Limiting Theorem of Empirical Process: An Example in Kolmogorov-Smirov Test and Brownian Bridge

Stochostic Process: Final Project

Wei-Chang Lee, Chi-Ning Chou
PROFESSOR RAOUL NORMAND
July 24, 2015

Abstract

In traditional statistical setting, we have the nice LLN and CLT, which give us an asymptotic sense in Gaussian distribution about the convergence behaviour of the statistics. When it comes to empirical process, however, the quadratic variation term breaks the beautiful structure and thus we have to come up with stronger sense about the asymptotic behaviour of the empirical process.

Kolmogorov statistics, which is the infinity norm of the empirical process, can tell us something about the asymptotic behaviour of the empirical process while it is the uniform upper bound of the whole process. With it's nice *distribution free* property and two strong theorem: Glivenko-Cantelli Theorem and Donsker Theorem, the uniform LLN and uniform CLT is guaranteed. Furthermore, the asymptotic behaviour of the Kolmogorov statistics will converge to the supremum norm of a special Gaussian process: Brownian Bridge. This is the beautiful analogy of asymptotic convergence to Gaussian random variable of traditional setting.

In the first two section, we will introduce the two limiting theorem of empirical process and the Gaussian process respectively. Then, the last section will prove the uniform CLT of empirical process and give a short tutorial on Kolmogorov test.

Keywords: Empirical Process, Glivenko-Cantelli Theorem, Donsker Theorem, Gaussian Process, Brownian Bridge, Kolmogorov-Smirov test

Contents

1 Empirical Process Theory **2**
- 1.1 Empirical Distribution 2
  - 1.1.1 Definition and Properties 2
  - 1.1.2 Kolmogorov Statistics 5
- 1.2 Asymptotic Convergence 6
  - 1.2.1 Glivenko-Cantelli Theorem: ULLN 6
  - 1.2.2 Donsker Theorem: UCLT 7

2 Gaussian Process **8**
- 2.1 Properties 8
- 2.2 Brownian Bridge 8

3 Donsker Theorem and Kolmogorov-Smirov Test **9**
- 3.1 Framework 9
- 3.2 Convergence of Kolmogorov Statistics 10

4 Appendix **12**

1 Empirical Process Theory

Empirical process theory focus on the distribution of the realization of unknown underlying distribution. With this approach, we can throw away the structural assumption on the distribution, making the theory highly suitable for non-parametric statistical inference.

1.1 Empirical Distribution

As observers, all we can see from a random experiment is the sampling results from an underlying distribution (if there exists one). In almost every case, we don't know the true probability distribution behind it. What we want to do is to make inferences about the underlying distribution.

1.1.1 Definition and Properties

As long as we only have the samples, it's intuitively to make a histogram and observe the structure. Furthermore, we can consider the *empirical distribution function*

$$\hat{F}_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}_{\{X_i \leq x\}}$$

where $X_1, X_2, ..., X_n$ are i.i.d. sample from a cumulative distribution function F. And $\mathbf{I}$ is the indicator function.

The intuition is that we record the number of occurrences from small sample value to large sample value and draw a cumulative function.

Also, we define *empirical process* according to the empirical distribution,

$$E_n(x) := \sqrt{n}[\hat{F}_n(x) - F(x)], \quad 0 \leq x \leq 1$$

We will discuss more details about empirical process later. For now, let's observe some properties about empirical distribution.

The most important issue after defining the empirical distribution is to find out whether it will converge to the real underlying distribution. And first, we need to know what kind of convergence we are looking for. Point-wise convergence is the most basic convergence, and in the following part of this section will show you the result. To go further, we need some stronger results in the convergence of empirical distribution, so that we can construct something like confidence interval, which can be utilized in many applications.

As a warm-up, let's consider the point-wise convergence of empirical distribution to the underlying distribution:

$$\hat{F}_n(z) \xrightarrow{P} F(z)$$

where $z \in [0, 1]$

The point-wise convergence is an immediate result of the following observation.

