Large Sample Techniques for Statistics
Empirical Process

Ma Ting Tin

Department of Statistics, Chinese University of Hong Kong

2021 Summer Lab Meeting

June 10, 2021

Introduction

Introduction

1. In Chapters before, we are studying random variables. In this Chapter, we study **Random Function** instead
2. In particular, study the limit theorem for **Empirical Process**

Suppose that $X_1, X_2, \cdots \overset{\text{IID}}{\sim} F$. The following quantity is known as the Empirical Cumulative Distribution Function (ECDF).

$$F_n(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x), \quad x \in \mathbb{R}$$

Noticing that **for a fixed** $\boldsymbol{x}$, $\mathbb{1}(X_i \leq x) \overset{\text{IID}}{\sim} \text{Bern}\{F(x)\}$ and hence for each fixed x,

- **(SLLN):** $F_n(x) \overset{\text{a.s.}}{\to} F(x)$.
- **(CLT):** $\sqrt{n}\left[F_n(x) - F(x)\right] \overset{\text{d}}{\to} N\big(0, F(x)[1 - F(x)]\big)$.

However, those result are **NOT** strong enough for our research purpose.

Motivation

In our research, we commonly want to study the limiting behaviour of a function of the data, eg. test statistics of some functional of the distribution.

Say, our quantity of interest is the **process** $\{Q_n(t) : t \in \mathbb{R}^+\}$. Suppose one have already developed SLLN or CLT for $Q_n(t)$ for each fixed t, eg: $Q_n(t) \xrightarrow{d} Q_t$ for some process Q_t for each fixed t.
QUESTION: What is the distribution of $\sum_{t=1}^{k} \left[Q_n(t) - Q_n(t-1)\right]^2$?
Unfortunately, we cannot answer this question by previous tool.

Remark 0.1

- *We do not know the structure between the process* $\{Q_t : t \in \mathbb{R}^+\}$.
- *The convergence for each* **fixed** t *does NOT implies the convergence of the whole process. It is because pointwise convergence does NOT implies uniform convergence (Refer to next page for the details).*

Supplement to the convergence issue (Additional)

Definition 1.1 (Pointwise and Uniform Convergence)

A sequence of function $\{f_n\}$ is said to be converges pointwisely (uniformly) to f on A if

- **[Pointwise Convergence]**
  $\forall x \in A,\ \forall \epsilon > 0,\ \exists N = N(\epsilon, x) \in \mathbb{N}$ s.t. whenever $n \geq N,\ |f_n(x) - f(x)| < \epsilon.$
- **[Uniform Convergence]**
  $\forall \epsilon > 0,\ \exists N = N(\epsilon) \in \mathbb{N}$ s.t. $\forall x \in A,$ whenever $n \geq N,\ |f_n(x) - f(x)| < \epsilon.$

Remark 0.2

- *Uniform convergence implies pointwise convergence but converse does NOT hold.*
- *An equivalent definition for uniform convergence is that* $\sup_{x \in A} |f_n(x) - f(x)| = 0,$ *which is commonly used for us.*
- **Interpretation:** *Uniform convergence is requiring the sequence of functions getting into its neigbourhood (like a tube).*

Simplification of Problem

As suggested by previous pages, we are interested in the behaviour of the **process** $\{\sqrt{n}[F_n(x) - F(x)] : x \in \mathbb{R}\}$, where $F_n(x)$ are ECDF of $X_1, \cdots, X_n \overset{\text{IID}}{\sim} F$.

Theorem 7.1 (Inverse Transform)

Define the pseudo inverse of F by $F^{-1}(t) := \inf\{x : F(x) \geq t\}$. Let $\xi \sim \text{UNIF}(0, 1)$, then $X := F^{-1}(\xi) \sim F$.

The above theorem implies that it suffices for us to only consider the empirical process induced by uniform R.V.s. Let $\xi_1, \cdots, \xi_n \overset{\text{IID}}{\sim} \text{UNIF}(0, 1)$, define

- $G_n(t) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}(\xi_i \leq t)$.
- $U_n(t) := \sqrt{n}\big[G_n(t) - t\big]$.

