)$ in distribution. We further note that the multivariate CLT asserts that for any fixed $S = \{s_1, ..., s_n\}$, $\pi_S(U_n(x))$ converges in distribution to the multivariate normal distribution $\mathcal{N}_n(\vec{0}, \Sigma_n)$. Here Σ_n is the covariance matrix with entries $(\Sigma)_{ij} = \min(s_i, s_j) - s_i s_j$.

There is a stochastic process indexed on $[0,1]$ which satisfies all properties above, that is the Brownian Bridge. Here we use the following definition of Brownian Bridge:

Definition 1.8: Brownian Bridge

A stochastic process $[U(t)](\omega)$ on $(\Omega, \mathcal{F}, \mathbb{P})$ indexed with $t \in [0,1]$ is a Brownian Bridge if

1. $U(0) = U(1) = 0$ a.s.
2. $[U(t)](\omega)$ is continuous w.r.t. t a.s.
3. For each finite subset S of $[0,1]$ the random vector $\pi_S U$ has a multivariate normal distribution with zero means and covariances given by $\mathbb{E}(U(s)U(t)) = s - st$ for all $s, t \in S$; $s < t$.

Remark. Recall that there is another stochastic process $[W(t)](\omega)$ from $(\Omega, \mathcal{F}, \mathbb{P})$ to $C^0([0,\infty)$ closely related to Brownian Bridge in the sense that it satisfies the following properties

1. $W(t) = 0$ a.s. with respect to ω.
2. $W(t)$ is continuous w.r.t. t a.s.
3. For all times $0 < t_1 < t_2 < ... < t_n$, the random variables $W(t_1), W(t_2) - W(t_1), ..., W(t_n) - W(t_{n-1})$ are independent, and

$$W(t_{i+1}) - W(t_i) \sim N(0, t_{i+1} - t_i)$$

Such a process is called a Brownian Motion. Notice from (3) we immediately establish that Brownian Motion is Markov.

See appendix for standard theorems of Brownian Bridge and Brownian Motion.

The Empirical Central Limit Theorem, as we will prove in Chapter 3, asserts that the empirical distribtuion $[F_n(x)](\omega)$ converges in distribution to the Brownian Bridge $[U(F_X(x))](\omega)$. Of course, for the case when $F_X(x)$ is continuous, $[U(F_X(x))](\omega)$ has the same distribution as $[U(x)](\omega)$.

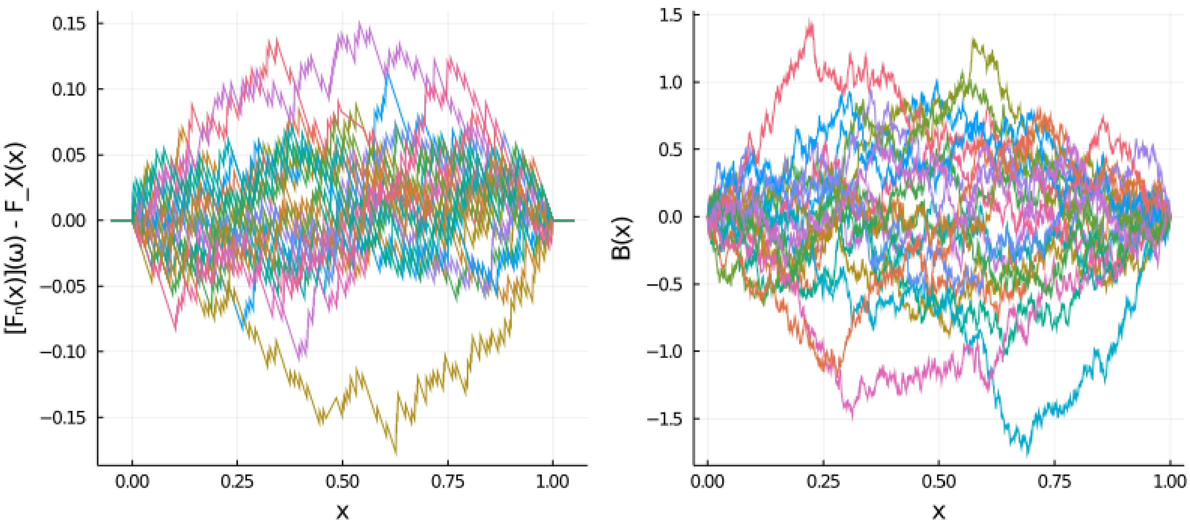

Figure 1.4: Simulations of the difference U_n and Brownian Bridge U

It is hard to compare the simulations. One should note that

$$\sqrt{n}\, \|[F_n(x)](\omega) - F_X(x)\|_{D([0,1])} \to \|[U(x)](\omega)\|_{D([0,1])} \tag{12}$$

This is a direct result of CMT (and the fact that the norm function is continuous). Here the random variable at RHS is *slightly* [1] easier to visualise, since we will prove that it can be written as infinite sum of exponentials. The following plot is a visualisation of the theorem.

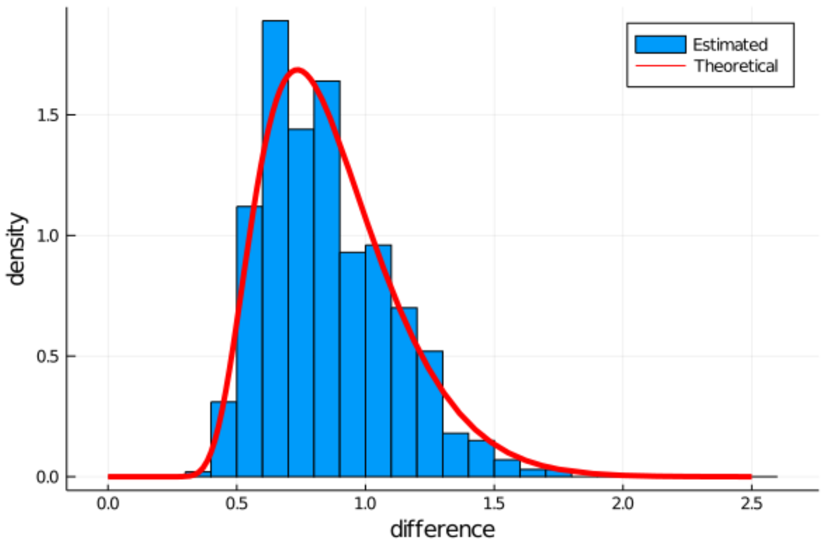

Figure 1.5: Simulations of the functionals $\|U_n\|_{D([0,1])}$ and Brownian Bridge $\|U\|_{D([0,1])}$

[1] but not significantly

2 Continuous Mapping Theorem

One of the steps we will be taking in proving the Empirical CLT is proving the Continuous Mapping Theorem (CMT) since an the empirical distribution process is a stochastic process and the CMT is necessary to show that stochastic processes converge.

