

SDS 384 11: Theoretical Statistics

Lecture 2: Stochastic Convergence

Purnamrita Sarkar
Department of Statistics and Data Science
The University of Texas at Austin
<https://psarkar.github.io/teaching>

Convergence of expectations: exchanging limit and integral

Theorem (Monotone convergence theorem)

If $0 \leq X_1 \leq X_2 \leq \dots \leq X_n \uparrow X$, a.s. then

$$E[X_n] \rightarrow E[X]$$

Lemma (Fatou's lemma)

If $X_n \geq Y \forall n$ for some random variable Y with $E|Y| < \infty$ then

$$\liminf_{n \rightarrow \infty} E[X_n] \geq E[\liminf_n X_n]$$

Theorem (Dominated convergence theorem)

If $X_n \xrightarrow{\text{a.s.}} X$ and $|X_n| \leq Y$ with $E[|Y|] < \infty$, then

$$E[X_n] \rightarrow E[X]$$

Convergence of expectations: exchanging limit and integral

- Another version of MCT requires $X_i > Y$ s.t. $EY > -\infty$.
- Consider $Z \sim U([0, 1])$
- $X_n = -\frac{1}{Z}1[z \in (0, 1/n)]$
- This is an increasing sequence, $X_n \xrightarrow{a.s.} 0$
- But $EX_n = -\int_0^{1/n} 1/z dz = -\infty$

Remember liminf and limsup

- $\liminf_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \inf_{m \geq n} a_m$
- $\limsup_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \sup_{m \geq n} a_m$

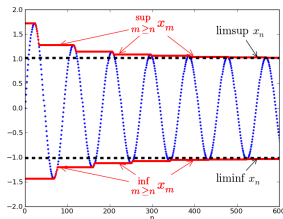


Figure 1: limsup and liminf always exist even though the sequence x_n is not converging. Courtest: wikipedia.

https://en.wikipedia.org/wiki/Limit_inferior_and_limit_superior

Proof.

- Consider the random variable $X_n - Y$. These are positive. For all $m \geq n$,

$$\inf_{k \geq n} (X_k - Y) \leq X_m - Y$$

$$E \left[\inf_{k \geq n} (X_k - Y) \right] \leq E[X_m - Y] \quad \text{Take } E[\cdot] \text{ of both sides}$$

$$\leq \inf_{m \geq n} E[X_m - Y]$$

$$\lim_{n \rightarrow \infty} E \left[\inf_{k \geq n} (X_k - Y) \right] \leq \lim_{n \rightarrow \infty} \inf_{m \geq n} E[X_m - Y] = \liminf_{n \rightarrow \infty} E[X_n - Y]$$

- All that is left, is to exchange limit and integral on LHS.
- Note that $\inf_{k \geq n} (X_k - Y)$ is an increasing positive sequence. This converges to $\liminf_{n \rightarrow \infty} (X_n - Y)$. Apply MCT.



We never used $E|Y| < \infty$

- Well, we are saying $E[Y]$ exists.
- Unless $E|Y| < \infty$, $E[Y]$ is not very well defined.
- Consider $1 - 1 + 1/2 - 1/2 + 1/3 - 1/3 + 1/4 - 1/4 + 1/5 - 1/5 \dots$
- This is clearly zero, right?
- But what if I permute it to have first two +ve and first -ve
- $1 + 1/2 - 1 + 1/3 + 1/4 - 1/2 + 1/5 + 1/6 - 1/3 \dots$ —this is $\ln 2$
- If you take a sum of the absolute values then that diverges.

Proof.

- Note that random variable $-Y \leq X_n \leq Y$.
- Also note that $E[\liminf_{n \rightarrow \infty} X_n] = E[X] = E[\limsup_{n \rightarrow \infty} X_n]$.
- Apply Fatou on X_n and $-X_n$.

$$E[\liminf_{n \rightarrow \infty} X_n] \leq \liminf_{n \rightarrow \infty} E[X_n] \leq \limsup_{n \rightarrow \infty} E[X_n] \leq E[\limsup_{n \rightarrow \infty} X_n]$$

- But both ends equal $E[X]$ and so the middle two quantities must be equal and hence proved.



Things you should know

Consider n i.i.d. random variables $X_i \sim F$.

Definition (Empirical distribution function)

The empirical distribution function is defined as:

$$F_n(x) = \frac{1}{n} \sum_i 1(X_i \leq x).$$

Theorem (Glivenko-Cantelli)

The random variable $\sup_x |F_n(x) - F(x)|$ almost surely converges to zero.

$$P\left(\sup_x |F_n(x) - F(x)| \rightarrow 0\right) = 1$$

Things you should know

Let X_1, \dots, X_n be i.i.d random variables with $E[|X_1|] \leq \infty$, mean μ .

Theorem (Weak law of large numbers)

$$\bar{X}_n \xrightarrow{P} \mu$$

Theorem (Strong law of large numbers)

$$\bar{X}_n \xrightarrow{a.s.} \mu$$

Theorem (Central limit theorem)

If $E[X_i^2] = \sigma^2$, $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$.

