

# SDS 384 11: Theoretical Statistics

## Lecture 1: Introduction

---

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

# Manegerial Stuff

- Instructor- Purnamrita Sarkar
- Course material and homeworks will be posted under <https://psarkar.github.io/teaching/sds384.html>
- Office hours: After class 30 mins. GDC 7.504
- TA: Yuguang Yue
- Homeworks are due Biweekly
- Grading - 4-5 homeworks (70% ), class participation (5%) Final Exam (25% )
- Books
  - Asymptotic Statistics, Aad van der Vaart. Cambridge. 1998.
  - Martin Wainwright's High dimensional statistics: A non-asymptotic view point

# Why do theory?

- Say you have estimated  $\hat{\theta}_n$  from data  $X_1, \dots, X_n$ . How do we know we have a “good” estimation method?
  - Does  $\hat{\theta}_n \rightarrow \theta$ ? This brings us to **Stochastic Convergence**.
- How do I know if one estimation method is better than another?
  - Does the estimate from one converge faster than the other?
  - Does one algorithm work under broader parameter regimes, or weaker assumptions?
  - What is the optimal rate for a given estimation problem?

# This class

Your instructor “hopes to cover”:

- Consistency of parameter estimates
  - Stochastic Convergence
  - Concentration inequalities
  - Asymptotic normality of estimators
- Empirical processes, VC classes, covering numbers
- Asymptotic testing
- Examples of network clustering with a bit of random matrix theory
- Bootstrap, Nonparametric regression and density estimation

# Stochastic Convergence

Assume that  $X_n, n \geq 1$  and  $X$  are elements of a separable metric space  $(S, d)$ .

## Definition (Weak Convergence)

A sequence of random variables converge in “law” or in “distribution” to a random variable  $X$ , i.e.  $X_n \xrightarrow{d} X$  if  $P(X_n \leq x) \rightarrow P(X \leq x) \forall x$  at which  $P(X \leq x)$  is continuous.

# Stochastic Convergence

Assume that  $X_n, n \geq 1$  and  $X$  are elements of a separable metric space  $(S, d)$ .

## Definition (Weak Convergence)

A sequence of random variables converge in “law” or in “distribution” to a random variable  $X$ , i.e.  $X_n \xrightarrow{d} X$  if  $P(X_n \leq x) \rightarrow P(X \leq x) \forall x$  at which  $P(X \leq x)$  is continuous.

## Definition (Convergence in Probability)

A sequence of random variables converge in “probability” to a random variable  $X$ , i.e.  $X_n \xrightarrow{P} X$  if  $\forall \epsilon > 0, P(d(X_n, X) \geq \epsilon) \rightarrow 0$ .

# Stochastic Convergence

Assume that  $X_n, n \geq 1$  and  $X$  are elements of a separable metric space  $(S, d)$ .

## Definition (Almost Sure Convergence)

A sequence of random variables converges almost surely to a random variable  $X$ , i.e.  $X_n \xrightarrow{a.s.} X$  if  $P\left(\lim_{n \rightarrow \infty} d(X_n, X) = 0\right) = 1$ .

- If you think about a (scalar) random variable as a function that maps events to a real number, almost sure convergence means 
$$P(\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)) = 1$$

## Definition (Convergence in quadratic mean)

A sequence of random variables converges in quadratic mean to a random variable  $X$ , i.e.  $X_n \xrightarrow{q.m.} X$  if  $E\left[d(X_n, X)^2\right] \rightarrow 0$ .

## Theorem

$$X_n \xrightarrow{a.s.} X, X_n \xrightarrow{q.m.} X \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$$
$$X_n \xrightarrow{d} c \Rightarrow X_n \xrightarrow{P} c$$



**Converses:**  $X_n \xrightarrow{d} X \not\Rightarrow X_n \xrightarrow{P} X$

- Convergence in law needs no knowledge of the joint distribution of  $X_n$  and the limiting random variable  $X$ .
- Convergence in probability does.

### Example

Consider  $X \sim N(0, 1)$ ,  $X_n = -X$ .  $X_n \xrightarrow{d} X$ . But how about  $X_n \xrightarrow{P} X$ ?

## Converses: $X_n \xrightarrow{d} X \not\Rightarrow X_n \xrightarrow{P} X$

- Convergence in law needs no knowledge of the joint distribution of  $X_n$  and the limiting random variable  $X$ .
- Convergence in probability does.