The CMT over Euclidean spaces goes as follows.

Theorem 2.1: Continuous Mapping Theorem (CMT), Euclidean Space

Let H be a measurable mapping from $\mathbb{R}^k$ into $\mathbb{R}^s$, and let C be the set of points in $\mathbb{R}^k$ at which H is continuous. If a sequence $\{X_n\}$ of random vectors taking values in $\mathbb{R}^k$ converges in distribution to a random vector X for which $\mathbb{P}\{X \in C\} = 1$, then $HX_n \xrightarrow{d} HX$.

In order to prove this theorem, we will need the Convergence Lemma.

Lemma 2.2: Convergence Lemma

Let h be a bounded, measurable, real-valued function of $\mathbb{R}^k$, continuous at each point of a measurable set C.

(i) Let $\{X_n\}$ be a sequence of random vectors such that $X_n \xrightarrow{d} X$. If $\mathbb{P}\{X \in C\} = 1$, then $\int h(X_n)\mathbb{P}(dx) \to \int h(X)\mathbb{P}(dx)$.

(ii) Let $\{P_n\}$ be a sequence of probability measures converging weakly to P. If $P(C) = 1$, then $\int_X h(x)P_n(dx) \to \int_X h(x)P(dx)$.

Remark. When proving the metric space form of CMT, we will be using a very similar Convergence Lemma.

Proof. We will show that there is an increasing sequence of bounded, continuous functions $f_i \leq h$ everywhere such that $f_i \uparrow h$ at each point of continuity of h. If there is such a sequence of F_i, then since $\{P_n\}$ weakly converges to P, we know that for each i we have

$$\int_X f_i(x)P_n(dx) \to \int_X f_i(x)P(dx)$$

Then, since f_i is bounded above by h, we have for each i,

$$\liminf_{n\to\infty} \int_X h(x)P_n(dx) \geq \int_X f_i(x)P_n(dx) \to \int_X f_i(x)P(dx)$$

Using the Monotone Convergence Theorem on f_i, and since $f_i \uparrow h$, then

$$\liminf_{n\to\infty} \int_X h(x)P_n(dx) \geq \int_X h(x)P(dx)$$

Repeating the same steps as above except with $\limsup$ and $-h$, we get

$$\limsup_{n\to\infty} \int_X h(x)P_n(dx) \leq \int_X h(x)P(dx)$$

which completes the proof once we show $\{f_i\}$ can be constructed.

Now, we will show how to construct the family of functions $\{f_i\}$.
First, we will assume $h > 0$ without loss of generality as we can add constants. Then, for each subset A of $\mathbb{R}^k$, we define the distance function d to be

$$d(x, A) = \inf\{|x - y| : y \in A\} \tag{13}$$

which is a continuous function of x for a fixed A. Then, let

$$f_{m,r}(x) = r^m d(x, \{h \leq r\}) \tag{14}$$

where m is a positive integer and r is a positive rational. Then, each $f_{m,r}$ is bounded and continuous and at most r if $h(x) > r$ and 0 if $h(x) \leq r$. Then, for each $x \in C$ and $\epsilon > 0$, we can choose an r such that $h(x) - \epsilon < r < h(x)$ and since h is continuous as x, it is greater than r is some neighborhood of x. Then, $d(x, \{h \leq r\}) > 0$ and $f_{m,r}(x) = r > h(x) - \epsilon$ for all m large enough. Then, we have a sequence f_i that satisfies the conditions we stated at the begining.

The proof for (i) is very similar except with a switch in the variables and functions. □

From this lemma, we get the following corollary that proves the CMT.

Corollary 2.3

If $\int_X f(x)P_n(dx) \to \int_X f(x)P(dx)$ for every bounded, uniformly continuous f, then $P_n \xrightarrow{d} P$. Similarly, for convergence in distribution of a sequence of random vectors.

From this, the CMT must be true as for each f in the class of all bounded, continuous, real functions on $\mathbb{R}^k$, $f \circ H$ is bounded and continuous at all point in C where H is continuous.

In order to move on to proving the general CMT, we need to introduce a few definitions as we will now be working with a general σ-algebra generated from a metric space.

The first thing we need to define is a $\mathcal{F}/\mathscr{A}$-measurable map as the Lebesgue integral of $f(U)$ is only well defined when $f(U)$ is measurable, and if U is on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then we can make sure $f(U)$ is measurable by taking a σ-field $\mathscr{P}$ such that U is $\mathcal{F}/\mathscr{P}$-measurable and f is $\mathscr{P}$-measurable. Taking the borel σ-field does not always lead to $f(U)$ being measurable, so we consider a more general case to work around this.

Definition 2.4: $\mathcal{F}/\mathscr{A}$-measurable map

An $\mathcal{F}/\mathscr{A}$-measurable map X from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ into a set $\mathscr{X}$ with a σ-field $\mathscr{A}$ is called a random element of $\mathscr{X}$.

The second thing we need to define are completely regular points as this is necessary for the proof of the general Convergence Lemma as we use a countable subfamily of functions whose pointwise supremum is h for all points in the set C. Before C was the set of continuinty points, but in order to have $\mathscr{A}$-measurability, we need C to be a set of completely regular points. Let us now define this term.

Definition 2.5: Completely regular points

A point x in $\mathscr{X}$ is completely regular with respect to the metric d and the σ-field $\mathscr{A}$ if for each neighborhood V of x, there exists a uniformly continuous, $\mathscr{A}$-measurable function g with $g(x) = 1$ and $g \leq V$.

Now, that all the terms are defined, let us look at the general Convergence Lemma.

Lemma 2.6: (General) Convergence Lemma

Let h be a bounded, $\mathscr{A}$-measurable, real-valued function on $\mathscr{X}$. If h is continuous at each point of some separable, $\mathscr{A}$-measurable set C of completely regular points, then

(i) $X_n \xrightarrow{d} X$ and $\mathbb{P}\{X \in C\} = 1 \implies \int h(X_n)\mathbb{P}(dx) \to \int h(X)\mathbb{P}(dx)$

(ii) $P_n \xrightarrow{d} P$ and $PC = 1 \implies \int_X h(x)P_n(dx) \to \int_X h(x)P(dx)$

The proof for this lemma follows the same steps as the previous Convergence Lemma, except it takes more steps to establish that an increasing sequence of bounded continuous functions f_i approximating h exists. For a detailed proof see [??].

Similarly to the earlier Convergence Lemma, the corollary that follows shows that the more general Continuous Mapping Theorem is true.

Here, we state a more general CMT as well as the corollary. The proof of this version of CMT is similar to the previous version so we present it without proof.

If $\int_{X_n} f(x)\mathbb{P}(dx) \to \int_X f(x)\mathbb{P}(dx)$ for every bounded, uniformly continuous, $\mathscr{A}$-measurable f, and if X concentrates on a separable set of completely regular points, then $X_n \xrightarrow{d} X$.

