0.1 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.

0.1.1 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*

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

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.

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.

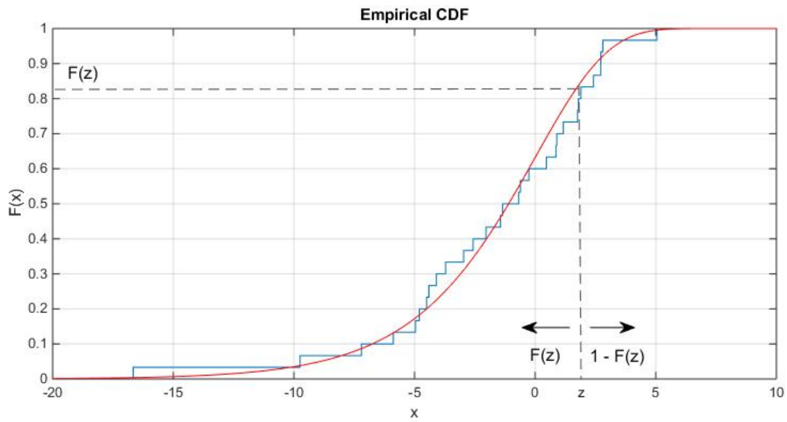

Figure 1: Empirical distribution and its point-wise convergence property.

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

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

.

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. 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 asymptotic behaviour of empirical distribution. And before introducing the advanced result about Uniform Law of Large Number(ULLN) and Uniform Central Limit Theorem(UCLT) about the uniform behabiour, let's consider an easier case: the Kolmogrov Statistics, which also draws a good intuition on the convergence of empirical distribution.

0.1.2 Kolmogrov Statistics

The Kolmogrov 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 Kolmogrov statistics D_n is the supremum pointwise distance between the empirical distribution and the target function. 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 Kolmogrov 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 Kolmogrov 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 Kolmogrov 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 Kolmogrov 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. $\square$

Apart from the amazing fact that the distribution of Kolmogrov statistics is distribution-free, the convergence is also guaranteed by the following Glivenko-Cantelli theorem. Also, the asymptotic behaviour of Kolmogrov 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.

0.2 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).

0.2.1 Glivenkp-Cantelli Theorem: ULLN

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

Theorem 4 (Glivenko-Cantelli)**.** *The Kolmogrov 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,

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

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

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

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

With triangle inequality and the above two results from ordered statistics, we can conclude that

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

Thus, we have the convergence of D_n for all n. $\square$

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.

0.2.2 Donsker Theorem: Uniform CLT