Large Sample Convergence Issues

Presentor: Chun-Jen Shih

Advisor: Jian-Jiun Ding

Table of Contents

Recap of the definitions of limits

1. $\lim_{x \to x_0} f(x) = L$

   $\forall \epsilon > 0,\ \exists \delta > 0$ s.t. $\forall x,\ 0 < |x - x_0| < \delta \Rightarrow |f(x) - L| < \epsilon$

2. $\lim_{x \to x_0} f(x) = \infty$

   $\forall M > 0,\ \exists \delta > 0$ s.t. $\forall x,\ 0 < |x - x_0| < \delta \Rightarrow f(x) > M$

3. $\lim_{x \to \infty} f(x) = L$

   $\forall \epsilon > 0,\ \exists N > 0$ s.t. $\forall x > N \Rightarrow |f(x) - L| < \epsilon$

4. $\lim_{x \to \infty} f(x) = \infty$

   $\forall M > 0,\ \exists N > 0$ s.t. $\forall x > N \Rightarrow f(x) > M$

Uniform and Pointwise Convergence

Consider a sequence of functions $\{f_k(x), x \in D\}_{k=1}^{\infty}$

1. Uniform Convergence: $f_k(x)$ converges uniformly to $f(x)$

   $\forall \epsilon > 0$, $\exists K > 0$ s.t. $\forall x \in D$, $k > K \Rightarrow |f_k(x) - f(x)| < \epsilon$

   - You have to find one K that works for all x to prove uniform convergence (K only depends on ϵ, thus uniform)

2. Pointwise Convergence: $f_k(x)$ converges pointwise to $f(x)$

   $\forall \epsilon > 0$, $\forall x \in D$, $\exists K > 0$ s.t. $\forall k > K \Rightarrow |f_k(x) - f(x)| < \epsilon$

   - You can find different values of K for different values of x to prove pointwise convergence (K can depend on x, thus pointwise)

Uniform and Pointwise Convergence

1. Uniform Convergence: $f_k(x)$ converges uniformly to $f(x)$

   $\forall \epsilon > 0,\ \exists K > 0$ s.t. $\forall x \in D,\ k > K \Rightarrow |f_k(x) - f(x)| < \epsilon$

   $\equiv \forall \epsilon > 0,\ \exists K > 0$ s.t. $\forall k > K \Rightarrow \sup_{x \in D} |f_k(x) - f(x)| < \epsilon$

   $\equiv \lim_{k \to \infty} \sup_{x \in D} |f_k(x) - f(x)| = 0$

2. Pointwise Convergence: $f_k(x)$ converges pointwise to $f(x)$

   $\forall \epsilon > 0,\ \forall x \in D,\ \exists K > 0$ s.t. $\forall k > K \Rightarrow |f_k(x) - f(x)| < \epsilon$

   $\equiv \forall x \in D,\ \forall \epsilon > 0,\ \exists K > 0$ s.t. $\forall k > K \Rightarrow |f_k(x) - f(x)| < \epsilon$

   $\equiv \forall x \in D,\ \lim_{k \to \infty} f_k(x) = f(x)$

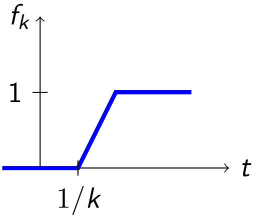
Uniform and Pointwise Convergence

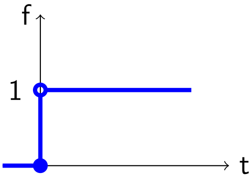
Properties

1. $f_k(x)$ is continuous $\not\Rightarrow$ $f(x)$ is continuous if only pointwise convergence is guaranteed.
2. $f_k(x)$ is continuous $\Rightarrow$ $f(x)$ is continuous if uniform convergence is guaranteed.
3. uniform convergence $\Rightarrow$ pointwise convergence
4. $\lim\limits_{k\to\infty} \int_{t_0}^{t_1} f_k(t)dt = \int_{t_0}^{t_1} f(t)dt$ if uniform convergence

Uniform and Pointwise Convergence

$$f_k(t) = \begin{cases} 0 & t \le 0 \\ kt & 0 < t < 1/k \\ 1 & t \ge 1/k \end{cases} \qquad f(t) = \begin{cases} 0 & t \le 0 \\ 1 & t > 0 \end{cases}$$





Uniform and Pointwise Convergence

$$f_k(t) = \begin{cases} 0 & t \le 0 \\ kt & 0 < t < 1/k \\ 1 & t \ge 1/k \end{cases} \qquad f(t) = \begin{cases} 0 & t \le 0 \\ 1 & t > 0 \end{cases}$$

1. It is clear that $\forall t$, $\lim\limits_{k\to\infty} f_k(t) = f(t)$ $\therefore$ pointwise convergence
2. $\sup\limits_t |f_k(t) - f(t)| = 1 \Rightarrow \lim\limits_{k\to\infty} \sup\limits_t |f_k(t) - f(t)| = 1 \neq 0$ $\therefore$ not uniform convergence

Uniform and Pointwise Convergence

Prove that $\lim_{k \to \infty} \int_{t_0}^{t_1} f_k(t)dt = \int_{t_0}^{t_1} f(t)dt$ if uniform convergence

$$
\begin{aligned}
|\int_{t_0}^{t_1} f_k(t)dt - \int_{t_0}^{t_1} f(t)dt| &\le \int_{t_0}^{t_1} |f_k(t) - f(t)|dt \\
&\le \int_{t_0}^{t_1} \sup_x |f_k(x) - f(x)|dt \\
&= \sup_x |f_k(x) - f(x)|(t_1 - t_0)
\end{aligned}
$$

$\because \forall \epsilon > 0\ \exists K > 0$ s.t. $\forall k > K$, $\sup_x |f_k(x) - f(x)| < \epsilon$

Let $\epsilon' = \epsilon(t_1 - t_0)$

$\therefore \forall \epsilon' > 0\ \exists K > 0$ s.t. $\forall k > K$, $| \int_{t_0}^{t_1} f_k(t)dt - \int_{t_0}^{t_1} f(t)dt| < \epsilon'$

Convergence in Distribution

Definition

A sequence of random variables $\{X_k, k = 1, 2, ...\}$ with CDF $\{F_{X_k}, k = 1, 2, ...\}$ converges in distribution to the random variable X with CDF F_X if

$$\lim_{k \to \infty} F_{X_k}(x) = F_X(x)$$

at all points x where $F_X(x)$ is continuous

- From the definition, we find that convergence in distribution has a lot to do with pointwise convergence. Indeed, if $\{F_{X_k}\}$ converges pointwise to F_X, then $\{X_k\}$ converges in distribution to X. However, the converse may not be true since convergence in distribution only requires convergence in the continuity points.