By defining $X_i := F^{-1}(\xi_i)$, $\mathbb{1}(\xi_i \leq F(t)) = \mathbb{1}(F^{-1}(\xi_i) \leq t) = \mathbb{1}(X_i \leq t)$. It follows

- $G_n(F) = F_n$
- $U_n(F) = \sqrt{n}(F_n - F)$

So it suffices to study only U_n, known as the uniform empirical process.

Glivenko-Cantelli theorem and statistical functionals

Uniform SLLN

- $G_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(\xi_i \leq t)$
- (Identity map) $I(t) := t$
- (Sup-norm) $||f|| := \sup_{x\in \mathrm{dom}(f)} |f(x)|$ defined on function space $C[0,1]$ or D,
  1. $C[0,1] := \{f : f \text{ cts on } [0,1],\ \mathrm{dom}(f) = [0,1]\}$
  2. $D := \{f : f \text{ is RCLL},\ \mathrm{dom}(f) = [0,1]\}$

Theorem 7.2 (Glivenko-Cantelli theorem)

1. $||G_n - I|| \overset{\text{a.s.}}{\to} 0$ *as* $n \to \infty$.
2. *(Example 7.1) Equivalently,* $\sup_x |F_n(x) - F(x)| \overset{\text{a.s.}}{\to} 0$ *as* $n \to \infty$.

Remark 0.3

- *The result of Glivenko-Cantelli shows a.s. uniform convergence instead of pointwise convergence, hence known as* **Uniform SLLN**.

Application of Glivenko-Cantelli Theorem

Example 2.1 (Sufficiency for studying Uniform Empirical Process)

Noticing that for $\widetilde{X}_i := F^{-1}(\xi_i)$, $\{\widetilde{X}_i\} \stackrel{d}{=} \{X_i\}$. So ECDF of those two sequence ($\widetilde{F}_n \& F_n$) have same probabilistic behavior. We have

$$\sup_{x\in\mathbb{R}} |\widetilde{F}_n(x) - F(x)| = \sup_{x\in\mathbb{R}} |G_n\{F(x)\} - F(x)| \leq \sup_{t\in[0,1]} |G_n(t) - t| = ||G_n - I|| \stackrel{\text{a.s.}}{\to} 0,$$

where the last convergence holds due to Glivenko-Cantelli Theorem.

Example 2.2 (Asymptotic properties of ordered-statistics)

For $\xi_1, \cdots, \xi_n \stackrel{\text{IID}}{\sim} \text{UNIF}(0,1)$, write $\xi_{(i)}$ as the i-th ordered statistics, i.e. the i-th smallest value within $\xi_1, \cdots, \xi_n$. Noticing that

1. $G_n^{-1}(t) = \sum_{i=1}^n \xi_{(i)} \mathbb{1}\left(\frac{i-1}{n} < t \leq \frac{i}{n}\right)$.
2. $||G_n^{-1} - I|| = ||G_n - I|| \stackrel{\text{a.s.}}{\to} 0$ by G.C.

with implies that $\max_{1\leq i\leq n} |\xi_{(i)} - i/n| \stackrel{\text{a.s.}}{\to} 0$, i.e. the sequence of h ordered statistics converges to the sequence $\{i/n\}_{i=1}^n$ uniformly in i.

Statistical Functionals

Definition 2.3 (Statistical Functionals)

Let $\mathscr{D} = \{F : F \text{ is a CDF}\}$, i.e. space of CDF. h is a statistical functionals if $h : \mathscr{D} \to \mathbb{R}$ is a function.

Example 2.4 (Common Statistical Functionals)

1. **(Fixed point a)** $h(F) := F(a)$
2. **(q-th Quantile)** $h(F) := F^{-1}(q)$
3. **p-th Moment** $h(F) := \mathrm{E}_F(X^p)$

Remark 0.4

Nearly all of our quantity of interest can be written as a statistical functional.