Observation 1. *The distribution of $n\hat{F}_n(z)$ for some $z \in \mathcal{R}$ is the same as binomial$(n, F(z))$.*

You can look deep into figure 1 to find more intuition. And the point-wise convergence can be easily deduced.

Theorem 2 (Point-wise Convergence of Empirical Distribution)**.** *Let $\hat{F}_n$ as the empirical distribution defined above from an underlying distribution F. Then $\forall z \in [0, 1]$,*

$$\hat{F}_n(z) \xrightarrow{P} F(z)$$

.

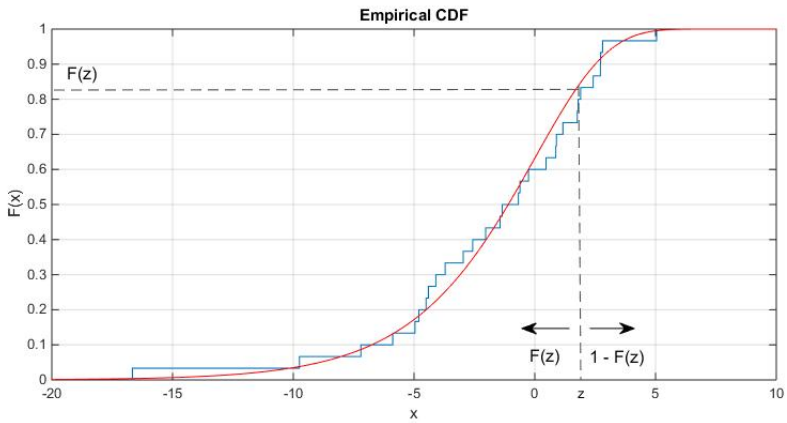

Figure 1: Empirical distribution and its point-wise convergence property.

Proof. First, because $n\hat{F}_n(z) \sim binomial(n, F(z))$

$$\begin{aligned} E[\hat{F}_n(z) - F(z)] &= E[\hat{F}_n(z)] = F(z) \\ &= \frac{1}{n}E[n\hat{F}_n(z)] - F(z) \\ &= \frac{nF(z)}{n} - F(z) = 0 \end{aligned}$$

Thus $\hat{F}_n$ is an unbiased estimator of F. Also, consider the variance

$$\begin{aligned} Var[\hat{F}_n(z)] &= \frac{1}{n^2}Var[n\hat{F}_n] \\ &= \frac{nF(z)[1-F(z)]}{n^2} = \frac{F(z)[1-F(z)]}{n} \end{aligned}$$

Applying Chebyshev inequality will lead to the result: $\hat{F}_n(z) \xrightarrow{P} F(z)$. $\square$

Now we have the point-wise convergence of empirical distribution and the corresponding asymptotic rate. Based on this, we can construct confidence interval for point-wise estimation. For example, the $(1-\alpha)$-level confidence interval of $F(z)$ is

$$\hat{F}_n(z) \pm z_{\alpha/2}\sqrt{\frac{\hat{F}_n(z)[1-\hat{F}_n(z)]}{n}}$$

where $z_{\alpha/2}$ is half the size of the $(1-\alpha)$-level confidence interval of standard Gaussian random variable.

However, what if we want to estimate the behaviour of two points or an interval? We need a stronger results about the asymptotic behaviour of empirical distribution so that we can make effective inferences.

As a result, our goal is to understand the global asymptotic behaviour of empirical distribution. And we can see that, the global convergence behaviour can be controlled by the supremum norm. In other words, to prove the asymptotic behaviour of the empirical process, it's sufficient to show that of the supremum norm of it. Namely, the Kolmogorov Statistics of the empirical process. In the following section, we will introduce you the Uniform Law of Large Number(ULLN) and the Uniform Central Limit Theorem(UCLT) of empirical process. Then, we can yield the similar asymptotic convergence behaviour as in the traditional setting.