Theorem 2.7: Continuous Mapping Theorem (CMT)

Let H be an $\mathscr{A}/\mathscr{A}'$-measurable mapping from $\mathscr{X}$ into $\mathscr{X}'$, and let C be a separable, $\mathscr{A}$-measurable set of completely regular points at which H is continuous. Then, if a sequence $\{X_n\}$ converges in distribution to a random vector X and $\mathbb{P}\{X \in C\} = 1$, then $HX_n \xrightarrow{d} HX$.

The reason why we have to introduce a new version of CLT is that it doesn't assumes that $\mathcal{A}'$ is Borel. One should note if we cannot work with the Borel sigma algebra $(D([0,1]), \mathcal{B}(D([0,1]))$, since our empirical process $[U_n(x)](\omega)$ is not $\mathcal{F}/\mathcal{B}(D([0,1])$ measurable. We can show that for all subsets of $A \in [0,1]$, not necessary Lebesgue measurable,

$$G_A = \{f \in D([0,1]) \,|\, x \text{ has a jump at some point of } A\}$$

is open. In fact, take $f \in G_A$ with a jump at $t^* \in A$ such that

$$|f(t^*) - f(t^*-)| = R > 0$$

Let $g \in B_{R/4}(f)$. Then we have $|f(t^*) - g(t^*)| < R/4$ and $|f(t^*-) - g(t^*-)| < R/4$. From this we have the estimate

$$|g(t^*) - g(t^*-)| \geq |f(t^*) - f(t^*-)| - |f(t^*) - g(t^*-)| - |g(t^*-) - f(t^*-)| > R/2 > 0$$

Therefore $g \in G_A$ and $B_{R/4}(f) \subseteq G_A$. It follows that if $[U_1(x)](\omega)$ (an empirical distribution) is Borel measurable, then the set $\{[U_1(x)](\omega) \in G_A\}$ is Lebesgue measurable! This is true even if A is not Lebesgue measurable (e.g. A is a Vitali set)

Instead, we work with the projective σ-algebra. Recall that we have defined the projection map π_S. The projective σ-algebra is the σ-algebra generated by all the projection maps. One can show that a random element X is $\mathcal{F}/\mathcal{A}$ measurable iff all projected maps $\pi_S X$ is Borel $\mathcal{B}(\mathbb{R})$ measurable.

3 Empirical CLT

3.1 Proof of Emperical CLT

We have established in section 1 that for a finite set of values $S = \{s_1, ..., s_n\}$, we have $\pi_S[U_n(x)](\omega) \to \pi_S[U(x)](\omega)$. One might think if we can establish convergence in distribution of process $U_n(x)$ to $U(x)$ by uniformly approximating $U_n(x)$ at a finite subset of values S. Unfortunately if $U_n(x)$ is too rough, we need to increase the size of S as n increases, and the argument breaks down.

The following theorem asserts that if we can choose a grid $0 = t_0 < t_1 < ... < t_m = 1$ such that $X(t)$ does not deviate too much from at least one of the grid points $X(t_i)$, then the approximation argument works. For the special case when $[U_n(x)](\omega)$ is the empirical process, such a grid *might* exist as suggested in the proof of Glivenko-Cantelli Theorem (Theorem 1.6).

Theorem 3.1: Necessary and Sufficient Condition for Convergence of Stochastic Processes in Distribution

Let $X(t), X_1(t), X_2(t), \dots$ be random elements of $D[0,1]$ from $(\Omega, \mathcal{F}, \mathbb{P})$ under its uniform metric and projection σ-field. Suppose $P\{X \in C\} = 1$ for some separable subset C of $D[0,1]$. The necessary and sufficient conditions for $\{X_n(t)\}$ to converge in distribution to X are:

(i) any finite-dimensional projections of X_n converge to the finite-dimensional projections of X

(ii) for each $\epsilon > 0$, $\delta > 0$, there corresponds a grid $0 = t_0 < t_1 < \cdots < t_m = 1$ such that

$$\limsup \mathbb{P}\left(\max_i \left(\sup_{t\in[t_i,t_{i+1})} |X_n(t) - X_n(t_i)| > \delta\right)\right) < \epsilon \tag{15}$$

Proof. We prove the sufficient direction only. For simplicity, we assume that $X(t)$ is in fact in $C^0([0,1])$ (the case when X is an increasing Catalag can be extended with a similar argument as stated in Theorem 1.6). Let $T = \{t_0, ..., t_m\}$. Define the approximation map sending $X \in D([0,1])$ as a *step* function, defined as

$$(A_T X)(t) = X(t_i), \quad t \in [t_i, t_{i+1}) \tag{16}$$

The condition (15) then asserts that

$$\limsup \mathbb{P}\left(\|X_n - A_T X_n\|_{D([0,1])} > \delta\right) < \epsilon \tag{17}$$

Continuity of X asserts that for all $\epsilon > 0$, there exists a set $S = \{s_0 = 1, s_1, ..., s_k = 1\}$ such that

$$\mathbb{P}(\|A_S X - X\| > \delta) \leq \epsilon \tag{18}$$

and without loss of generality we may assum $S = T$ because given S we can always refine T without violating (14) (at least for the case of empirical process when they are increasing). At worst the δ is replaced by 2δ.

We may then utilise a generalised version of Corollary 2.3: let f be a bounded, uniformly continuous function. We may write $f \circ A_T = g \circ \pi_T$, since $f \circ A_T(X(t))$ only depends on the grid points (t_i). For a fix ϵ choose δ such that $|f(x) - f(y)| < \epsilon$ whenever $|x - y| \leq \delta$, $x, y \in \mathbb{R}$,

$$\begin{aligned}
&\left|\int f(X_n)\, d\mathbb{P} - \int f(X_n)\, d\mathbb{P}\right| \\
&\leq \int |f(X_n) - f(A_T X_n)|\, d\mathbb{P} + \int |f(A_T X_n) - f(A_T X)|\, d\mathbb{P} + \int |f(A_T X) - f(X)|\, d\mathbb{P}
\end{aligned}$$

The first term can be controlled by

$$\begin{aligned}
&\int |f(X_n) - f(A_T X_n)|\, d\mathbb{P} \\
&= \int |f(X_n) - f(A_T X_n)|(\mathbb{I}_{|X_n - A_T X_n| \leq \delta} + \mathbb{I}_{|X_n - A_T X_n| \geq \delta})\, d\mathbb{P} \\
&\leq \int \Bigg(\underbrace{\epsilon \mathbb{I}_{|X_n - A_T X_n| \leq \delta}}_{\text{continuity of } f} + 2\,\|f\|\, \mathbb{I}_{|X_n - A_T X_n| \geq \delta}\Bigg)\, d\mathbb{P} \\
&< \epsilon(1 + 2\,\|f\|)
\end{aligned}$$

And similarly for the third term. (Here $\|.\|$ is again the uniform norm. The second term vanishes, utilising CMT with the assumption $\pi_T X_n \to \pi_T X$. Therefore the difference is bounded by $2\epsilon(1 + 2\,\|f\|)$, indicating convergence of distribution. □

It remains for us to construct a suitable grid to prove the empirical CLT. We need an intermediate lemma.