Convergence of Statistical Functionals

Definition 2.5 (Continuous Statistical Functional)

Let h be a statistical functional. h is said to be continuous at F if for any **sequence of CDF** $\{H_n\}$, $||H_n - F|| \to 0$ implies that $h(H_n) \to h(F)$

Remark 0.5

It is somehow mimicking the sequential criterion for describing the continuity of function, i.e. a function f is continuous at $c \in \mathbb{R}$ iff $\lim a_n = c$ implies $\lim f(a_n) = f(c)$.

Theorem 7.3

Let $X_1, \cdots X_n \overset{\text{IID}}{\sim} F$ with F_n as the ECDF. If h is continuous at F, then $h(F_n) \overset{\text{a.s.}}{\to} h(F)$.

The proof of Thereom 7.3 is trivial. By Glivenko-Cantelli Theorem, we have $||F_n - F|| \overset{\text{a.s.}}{\to} 0$ and hence according to definition of continuous functional, $h(F_n) \overset{\text{a.s.}}{\to} h(F)$.

Convergence of Statistical Functionals (Continued)

There are several comments for the Theorem 7.3.

- 👎: The condition proposed is a sufficient condition, the convergence might hold even though the condition is NOT satisfied (See thirt item of following example).
- 👎: Not efficient as checking the continuity of h is tedious, even harder for directly checking the a.s. convergence.
- 👍: Some common functional h might have been proven to be continuous, then one can directly apply Theorem 7.3.

Example 2.6 (Continuity of Common Statistical Functionals)

1. **(Fixed point a)** $h(F) := F(a)$ is cts
2. **(q-th Quantile)** $h(F) := F^{-1}(q)$ is cts
3. **p-th Moment** $h(F) := \mathrm{E}_F(X^p)$ NOT cts

Weak convergence of empirical processes

Preliminary

Definition 3.1 (Brownian Motion)

A process $\{W(t)\}$ is a Brownian Motion if $W(0) = 0$ and

1. (Independent increment) For $t_1 < t_2 \leq t_3 < t_4$, $(W_{t_2} - W_{t_1}) \perp\!\!\!\perp (W_{t_4} - W_{t_3})$.
2. (Normality) For $t > s \geq 0$, $W_t - W_s \sim N(0, (t-s)\sigma^2)$ for some $\sigma^2 > 0$.

Definition 3.2 (Gaussian Process)

A process $\{X(t) : 0 \leq t \leq 1\}$ is a Gaussian Process if for any $t_1 < \cdots < t_k$, $[X(t_1), \cdots, X_(t_k)]$ follows multivariate normal distribution.

Definition 3.3 (Brownian Bridge)

A process $\{U(t)\}$ defined by $U(t) := W(t) - tW(1)$ is a brownian bridge.

- It is a Gaussian process with $\mathrm{E}(U(t)) = 0$ and $\mathrm{Cov}(U(s), U(t)) = \min(s, t) - st$.

Convergence in distribution (Additional)

Let $\mathscr{D}$ be the σ-field generated by the finite-dimensional subsets of D.

Definition 3.4 (Convergence on $(D, \mathscr{D}, ||\cdot||)$)

$\xi_n \xrightarrow{d} \xi$ on $(D, \mathscr{D}, ||\cdot||)$ if $\forall B \in \mathscr{D}$, $\lim_{n\to\infty} \mathrm{P}(\xi_n \in B) = \mathrm{P}(\xi \in B)$.

Theorem (Portmanteau)

For any random vectors X_n and X, the following are equivalent:

1. $\lim \mathrm{P}(X_n \leq x) = \mathrm{P}(X \leq x)$ *for all continuity points of* $x \mapsto \mathrm{P}(X \leq x)$.
2. $\lim \mathrm{E}\{f(X_n)\} = \mathrm{E}\{f(X)\}$ *for all bounded, continuous functions* f.
3. $\lim \mathrm{E}\{f(X_n)\} = \mathrm{E}\{f(X)\}$ *for all bounded, lipschitz functions* f.
4. $\liminf \mathrm{E}\{f(X_n)\} \geq \mathrm{E}\{f(X)\}$ *for all nonnegative, continuous function* f.
5. $\liminf \mathrm{P}(X_n \in G) \geq \mathrm{P}(X \in G)$ *for all open set* G.
6. $\limsup \mathrm{P}(X_n \in F) \leq \mathrm{P}(X \in F)$ *for all closed set* F.
7. $\lim \mathrm{P}(X_n \in B) = \mathrm{P}(X \in B)$ *for all borel sets* B *with* $\mathrm{P}(X \in \delta B) = 0$.