### Example

Consider  $X \sim N(0, 1)$ ,  $X_n = -X$ .  $X_n \xrightarrow{d} X$ . But how about  $X_n \xrightarrow{P} X$ ?

- $P(|X_n - X| \geq \epsilon) = P(2|X| \geq \epsilon) \not\rightarrow 0 \forall \epsilon > 0$ . So  $X_n$  does not converge in probability to  $X$ .

## Example

### Example

Let  $Z \sim U(0, 1)$  and for  $n = 2^k + m$  for  $k \geq 0, 0 \leq m < 2^k$   
 $X_n = 1(Z \in [m2^{-k}, (m+1)2^{-k}])$ , i.e.  $X_1 = 1, X_2 = 1(Z \in [0, 1/2))$ ,  
 $X_3 = 1(Z \in [1/2, 1))$ ,  $X_4 = 1(Z \in [0, 1/4))$ ,  $X_5 = 1(Z \in [1/4, 1/2))$ .

## Example

### Example

Let  $Z \sim U(0, 1)$  and for  $n = 2^k + m$  for  $k \geq 0, 0 \leq m < 2^k$   
 $X_n = 1(Z \in [m2^{-k}, (m+1)2^{-k}])$ , i.e.  $X_1 = 1$ ,  $X_2 = 1(Z \in [0, 1/2))$ ,  
 $X_3 = 1(Z \in [1/2, 1))$ ,  $X_4 = 1(Z \in [0, 1/4))$ ,  $X_5 = 1(Z \in [1/4, 1/2))$ .

- For any  $Z \in (0, 1)$ , the sequence  $\{X_n(Z)\}$  does not converge. So  $X_n \not\overset{a.s.}{\rightarrow} 0$ .

## Example

### Example

Let  $Z \sim U(0, 1)$  and for  $n = 2^k + m$  for  $k \geq 0, 0 \leq m < 2^k$   
 $X_n = 1(Z \in [m2^{-k}, (m+1)2^{-k}])$ , i.e.  $X_1 = 1$ ,  $X_2 = 1(Z \in [0, 1/2))$ ,  
 $X_3 = 1(Z \in [1/2, 1))$ ,  $X_4 = 1(Z \in [0, 1/4))$ ,  $X_5 = 1(Z \in [1/4, 1/2))$ .

- For any  $Z \in (0, 1)$ , the sequence  $\{X_n(Z)\}$  does not converge. So  $X_n \not\overset{a.s.}{\rightarrow} 0$ .
- $X_n$  are a sequence of bernoulli's with probabilities  $p_n = 1/2^k$  where  $k = \lfloor \log n \rfloor$ .

## Example

### Example

Let  $Z \sim U(0, 1)$  and for  $n = 2^k + m$  for  $k \geq 0, 0 \leq m < 2^k$   
 $X_n = 1(Z \in [m2^{-k}, (m+1)2^{-k}])$ , i.e.  $X_1 = 1$ ,  $X_2 = 1(Z \in [0, 1/2))$ ,  
 $X_3 = 1(Z \in [1/2, 1))$ ,  $X_4 = 1(Z \in [0, 1/4))$ ,  $X_5 = 1(Z \in [1/4, 1/2))$ .

- For any  $Z \in (0, 1)$ , the sequence  $\{X_n(Z)\}$  does not converge. So  $X_n \not\overset{a.s.}{\rightarrow} 0$ .
- $X_n$  are a sequence of bernoulli's with probabilities  $p_n = 1/2^k$  where  $k = \lfloor \log n \rfloor$ .
- So  $X_n \overset{P}{\rightarrow} 0$  and  $X_n \overset{qm}{\rightarrow} 0$

## Example

### Example

Let  $Z \sim U([0, 1])$  and  $X_n = 2^n 1(Z \in [0, 1/n])$ . Does  $X_n$  converge to  $X$  in quadratic mean, almost surely or in probability?

## Example

### Example

Let  $Z \sim U([0, 1])$  and  $X_n = 2^n 1(Z \in [0, 1/n])$ . Does  $X_n$  converge to  $X$  in quadratic mean, almost surely or in probability?

- $P(\lim_{n \rightarrow \infty} X_n = X) = P(Z > 0) = 1$ . So  $X_n \xrightarrow{a.s.} X$ .



## Example

### Example

Let  $Z \sim U([0, 1])$  and  $X_n = 2^n 1(Z \in [0, 1/n])$ . Does  $X_n$  converge to  $X$  in quadratic mean, almost surely or in probability?