Convergence in Distribution

Definition

A sequence of random variables $\{X_k, k = 1, 2, ...\}$ with CDF $\{F_{X_k}, k = 1, 2, ...\}$ converges in distribution to the random variable X with CDF F_X if

$$\lim_{k \to \infty} F_{X_k}(x) = F_X(x)$$

at all points x where $F_X(x)$ is continuous

- Convergence in distribution is the weakest form of convergence. However, it is involved in the central limit theorem.
- Convergence in distribution is denoted as $X_k \xrightarrow{d} X$

Convergence in Probability

Definition

A sequence of random variables $\{X_k, k = 1, 2, ...\}$ converges in probability towards the random variable X if $\forall \epsilon > 0$

$$\lim_{k \to \infty} P(|X_k - X| > \epsilon) = 0$$

- $\forall \epsilon > 0$, $\forall \epsilon' > 0$, $\exists K > 0$ s.t. if $k > K$ $|P(|X_k - X| > \epsilon) - 0| < \epsilon'$, i.e. $P(|X_k - X| < \epsilon) < 1 - \epsilon'$ $\rightarrow$ the probability that $\{X_k\}$ equals X is asymptotically increasing and approaches 1 but never actually reaches 1.

Convergence in Probability

Definition

A sequence of random variables $\{X_k, k = 1, 2, ...\}$ converges in probability towards the random variable X if $\forall \epsilon > 0$

$$\lim_{k \to \infty} P(|X_k - X| > \epsilon) = 0$$

- Convergence in probability is widely encountered in statistics. For example, an estimator is called consistent if it converges in probability to the quantity being estimated. Convergence in probability is also involved in the weak law of large numbers.
- Convergence in probability is denoted as $X_k \xrightarrow{p} X$

Convergence in Probability

Definition

A sequence of random variables $\{X_k, k = 1, 2, ...\}$ converges in probability towards the random variable X if $\forall \epsilon > 0$

$$\lim_{k \to \infty} P(|X_k - X| > \epsilon) = 0$$

- The continuous mapping theorem states that for every continuous function g, if $X_k \xrightarrow{p} X$, then $g(X_k) \xrightarrow{p} g(X)$

Almost Sure Convergence

Definition

A sequence of random variables $\{X_k, k = 1, 2, ...\}$ converges almost surely towards the random variable X if

$$P(\lim_{k\to\infty} X_k = X) = 1$$

- $P(\forall\epsilon > 0, \exists K > 0 \textit{ s.t. if } k > K \Rightarrow |X_k - X| < \epsilon) = 1$
  $\therefore \forall\epsilon > 0$, $\exists K > 0$, s.t. if $k > K \Rightarrow P(|X_k - X| < \epsilon) = 1$
  $\rightarrow$ the probability that $\{X_k\}$ equals X is asymptotically increasing and will eventually reach 1.

Almost Sure Convergence

Definition

A sequence of random variables $\{X_k, k = 1, 2, ...\}$ converges almost surely towards the random variable X if

$$P(\lim_{k\to\infty} X_k = X) = 1$$

- Almost sure convergence is involved in the strong law of large numbers.
- Almost sure convergence is denoted as $X_k \stackrel{a.s.}{\to} X$

Sure Convergence

Definition

A sequence of random variables $\{X_k, k = 1, 2, ...\}$ converges surely or everywhere or pointwise towards the random variable X if

$$\lim_{k \to \infty} X_k = X$$

- $\forall \epsilon > 0$, $\exists K > 0$ s.t. if $k > K \Rightarrow |X_k - X| < \epsilon$
- Sure convergence of a random variable is the strongest. However, almost sure convergence has already been almost identical to sure convergence. The difference between the two only exists on sets with probability zero. Thus, this is why sure convergence is very rarely used.

Convergence in the r^{th} Mean

Definition

Given a real number $r \geq 1$, a sequence of random variables $\{X_k, k = 1, 2, ...\}$ converges in the r^{th} mean towards the random variable X if

1. $\mathbb{E}[|X_k|^r]$ of X_k and $\mathbb{E}[|X|^r]$ of X exist
2. $\lim\limits_{k \to \infty} \mathbb{E}[|X_k - X|^r] = 0$

- When $r = 1$, we say that X_k converges in mean to X. When $r = 2$, we say that X_k converges in mean square to X.
- Convergence in mean square is involved in the Karhunen-Loéve expansion

Convergence in the r^{th} Mean

Definition

Given a real number $r \geq 1$, a sequence of random variables $\{X_k, k = 1, 2, ...\}$ converges in the r^{th} mean towards the random variable X if

1. $\mathbb{E}[|X_k|^r]$ of X_k and $\mathbb{E}[|X|^r]$ of X exist
2. $\lim_{k \to \infty} \mathbb{E}[|X_k - X|^r] = 0$

- Convergence in the r^{th} mean is denoted as $X_k \xrightarrow{L^r} X$
- $X_k \xrightarrow{L^r} X \Rightarrow \lim_{k \to \infty} \mathbb{E}[|X_k|^r] = \mathbb{E}[|X|^r]$

Properties of Convergence of Random Variables

$$
\begin{array}{ccccc}
\xrightarrow{L^s} & \underset{s>r\geq 1}{\Rightarrow} & \xrightarrow{L^r} & & \\
 & & \Downarrow & & \\
\xrightarrow{a.s.} & \Rightarrow & \xrightarrow{p} & \Rightarrow & \xrightarrow{d}
\end{array}
$$

Table of Contents

A General Inequality

$$P(|X| \geq a) \leq \frac{\mathbb{E}[|X|^p]}{a^p} \quad \textit{for any } p = 1, 2, \ldots \textit{ and } a > 0$$

Proof.

$$\begin{aligned}
\mathbb{E}[|X|^p] &= \int_{\mathbb{R}} |x|^p f_X(x) dx \\
&\geq \int_{|X| \geq a} |x|^p f_X(x) dx \\
&\geq a^p \int_{|X| \geq a} f_X(x) dx = a^p P(|X| \geq a)
\end{aligned}$$

$\square$

The Markov Inequality

Apply $p = 1$ to the general inequality and we can get the Markov inequality

$$P(|X| \geq a) \leq \frac{\mathbb{E}[|X|]}{a}$$

We can have two natural intuitions from this inequality.

- When a gets larger, it is harder for $|X|$ to take values that exceed a
- When a is fixed, there will be more probability mass above a or below $-a$ if $\mathbb{E}[|X|]$ is larger. Thus, it is easier for $|X|$ to take values exceeding a.

The Chebyshev Inequality

Apply $p = 2$ to the general inequality and we can get the Chebyshev inequality

$$P(|X - \mathbb{E}[X]| \geq a) \leq \frac{Var(X)}{a^2} \quad a > 0$$

We can have two natural intuitions from this inequality.

- When a gets larger, it is harder for X to take values that are a units away from the mean
- The variance quantifies how dispersed X is around its mean. Thus, it is easier for X to take values that are a units away from the mean if $Var(X)$ is larger.