Doob-Donsker Theorem: Generalization of CLT

Recall that $\xi_1, \cdots, \xi_n \overset{\text{IID}}{\sim} \text{UNIF}(0,1)$, with the following notions

- $G_n(t) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}(\xi_i \leq t)$.
- $U_n(t) := \sqrt{n}\big[G_n(t) - t\big]$.

Theorem 7.3 (Doob-Donsker)

$U_n \xrightarrow{\text{d}} U$ on $(D, \mathscr{D}, ||\cdot||)$ as $n \to \infty$, where U is the Brownian bridge.

Example 3.5 (Asymptotics of Kolmogorov-Smirnov Statistics)

Let $X_1, \cdots, X_n \overset{\text{IID}}{\sim} F$. Then we are testing $H_0 : F = F_0$ against $H_1 : F \geq F_0$ with the test statistics $T := \sqrt{n}\sup_x(F_n(x) - F_0(x))$. Then

$$\mathrm{P}(T \leq \lambda) = \mathrm{P}\Big(\sup_x \sqrt{n}(F_n(x) - F(x)) \leq \lambda\Big) = \mathrm{P}\Big(\sup_x U_n\{F(x)\} \leq \lambda\Big)$$

Noticing that $\sup_x U_n\{F(x)\} = \sup_{0\leq t\leq 1} U_n(t)$ and that $h(f) := \sup_{0\leq t\leq 1} f(t)$ is a continuous functional on $(D, ||\cdot||)$ and hence by C.M.T and the Doob-Donsker, $\sup_{0\leq t\leq 1} = g(U_n) \to g(U) = \sup_{0\leq t\leq 1} U(t)$.

LIL and strong Approximation

Law of Iterated Logarithm (LIL)

Theorem 7.5 & 7.6

1. $\limsup ||U_n||/\sqrt{2\log\log n} = 1/2$ *a.s.*
2. $\liminf \sqrt{2\log\log n}||U_n|| = \pi/2$ *a.s.*

Remark 0.6

- *Those are concerning upper limit and lower limit related to $||U_n||$, which describes the precise a.s. convergence rate of the empirical process.*
- *It is natural to extend the result to a functional one.*

The following are implied by theorem 7.5 and 7.6 (Sufficient to study U_n):

1. $\limsup ||\sqrt{n}(F_n - F)||/\sqrt{2\log\log n} \leq 1/2$ a.s.
2. $\liminf \sqrt{2\log\log n}||\sqrt{n}(F_n - F)|| = \pi/2$ a.s.

Functional LIL

Let (M, ρ) be a metric space and $S \subset M$. Let $\{\xi_n\}$ be a sequence of M-valued random variables on $(\Omega, \mathscr{A}, \mathrm{P})$.

Definition 4.1 (Relatively Compact)

We say $\{\xi_n\}$ is relatively compact w.r.t. ρ on M with limit set S, denoted by ξ_n r.c. S w.r.t. ρ on M a.s, if $\exists A \in \mathscr{A}$ with $\mathrm{P}(A) = 1$ s.t.

1. **(Bolzano-Weierstrass Property)** For each $\omega \in \Omega$, $\{\xi_n(\omega)\}$ is Cauchy w.r.t ρ.
2. **(Closedness of S)** All of the ρ-limit points of $\xi_n(\omega)$ belong to S.
3. **(Limit set)** For all $s \in S$, $\exists (n_k) \subset \mathbb{N}$ s.t. $\rho\{\xi_{n_k}(\omega), s\} \to 0$.