Lemma 3.2

Let $(([Z(t)](\omega))_{t\in[0,b]}$ be a process on $(\Omega, \mathcal{F}, \mathbb{P})$ with sample paths in $D([0, b])$ and satisfy $Z(0) = 0$, and let $(\mathscr{E}_t)_{t\in[0,b]}$ be a filtration so that $Z(t)$ is adapted in this filtration (e.g. natural filtration). If at each point of $\{|Z(t)| > \delta\}$

$$\mathbb{P}\left(|Z(b) - Z(t)| \leq \frac{1}{2}|Z(t)| \,\middle|\, \mathscr{E}_t\right) \geq \beta \tag{19}$$

where β is a positive number depending on δ, then

$$\mathbb{P}\left(\sup_{0\le t\le b}|Z(t)|>\delta\right)\le\beta^{-1}\mathbb{P}\left(|Z(b)|>\frac{1}{2}\delta\right) \tag{20}$$

Proof. There are neater proofs if we assume that $Z(t)$ is Markov or being a martingale, but we proof the general case here. Again we consider S a finite subset of $[0,b]$ containing b. By right continuity, for any sample path of $Z(t)$,

$$\max_S|Z(t)|\to\sup_{[0,b]}|Z(t)|$$

as S expands up to a countable, dense subset of $[0,b]$.

It is sufficient to bound the probability for the event $\{\max_S|Z(t)|\ge\delta\}$. Define the stopping time $\tau=\inf t\,|\,|Z(t)|>\delta$. By law of total probability,

$$\begin{aligned}
\mathbb{P}\left(\max_S|Z(t)|\ge\delta\right)&=\sum_S\mathbb{P}(\tau=t)\\
&\le\beta^{-1}\sum_S\mathbb{P}(\tau=t)\mathbb{P}\left(|Z(b)-Z(t)|\le\frac{1}{2}|Z(t)|\,\middle|\,\mathscr{E}_t\right)\\
&=\beta^{-1}\sum_S\mathbb{P}\left(\tau=t,\,|Z(b)-Z(t)|\le\frac{1}{2}|Z(t)|\right)\\
&=\beta^{-1}\sum_S\mathbb{P}\left(\tau=t,\,|Z(b)|>\delta/2\right)\\
&=\beta^{-1}\mathbb{P}(|Z(b)|>\delta/2)
\end{aligned}$$

□

We are therefore ready to prove the Empirical CLT for the uniform case.

Theorem 3.3: Empirical CLT, Uniform Case

As in equation (8), $[U_n(x)](\omega)$ converges in distribution to the Brownian Bridge $[U(t)](\omega)$.

Proof. The idea is to find a grid $0=t_0<t_1<\cdots<t_m=1$ that satisfies

$$\limsup\mathbb{P}\left(\max_i\left(\sup_{t\in[t_i,t_{i+1})}|U_n(t)-U_n(t_i)|>\delta\right)\right)<\epsilon$$

as from the theorem above, this is sufficient to show that the Empirical CLT is true. This can be done by finding a grid that makes

$$\sum_{i=0}^{m-1}\mathbb{P}\left(\sup_{J_i}|U_n(t)-U_n(t_i)|>\delta\right)$$

small, and to make this easier, we will take equally spaced t_i and find an m that satisfies the inequality. Taking evenly spaced t_i simplifies the sum above to be

$m\mathbb{P}(\sup_{[0,1/m]} |U_n(t)| > \delta)$ by symmetry.

Then, if take the filtration, $\mathscr{E}_t$, generated by $U_n(s)$ for $0 \le s \le t$, we would know how many observations lie in the interval $[0,t]$. Suppose there are k observations in the interval $[0,t]$. Then, on $\{U_n(t) = n^{-1/2}(k - nt)\}$, the conditional distribution of $U_n(1/m) - U_n(t)$ given $\mathscr{E}_t$ is

$$n^{-1/2}\left(\text{Binomial}\left(n-k, \frac{\frac{1}{m}-t}{1-t}\right) - n(b-t)\right)$$

which depends on t only (so the empirical process is Markov!) Applying Chebyshev's Inequality to the set where $|U_n(t)| > \delta$ gives

$$\mathbb{P}\left(|U_n(1/m) - U_n(t)| > \frac{1}{2}|U_n(t)| \,\middle|\, \mathcal{E}_t\right) \le \frac{1}{2}$$

for m large enough. We may therefore utilise Lemma 3.3 with $\beta = 1/2$. For sufficiently large m, we have

$$m\mathbb{P}\left(\sup_{0\le t\le b} |U_n(t)| \ge \delta\right) \le 2m\mathbb{P}\left(|U_n(b)| > \frac{1}{2}\delta\right) \tag{21}$$

Then fixing m and taking n to infinity and using finite dimensional convergence, the inequality converges to

$$m\mathbb{P}\left(\sup_{0\le t\le b} |U_n(t)| \ge \delta\right) \le 2m\mathbb{P}\left(|U_n(b)| > \frac{1}{2}\delta\right) \le 32m^{-1}\delta^{-4}K \tag{22}$$

where K represents the 4-th moment of a standard normal distribution, which is finite. The sum tends to 0 as $m \to \infty$. This completes the proof. □

The proof for the general case of the Empirical CLT builds on the uniform case of the Empirical CLT and requires the CMT proven in the last section.

Theorem 3.4: Empirical CLT

For the general case when $[F_n(x)](x)$ is obtained by samples from a distribution F_X, in distribution

$$\sqrt{n}[F_n(x)](\omega) - F_X(x) \to U(F_X(x)) \tag{23}$$

Proof. The theorem is proven by defining a map H from $D[0,1]$ to $D[-\infty,\infty]$ by sending a path X to $(HX)(r) = X(F_X(r))$. Then, as H is measurable and uniformly continuous, we can use the Uniform Empirical CLT as well as the CMT to find that $HU_n \xrightarrow{d} HU$ where U_n are empirical processes constructed by independent sampling from Uniform(0, 1). We conclude by noting

$$U_n(F_X(x)) = [F_n(x)](\omega) - F_X(x)$$

□

3.2 Towards Kolmogorov Theorem

By CMT we have

$$\sqrt{n}\,\|[F_n(x)](\omega)-F_X(x)\|_{D([0,1])}\to\|[U(x)](\omega)\|_{D([0,1])}$$

Therefore we digress and study the distribution of $\|[U(x)](\omega)\|_{D([0,1]}$, known as the Kolmogorov-Smirnov (KS) distribution. We will prove that