Goal (Uniform Convergence of Empirical Distribution)**.** *The empirical distribution E_n will converge in distribution to the underlying distribution.*

1.1.2 Kolmogorov Statistics

The Kolmogorov statistics is defined on an empirical distribution function $\hat{F}_n$ and a cumulative objective function F as follow:

$$D_n := \sup_{x \in \mathcal{R}} |\hat{F}_n(x) - F(x)|$$

where n is the number of samples.

We can see that the Kolmogorov statistics D_n is the supremum over empirical process E_n defined in the previous subsection. The smaller the D_n is we can some how think of that the closer the two distribution are.

As long as we consider the Kolmogorov statistics between the empirical distribution and its underlying distribution, there are some nice convergence behaviours.

The first one is *distribution-free property.* It means that no matter what underlying property is, the behaviour of the Kolmogorov statistics will be the same! Concretely, the distribution will only in some sense related to the uniform distribution.

Theorem 3 (Distribution-Free Property)**.** *The distribution of the Kolmogorov statistics D_n is the same for all continuous underlying cumulative distribution.*

Proof. For the simplicity, let's consider the case where F is strictly increasing. Namely, F^{-1} exists. Thus, $\forall x \in \mathcal{R}, \exists y \in [0,1]$ *s.t.* $x = F(y)$. Consider the Kolmogorov statistics:

$$\begin{aligned} D_n &= \sup_{x \in \mathcal{R}} |\hat{F}_n(x) - F(x)| = \sup_{y \in [0,1]} |\hat{F}_n(F^{-1}(y)) - F(F^{-1}(y))| \\ &= \sup_{y \in [0,1]} |\hat{F}_n(F^{-1}(y)) - y| \end{aligned}$$

Observe the term $\hat{F}_n(F^{-1}(y))$

$$\hat{F}_n(F^{-1}(y)) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{I}_{\{X_i \le F^{-1}(y)\}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{I}_{\{F(X_i)\le y\}}$$

From Statistics101, we know that $F(X_i)$ has the same distribution as $Uni[0,1]$. As a result, the supremum will not differ from distribution to distribution. Actually, the distribution of D_n will related to that of the ordered statistics of uniform distribution. □

Apart from the amazing fact that the distribution of Kolmogorov statistics is distribution-free, the convergence is also guaranteed by the following Glivenko-Cantelli theorem. Also, the asymptotic behaviour of Kolmogorov statistics is proved to be the same as the distribution of Brownian Bridge in another important theorem: Donsker Theorem. Both theorems will be discussed in details in the following subsection, and the definition and properties of Brownian Bridge will also be introduced in the Gaussian Process section.

1.2 Asymptotic Convergence

Our goal is to use empirical distribution to draw inference on the unknown. In the first part of the section we proved the point-wise convergence. In this part, we are going to explore two stronger results: Uniform Law of Large Number(ULLN) and Uniform Central Limit Theorem(UCLT).

1.2.1 Glivenko-Cantelli Theorem: ULLN

ULLN consider the universal convergence of the empirical distribution. And we can see that the convergence of Kolmogorov statistics, the supremum difference, is sufficient for the result. And it's guaranteed by the following Glivenko-Cantelli Theorem.

Theorem 4 (Glivenko-Cantelli)**.** *The Kolmogorov statistics will converge to zero almost surely as the number of samples grows to infinity. That is,*

$$D_n \xrightarrow{a.s.} 0$$

, as $n \to \infty$

Proof. First, we consider the *ordered statistics* of the samples: $X_{1:n}, X_{2:n}, ..., X_{n:n}$ instead of the sample itself: $X_1, X_2, ..., X_n$. And it immediately follows that, $\hat{F}_n(X_{i:n}) = \frac{i}{n}$. Thus,

$$D_n = max_{1\le i\le n}|\hat{F}_n(X_{i:n}) - F(X_{i:n})| = max_{1\le i\le n}|\frac{i}{n} - X_{i:n}|$$

Next, we use the two properties of the ordered statistics of uniform distribution:

$$\text{(i)}\ \ max_{1\leq i\leq n}|X_{i:n} - E[X_{i:n}]| \to 0$$
$$\text{(ii)}\ \ max_{1\leq i\leq n}|\tfrac{i}{n} - E[X_{i:n}]| \to 0$$