The Chebyshev Inequality

Apply $p = 2$ to the general inequality and we can get the Chebyshev inequality

$$P(|X - \mathbb{E}[X]| \geq a) \leq \frac{Var(X)}{a^2} \quad a > 0$$

It is useful to express a as $k\sigma$, where $k > 0$ and σ is the standard deviation of X. We can get

$$P(|X - \mathbb{E}[X]| \geq k\sigma) \leq \frac{1}{k^2}$$

The Chebyshev Inequality

$$P(|X - \mathbb{E}[X]| \geq k\sigma) \leq \frac{1}{k^2}$$

- The probability of encountering an observation that is at least k standard deviations away from the mean is bounded above, and the upper bound is inversely proportional to k^2. It is strikingly unlikely to run into values that are several standard deviations away from the mean value.

The Chebyshev Inequality

$$P(|X - \mathbb{E}[X]| \geq k\sigma) \leq \frac{1}{k^2}$$

- Consider there are N i.i.d. samples X_i, $i = 1, 2, ..., N$, of X with mean $\mathbb{E}[X] = \mu$ and variance $Var(X) = \sigma^2$. Let $\overline{X}_N = \frac{X_1 + X_2 + ... + X_N}{N}$.

$$P(|\overline{X}_N - \mathbb{E}[\overline{X}_N]| \geq kStd(\overline{X}_N)) \leq \frac{1}{k^2}$$

$$\Rightarrow P((\overline{X}_N - \mu)^2 \geq k^2 \frac{\sigma^2}{N}) \leq \frac{1}{k^2}$$

$$\Rightarrow P((\overline{X}_N - \mu)^2 \geq k^2 \sigma^2) \leq \frac{1}{k^2 N}$$

When the number of samples N gets larger, it is more likely that the sample mean lies near the true mean.

Table of Contents

For N i.i.d. samples X_i, $i = 1, 2, ..., N$, of X with mean $\mathbb{E}[X] = \mu$ and variance $Var(X) = \sigma^2$. Let $\overline{X}_N = \frac{X_1 + X_2 + ... + X_N}{N}$. WLLN states that $\overline{X}_N \xrightarrow{p} \mu$ when $N \to \infty$, i.e., $\forall \epsilon > 0 \ \lim_{N \to \infty} P(|\overline{X}_N - \mu| < \epsilon) = 1$

Proof.

Apply the Chebyshev inequality on $\overline{X}_N$ and we can get

$$P(|\overline{X}_N - \mu| \geq \epsilon) \leq \frac{\sigma^2 / N}{\epsilon^2}$$

$$\Rightarrow \lim_{N \to \infty} P(|\overline{X}_N - \mu| \geq \epsilon) \leq \lim_{N \to \infty} \frac{\sigma^2}{N \epsilon^2} = 0$$

□

Table of Contents

Limit Supremum / Infimum of A Set

To talk about the limit supremum or infimum of a set, we should introduce the concept of limit points first.

Definition

A point is a limit point or a cluster point or an accumulation point of a set if any neighborhood of that point contains an element of that set other than the point itself.

Limit Supremum / Infimum of A Set

Definition

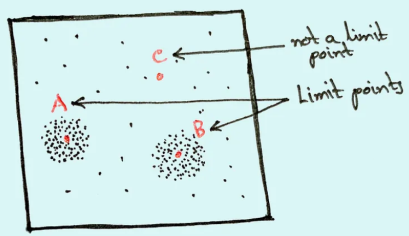
A point is a limit point or a cluster point or an accumulation point of a set if any neighborhood of that point contains an element of that set other than the point itself.

- Within any neighborhood of a limit point, we can get further smaller neighborhoods containing an element of that set other than the limit point. Thus, any neighborhood of a limit point contains infinitely many elements of the set. So, for a set to have limit points, it must have infinitely many elements.

Limit Supremum / Infimum of A Set

Definition

A point is a limit point or a cluster point or an accumulation point of a set if any neighborhood of that point contains an element of that set other than the point itself.



Limit Supremum / Infimum of A Set

Definition

The limit supremum of a set is the supremum of the limit points of the set.

The limit infimum of a set is the infimum of the limit points of the set.

- Limit points are points around which a set accumulates or clusters around. The largest and the smallest of such limit points are the limit supremum and the limit infimum of a set.

Limit Supremum / Infimum of A Sequence of Real Numbers

Definition

The limit supremum / infimum of a sequence of real numbers is the smallest / largest real number that is greater / smaller than or equal to infinitely many members of the sequence.

Comparison:

- Supremum / infimum is the least / greatest of all upper / lower bounds of the whole sequence.

  Limit supremum / limit infimum is the least / greatest of all upper / lower bounds of infinitely many members of the sequence

Limit Supremum / Infimum of A Sequence of Real Numbers

Definition

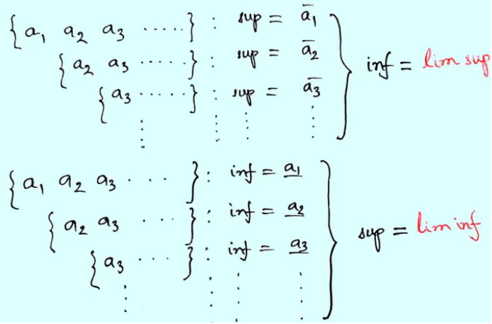
The limit supremum / infimum of a sequence of real numbers is the smallest / largest real number that is greater / smaller than or equal to infinitely many members of the sequence.

How do we translate this literal definition to mathematical definition?

Limit Supremum / Infimum of A Sequence of Real Numbers

For a sequence of real numbers $\{a_n\}$, the limit supremum and the limit infimum are respectively defined as

$$\limsup_{n\to\infty} a_n = \inf_{n>0} \sup_{m\geq n} a_m \qquad \liminf_{n\to\infty} a_n = \sup_{n>0} \inf_{m\geq n} a_m$$



Limit Supremum / Infimum of A Sequence of Real Numbers

For a sequence of real numbers $\{a_n\}$, the limit supremum and the limit infimum are respectively defined as

$$\limsup_{n\to\infty} a_n = \inf_{n>0} \sup_{m\geq n} a_m \quad \liminf_{n\to\infty} a_n = \sup_{n>0} \inf_{m\geq n} a_m$$

- Any tail of the sequence $(\{a_m, m \geq n, n > 0\})$ is infinitely long. Hence, the supremum / infimum of any tail $(\sup_{m\geq n} a_m, n > 0)$ is greater / smaller than or equal to infinitely many members of the sequence. Thus, the supremum / infimum of any tail is a candidate for limit supremum / limit infimum.

Limit Supremum / Infimum of A Sequence of Real Numbers

For a sequence of real numbers $\{a_n\}$, the limit supremum and the limit infimum are respectively defined as

$$\limsup_{n\to\infty} a_n = \inf_{n>0} \sup_{m\geq n} a_m \quad \liminf_{n\to\infty} a_n = \sup_{n>0} \inf_{m\geq n} a_m$$

- The infimum / supremum of all candidates is indeed the smallest / largest, which matches the mathematical definition. However, whether can we construct a sub-sequence with infinitely many elements and different from any tail of the sequence, and supremum / infimum of this sub-sequence is smaller / larger than $\inf_{n>0} \sup_{m\geq n} a_m$ / $\sup_{n>0} \inf_{m\geq n} a_m$?