Theorem 3.5: CDF of KS distribution

$$K(\alpha)=\mathbb{P}(\|U(x)\|_{D([0,1])}\le\alpha)=1-2\sum_{n=1}^{\infty}(-1)^{n+1}e^{-2n^2\alpha^2} \tag{24}$$

Note that if U is a Brownian Bridge then $W(x)=(1+x)\,U\left(\frac{x}{1+x}\right)$ is a Brownian Motion starting at 0. Notice that $B(1)=0$ so $x=1$ can be omitted in our discussion. Also note that $t\mapsto t/(1+t)$ is a bijection from $[0,\infty)$ to $[0,1)$. Therefore for $\alpha\ge 0$,

$$\begin{aligned}\mathbb{P}(\|U(x)\|_{D([0,1])}\ge\alpha)&=\mathbb{P}\left(\sup_{t\in[0,\infty)}\left|U\left(\frac{t}{1+t}\right)\right|\ge\alpha\right)\\&=\mathbb{P}\left(|W(t)|\text{ hits }\alpha(1+t)\right)\\&=\mathbb{P}\left(W(t)\text{ hits }\alpha(1+t)\text{ or }W(t)\text{ hits }-\alpha(1+t)\right)\\&=\mathbb{P}\left(W(t)\text{ hits I or }W(t)\text{ hits II}\right)\end{aligned}$$

where I represents the barrier $\alpha(1+t)$ and II represents the barrier $-\alpha(1+t)$. We try to split the RHS. It is clear that

$$\mathbb{P}(\|U(x)\|_{D([0,1])}\ge\alpha)=\mathbb{P}\left(W(t)\text{ hits I}\right)+\mathbb{P}\left(W(t)\text{ hits II}\right)-\mathbb{P}\left(W(t)\text{ hits I and II}\right)$$

And

$$\begin{aligned}\mathbb{P}\left(W(t)\text{ hits I and II}\right)&=\mathbb{P}\left(W(t)\text{ hits I then II}\right)+\mathbb{P}\left(W(t)\text{ hits II then I}\right)\\&\quad-\mathbb{P}\left(W(t)\text{ (hits I then II) and (hits II then I)}\right)\end{aligned}$$

The last term can be further splitted in similar manner. At the very end one might conjecture the following

$$\begin{aligned}\mathbb{P}(\|U(x)\|_{D([0,1])}\ge\alpha)&=\mathbb{P}\left(W(t)\text{ hits I}\right)+\mathbb{P}\left(W(t)\text{ hits II}\right)\\&\quad-\mathbb{P}\left(W(t)\text{ hits I then II}\right)+\mathbb{P}\left(W(t)\text{ hits II then I}\right)\\&\quad+\mathbb{P}\left(W(t)\text{ hits I then II then I}\right)+\mathbb{P}\left(W(t)\text{ hits II then I then II}\right)-\dots\end{aligned} \tag{25}$$

At this stage we do not know whether the series at the RHS is summable. The best thing to do is to compute each term and to conclude that the RHS is summable. Notice the reflection principle implies:

$$\begin{aligned}\mathbb{P}\left(W(t)\text{ hits I}\right)&=\mathbb{P}\left(W(t)\text{ hits II}\right)\\\mathbb{P}\left(W(t)\text{ hits I then II}\right)&=\mathbb{P}\left(W(t)\text{ hits II then I}\right)\end{aligned}$$

so on and so forth. Therefore if RHS of equation (25) converges then we have

$$\mathbb{P}(\|U(x)\|_{D([0,1])} \geq \alpha) = 2\left(\mathbb{P}\left(W(t) \text{ hits I}\right) - \mathbb{P}\left(W(t) \text{ hits I then II}\right) + ...\right) \tag{26}$$

Let us be reminded the strong Markov property of the Brownian motion $W(t)$. For a stopping time τ (w.r.t. natural filtration of W), define the time-shifted process

$$W^{(\tau)}(t) = W(t+\tau) - W(\tau)$$

The strong Markov property asserts that $W^{(\tau)}(t)$ is independent to $\mathcal{F}_\tau$ when conditioned on $\{\tau < \infty\}$. Here the σ-algebra$\mathcal{F}_\tau$ contains information of $W(t)$ when '$t \leq \tau$'. Moreover, $W^{(\tau)}(t)$ has same distribution as the original Brownian Motion $W(t)$.

Computing the first term: We attempt to compute the probability of hitting one barrier $\alpha + \beta t$. We define

$$\phi(\alpha, \beta) = \mathbb{P}(W(t) \text{ hits } \alpha + \beta t), \quad \alpha, \beta \geq 0$$

Then $\mathbb{P}(W(t) \text{ hits I})$ is $\phi(\alpha, \alpha)$. We first establish that fixing β, we have

$$\phi(\alpha_1 + \alpha_2, \beta) = \phi(\alpha_1, \beta)\phi(\alpha_2, \beta) \tag{27}$$

This can be seen by the fact that if $W(t)$ hits $(\alpha_1 + \alpha_2) + \beta t$ then it must hit $\alpha_1 + \beta t$. Define the stopping time $T_{\alpha,\beta} = \inf\{t \,|\, W(t) = \alpha + \beta t\}$. Then we have

$$\begin{aligned}
\phi(\alpha_1 + \alpha_2, \beta) &= \mathbb{P}(W(t) \text{ hits } \alpha_1 + \alpha_2 + \beta t) \\
&= \mathbb{P}(\{T_{\alpha_1,\beta} < \infty\} \cap W(t) \text{ hits } \alpha_1 + \alpha_2 + \beta t) \\
&= \mathbb{P}(\{T_{\alpha_1,\beta} < \infty\} \cap W^{(T_{\alpha,\beta})}(t) \text{ hits } \alpha_2 + \beta t) \\
&= \mathbb{P}(\{T_{\alpha_1,\beta} < \infty\})\mathbb{P}(W(t) \text{ hits } \alpha_2 + \beta t) \\
&= \phi(\alpha_1, \beta)\phi(\alpha_2, \beta)
\end{aligned}$$