Inequality (i) is the result after applying the extension of Chebyshev inequality: $P[|X - E(X)| > \epsilon] \leq \frac{E[|X-E(X)|^k]}{\epsilon^k}$ up to fourth moment. And inequality (ii) is just the result of LLN.
With triangle inequality and the above two results from ordered statistics, we can conclude that

$$max_{1\leq i\leq n}|X_{i:n} - \frac{i}{n}]| \to 0$$

Thus, D_n will converge to 0 as n grows to infinity. □

With this theorem, we have the uniform convergence of empirical distribution. Namely, for any $\epsilon > 0$ there exists a N such that for all $n > N$, the underlying distribution will lies in the ϵ-neighborhood of the empirical distribution.

1.2.2 Donsker Theorem: UCLT

Finally, we come to the most important theorem in empirical process theory: the *Donsker Theorem.*

Theorem 5 (Donsker)**.** *Let $E_n(x) = \hat{F}_n(x) - F(x)$ be the empirical process of F, which is a cumulative distribution function. Then, $En(x)$ will converge in distribution to a Gaussian process: Brownian Bridge. Thus, the limit of the empirical process $G(x)$ can be written as $B(F(x))$, where B is the standard Brownian Bridge.*

With Donsker theorem, we can easily construct the confidence interval of the Kolmogorov statistics D_n. And actually this is what Kolmogorov-Smirov test is doing.

To understand why Donsker theorem works, we need to know more about the distribution of the order statistics of uniform distribution. Also, we have to learn some basic concepts about Gaussian process, which will be introduced in the next section. As a result, the proof of Donsker theorem will be left until the last section.

Remark. *Formally, the function that can be applied to Donsker theorem is in the Skorokhod space.*

2 Gaussian Process

2.1 Properties

Definition (Gaussian Process)**.** *A Gaussuan process $\{X_t, t \in T\}$ is a stochastic process that any finite linear combination of samples has a joint Gaussian distribution.*

The most common example is the Brownian motion $\{B_t, t \in T\}$. It starts from 0 with probability 1. The variance is t and the covariance between B_t and B_s is $t \wedge s$.

One of the most important property of Gaussian process is that a the behaviour of a Gaussian process is fully determined by its mean and covariance function.

Property (Uniqueness Property)**.** *The covariance function of a Gaussian process completely determine its behaviour.*

In the proof of Donsker theorem, we will apply this result to identify the behaviour of the statistics that we are desired. As a little remark, to prove this property, you can consider the characteristic function of the Gaussian process then you will find something interesting.

2.2 Brownian Bridge

Brownian bridge is one the the common Gaussian process in real life. It is obtained by taking a standard Brownian motion restricted on [0, 1] and conditioning on the event that $X_0 = X_1 = 0$. This definition gives us an intuition about the distribution of a Brownian bridge. First, we can see that the future of the Brownian bridge is not independent to the past. Moreover, with this operational definition, we can have a picture about why this process is called *bridge*.

Before introducing you more properties about Brownian bridge, let's take a look at its formal definition in advance.

Definition. *A Brownian Bridge is a Gaussian process with covariance function $Cov(X_s, X_t) = s \wedge t - st$.*

As you can see, Brownian bridge is strongly related to Brownian motion. And actually, they can generate each other!That is, given a Brownian motion, then you can construct a Brownian bridge and vice versa.

In the following we will list some constructions of Brownian bridge via Brownian motion without proving it since they are not very important for our goal: Donsker theorem.

Property (Construction of Brownian Bridge)**.** *Suppose* $\mathbf{Z} = \{Z_t : t \in [0,1]\}$ *is a standard Brownian motion on [0,1]. Then* $\mathbf{X} = \{X_t : t \in [0,1]\}$ *is a Brownian bridge if one of the following is true*

1. $X_t = Z_t - tZ_1$
2. $X_t = (1-t)Z(\frac{t}{1-t})$
3. $X_t = (1-t)\int_0^t \frac{1}{1-s} dZ_s$
4. $dX_t = \frac{X_t}{1-t}dt + dZ_t$

To verify the above constructions, you can simply check that the covariance function of X_t and X_s is $t \wedge s - ts$.