- $P(\lim_{n \rightarrow \infty} X_n = X) = P(Z > 0) = 1$ . So  $X_n \xrightarrow{a.s.} X$ .
- $E|X_n|^2 = 2^{2n}/n \rightarrow \infty$ . So  $X_n \not\xrightarrow{qm} 0$

## Example

### Example

Let  $Z \sim U([0, 1])$  and  $X_n = 2^n 1(Z \in [0, 1/n])$ . Does  $X_n$  converge to  $X$  in quadratic mean, almost surely or in probability?

- $P(\lim_{n \rightarrow \infty} X_n = X) = P(Z > 0) = 1$ . So  $X_n \xrightarrow{a.s.} X$ .
- $E|X_n|^2 = 2^{2n}/n \rightarrow \infty$ . So  $X_n \not\xrightarrow{qm} 0$
- $P(|X_n| \geq \epsilon) = P(X_n = 2^n) = P(Z \in [0, 1/n]) = 1/n \rightarrow 0$

# Continuous Mapping Theorem

## Theorem

*Let  $g$  be continuous on a set  $C$  where  $P(X \in C) = 1$ . Then,*

$$X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$$

$$X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X)$$

$$X_n \xrightarrow{a.s.} X \Rightarrow g(X_n) \xrightarrow{a.s.} g(X)$$

## Example

Let  $X_n \xrightarrow{d} X$  where  $X \sim N(0, 1)$ . Then  $X_n^2 \xrightarrow{d} ?$

## Example

Let  $X_n \xrightarrow{d} X$  where  $X \sim N(0, 1)$ . Then  $X_n^2 \xrightarrow{d} ?$

- Use  $g(x) = x^2$ .

## Example

Let  $X_n \xrightarrow{d} X$  where  $X \sim N(0, 1)$ . Then  $X_n^2 \xrightarrow{d} ?$

- Use  $g(x) = x^2$ .
- Use  $X^2 \sim \chi_1^2$ .

## Example

Let  $X_n \xrightarrow{d} X$  where  $X \sim N(0, 1)$ . Then  $X_n^2 \xrightarrow{d} ?$

- Use  $g(x) = x^2$ .
- Use  $X^2 \sim \chi_1^2$ .
- So  $X_n^2 \xrightarrow{d} \chi_1^2$

## Example-continuity points

Let  $X_1, \dots, X_n$  be i.i.d. with mean  $\mu$  and variance  $\sigma^2$ . We have  $\bar{X}_n - \mu \xrightarrow{d} 0$ . Consider  $g(x) = 1_{x>0}$ . Then  $g((\bar{X}_n - \mu)^2) \xrightarrow{d} ?$



## Example-continuity points

Let  $X_1, \dots, X_n$  be i.i.d. with mean  $\mu$  and variance  $\sigma^2$ . We have  $\bar{X}_n - \mu \xrightarrow{d} 0$ . Consider  $g(x) = 1_{x>0}$ . Then  $g((\bar{X}_n - \mu)^2) \xrightarrow{d} ?$

- Using Continuous Mapping Theorem,  $(\bar{X}_n - \mu)^2 \xrightarrow{d} 0$

## Example-continuity points

Let  $X_1, \dots, X_n$  be i.i.d. with mean  $\mu$  and variance  $\sigma^2$ . We have  $\bar{X}_n - \mu \xrightarrow{d} 0$ . Consider  $g(x) = 1_{x>0}$ . Then  $g((\bar{X}_n - \mu)^2) \xrightarrow{d} ?$

- Using Continuous Mapping Theorem,  $(\bar{X}_n - \mu)^2 \xrightarrow{d} 0$
- Can we use Continuous Mapping Theorem to claim that  $g(\bar{X}_n - \mu)^2 \xrightarrow{d} 0$ ?

## Example-continuity points

Let  $X_1, \dots, X_n$  be i.i.d. with mean  $\mu$  and variance  $\sigma^2$ . We have  $\bar{X}_n - \mu \xrightarrow{d} 0$ . Consider  $g(x) = 1_{x>0}$ . Then  $g((\bar{X}_n - \mu)^2) \xrightarrow{d} ?$

- Using Continuous Mapping Theorem,  $(\bar{X}_n - \mu)^2 \xrightarrow{d} 0$
- Can we use Continuous Mapping Theorem to claim that  $g(\bar{X}_n - \mu)^2 \xrightarrow{d} 0$ ?
- NO. Because, 0 is a random variable whose mass is at 0, where  $g$  is discontinuous.