Limit Supremum / Infimum of A Sequence of Real Numbers

In the following, we discuss the limit supremum since the limit infimum just parallels the discussion. Consider there exists a sub-sequence S with infinitely many elements. The elements of S are all from $\{a_n\}$. Suppose the supremum of S is a_j, which is certainly an element in $\{a_n\}$. Also suppose $\inf_{n>0} \sup_{m\geq n} a_m = a_k$.

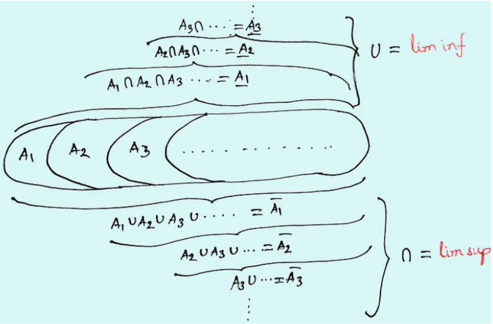
Since $a_k = \inf_{n>0} \sup_{m\geq n} a_m$, $a_k = \sup_{m\geq k} a_m$ and $a_k \leq \inf_{n>k} \sup_{m\geq n} a_m$, which implies $a_k = a_{k+1} = \ldots$ Now since S contains infinitely many elements, S must contain a_ℓ, where $\ell \geq k$. That the supremum of S is a_j implies $a_j \geq a_\ell = a_k$.

Thus, we verify that we cannot construct a sub-sequence S with supremum smaller than $\inf_{n>0} \sup_{m\geq n} a_m$.

Limit Supremum / Infimum of A Sequence of Sets

For a sequence of sets $(A_n)_{n=1}^{\infty}$, the limit supremum and the limit infimum are sets given by

$$\limsup_{n\to\infty} A_n = \bigcap_{n\geq 1} \bigcup_{j\geq n} A_j \qquad \liminf_{n\to\infty} A_n = \bigcup_{n\geq 1} \bigcap_{j\geq n} A_j$$



Limit Supremum / Infimum of A Sequence of Sets

For a sequence of sets $(A_n)_{n=1}^{\infty}$, the limit supremum and the limit infimum are sets given by

$$\limsup_{n\to\infty} A_n = \bigcap_{n\geq 1} \bigcup_{j\geq n} A_j \quad \liminf_{n\to\infty} A_n = \bigcup_{n\geq 1} \bigcap_{j\geq n} A_j$$

- The limit supremum embodies an eliminative process. The initial set (the first union with n=1) has already contained all the elements of the limit supremum. During the process, finitely-occurring elements are successively eliminated (by taking intersections). When $n \to \infty$, what are left are elements that occur infinitely often (i.o) and those elements compose the limit supremum.

Limit Supremum / Infimum of A Sequence of Sets

For a sequence of sets $(A_n)_{n=1}^{\infty}$, the limit supremum and the limit infimum are sets given by

$$\limsup_{n\to\infty} A_n = \bigcap_{n\geq 1} \bigcup_{j\geq n} A_j \quad \liminf_{n\to\infty} A_n = \bigcup_{n\geq 1} \bigcap_{j\geq n} A_j$$

- The limit infimum embodies a constructive process whereby the limit infimum is constructed by successively adding (through union) infinitely-occurring elements (intersection of an infinite number of sets).

Limit Supremum / Infimum of A Sequence of Sets

For a sequence of sets $(A_n)_{n=1}^{\infty}$, the limit supremum and the limit infimum are sets given by

$$\limsup_{n\to\infty} A_n = \bigcap_{n\geq 1} \bigcup_{j\geq n} A_j \quad \liminf_{n\to\infty} A_n = \bigcup_{n\geq 1} \bigcap_{j\geq n} A_j$$

Comparison: The elements of the limit supremum and the limit infimum both occur in infinitely many sets. For the limit supremum, the elements are not necessary to be in consecutive sets. That is, an element x can be in A_j and A_k but not in A_ℓ, where $j < \ell < k$. However, for the limit infimum, once an element x is in A_j, x must also be in $A_{j+1}, A_{j+2}, \ldots$

Limit Supremum / Infimum of A Sequence of Sets

For a sequence of sets $(A_n)_{n=1}^{\infty}$, the limit supremum and the limit infimum are sets given by

$$\limsup_{n\to\infty} A_n = \bigcap_{n\geq 1} \bigcup_{j\geq n} A_j \quad \liminf_{n\to\infty} A_n = \bigcup_{n\geq 1} \bigcap_{j\geq n} A_j$$

In the probability theory, a sequence of events is a sequence of sets and each element of an event is an outcome.

- The limit supremum of a sequence of events is the collection of all outcomes that occur in the sequence infinitely often (i.o), i.e., that never leave the sequence for good.
- The limit infimum of a sequence of events is the collection of all outcomes that occur in the sequence eventually, i.e., that from some point onward stays in the sequence for good.

Limit Supremum / Infimum of A Function

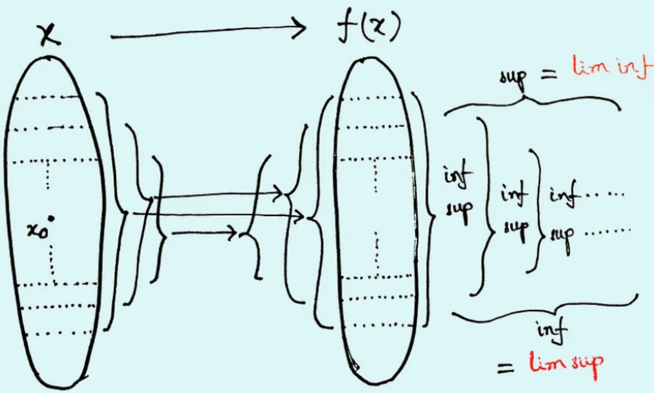
The limit supremum and limit infimum of a function, $f(x)$, at a given point x_0 are defined as

$$\limsup_{x \to x_0} f(x) = \inf_{\epsilon > 0} \sup\{f(x) : x \in (x_0 - \epsilon, x_0 + \epsilon) \backslash x_0\}$$

$$\liminf_{x \to x_0} f(x) = \sup_{\epsilon > 0} \inf\{f(x) : x \in (x_0 - \epsilon, x_0 + \epsilon) \backslash x_0\}$$

- The limit supremum / infimum of a function at a given point is the limit (infimum / supremum) of the supremums / infimums of the function over successively decreasing neighborhoods of the point excluding the point itself.