Next, let's consider the distribution of the supremum of Brownian bridge.

Property (Supremum of Brownian bridge)**.** *Let* $\mathbf{X} = \{X_t : t \in [0,1]\}$ *be a Brownian bridge and* $S^+ = \sup_{0 \le t \le 1} X(t)$ *and* $S = \sup_{0 \le t \le 1} |X(t)|$*. Then,*

1. $F(S^+ \le x) = 1 - \exp(-2x^2)$
2. $F(S \le x) = 1 + 2\sum_{k=1}^{\infty}(-1)^k \exp(-2k^2x^2)$

3 Donsker Theorem and Kolmogorov-Smirov Test

With Glivenko-Cantelli theorem, the uniform LLN is guaranteed through the convergence of Kolmogorov statistics. With Donsker theorem, the uniform CLT is guaranteed and thus we know the asymptotic behaviour of Kolmogorov statistics. In the previous discussion, we consider the setting that the underlying distribution is known. However, in real life, this is not the case! What we have is the empirical distribution and some assumption about the random source. And our goal is to find out the underlying distribution. With the help of Donsker theorem, we can construct a confidence interval of Kolmogorov statistics and draw inference about the empirical process. And finally test whether our assumption on the true distribution is correct with high probability.

3.1 Framework

Kolmogorov-Smirov test is a famous non-parametric goodness of fitting test. The test consider the Kolmogorov statistics: $D_n = \sup_{x \in \mathcal{R}} |\hat{F}_n(x) - F(x)|$, which is a *distribution-free* statistics. The convergence of D_n provides

us a way to see that whether a source is sampled from the guessing distribution or not. Moreover, since the probability distribution of D_n will converge to that of a Brownian Bridge, the confidence interval can be calculated.

3.2 Convergence of Kolmogorov Statistics

Donsker theorem says that the Kolmogorov statistics will converge in distribution to the spremum of a Brownian bridge. To prove this, we prove a stronger result:

Theorem 6. *Let E_n be an empirical process with n samples and $\mathbf{B} = \{B_t : t \in [0,1]\}$ be a Brownian bridge. Then for all $t \in [0,1]$, $E_n(t) \to B_t$ almost surely as $n \to \infty$.*

To prove this theorem, we need three steps:

1. E_n converge to $\mathbf{B}$ almost surely on **finite** many of points in [0,1] as $n \to \infty$.
2. E_n with **finite** many of points in [0,1] will define the original empirical process as the number of points grows to infinity.
3. $\mathbf{B}$ with **finite** many of points in [0,1] will define the original Brownian bridge as the number of points grows to infinity.

For 2 and 3, Daniell-Kolmogorov extension theorem guaranteed the convergence. Since this is out of the discussion of this project, so we will not prove it here. However, if you are interested with it, you can find lots of resource on the internet.

As a result, our goal here is to show that the empirical process E_n will converge to the Brownian bridge $\mathbf{B}$ almost surely on **finite** many of points in [0,1] as $n \to \infty$.

Proof. First, recall the *Probability integral transform theorem* that the distribution of the CDF(cumulative distribution function) transform of any random variable is uniform on [0,1]. As a result, we can consider the empirical process of uniform distribution case first. As to the empirical process of general distribution function F, we can simply compose the $F(x)$ term to the result of uniform case. Details will be explained later.