Let K be the function space that contain all absolutely continuous function f on $[0, 1]$ with $f(0) = 0$ and $\int_0^1 f^2(t)dt < 1$.

Theorem 7.7

$U_n/\sqrt{2 \log \log n}$ r.c. K w.r.t. $||\cdot||$ on D a.s.

Bounds and Large deviations

Inequalities for maximum type quantities

Let's recall and introduce some notations:

- $||f|| = \sup_x |f(x)|$
- $||f^+|| = \sup_x f(x)$
- $||f^-|| = \sup_x -f(x)$

Theorem 7.9 - 7.12

1. **(7.9)** *For* $0 < p \le 1/2$ *and* $\lambda > 0$,
   $$P\left(\left|\left|\frac{U_n^+}{1-I}\right|\right|_0^p \ge \frac{\lambda}{1-p}\right) \le \exp\left\{-\frac{\lambda^2}{2p(1-p)}\psi\left(\frac{\lambda}{p\sqrt{n}}\right)\right\},$$ *where*
   $||x||_a^b := \sup_{a\le t\le b}|x(t)|$ *and* $\psi(u) := 2u^{-2}[(1+u)\{\log(1+u)-1\}+1]$.
2. **(7.10: Shorack)** *For* $0 < p \le 1/2$ *and* $0 < \lambda \le \sqrt{np}$, *we have*
   $$P\left(\left|\left|\frac{U_n^-}{1-I}\right|\right|_0^p \ge \frac{\lambda}{1-p}\right) \le \exp\left\{-\frac{\lambda^2}{2p}\psi\left(-\frac{\lambda}{p\sqrt{n}}\right)\right\}\exp\left(-\frac{\lambda^2}{2p(1-p)}\right)$$
3. **(7.11)** *For* $0 \le s \le t \le 1$, $0 < a \le b \le 1/2$ *and* $\lambda > 0$, *we have*
   $$P\left(\sup_{t,s:|t-s|\le a}|U_n(t)-U_n(s)| \ge \lambda\sqrt{a}\right) \le \frac{20}{ab^3}\exp\left\{-(1-b)^4\frac{\lambda^2}{2}\psi\left(\frac{\lambda}{\sqrt{an}}\right)\right\}.$$
4. **(7.12: DKW)** *For* $\lambda \ge \sqrt{\ln\sqrt{2}}$, $\frac{1}{2}P(||U_n|| \ge \lambda) \le P(||U_n^-|| \ge \lambda) \le e^{-2\lambda^2}$.

Large Deviation and Application of inequalities

Definition 5.1 (Large Deviation)

The word "Large deviation" refer to probability that the sample mean of $X_1, \cdots, X_n \overset{\text{IID}}{\sim} F$ deviates from $\mathrm{E}(X)$. Let $S_n := \sum_{i=1}^n X_i$. There are two type of results

1. **(WLLN type deviations): Lv1:** Distribution of $n^{-1}S_n$;
   Lv2: Empirical process; **Lv3:** Process generated by iid sequence.
2. **(CLT type deviation):** Study of convergence rate of $F_n(x_n)$, where $\lim x_n = \infty$.

Example 5.2

Let $X_1, \cdots, X_n \overset{\text{IID}}{\sim} F$, where F is unknown and continuous.
Target: Find n s.t. $\mathrm{P}\left(\sup_x |F_n(x) - F(x)| < 0.1\right) \geq 0.95$. Noticing that

$$\mathrm{P}\left(\sup_x |G_n\{F(x)\} - F(x)| < 0.1\right) = \mathrm{P}\left\{\sup_{0 \leq t \leq 1} |G_n(t) - t| < 0.1\right\}$$

$$= \mathrm{P}(||U_n|| < 0.1\sqrt{n}) = 1 - \mathrm{P}(||U_n|| \geq 0.1\sqrt{n}) \geq 1 - 2e^{-0.02n}$$

where the last inequalities hold due to DKW. So it suffices to take $n \geq 185$.

