THE UNIVERSITY OF TEXAS AT AUSTIN
Department of Statistics and Data Sciences
College of Natural Sciences

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

Lemma (Fatou's lemma)

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

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

Theorem (Monotone convergence theorem)

If $0 \leq X_1 \leq X_2 \leq \cdots \leq X_n \uparrow X$, then

$$E[X_n] \to E[X]$$

Theorem (Dominated convergence theorem)

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

$$E[X_n] \to E[X]$$

Fatou to MCT

- Since $X_n \geq 0$ and $E[\liminf_n X_n] = E[X]$

Fatou to MCT

- Since $X_n \geq 0$ and $E[\liminf_n X_n] = E[X]$
- $\liminf_n E[X_n] \geq E[\liminf_n X_n] = E[X]$

Fatou to MCT

- Since $X_n \geq 0$ and $E[\liminf_n X_n] = E[X]$
- $\liminf_n E[X_n] \geq E[\liminf_n X_n] = E[X]$
- Since $X_n \uparrow X$ and integration preserves monotonicity, $E[X_1] \leq \cdots \leq E[X_n] \leq E[X]$

Fatou to MCT

- Since $X_n \geq 0$ and $E[\liminf_n X_n] = E[X]$
- $\liminf_n E[X_n] \geq E[\liminf_n X_n] = E[X]$
- Since $X_n \uparrow X$ and integration preserves monotonicity, $E[X_1] \leq \cdots \leq E[X_n] \leq E[X]$
- So $\limsup_n E[X_n] \leq E[X]$

Fatou to MCT

- Since $X_n \geq 0$ and $E[\liminf_n X_n] = E[X]$
- $\liminf_n E[X_n] \geq E[\liminf_n X_n] = E[X]$
- Since $X_n \uparrow X$ and integration preserves monotonicity, $E[X_1] \leq \cdots \leq E[X_n] \leq E[X]$
- So $\limsup_n E[X_n] \leq E[X]$
- $E[X] \geq \limsup_n E[X_n] \geq \liminf_n E[X_n] \geq E[\liminf_n X_n]$

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.*

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

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 $E[|X_1|] \leq \infty$, mean μ.

Theorem (Berry Esseen)

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

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

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 that 0.7975.

Lindeberg-feller CLT for triangular arrays

$$
\begin{aligned}
&X_{11} \\
&X_{21}, X_{22} \\
&X_{21}, X_{22}, X_{23} \\
&\ldots
\end{aligned}
$$

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 $B_n^2 = var(Z_n)$. 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 $\epsilon > 0$,

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

Converse: If $\dfrac{\sigma_{nj}^2}{B_n^2} \to 0$ as $n \to \infty$, 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 $\dfrac{\sigma_{nj}^2}{B_n^2} \to 0$, the the Lindeberg condition is necessary and sufficient to show the CLT.

Example

Let $X_1, \ldots, X_n$ 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 $X_1, \ldots, X_n$ 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 $X_1, \ldots, X_n$ 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 $X_1, \ldots, X_n$ 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 $\epsilon = 1$.

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

Example

Let $X_1, \ldots, X_n$ 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 $\epsilon = 1$.

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

- Since $\sigma_{nj}^2 / B_n^2 = 1/n \to 0$, 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, $(Z_1, \ldots, Z_n)$ has 2^n possible values $(\pm|z_1|, \ldots, \pm|z_n|)$.
- Conditioned on the magnitudes of the differences, $B_n^2 = \sum_i z_i^2$.

  Assume that $\max_i z_i^2 / B_n^2 \to 0$. 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:

$$\frac{\sum_{j=1}^{n} E[Z_j^2 1(|Z_j| \geq \epsilon B_n) || Z_1|, \ldots, |Z_n|]}{B_n^2} = \frac{\sum_j Z_j^2 1(|Z_j| \geq \epsilon B_n)}{B_n^2}$$
$$\leq \frac{(\sum_j Z_j^2) 1(\max_j |Z_j| \geq \epsilon B_n)}{B_n^2}$$
$$= 1(\max_j |Z_j| \geq \epsilon B_n)$$

- Since $\max_i z_i^2 / B_n^2 \to 0$, 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