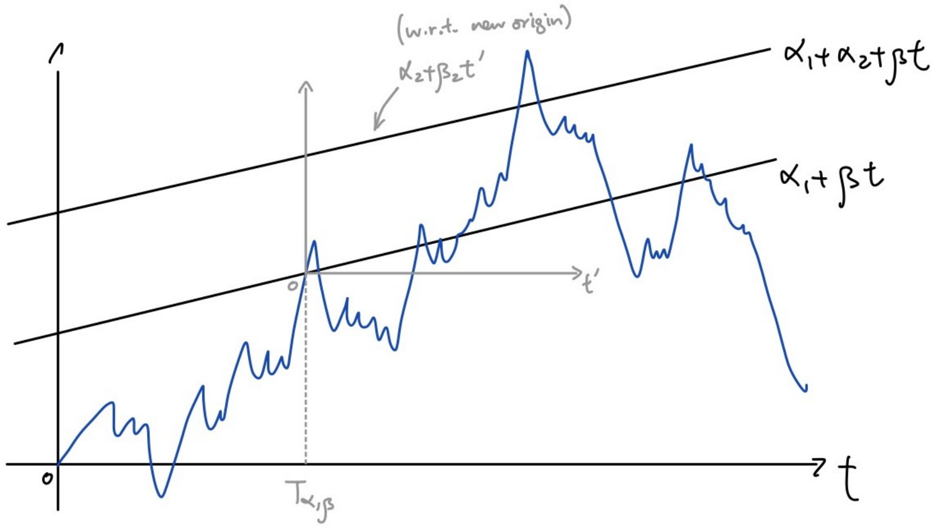

Figure 3.1

Equation (27) informs us that, fixing β, $\phi(\,.\,,\beta)$ must be an exponential function. Notice that $\phi(0,\beta) = 1$ (since $W(0) = 0$ almost surely and therefore $W(t)$ hits βt at $t = 0$) and $\phi(\alpha,\beta)$ decreases as α increases, one must have $\phi(\alpha) = \exp(-C(\beta)\alpha)$, where $C(\beta)$ is always positive and is independent of α.

To obtain $C(\beta)$ we turn our focus to the stopping time $\tau_\alpha = \inf\{t \mid W(t) = \alpha\}$. We utilise the Strong Markov Property again:

$$\begin{aligned}
\mathbb{P}(W(t) \text{ hits } \alpha+\beta t) &= \mathbb{P}(\{\tau_\alpha < \infty\} \cap W(t) \text{ hits } \alpha+\beta t) \\
&= \mathbb{P}(\{\tau_\alpha < \infty\} \cap W^{(\tau_\alpha)}(t) \text{ hits } \beta\tau+\beta t) \\
&= \mathbb{E}\left(\mathbb{E}\left(\mathbb{I}_{\{\tau_\alpha<\infty\}}\mathbb{I}_{\left\{W^{(\tau_\alpha)}(t) \text{ hits } \beta\tau+\beta t\right\}}\middle|\mathcal{F}_{\tau_\alpha}\right)\right) \\
&= \mathbb{E}\left(\mathbb{I}_{\{\tau_\alpha<\infty\}}\mathbb{E}\left(\mathbb{I}_{\left\{W^{(\tau_\alpha)}(t) \text{ hits } \beta\tau+\beta t\right\}}\middle|\mathcal{F}_{\tau_\alpha}\right)\right) \\
&= \mathbb{E}\left(\mathbb{I}_{\{\tau_\alpha<\infty\}}\phi(\beta\tau_\alpha)\right) \\
&= \mathbb{E}(\exp(-(\beta C(\beta))\tau_\alpha))
\end{aligned}$$

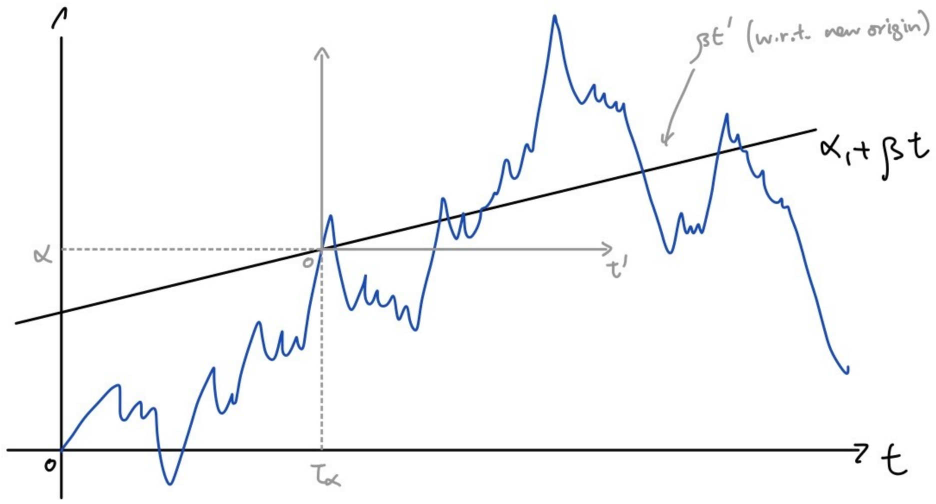

Figure 3.2

It is important to note the following result

$$\mathbb{E}(\exp(-\lambda\tau_\alpha)) = \exp(-\alpha\sqrt{2\lambda}) \tag{28}$$

Therefore we have the following equality

$$\exp(-C(\beta)\alpha) = \exp(-(2\beta C(\beta)^{1/2})\alpha)$$

which yields $C(\beta) = 2\beta$. Therefore

$$\phi(\alpha,\beta) = \mathbb{P}(W(t) \text{ hits } \alpha+\beta t) = \exp(-2\alpha\beta) \tag{29}$$

Computing other term: Let's say we want to compute the probability

$$\mathbb{P}(W(t) \text{ hits I then II then ... then I then II})$$

We may consider a generalised problem of computing

$$\mathbb{P}(W(t)\ \underbrace{\text{hits } \alpha_1+\beta_1 t \text{ then } -\alpha_2-\beta_2 t \text{ then ...}}_{n-1 \text{ hits}} \text{ then } -\alpha_n-\beta_n)$$

where n is even and $\alpha_i, \beta_i > 0$ for all i. We denote this probability

$$\pi_n(\alpha_1,\ \beta_1;\ \alpha_2,\ \beta_2;\ ...;\ \alpha_n,\ \beta_n)$$

We iteratively define the following stopping times:

$$\tau_1 = \inf\{t \,|\, W(t) = \alpha_1 + \beta_1 t\}$$
$$\tau_i = \inf\left\{t \geq \tau_{i-1} \,|\, W(t) = (-1)^{i-1}(\alpha_i + \beta_i t)\right\}, \quad i \geq 2$$

Then

$$\begin{aligned}
&\pi_n(\alpha_1, \beta_1;\ ..., \alpha_{n-1},\ \beta_{n-1};\ \alpha_n,\ \beta_n)\\
&= \mathbb{E}(\mathbb{E}(\mathbb{I}_{\{\tau_n<\infty\}} \,|\, \{\tau_{n-1} < \infty\}))\\
&= \mathbb{E}(\mathbb{I}_{\{\tau_{n-1}<\infty\}}\mathbb{P}(W^{(\tau_{n-1})}(t) + (\alpha_{n-1} + \beta_{n-1}\tau_{n-1}) \text{ hits } -(\alpha_n + \beta_n t) \mid \{\tau_{n-1} < \infty\}))
\end{aligned}$$