Non-i.i.d. observations

Extension 1: Triangular arrays

Let $X_{n1}, \cdots, X_{nn}$ be independent and $X_{ni} \sim F_{ni}$ (may not be identical). Define

$$\overline{F}_n(x) := \frac{1}{n}\sum_{i=1}^{n} F_{ni}(x) \quad \text{and} \quad F_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_{ni} \le x), \quad x \in \mathbb{R}.$$

Theorem 7.14

$||F_n - \overline{F}_n|| \xrightarrow{\text{a.s.}} 0$ *as* $n \to \infty$.

Extension 2: Stationary ergodic sequences

Definition 6.1

1. **(Strictly Stationary)** A sequence of R.V.s is strictly stationary if for all $k \in \mathbb{N}$, $(X_{k+1}, X_{k+1}, \cdots) \overset{d}{=} (X_0, X_1, \cdots)$.
2. **(Measure-preserving)** Let $(\Omega, \mathscr{A}, \mathrm{P})$ be a probability space, a measurable map $T : \Omega \to \Omega$ is *measure-preserving* if $\mathrm{P}(A) = \mathrm{P}\{T^{-1}(A)\}$ for all $A \in \mathscr{A}$.

Theorem (Additional)

A sequence $\{X_i\}$ is stationary iff $\exists$ measure-preserving transformation T s.t. $X_i(\omega) = X\{T^i(\omega)\}$ for all $\omega \in \Omega$.

Remark 0.7

The theorem shows the linkage between the measure-preserving transformation and strictly stationary process.

Extension 2: Stationary ergodic sequences

Definition 6.2 (Ergodic)

A stationary sequence $\{X_i\}$ generated by transformation T is **ergodic** if T satisfies the following: For all $A \in \mathscr{A}$, $T^{-1}A = A$ implies $\mathrm{P}(A) = 0$ or 1.

Theorem

If T is measure-preserving and $\{X_i\}$ is the sequence generated by T with $\mathrm{E}(|X_0|) < \infty$, then

$$\frac{1}{n}\sum_{i=0}^{n-1} X_i \overset{\text{a.s.}}{\to} \mathrm{E}(X_0|\mathscr{I}),$$

where $\mathscr{I} = \{A \in \mathscr{A} : T^{-1}A = A\}$, i.e. the invariant σ-field w.r.t. T.

Theorem 7.15

Let $\{X_i\}$ be a stationary ergodic sequence with common CDF F and $F_n(x) := n^{-1}\sum_{i=0}^{n-1} \mathbb{1}(X_i \leq x)$. Then $||F_n - F|| \overset{\text{a.s.}}{\to} 0$ as $n \to \infty$.

Extension 2: Stationary ergodic sequences (Additional)

There is alternative definition for Ergodicity and Ergodic Theorem.

Definition 6.3 (Ergodic)

A stationary sequence $\{X_i\}$ is ergodic if for all bounded functions $f : \mathbb{R}^a \to \mathbb{R}$ and $g : \mathbb{R}^b \to \mathbb{R}$,

$$\begin{aligned}\lim_{k\to\infty} & \left| \mathrm{E}\left\{ f\left(X_i, X_{i+1}, \cdots, X_{i+a-1}\right) g\left(X_{i+k}, X_{i+k+1}, \cdots, X_{i+k+b-1}\right) \right\} \right| \\ = & \left| \mathrm{E}\left\{ f\left(X_i, X_{i+1}, \cdots, X_{i+a-1}\right) \right\} \mathrm{E}\left\{ g\left(X_{i+k}, X_{i+k+1}, \cdots, X_{i+k+b-1}\right) \right\} \right|\end{aligned}$$

Theorem (Additional)

Let $\{X_i\}$ be a stationary ergodic sequence with $\mathrm{E}|X_i|$ being finite constant. Then

$$\bar{X}_n \xrightarrow{\text{a.s.}} \mathrm{E}(X_i)$$

Extension 3: Stationary φ-mixing sequences

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a stationary time series. Denote $\sigma(X_i, i \in I)$ as the σ-field generated by $X_i, i \in I$. Write $\mathscr{F}_{-\infty}^{k} := \sigma(X_i, i \leq k)$ and $\mathscr{F}_{k}^{\infty} := \sigma(X_i, i \geq k)$.