Limit Supremum / Infimum of A Function



Limit Supremum / Infimum of A Sequence of Functions

The (pointwise) limit supremum and limit infimum of a sequence of functions, $(f_n(.))_1^\infty$ are functions defined as

$$\bar{f}(x) = \limsup_{n\to\infty}\{f_n(x)\} \quad \forall x$$

$$\underline{f}(x) = \liminf_{n\to\infty}\{f_n(x)\} \quad \forall x$$

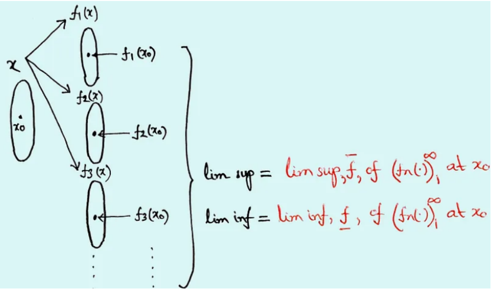


Table of Contents

The First Borel-Cantelli Lemma

Let $\{E_n\}$ be a sequence of events in some probability space. If the sum of the probabilities of all events is finite, i.e., $\sum\limits_{n=1}^{\infty} P(E_n) < \infty$, then the probability that an outcome occurs infinitely many times is 0, i.e. $P(\limsup\limits_{n\to\infty} E_n) = 0$

Proof.

$$
\begin{aligned}
P(\limsup_{n\to\infty} E_n) = P(\bigcap_{n\geq 1}\bigcup_{k\geq n} E_k) \leq \inf_{n\geq 1} P(\bigcup_{k\geq n} E_k) &\leq \lim_{n\to\infty} P(\bigcup_{k\geq n} E_k) \\
&\leq \lim_{n\to\infty} \sum_{k=n}^{\infty} P(E_k)
\end{aligned}
$$

$$
\because \sum_{n=1}^{\infty} P(E_n) < \infty \therefore \lim_{n\to\infty} \sum_{k=n}^{\infty} P(E_k) = 0 \qquad \square
$$

The Second Borel-Cantelli Lemma

If $\sum_{n\geq 1} P(E_n) = \infty$ and the events $(E_n)_{n=1}^{\infty}$ are independent, then $P(\limsup_{n\to\infty} E_n) = 1$

Proof.

It is sufficient to prove that $P((\bigcap_{n\geq 1}\bigcup_{k\geq n} E_k)^c) = 0$

$$P((\bigcap_{n\geq 1}\bigcup_{k\geq n} E_k)^c) = P(\bigcup_{n\geq 1}\bigcap_{k\geq n} E_k^c) \leq \sum_{n\geq 1} P(\bigcap_{k\geq n} E_k^c)$$

$\because$ $(E_n)_{n=1}^{\infty}$ are independent $\therefore$ $P(\bigcap_{k\geq n} E_k^c) = \prod_{k\geq n}(1-P(E_k))$

$\because 1 - x \leq e^{-x}, x \geq 0$

$$\therefore \prod_{k\geq n}(1-P(E_k)) \leq \prod_{k\geq n} e^{-P(E_k)} = e^{-\sum_{k\geq n} P(E_k)} = e^{-\infty} = 0 \ \forall n \qquad \square$$

The Second Borel-Cantelli Lemma

A famous application is the infinite monkey theorem: a monkey hitting keys at random on a typewriter keyboard for an infinite amount of time will almost surely type any string, including the complete works of William Shakespeare.

Proof.

Divide the random infinite string, typed by the monkey, into non-overlapping blocks whose size match that of the desired string. Let E_n denote the event where the n^{th} block equals the desired string. $P(E_n) \ll 1$ but is non-zero. Thus $\sum_{n=1}^{\infty} P(E_n) = \infty$. Furthermore, $(E_n)_{n=1}^{\infty}$ are independent since the monkey hits keys at random. Applying the second Borel-Cantelli lemma can prove this theorem. $\square$

Table of Contents

For N i.i.d. samples X_i, $i = 1, 2, ..., N$ of X with mean $\mathbb{E}[X] = \mu$ and variance $Var(X) = \sigma^2$. Let $\overline{X}_N = \frac{X_1 + X_2 + ... + X_N}{N}$. SLLN states that $\overline{X}_N \overset{a.s.}{\to} \mu$ when $N \to \infty$, i.e., $P(\lim_{N \to \infty} \overline{X}_N = \mu) = 1$

We prove under the additional constraint that $\sigma^2 = \mathbb{E}[X^2] < \infty$ and $\mathbb{E}[X^4] < \infty$. Without loss of generality, we assume $\mu = 0$.

$P(\lim_{N\to\infty} \overline{X}_N = 0) = 1$ is equivalent to $P(\lim_{N\to\infty} \overline{X}_N \neq 0) = 0$.

$\lim_{N\to\infty} \overline{X}_N \neq 0$ means that $\exists \epsilon > 0$ such that for infinitely many N $|\overline{X}_N| \geq \epsilon$, i.e. $|S_N| \geq N\epsilon$, where $S_N = \sum_{i=1}^{N} X_i$. Thus, it is sufficient to prove that $\exists \epsilon > 0$ such that $P(|S_N| \geq N\epsilon \ i.o.) = 0$. Indeed, we will prove a stricter version -

$$\forall \epsilon > 0, P(|S_N| \geq N\epsilon \ i.o.) = 0$$

For each N, we want to bound $P(|S_N| \geq N\epsilon)$. Apply $p = 4$ to the general inequality, we can get

$$P(|S_N| \geq N\epsilon) \leq \frac{\mathbb{E}[S_N^4]}{N^4\epsilon^4}$$

$\mathbb{E}[S_N^4] = \mathbb{E}[\sum_{1 \le i,j,k,\ell \le N} X_i X_j X_k X_\ell]$. When the sums are multiplied out there will be terms of the form: $\mathbb{E}[X_i^3 X_j]$, $\mathbb{E}[X_i^2 X_j X_k]$, $\mathbb{E}[X_i X_j X_k X_\ell]$, $\mathbb{E}[X_i^4]$ and $\mathbb{E}[X_i^2 X_j^2] = (\mathbb{E}[X^2])^2$. The first three terms are all zero since $\mathbb{E}[X_i] = 0$ and the random variables are independent. The latter two terms are non-zero. There are N terms of the form $\mathbb{E}[X_i^4]$ and $\binom{N}{2}\binom{4}{2} = 3N(N-1)$ terms of the form $\mathbb{E}[X_i^2 X_j^2]$. Thus

$$
\begin{aligned}
\mathbb{E}[S_N^4] &= N\mathbb{E}[X^4] + 3N(N-1)\sigma^4 \\
&= 3\sigma^4 N^2 + (\mathbb{E}[X^4] - 3\sigma^4)N \\
&\le CN^2 \quad \textit{for sufficiently large } N \ (\textit{say } N_0) \textit{ and } C \\
&\qquad\qquad \textit{can be chosen to be } 3\sigma^4 + 1
\end{aligned}
$$

Hence, up to now we have proved that

$$P(|S_N| \geq N\epsilon) \leq \frac{\mathbb{E}[S_N^4]}{N^4\epsilon^4} \leq \frac{C}{N^2\epsilon^4} \quad \forall N \geq N_0$$

Thus

$$\begin{aligned}\sum_{N=1}^{\infty} P(|S_N| \geq N\epsilon) &\leq (N_0 - 1) + \sum_{N \geq N_0} P(|S_N| \geq N\epsilon) \\ &\leq (N_0 - 1) + \sum_{N \geq N_0} \frac{C}{N^2\epsilon^4} < \infty\end{aligned}$$

By the first Borel-Cantelli lemma, we can derive that

$$\forall \epsilon > 0, P(|S_N| \geq N\epsilon \ i.o.) = 0$$

which proves the SLLN under the two additional constraints.