## How about convergence in q.m.?

- If  $X_n \xrightarrow{qm} X$ , then is it true that for continuous  $f$  (discontinuous only at a measure zero set),  $f(X_n) \xrightarrow{qm} f(X)$ ?

## How about convergence in q.m.?

- If  $X_n \xrightarrow{qm} X$ , then is it true that for continuous  $f$  (discontinuous only at a measure zero set),  $f(X_n) \xrightarrow{qm} f(X)$ ?
- Consider an  $L$ -Lipschitz function  $f(X)$ .  $|f(x) - f(y)| \leq L|x - y|$ .

## How about convergence in q.m.?

- If  $X_n \xrightarrow{qm} X$ , then is it true that for continuous  $f$  (discontinuous only at a measure zero set),  $f(X_n) \xrightarrow{qm} f(X)$ ?
- Consider an  $L$ - Lipschitz function  $f(X)$ .  $|f(x) - f(y)| \leq L|x - y|$ .
- $E[|f(X_n) - f(X)|^2] \leq L^2 E[|X_n - X|^2] \rightarrow 0$ . So for Lipschitz functions quadratic mean convergence goes through.

## How about convergence in q.m.?

- If  $X_n \xrightarrow{qm} X$ , then is it true that for continuous  $f$  (discontinuous only at a measure zero set),  $f(X_n) \xrightarrow{qm} f(X)$ ?
- Consider an  $L$ -Lipschitz function  $f(X)$ .  $|f(x) - f(y)| \leq L|x - y|$ .
- $E[|f(X_n) - f(X)|^2] \leq L^2 E[|X_n - X|^2] \rightarrow 0$ . So for Lipschitz functions quadratic mean convergence goes through.
- Can you come up with a non-Lipschitz function and a sequence  $\{X_n\}$  where  $f(X_n) \not\xrightarrow{qm} 0$ ?

# Portmanteau Theorem

## Theorem

*The following are equivalent.*

- $X_n \xrightarrow{d} X$
- $E[f(X_n)] \rightarrow E[f(X)]$  for all continuous  $f$  that vanish outside a compact set.
- $E[f(X_n)] \rightarrow E[f(X)]$  for all bounded and continuous  $f$ .
- $E[f(X_n)] \rightarrow E[f(X)]$  for all bounded measurable functions  $f$  s.t.  $P(X \in C(f)) = 1$ , where  $C(f) = \{x : f \text{ is continuous at } x\}$  is called the continuity set of  $f$ .



## Example-bounded

Consider  $f(x) = x$  and

$$X_n = \begin{cases} n & \text{w.p. } 1/n \\ 0 & \text{w.p. } 1 - 1/n \end{cases}$$

## Example-bounded

Consider  $f(x) = x$  and

$$X_n = \begin{cases} n & \text{w.p. } 1/n \\ 0 & \text{w.p. } 1 - 1/n \end{cases}$$

- $X_n \xrightarrow{d} 0$ , but  $E[X_n] \rightarrow ?$

## Example-bounded

Consider  $f(x) = x$  and

$$X_n = \begin{cases} n & \text{w.p. } 1/n \\ 0 & \text{w.p. } 1 - 1/n \end{cases}$$

- $X_n \xrightarrow{d} 0$ , but  $E[X_n] \rightarrow ?$
- $E[X_n] = 1$ . What went wrong?

## Example-bounded

Consider  $f(x) = x$  and

$$X_n = \begin{cases} n & \text{w.p. } 1/n \\ 0 & \text{w.p. } 1 - 1/n \end{cases}$$

- $X_n \xrightarrow{d} 0$ , but  $E[X_n] \rightarrow ?$
- $E[X_n] = 1$ . What went wrong?
- $f(x) = x$  is not bounded.