Definition 6.4 (φ-mixing)

Let $\varphi : \mathbb{N} \to \mathbb{R}$ be function with $\lim_{n\to\infty} \varphi(n) = 0$. The sequence $\{X_t\}_{t\in\mathbb{Z}}$ is φ-mixing if for $k \in \mathbb{Z}$ and $n \in \mathbb{N}^+$,

$$\Big|\mathrm{P}(E_1 \cap E_2) - \mathrm{P}(E_1)\mathrm{P}(E_2)\Big| \leq \varphi(n)\mathrm{P}(E_1)$$

for all $E_1 \in \mathscr{F}_{-\infty}^{k}, E_2 \in \mathscr{F}_{k+n}^{\infty}$.

Remark 0.8

- *If* $\mathrm{P}(E_1) > 0$, *the condition can be re-written as* $|\mathrm{P}(E_2|E_1) - \mathrm{P}(E_2)| \leq \varphi(n)$.
- *From the last bullet, we can see that the mixing condition means that*
  1. *There is a decay in dependency as the random variables in the sequence are further apart.*
  2. *The rate of decay is controlled by the function* $\varphi(n)$.

Extension 3: Stationary φ-mixing sequences

Theorem 7.17

Let $\{X_i\}$ be stationary φ-mixing and $X_i \in [0,1]$. Let F be the CDF of X_i and let F_n be the ECDF. If F is continuous with $\sum_{n=1}^{\infty} n^2 \sqrt{\varphi(n)} < \infty$, then $\sqrt{n}(F_n - F) \xrightarrow{\mathrm{d}} Z$ on $(D, \mathscr{D}, ||\cdot||)$, where Z is a Gaussian process (NOT Brownian Motion) satifying $\mathrm{E}[Z(t)] = 0$ and

$$\mathrm{Cov}\Big(Z(s), Z(t)\Big) = \mathrm{E}\big\{g_s(X_0)g_t(X_0)\big\} + \sum_{i=1}^{\infty} \Big[\mathrm{E}\big\{g_s(X_0)g_t(X_i)\big\} + \mathrm{E}\big\{g_s(X_i)g_t(X_0)\big\}\Big]$$

with $g_t(x) := \mathbb{1}(0 \le x \le t) - F(x)$ and the path of Z is continuous a.s.

Empirical processes indexed by functions

Motivation

We can rewrite the ECDF of $X_1, \cdots, X_n$ as a functional form by

$$P_n(f) := \frac{1}{n}\sum_{i=1}^{n} f(X_i),$$

where $f(y) = \mathbb{1}(y \leq x)$.
What we are studying before is the process $\{P_n(\mathbb{1}(-\infty, x]), x \in \mathbb{R}\}$. Define $P(f) = \int f d\mathrm{P} = F(x)$ and $\mathscr{F} := \{\mathbb{1}(-\infty, x), x \in \mathbb{R}\}$. So the Glivenko-Cantelli theorem can be written as

$$\sup_{f\in\mathscr{F}} |P_n(f) - P(f)| \xrightarrow{\text{a.s.}} 0 \quad \text{as } n \to \infty$$

A natural generalization is to consider an arbitrary class of function $\mathscr{F}$, i.e.

- For each fixed $f \in \mathscr{F}$, $P_n(f)$ is a Random Variable.
- For $\{P_n(f) : f \in \mathscr{F}\}$ is an empirical process **indexed by functions**.

So the question comes: Whether $\sup_{f\in\mathscr{F}} |P_n(f) - P(f)| \xrightarrow{\text{a.s.}} 0$ holds?

P-Glivenko-Cantelli class

Definition 7.1

A class of function $\mathscr{F}$ is said to be a P-Glivenko-Cantelli class if $\sup_{f \in \mathscr{F}} |P_n(f) - P(f)| \overset{\text{a.s.}}{\to} 0$.