Things you should know

Let X_1, \dots, X_n be i.i.d random variables with mean μ .

Theorem (Berry Esseen)

If $E[X_i^2] = \sigma^2$, and $E[|X_i|^3] = \rho < \infty$,

$$\sup_x \left| P \left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq x \right) - \Phi(x) \right| \leq \frac{C\rho}{\sigma^3\sqrt{n}} \quad \forall n,$$

where $\Phi(x)$ is the CDF of the standard normal and c is an universal constant known to be greater than 0.4097 and less than 0.7975.

Lindeberg-feller CLT for triangular arrays

X_{11}

X_{21}, X_{22}

X_{31}, X_{32}, X_{33}

...

Theorem

For each n let $(X_{ni})_{i=1}^n$ be independent random variables with mean zero

and variance σ_{ni}^2 . Let $Z_n = \sum_{i=1}^n X_{ni}$ and $B_n^2 = \text{var}(Z_n)$. Then

$Z_n/B_n \xrightarrow{d} N(0,1)$, as long as the **Lindeberg condition** holds.

The Lindeberg condition

Definition (Lindeberg condition)

For every $\epsilon > 0$,

$$\frac{1}{B_n^2} \sum_{j=1}^n E[X_{nj}^2 1(|X_{nj}| \geq \epsilon B_n)] \rightarrow 0 \text{ as } n \rightarrow \infty \quad (1)$$

Converse: If $\frac{\sigma_{nj}^2}{B_n^2} \rightarrow 0$ as $n \rightarrow \infty$, i.e. no one variance plays a significant role in the limit, and if $Z_n/B_n \xrightarrow{d} N(0,1)$, then the Lindeberg condition holds.

Necessary and Sufficient: If $\frac{\sigma_{nj}^2}{B_n^2} \rightarrow 0$, the the Lindeberg condition is necessary and sufficient to show the CLT.

Example

Let X_1, \dots, X_n be independent random variables with mean zero and variance one. Do you think $\sqrt{n}\bar{X}_n \xrightarrow{d} N(0, 1)$?

Example

Let X_1, \dots, X_n be independent random variables with mean zero and variance one. Do you think $\sqrt{n}\bar{X}_n \xrightarrow{d} N(0, 1)$?

•

$$X_{nj} = \begin{cases} 2j & \text{w.p. } \frac{1}{8j^2} \\ 0 & \text{w.p. } 1 - \frac{1}{4j^2} \\ -2j & \text{w.p. } \frac{1}{8j^2} \end{cases}$$

Example

Let X_1, \dots, X_n be independent random variables with mean zero and variance one. Do you think $\sqrt{n}\bar{X}_n \xrightarrow{d} N(0, 1)$?

-

$$X_{nj} = \begin{cases} 2j & \text{w.p. } \frac{1}{8j^2} \\ 0 & \text{w.p. } 1 - \frac{1}{4j^2} \\ -2j & \text{w.p. } \frac{1}{8j^2} \end{cases}$$

- $E[X_{nj}] = 0$ and $\text{var}(X_{nj}) = 1$. $B_n^2 = n$.

Example

Let X_1, \dots, X_n be independent random variables with mean zero and variance one. Do you think $\sqrt{n}\bar{X}_n \xrightarrow{d} N(0, 1)$?

•

$$X_{nj} = \begin{cases} 2j & \text{w.p. } \frac{1}{8j^2} \\ 0 & \text{w.p. } 1 - \frac{1}{4j^2} \\ -2j & \text{w.p. } \frac{1}{8j^2} \end{cases}$$

- $E[X_{nj}] = 0$ and $\text{var}(X_{nj}) = 1$. $B_n^2 = n$.
- Lets check the Lindeberg condition with $\epsilon = 1$.

$$\frac{1}{n} \sum_j E[X_{nj}^2 1(|X_{nj}| \geq \sqrt{n})] = \frac{1}{n} \sum_j 2 \times 4j^2 1(2j \geq \sqrt{n}) \frac{1}{8j^2} = \frac{1}{n} \sum_{j \geq \sqrt{n}/2} 1 \rightarrow 1$$

Example

Let X_1, \dots, X_n be independent random variables with mean zero and variance one. Do you think $\sqrt{n}\bar{X}_n \xrightarrow{d} N(0, 1)$?

•

$$X_{nj} = \begin{cases} 2j & \text{w.p. } \frac{1}{8j^2} \\ 0 & \text{w.p. } 1 - \frac{1}{4j^2} \\ -2j & \text{w.p. } \frac{1}{8j^2} \end{cases}$$

- $E[X_{nj}] = 0$ and $\text{var}(X_{nj}) = 1$. $B_n^2 = n$.
- Lets check the Lindeberg condition with $\epsilon = 1$.

$$\frac{1}{n} \sum_j E[X_{nj}^2 1(|X_{nj}| \geq \sqrt{n})] = \frac{1}{n} \sum_j 2 \times 4j^2 1(2j \geq \sqrt{n}) \frac{1}{8j^2} = \frac{1}{n} \sum_{j \geq \sqrt{n}/2} 1 \rightarrow 1$$

- Since $\sigma_{nj}^2/B_n^2 = 1/n \rightarrow 0$, this implies that the CLT does not hold for the sum.