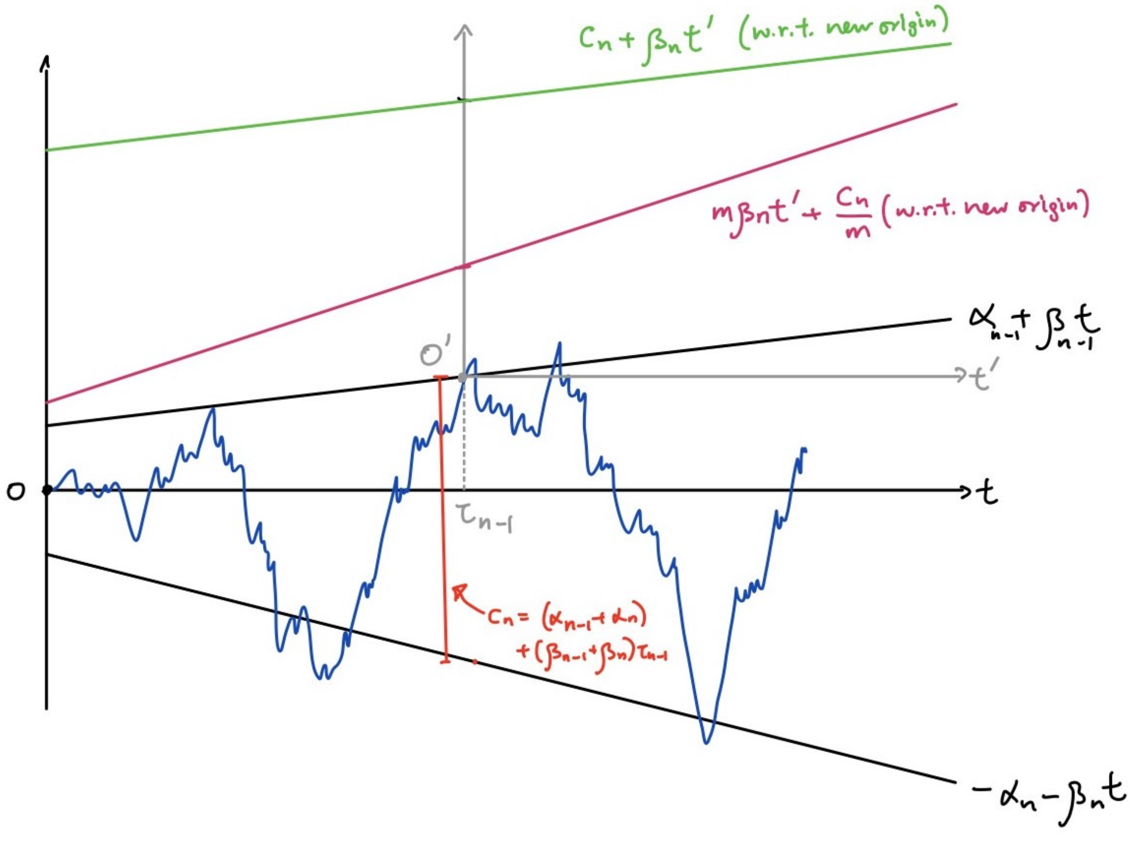

Figure 3.3: Notice the equations for the barriers.

The strong Markov property and tower law asserts that

$$
\begin{aligned}
&\mathbb{P}(W^{(\tau_{n-1})}(t) + (\alpha_{n-1} + \beta_{n-1}\tau_{n-1}) \text{ hits } -(\alpha_n + \beta_n t) \mid \{\tau_{n-1} < \infty\}) \\
&= \mathbb{P}(W(t) + (\alpha_{n-1} + \beta_{n-1}\tau_{n-1}) \text{ hits } -(\alpha_n + \beta_n t) \mid \{\tau_{n-1} < \infty\}) \\
&= \mathbb{P}(W(t) \text{ hits } -(\underbrace{(\alpha_{n-1} + \alpha_n) + (\beta_{n-1} + \beta_n)\tau_{n-1})}_{c_n(\tau_{n-1})} - \beta_n t \mid \{\tau_{n-1} < \infty\}) \quad \text{(shifting origin)} \\
&= \mathbb{P}(W(t) \text{ hits } c_n(\tau_{n-1}) + \beta_n t \mid \{\tau_{n-1} < \infty\}) \quad \text{(reflection)} \\
&= \exp(-2\, c_n\beta_n)
\end{aligned}
$$

It is important to note that the probability of $W(t)$ hitting the barrier $c_n + \beta_n t$ is equal to that of hitting $c_n/m + m\beta_n t$ for $m > 0$ (all expressed w.r.t. shifted coordinates). The equation of this new barrier in the original coordinates is

$$
m\beta_n(t - \tau_{n-1}) + c_n/m + (\alpha_{n-1} + \beta_{n-1}\tau_{n-1})
$$

which has intercept

$$
\begin{aligned}
&- m\beta_n\tau_{n-1} + \frac{(\alpha_{n-1} + \alpha_n) + (\beta_{n-1} + \beta_n)\tau_{n-1}}{m} + (\alpha_{n-1} + \beta_{n-1}\tau_{n-1}) \\
&= \frac{\alpha_{n-1} + \alpha_n}{m} + \alpha_{n-1} + \left(\left(\frac{1}{m} - 1\right)\beta_{n-1} + \left(\frac{1}{m} - m\right)\beta_n\right)\tau_{n-1}
\end{aligned}
$$

We choose $m > 0$ so that the intercept to be independent from τ_{n-1}. In such case m must satisfies

$$
\begin{aligned}
&\left(\frac{1}{m} + 1\right)\beta_{n-1} + \left(\frac{1}{m} - m\right)\beta_n = 0 \\
\implies &\beta_n m^2 - \beta_{n-1}m - (\beta_{n-1} + \beta_n) = 0 \\
\implies &(m+1)(\beta_n m - (\beta_n + \beta_{n-1})) = 0 \\
&\implies m = \frac{\beta_n + \beta_{n-1}}{\beta_n}
\end{aligned}
$$

The new barrier therefore has equation $\alpha_n' + \beta_n'$ in original coordinates:

$$
\alpha_n' + \beta_n' t := \frac{2\alpha_{n-1}\beta_n + \alpha_n\beta_n + \alpha_{n-1}\beta_{n-1}}{\beta_n + \beta_{n-1}} + (\beta_{n-1} + \beta_n)t
$$

which means we have

$$
\pi_n(\alpha_1, \beta_1;\, ..., \alpha_{n-1},\, \beta_{n-1};\, \alpha_n,\, \beta_n) = \pi_n(\alpha_1, \beta_1;\, ..., \alpha_{n-1},\, \beta_{n-1};\, \alpha_n',\, \beta_n')
$$

But note that this new barrier has a larger intercept and slope than the barrier $\alpha_{n-1} + \beta_{n-1}t$ (and also $\alpha_n + \beta_n t$). Therefore we have establish the following recurrence relation.