Let $||f||_{r,P} := \left(\int |f|^r dP\right)^{1/2}$ for $1 \leq r < \infty$ and $\mathscr{L}_r(P) = \{f : ||f||_{r,P} < \infty\}$. Then

Definition 7.2 (ϵ-bracket & bracketing number)

1. $[g, h]$ is an ϵ-bracket for f if $g, h \in \mathscr{L}_r(P)$, $\mathrm{P}\{g(X) \leq f(X) \leq h(X)\} = 1$ and $||h - g||_{r,P} \leq \epsilon$.
2. Bracketing number $N\{\epsilon, \mathscr{F}, \mathscr{L}_r(P)\}$ is the minimum number of ϵ-brackets in $\mathscr{L}_r(P)$ needed to cover $\mathscr{F}$.

Theorem 7.18

If $N\{\epsilon, \mathscr{F}, \mathscr{L}_r(P)\} < \infty$ for $\epsilon > 0$, then $\mathscr{F}$ is a P-Glivenko-Cantelli class.

P-Donsker

- Define the entropy with bracketing number as $E\{\epsilon, \mathscr{F}, \mathscr{L}_r(P)\} := \log N\{\epsilon, \mathscr{F}, \mathscr{L}_r(P)\}$.
- Let $l^\infty(\mathscr{F})$ denote the collection of all bounded functionals $P : \mathscr{F} \mapsto \mathbb{R}$

Definition 7.3 (P-Donsker)

$\mathscr{F}$ is a P-Donsker if $\sqrt{n}(P_n - P) \xrightarrow{d} G$ on $\left(l^\infty(\mathscr{F}), \sigma\{l^\infty(\mathscr{F})\}, \varphi\right)$, where

1. $\varphi(P, Q) := \sup_{f \in \mathscr{F}} |P(f) - Q(f)|$.
2. G is a Gaussian process indexed by $f \in \mathscr{F}$ with mean 0 and $\mathrm{Cov}\{G(f_1), G(f_2)\} = \mathrm{Cov}\{f_1(X), f_2(X)\}$, where $f_1, f_2 \in \mathscr{F}$.

Theorem 7.19

If $\int_0^1 \sqrt{E\{\epsilon, \mathscr{F}, \mathscr{L}_2(P)\} d\epsilon} < \infty$, *then* $\mathscr{F}$ *is* P*-Donsker.*

Example of P-Donsker/Glivenko-Cantelli (Additional)

Example 7.4

Let $\mathscr{F} = \{f_t := \mathbb{1}(-\infty, t) : t \in \mathbb{R}\}$. Then it is obviously P-Donsker and P-Glivenko-Cantelli class as these are just the classical empirical process. However, we attempt to proceed by applying Theorem 7.18 and 7.19.

Fix $\epsilon > 0$. There exists partition of $\mathbb{R}$, $-\infty = t_0 < t_1 < \cdots < t_k = \infty$ s.t. $F(t_i) - F(t_{i-1}) < \epsilon$ for all i. Noticing that

- As $|F(x)| \leq 1$, so k can be chosen as number smaller tham $2/\epsilon$.
- $L_1(F)$ size of bracket is less than ϵ and so its $L_2(F)$ size is bounded above by $\sqrt{\epsilon}$.

Therefore. we have

$$N\{\sqrt{\epsilon}, \mathscr{F}, \mathscr{L}_2(P)\} < 2/\epsilon \quad \Rightarrow \quad N\{\epsilon, \mathscr{F}, \mathscr{L}_2(P)\} < 2/\epsilon^2$$

- By Theorem 7.18, $N\{\epsilon, \mathscr{F}, \mathscr{L}_2(P)\} < 2/\epsilon^2 < \infty$ for all ϵ and hence $\mathscr{F}$ is P-Glivenko-Cantelli class.
- By Theorem 7.19, as $\int_0^1 E\{\epsilon, \mathscr{F}, \mathscr{L}_2(P)\} < \infty$, $\mathscr{F}$ is P-Donsker class.

Thank You!