Permutation Tests

Consider $2n$ paired experimental units with measurement $(X_i, Y_i)_{i=1}^n$ in which X_j is the result of the treatment and Y_j is the result of control.

- H_0 is that the treatment has had no effect, i.e. $Z_j = X_j - Y_j$ conditioned on the magnitude $|Z_j|$ is symmetric, i.e.
 $P(Z_j = |z_j|) = P(Z_j = -|z_j|) = 1/2$.

Permutation Tests

Consider $2n$ paired experimental units with measurement $(X_i, Y_i)_{i=1}^n$ in which X_j is the result of the treatment and Y_j is the result of control.

- H_0 is that the treatment has had no effect, i.e. $Z_j = X_j - Y_j$ conditioned on the magnitude $|Z_j|$ is symmetric, i.e.
 $P(Z_j = |z_j|) = P(Z_j = -|z_j|) = 1/2$.
- Thus, under H_0 , (Z_1, \dots, Z_n) has 2^n possible values $(\pm|z_1|, \dots, \pm|z_n|)$.

Permutation Tests

Consider $2n$ paired experimental units with measurement $(X_i, Y_i)_{i=1}^n$ in which X_j is the result of the treatment and Y_j is the result of control.

- H_0 is that the treatment has had no effect, i.e. $Z_j = X_j - Y_j$ conditioned on the magnitude $|Z_j|$ is symmetric, i.e.
 $P(Z_j = |z_j|) = P(Z_j = -|z_j|) = 1/2$.
- Thus, under H_0 , (Z_1, \dots, Z_n) has 2^n possible values $(\pm|z_1|, \dots, \pm|z_n|)$.
- Conditioned on the magnitudes of the differences, $B_n^2 = \sum_i z_i^2$.

Assume that $\max_i z_i^2 / B_n^2 \rightarrow 0$. Then $\sum_i Z_i / B_n \xrightarrow{d} N(0, 1)$ using the Lindeberg-feller theorem.

Permutation tests: proof

Proof.

- Lets check the Lindeberg condition:

$$\begin{aligned} \frac{\sum_{j=1}^n E[Z_j^2 1(|Z_j| \geq \epsilon B_n) | Z_1 = z_1, \dots, Z_n = z_n]}{B_n^2} &= \frac{\sum_j z_j^2 1(z_j \geq \epsilon B_n)}{B_n^2} \\ &\leq \frac{(\sum_j z_j^2) 1(\max_j z_j \geq \epsilon B_n)}{B_n^2} \\ &= 1(\max_j z_j \geq \epsilon B_n) \end{aligned}$$

- Since $\max_i z_i^2 / B_n^2 \rightarrow 0$, the above is zero for all sufficiently large n .



Tail probabilities

- Most often, we are interested in the question, how far is an empirical quantity from its “population variant”?
- This empirical quantity can be an eigenvalue of a matrix, or the weight vector learned using linear regression and so on.
- For rest of today, and a few more lectures we will brush up on tail inequalities.
- Lets start with the mean?

Concentration inequalities

- How will you bound $P(|\bar{X}_n - \mu| \geq t)$? Central limit theorem works under regularity conditions, but its only asymptotic.
- We will look at three methods:
 - Moment based:

Concentration inequalities

- How will you bound $P(|\bar{X}_n - \mu| \geq t)$? Central limit theorem works under regularity conditions, but its only asymptotic.
- We will look at three methods:
 - Moment based: Markov (first), Chebyshev (second)

Concentration inequalities

- How will you bound $P(|\bar{X}_n - \mu| \geq t)$? Central limit theorem works under regularity conditions, but its only asymptotic.
- We will look at three methods:
 - Moment based: Markov (first), Chebyshev (second)
 - Moment generating function based bounds:

Concentration inequalities

- How will you bound $P(|\bar{X}_n - \mu| \geq t)$? Central limit theorem works under regularity conditions, but its only asymptotic.
- We will look at three methods:
 - Moment based: Markov (first), Chebyshev (second)
 - Moment generating function based bounds: Hoeffding, Chernoff, Bernstein, subgaussian and subexponential random variables

Concentration inequalities

- How will you bound $P(|\bar{X}_n - \mu| \geq t)$? Central limit theorem works under regularity conditions, but its only asymptotic.
- We will look at three methods:
 - Moment based: Markov (first), Chebyshev (second)
 - Moment generating function based bounds: Hoeffding, Chernoff, Bernstein, subgaussian and subexponential random variables
 - Martingale based methods:

Concentration inequalities

- How will you bound $P(|\bar{X}_n - \mu| \geq t)$? Central limit theorem works under regularity conditions, but its only asymptotic.
- We will look at three methods:
 - Moment based: Markov (first), Chebyshev (second)
 - Moment generating function based bounds: Hoeffding, Chernoff, Bernstein, subgaussian and subexponential random variables
 - Martingale based methods: Azuma-Hoeffding, McDiarmid