Table of Contents

For N i.i.d. samples X_i, $i = 1, 2, ..., N$ of X with mean $\mathbb{E}[X] = \mu$ and variance $Var(X) = \sigma^2 < \infty$. Let $\overline{X}_N = \frac{X_1 + X_2 + ... + X_N}{N}$. CLT states that

$$\sqrt{N}\left(\frac{\overline{X}_N - \mu}{\sigma}\right) \xrightarrow{d} Normal(0, 1)$$

Let $Y_N = \sqrt{N}(\frac{\overline{X}_N - \mu}{\sigma}) = \sum_{i=1}^{N}(\frac{X_i - \mu}{\sqrt{N}\sigma})$. The moment generating function $M_{Y_N}(t)$ is

$$\begin{aligned}
M_{Y_N}(t) = \mathbb{E}[e^{Y_N t}] &= \mathbb{E}[e^{(\sum_{i=1}^{N}(\frac{X_i-\mu}{\sqrt{N}\sigma}))t}] \\
&= \mathbb{E}[\prod_{i=1}^{N} e^{(\frac{X_i-\mu}{\sqrt{N}\sigma})t}] \\
&= \prod_{i=1}^{N} \mathbb{E}[e^{(\frac{X_i-\mu}{\sqrt{N}\sigma})t}] \quad \because \textit{ } X_i' s \textit{ are independent} \\
&= (\mathbb{E}[e^{(\frac{X-\mu}{\sqrt{N}\sigma}t)}])^N \quad \because \textit{ } X_i' s \textit{ are identically distributed} \\
&= (M_{X-\mu}(\frac{t}{\sqrt{N}\sigma}))^N
\end{aligned}$$

$$
\begin{aligned}
M_{X-\mu}\left(\frac{t}{\sqrt{N}\sigma}\right) &= M_{X-\mu}\left(\frac{t}{\sqrt{N}\sigma}=0\right) + M'_{X-\mu}\left(\frac{t}{\sqrt{N}\sigma}=0\right)\frac{t}{\sqrt{N}\sigma} \\
&\quad + \frac{M''_{X-\mu}\left(\frac{t}{\sqrt{N}\sigma}=0\right)}{2}\left(\frac{t}{\sqrt{N}\sigma}\right)^2 \\
&\quad + \frac{M'''_{X-\mu}\left(\frac{t}{\sqrt{N}\sigma}=0\right)}{3!}\left(\frac{t}{\sqrt{N}\sigma}\right)^3 + \ldots \\
&= 1 + 0 * \frac{t}{\sqrt{N}\sigma} + \frac{\sigma^2}{2}\left(\frac{t}{\sqrt{N}\sigma}\right)^2 + O\left(\frac{\frac{1}{\sqrt{N}}t^3}{N}\right) + O\left(\frac{\frac{1}{N}t^4}{N}\right) + \ldots \\
&= 1 + \frac{t^2 + O\left(\frac{1}{\sqrt{N}}t^3\right) + O\left(\frac{1}{N}t^4\right) + \ldots}{2N}
\end{aligned}
$$

$\star$ If $\lim_{N\to\infty} a_N = a$, then $\lim_{N\to\infty} (1 + \frac{a_N}{N})^N = e^a$

$\therefore \lim_{N\to\infty} M_{Y_N}(t) = e^{\frac{1}{2}t^2}$, which is the moment generating function of Normal(0,1)

$\because$ Pointwise convergence of CDFs implies convergence in distribution of random variables

$\because$ Pointwise convergence of moment generating functions also implies convergence in distribution of random variables

$\because Y_N = \sqrt{N}(\frac{\overline{X}-\mu}{\sigma}) \xrightarrow{d}$ Normal(0,1)

Two Problems about The CLT

One key assumption of the CLT is that each member of a sample should be independent. For a time-series data $\{x_t\}$, what if x_i and x_j are dependent for some i and j?

- We consider a particular kind of dependence: M-dependence.
  - A time-series x_t is M-dependent if the set of values x_s, $s \leq t$ is independent of the set of values x_s, $s \geq t + M + 1$
    $\Rightarrow$ time points separated by more than M units are independent.

Two Problems about The CLT

Theorem (M-Dependent Central Limit Theorem)

If x_t is a strictly stationary M-dependent sequence of random variables with mean zero and autocovariance function $\gamma(.)$ and if

$$Vm = \sum_{u=-M}^{M} \gamma(u),$$

where $V_m \neq 0$,

$$\frac{1}{n}\sum_{i=1}^{n} x_i \xrightarrow{d} Normal(0, Vm/n)$$

- This theorem can be proved using the CLT and the Basic Approximation Theorem.

Two Problems about The CLT

The other key assumption of the CLT is that X should have finite mean and variance. What if this is not the case?

$\rightarrow$ e.g. The Cauchy random variable: $f(x) = \frac{1}{\pi}\frac{1}{1+(x-\mu)^2}$

$\rightarrow$ the mean and variance are both undefined.

- Many financial models assume that the price changes are drawn from the Cauchy distribution.
- The median of the Cauchy random variable exists and is μ.
  $\rightarrow$ consider the median instead?
- There are situations where the sample median converges to a nice distribution while the sample mean does not.
  $\rightarrow$ the Median theorem

Table of Contents

Definition

The median $\tilde{\mu}$ of a random variable X are all points x such that

$$Pr(X \leq x) \geq \frac{1}{2} \quad and \quad Pr(X \geq x) \geq \frac{1}{2}$$

- The median may not be unique but an interval of values.
- The median may be unique.
  - For continuous distributions, as long as the density never vanishes except on an interval of the form $(-\infty, a)$, $(-\infty, a]$, $[b, \infty)$, or (b, ∞), there is a unique median.
- Let X be a random variable with density p that is symmetric about its mean μ. Then the median $\tilde{\mu}$ equals the mean μ. $\rightarrow$ we can get the estimation for the mean once we can estimate the median.

Table of Contents

Suppose that the random variables $X_1, X_2, ..., X_n$ form a sample of size n from an infinite population with the continuous density. It is often useful to reorder them from the smallest to the largest and produce $Y_1, Y_2, ... Y_n$.

- $Y_1 = \min_i(X_1, ..., X_n)$
- $Y_n = \max_i(X_1, ..., X_n)$
- Y_r is called the r^{th} order statistic of the sample.
- If n is odd, the median is $Y_{(n+1)/2}$, while if n is even, the median is any value in $[Y_{n/2}, Y_{n/2+1}]$.