So let's consider n uniform *i.i.d.* random variables $X_1, X_2, ..., X_n$ from $F(x) = x$ on [0,1] and use them to construct an empirical distribution $\hat{F}_n$. Now take arbitrary k points $x_1, x_2, ..., x_k$ in [0,1]. We observe the behaviour on these k finite points. Represent them in a random vector:

$$\sqrt{n}\begin{bmatrix} \hat{F}_n(x_1) - F(x_1) \\ \vdots \\ \hat{F}_n(x_k) - F(x_k) \end{bmatrix} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{bmatrix} \mathbf{I}\{X_i \leq x_1\} - F(x_1) \\ \vdots \\ \mathbf{I}\{X_i \leq x_k\} - F(x_k) \end{bmatrix} := \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{Z}_i$$

where $\mathbf{Z}_1, \mathbf{Z}_2, ..., \mathbf{Z}_n$ are *i.i.d.* k-dimensional random vectors with $E[\mathbf{Z}_i] = \mathbf{0}$ and $Cov[\mathbf{Z_i}] = \mathbf{Q}$ $\forall i$, and $\mathbf{Q}$ is the covariance matrix such that

$$\mathbf{Q}_{i,i} = F(x_i)[1 - F(x_i)]$$

and

$$\mathbf{Q}_{i,j} = F(x_i \wedge x_j) - F(x_i)F(x_j) = x_i \wedge x_j - x_i x_j$$

Note that the random part is in the random variables $X_1, ..., X_n$, that is, the empirical distribution $\hat{F}_n$. Not the arbitrary points $x_1, x_2, ..., x_k$.

Then, by the multinomial central limit theorem, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{Z}_i$ converges in distribution to $N_k(\mathbf{0}, \mathbf{Q})$. In other words,

$$\begin{bmatrix} E_n(x_1) \\ \vdots \\ E_n(x_k) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} W(x_1) \\ \vdots \\ W(x_k) \end{bmatrix}$$

where $\mathbf{W} = (W_1, W_2, ..., W_k)' \sim N_k(\mathbf{0}, \mathbf{Q})$.

Also, by the *Uniqueness Property* of Gaussian process and the property of $\mathbf{Q}$, we can know that $W(x_i) = B(x_i)$ $\forall i$, where B is a Brownian bridge. That is, the empirical process will converge to a Brownian bridge at finite points $x_1, x_2, ..., x_k$. Thus, 1 has been proved!

To sum up, the empirical process of uniform distribution will converge to the Brownian bridge in finite many points. Also, by Daniell-Kolmogorov extension theorem, finite many points of distributions will define a stochastic process. And for $(E_n(x_1), E_n(x_2), ..., E_n(x_k))$ and $(B(x_1)mB(x_2), ..., B(x_k))$ will define the empirical process of uniform distribution and Brownian bridge respectively. The following $\square$

$$\begin{array}{ccc} \begin{bmatrix} E_n(x_1) \\ \vdots \\ E_n(x_k) \end{bmatrix} & \xrightarrow{d} & \begin{bmatrix} B(x_1) \\ \vdots \\ B(x_k) \end{bmatrix} \\ \downarrow & & \downarrow \\ E_n & & B \end{array}$$

Figure 2: Proof flow of Theorem 6

Finally, let's put everything together. First, since $E_n \to B$, it's clearly that

$$\sup_{0\le x\le 1} |E_n(x)| \xrightarrow{d} \sup_{0\le x\le 1} |B(x)|$$

Namely, the Kolmogorov statistics of the uniform distribution on [0,1] will converge in distribution to the absolute supremum norm of a Brownian bridge.

As to general distribution F. Let G denotes the uniform distribution on [0, 1]. Consider the Kolmogorov statistics

$$\begin{aligned} D_n :&= \sqrt{n} \sup_{0\le x\le 1} |\hat{F}_n(x) - F(x)| \\ &= \sqrt{n} \sup_{0\le x\le 1} |G[\hat{F}_n(x)] - G[F(x)]| \\ &\xrightarrow{d} \sup_{0\le x\le 1} |B(F(x))| \end{aligned}$$

This is the Donsker Theorem, the uniform central limit theorem for empirical process.

4 Appendix

It's welcome to discuss the code with us!

References

1. *Brownian Bridge*, http://www.math.uah.edu/stat/brown/Bridge.html