$$
\pi_n(\alpha_1, \beta_1;\, ..., \alpha_{n-1},\, \beta_{n-1};\, \alpha_n,\, \beta_n) = \pi_{n-2}(\alpha_1, \beta_1;\, ..., \alpha_{n-2},\, \beta_{n-2};\, \alpha_n',\, \beta_n') \tag{30}
$$

In fact this recurrence relation is also true for odd n. In the special case when $\alpha_i \equiv \beta_i \equiv \alpha$ the recurrence relation simplifies. Let $\alpha^{(i)}, \beta^{(i)}$ be the intercept and slope of the new Barrier after performing $i-1$ steps of iteration from equation (19), where $\alpha^{(1)} = \beta^{(1)} = \alpha$. Then we have the relation

$$\beta^{(i+1)} = \beta^{(i)} + \alpha \implies \beta^{(i)} = i\alpha$$
$$\alpha^{(i+1)} = \frac{2\alpha\beta^{(i)} + \alpha^{(i)}\beta^{(i)} + \alpha^2}{(i+1)\alpha} = \frac{2i\alpha^2 + \alpha^{(i)}(i\alpha) + \alpha^2}{(i+1)\alpha} = \frac{\alpha^2(2i+1) + i\alpha\alpha^{(i)}}{(i+1)\alpha}$$

By induction we will see that $\alpha^{(i)} = i\alpha$. Therefore $\pi_n(...) = \pi_1(n\alpha, n\alpha) = e^{-2n^2\alpha^2}$. We note that in such case the RHS is summable and therefore

$$K(\alpha) = \mathbb{P}(\|U(x)\|_{D([0,1])} \leq \alpha) = 1 - \mathbb{P}(\|U(x)\|_{D([0,1])} \geq \alpha) = 1 - 2\sum_{n=1}^{\infty}(-1)^{n+1}e^{-2n^2\alpha^2}$$

which concludes our proof.

4 Application - Goodness of Fit Test

One application of the theorem is to determine how likely a sequence of numbers $\vec{s} = (x_1, x_2, ..., x_n) \in \mathbb{R}^n$ (with n fixed) is actually a sample from a random variable X with distribution $F_X(x)$ with $F_X(x)$ continuous. Formally, we want to devise a *hypothesis test* to test the hypothesis

$$H_0 : \vec{S} \text{ is a sample from a random variable } X.$$

against

$$H_1 : \vec{S} \text{ is not a sample from a random variable } X.$$

A hypothesis test is an indicator function $\mathbb{I}_R(x)$ from $\mathbb{R}^n$ to $\{0, 1\}$ with $R \in \mathbb{R}^n$. We say that R is the *rejection region*. Intuitively speaking, if you have performed an experiment (e.g. simulate from certain fancy algorithms) and obtain a *deterministic* vector $\vec{s}$, then

- If $\vec{s} \in R$ the experiment suggests that H_0 *might be* false, or *there is sufficient evidence to reject* H_0.
- If $\vec{s} \in R^C$ the experiment suggests that H_0 *might be* true, or *there is insufficient evidence to reject* H_0.

It is important to note that 'there is insufficient evidence to reject H_0' doesn't really mean H_0 is itself true, and vice versa. In fact, one would like to look at the probability of 'false-negative' (known as *size* or *Type I error*), or in our context, the probability

$$\alpha = \mathbb{P}((X_1(\omega), .., X_n(\omega)) \in R) \tag{31}$$

where $X_1, ..., X_n$ iid with X.

> *Remark.* One might wish to talk about 'false-positive', or the probability
>
> $$\beta = \mathbb{P}((X_1(\omega), .., X_n(\omega)) \notin R)$$
>
> if X follows 'another distribution'. But this depends on the actual distribution X is falling in, so is meaningless to be discussed in our context.

The task is to construct a R which satisfies (31). Here we outline how Kolmogorov constructed R from the theorems we have proven. First note that the map

$$T : (x_1, ..., x_n) \mapsto \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{I}_{(-\infty, F_X(x_i)]}(x) - x \right\|_{D([0,1])} \tag{32}$$

is a well-defined map from $\mathbb{R}^n$ to $D([0, 1])$. Moreover we have established for n large, the random variable $T(X_1(\omega), ..., X_n(\omega))$ (with $X_1, ..., X_n$ iid with X) can be approximated by the Kolmogorov distribution as derived in previous section. One can define $R = T^{-1}(K^{-1}((\alpha, 1]))$, then (31) is immediately satisfied. This is known as an one-sided Kolmogorov test.

The test is best for determining whether a psuedo-random number generator can successfully generate 'samples' from a distribution. However, there are a number of limitations for the test. Firstly, the function $K(\alpha)$ is really hard to be computed and/or inverted. Several numerical methods have been devised by considering a relevant matrix problem. A more important problem would be the fact that it is not applicable when $F_X(x)$ is not known. Consider the following example

Example 4.1: Linear Regression

Consider you have collected a set of data $\{(\vec{x}_1, y_1), ..., (\vec{x}_n, y_n)\}$ with $(\vec{x}_i, y_i) \in \mathbb{R}^{p+1}$, and you have hypothesised that they are a sample of the distribution $(\vec{X}, Y)$, where

$$Y|\vec{X} \sim \mathrm{N}(\vec{X}^T\vec{\beta}, \sigma^2)$$

and $\vec{\beta} \in \mathbb{R}^p$ and $\sigma^2 \in \mathbb{R}_{>0}$ are unknown. The one sided KS-test cannot be used.

The method of estimating $\vec{\beta}$ from the data is well known. Let's say you have obtained an estimated $\vec{\beta}$ from the data, say $\hat{\beta}$. One might transform the data to the new data set $\{E_i\}_{i=1}^n$

$$\forall i, (\vec{x}_i, y_i) \mapsto E_i := y_i - \hat{\beta}^T\vec{x}_i$$

then normalise to $\{e_i\}_{i=1}^n$ so that it has sample variance 1. We consider this process as a map M from $\mathbb{R}^n \times \mathbb{R}^{p+1}$ to $\mathbb{R}^n$. We finally obtain map this dataset in $\mathbb{R}^n$ by T and hope we can peform an one-sided KS test. Unfortunately it has been shown that if $(\vec{X}_i, Y_i)_{i=1}^n$ is iid with $(\vec{X}, Y)$, the random variable

$$T(M((\vec{X}_i, Y_i)_{i=1}^n))$$

in no way converges to the Kolmogorov distribution.

Over the years there are new methods to tackle these problems, for instance the Kolomogorov-Lillefors test. However, the analysis of such methods often require further probabilistic techniques, including the introduction of new metric, therefore is beyond our scope of studies.

Declaration and Acknowledgements: We pledge that this essay and the work reported herein was our original work. Wherever works of others are involved, they have been cited clearly. We are grateful to Yilin Wang for her guidance.