Theorem

For a random sample of size n from an infinite population having values x and continuous density $f(x)$, the probability density of the r^{th} order statistic Y_r is given by

$$g_r(y_r) = \frac{n!}{(r-1)!(n-r)!}\left[\int_{-\infty}^{y_r} f(x)dx\right]^{r-1} f(y_r)\left[\int_{y_r}^{\infty} f(x)dx\right]^{n-r}$$

- If $X_1, X_2, ..., X_n$ are i.i.d. uniform random variables on [0,1], then
  - $g_1(y_1) = n(1-y_1)^{n-1} \Rightarrow \mathbb{E}[g_1(y_1)] = \frac{1}{n+1}$
  - $g_n(y_n) = ny_n^n \Rightarrow \mathbb{E}[g_n(y_n)] = \frac{n}{n+1}$

Table of Contents

Theorem (The Median Theorem)

Let a sample of size $n = 2m + 1$ with n large be taken from an infinite population with a density function $f(x)$ that is nonzero at the population median $\tilde{\mu}$ and continuously differentiable in a neighborhood of $\tilde{\mu}$. The sample distribution of the median is approximately normal with mean $\tilde{\mu}$ and variance $\frac{1}{8mf(\tilde{\mu})^2}$.

1. Let the median random variable $\tilde{X}$ have values $\tilde{x}$ and density $g(\tilde{x})$. The median is the $(m+1)^{th}$ statistic, so

$$g(\tilde{x}) = \frac{(2m+1)!}{m!m!}\left[F(\tilde{x})\right]^m f(\tilde{x})\left[(1-F(\tilde{x}))\right]^m$$

2. Using the Stirling's formula: $n! = (\frac{n}{e})^n\sqrt{2\pi n}(1+O(n^{-1}))$

$$g(\tilde{x}) \approx \frac{(2m+1)4^m}{\sqrt{\pi m}}\left[F(\tilde{x})\right]^m f(\tilde{x})\left[(1-F(\tilde{x}))\right]^m$$

3. Taylor's series expansion of $F(\tilde{x})$ about $\tilde{\mu}$

$$\begin{aligned}F(\tilde{x}) &= F(\tilde{\mu}) + F'(\tilde{\mu})(\tilde{x}-\tilde{\mu}) + \frac{F''(\tilde{\mu})}{2}(\tilde{x}-\tilde{\mu})^2 + O((\tilde{x}-\tilde{\mu})^3)\\ &= \frac{1}{2} + f(\tilde{\mu})(\tilde{x}-\tilde{\mu}) + \frac{f'(\tilde{\mu})}{2}(\tilde{x}-\tilde{\mu})^2 + O((\tilde{x}-\tilde{\mu})^3)\end{aligned}$$

   This expansion is useful only when $(\tilde{x}-\tilde{\mu})$ is small; that is, we need $\lim_{m\to\infty}|\tilde{x}-\tilde{\mu}| = 0$ (see appendix)

4. Plug the Taylor's series expansion into $g(\tilde{x})$ and let $t = \tilde{x} - \tilde{\mu}$

$$
\begin{aligned}
g(\tilde{x}) &\approx \frac{(2m+1)4^m}{\sqrt{\pi m}} f(\tilde{x}) \left[\frac{1}{4} - (f(\tilde{\mu})t)^2 + O(t^3)\right]^m \\
&= \frac{(2m+1)f(\tilde{x})}{\sqrt{\pi m}} \left[1 - 4(f(\tilde{\mu})t)^2 + O(t^3)\right]^m
\end{aligned}
$$

5. Simplify the term $\left[1 - 4(f(\tilde{\mu})t)^2 + O(t^3)\right]^m$

$$
\left[1 - 4m(f(\tilde{\mu})t)^2 + O(t^3)\right]^m = \exp(m \log(1 - 4(f(\tilde{\mu})t)^2) + O(t^3))
$$

Use the Taylor series expansion: $\log(1-x) = -x + O(x^2)$.
Hence,

$$
\begin{aligned}
\left[1 - 4(f(\tilde{\mu})t)^2 + O(t^3)\right]^m &= \exp(-m * 4(f(\tilde{\mu})t)^2 + O(mt^3)) \\
&= \exp(-\frac{(\tilde{x} - \tilde{\mu})^2}{1/(4m(f(\tilde{\mu})^2))}) * \exp(O(mt^3))
\end{aligned}
$$

5.

$$\left[1-4(f(\tilde{\mu})t)^2+O(t^3)\right]^m \approx \exp(-\frac{(\tilde{x}-\tilde{\mu})^2}{1/(4m(f(\tilde{\mu}))^2)})$$

$\because mt^3 \to 0$ as $m \to \infty$ (see appendix)

6.

$$g(\tilde{x}) \approx \frac{(2m+1)f(\tilde{x})}{\sqrt{\pi m}} \exp(-\frac{(\tilde{x}-\tilde{\mu})^2}{1/(4m(f(\tilde{\mu}))^2)})$$

Since $f(x)$ is continuously differentiable in a neighborhood of $\tilde{\mu}$, using the mean value theorem: $\frac{f(\tilde{x})-f(\tilde{\mu})}{\tilde{x}-\tilde{\mu}} = f'(c_{\tilde{x},\tilde{\mu}})$, where $c_{\tilde{x},\tilde{\mu}}$ is some constant between $\tilde{x}$ and $\tilde{\mu}$. Thus,

$$\begin{aligned} g(\tilde{x}) &\approx \frac{(2m+1)(f(\tilde{\mu})+f'(c_{\tilde{x},\tilde{\mu}})(\tilde{x}-\tilde{\mu}))}{\sqrt{\pi m}} \exp(-\frac{(\tilde{x}-\tilde{\mu})^2}{1/(4m(f(\tilde{\mu}))^2)}) \\ &\approx \frac{(2m+1)f(\tilde{\mu})}{\sqrt{\pi m}} \exp(-\frac{(\tilde{x}-\tilde{\mu})^2}{1/(4m(f(\tilde{\mu}))^2)}) \end{aligned}$$

$\because \lim_{m\to\infty} |\tilde{x}-\tilde{\mu}| = 0$ (see appendix)

7.

$$g(\tilde{x}) \approx \frac{(2m+1)f(\tilde{\mu})}{\sqrt{\pi m}} \exp(-\frac{(\tilde{x}-\tilde{\mu})^2}{1/(4m(f(\tilde{\mu}))^2)})$$

Let $2\sigma^2 = \frac{1}{4m(f(\tilde{\mu}))^2}$

$$\begin{aligned} g(\tilde{x}) &\approx \frac{(2m+1)f(\tilde{\mu})}{\sqrt{\pi m}} \frac{\sqrt{\pi}}{\sqrt{4m(f(\tilde{\mu}))^2}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(\tilde{x}-\tilde{\mu})^2}{2\sigma^2}) \\ &= \frac{2m+1}{2m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(\tilde{x}-\tilde{\mu})^2}{2\sigma^2}) \\ &\approx \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(\tilde{x}-\tilde{\mu})^2}{2\sigma^2}) \quad \text{as } m \to \infty \end{aligned}$$

Q.E.D

Appendix

Lemma

Suppose f(x) is continuously differentiable in some neighborhood of $\tilde{\mu}$. Then for any $c > 0$, we have

$$\lim_{m\to\infty} P(|\tilde{X} - \tilde{\mu}| \geq c) = 0$$

That is, $\tilde{X}$ converges in probability to $\tilde{\mu}$ ($\tilde{X} \xrightarrow{p} \tilde{\mu}$)

This is equivalent to proving that $\lim_{m\to\infty} P(\tilde{X} \leq \tilde{\mu} - c) = 0$ and $\lim_{m\to\infty} P(\tilde{X} \geq \tilde{\mu} + c) = 0$. We focus on the former because the proof of the latter is similar.

1.
$$g(\tilde{x}) \approx \frac{(2m+1)4^m f(\tilde{x})}{\sqrt{\pi m}} F(\tilde{x})^m (1 - F(\tilde{x}))^m$$

Consider the function $h(\theta) = \theta(1-\theta)$

- $h(\theta)$ is max when $\theta = 1/2$
  - $\max h(F(\tilde{x})) = 1/2$ holds only when $\tilde{x} = \tilde{\mu}$
- $h(\theta)$ is increasing for $\theta < 1/2$
  - $\because$ $F(\tilde{x})$ is also increasing, $h(F(\tilde{x}))$ increases as $\tilde{x}$ increases.

$$\begin{aligned} F(\tilde{x})^m (1 - F(\tilde{x}))^m &\leq F(\tilde{\mu} - c)^m (1 - F(\tilde{\mu} - c))^m \\ &< F(\tilde{\mu})^m (1 - F(\tilde{\mu}))^m = (\frac{1}{4})^m \end{aligned}$$

Let $\frac{\alpha}{4} = F(\tilde{\mu} - c)(1 - F(\tilde{\mu} - c))$. $(h(F(\tilde{x})))^m \leq (\frac{\alpha}{4})^m < (\frac{1}{4})^m$

2.

$$
\begin{aligned}
P(\tilde{X} \leq \tilde{\mu} - c) &= \int_{-\infty}^{\tilde{\mu}-c} g(\tilde{x}) d\tilde{x} \\
&\approx \int_{-\infty}^{\tilde{\mu}-c} \frac{(2m+1)4^m}{\sqrt{\pi m}} f(\tilde{x}) F(\tilde{x})^m (1 - F(\tilde{x}))^m d\tilde{x} \\
&< \frac{(2m)4^m}{\sqrt{m}} \int_{-\infty}^{\tilde{\mu}-c} f(\tilde{x}) F(\tilde{x})^m (1 - F(\tilde{x}))^m d\tilde{x} \\
&\quad \because\ 2m > (2m+1)/\sqrt{\pi} \textit{ for } m \textit{ sufficiently large} \\
&= \frac{(2m)4^m}{\sqrt{m}} \left(\frac{\alpha}{4}\right)^m \int_{-\infty}^{\tilde{\mu}-c} f(\tilde{x}) d\tilde{x} \\
&< 2\alpha^m \sqrt{m} \int_{-\infty}^{\tilde{\mu}} f(\tilde{x}) d\tilde{x} \\
&= \alpha^m \sqrt{m} \to 0 \textit{ as } m \to \infty \ \because\ \alpha < 1
\end{aligned}
$$

If we want to argue that $\lim_{m \to \infty} P(m^a(|\tilde{X} - \tilde{\mu}|)^b \geq c) = 0$ for $a > 0$, $b > 0$, we have to deal with $\lim_{m \to \infty} P(\tilde{X} \leq \tilde{\mu} - (\frac{c}{m^a})^{1/b})$ and $\lim_{m \to \infty} P(\tilde{X} \geq \tilde{\mu} + (\frac{c}{m^a})^{1/b})$. Because $\tilde{\mu} - (\frac{c}{m^a})^{1b}$ is also smaller than $\tilde{\mu}$, we can come up with the same result by following the same proving argument. Specifically, we obtain

$$\lim_{m \to \infty} P(m|\tilde{X} - \tilde{\mu}|^3 \geq c) = 0$$

Comparison with The CLT

Consider the normal distribution with mean μ and variance σ^2.

- The median theorem:
  - median $\tilde{\mu}$ is also μ
  - variance is $\frac{1}{(8mf(\tilde{\mu})^2} = \frac{1}{8mf(\mu)^2} = \frac{\pi\sigma^2}{4m}$

  Hence, the sample median is asymptotically normal with mean μ and variance $\frac{\pi\sigma^2}{4m}$
- The CLT:
  - mean is μ
  - variance is $\frac{\sigma^2}{n} = \frac{\sigma^2}{2m+1}$

  Hence, the sample mean is normal with mean μ and variance $\frac{\sigma^2}{2m+1}$

Comparison with The CLT

$$\frac{\textit{variance of sample median}}{\textit{variance of sample mean}} = \frac{\pi\sigma^2/4m}{\sigma^2/(2m+1)} = \frac{\pi}{2}\frac{2m+1}{2m}$$

For large m, this ratio is $\frac{\pi}{2} \approx 1.57$

$\Rightarrow$ Although the sample median and sample mean have the same expected value, the sample median has larger fluctuations

$\Rightarrow$ The median value theorem cannot replace the CLT whenever the mean and the median are equal. The CLT can give better results. The median theorem is useful when the CLT does not work.

Table of Contents

Reference

- [數學分析] 逐點收斂與均勻收斂 (1)- sequence of functions https://ch-hsieh.blogspot.com/2014/04/pointwise-convergence-uniformly.html#google_vignette
- Convergence of random variables https://en.wikipedia.org/wiki/Convergence_of_random_variables
- The Markov and the Bienaymé–Chebyshev Inequalities https://towardsdatascience.com/the-markov-and-the-bienaym%C3%A9-chebyshev-inequalities-cbf7ccc856f9
- Intuitions: limit supremum and limit infimum of sets, sequences and functions https://nilotpalsanyal.medium.com/intuitions-limit-supremum-and-limit-infimum-of-sets-sequences-functions-a24b4f2e0154

Reference

- Borel–Cantelli lemma
  https://en.wikipedia.org/wiki/Borel%E2%80%93Cantelli_lemma
- Infinite monkey theorem
  https://en.wikipedia.org/wiki/Infinite_monkey_theorem
- Laws of Large Numbers
  https://www.math.ucdavis.edu/ tracy/courses/math135A/UsefullCourseMaterial/lawLargeNo.pdf
- R. H. Shumway, D. S. Stoffer, and D. S. Stoffer, Time series analysis and its applications. Springer, 2000, vol. 3.
- S. J. Miller, "The probability lifesaver: Order statistics and the median theorem," 2017.