# Putting everything together

## Theorem

$$X_n \xrightarrow{d} X \text{ and } d(X_n, Y_n) \xrightarrow{P} 0 \Rightarrow Y_n \xrightarrow{d} X \quad (1)$$

$$X_n \xrightarrow{d} X \text{ and } Y_n \xrightarrow{d} c \Rightarrow (X_n, Y_n) \xrightarrow{d} (X, c) \quad (2)$$

$$X_n \xrightarrow{P} X \text{ and } Y_n \xrightarrow{P} Y \Rightarrow (X_n, Y_n) \xrightarrow{P} (X, Y) \quad (3)$$

# Putting everything together

## Theorem

$$X_n \xrightarrow{d} X \text{ and } d(X_n, Y_n) \xrightarrow{P} 0 \Rightarrow Y_n \xrightarrow{d} X \quad (1)$$

$$X_n \xrightarrow{d} X \text{ and } Y_n \xrightarrow{d} c \Rightarrow (X_n, Y_n) \xrightarrow{d} (X, c) \quad (2)$$

$$X_n \xrightarrow{P} X \text{ and } Y_n \xrightarrow{P} Y \Rightarrow (X_n, Y_n) \xrightarrow{P} (X, Y) \quad (3)$$

- Eq 3 does not hold if we replace convergence in probability by convergence in distribution.

# Putting everything together

## Theorem

$$X_n \xrightarrow{d} X \text{ and } d(X_n, Y_n) \xrightarrow{P} 0 \Rightarrow Y_n \xrightarrow{d} X \quad (1)$$

$$X_n \xrightarrow{d} X \text{ and } Y_n \xrightarrow{d} c \Rightarrow (X_n, Y_n) \xrightarrow{d} (X, c) \quad (2)$$

$$X_n \xrightarrow{P} X \text{ and } Y_n \xrightarrow{P} Y \Rightarrow (X_n, Y_n) \xrightarrow{P} (X, Y) \quad (3)$$

- Eq 3 does not hold if we replace convergence in probability by convergence in distribution.
- Example:  $X_n \sim N(0, 1)$ ,  $Y_n = -X_n$ .  $X \perp Y$  and  $X, Y$  are independent standard normal random variables.

# Putting everything together

## Theorem

$$X_n \xrightarrow{d} X \text{ and } d(X_n, Y_n) \xrightarrow{P} 0 \Rightarrow Y_n \xrightarrow{d} X \quad (1)$$

$$X_n \xrightarrow{d} X \text{ and } Y_n \xrightarrow{d} c \Rightarrow (X_n, Y_n) \xrightarrow{d} (X, c) \quad (2)$$

$$X_n \xrightarrow{P} X \text{ and } Y_n \xrightarrow{P} Y \Rightarrow (X_n, Y_n) \xrightarrow{P} (X, Y) \quad (3)$$

- Eq 3 does not hold if we replace convergence in probability by convergence in distribution.
- Example:  $X_n \sim N(0, 1)$ ,  $Y_n = -X_n$ .  $X \perp Y$  and  $X, Y$  are independent standard normal random variables.
- Then  $X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{d} Y$ . But  $(X_n, Y_n) \xrightarrow{d} (X, -X)$ , not  $(X_n, Y_n) \xrightarrow{d} (X, Y)$ .



# Putting everything together

## Theorem (Slutsky's theorem)

$X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{d} c$  imply that

$$X_n + Y_n \xrightarrow{d} X + c$$

$$X_n Y_n \xrightarrow{d} cX$$

$$X_n / Y_n \xrightarrow{d} X / c$$

# Putting everything together

## Theorem (Slutsky's theorem)

$X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{d} c$  imply that

$$X_n + Y_n \xrightarrow{d} X + c$$

$$X_n Y_n \xrightarrow{d} cX$$

$$X_n / Y_n \xrightarrow{d} X / c$$

- Does  $X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{d} Y$  imply  $X_n + Y_n \xrightarrow{d} X + Y$ ?

# Putting everything together

## Theorem (Slutsky's theorem)

$X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{d} c$  imply that

$$X_n + Y_n \xrightarrow{d} X + c$$

$$X_n Y_n \xrightarrow{d} cX$$

$$X_n / Y_n \xrightarrow{d} X / c$$

- Does  $X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{d} Y$  imply  $X_n + Y_n \xrightarrow{d} X + Y$ ?
- Take  $Y_n = -X_n$ , and  $X, Y$  as independent standard normal random variables.  $X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{d} Y$  but  $X_n + Y_n \xrightarrow{d} 0$ .

## Using all this

If  $X_1, \dots, X_n$  are i.i.d. random variables with mean  $\mu$  and variance  $\sigma^2$ ,  
prove that  $\sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \xrightarrow{d} N(0, 1)$ .

## Using all this

If  $X_1, \dots, X_n$  are i.i.d. random variables with mean  $\mu$  and variance  $\sigma^2$ , prove that  $\sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \xrightarrow{d} N(0, 1)$ .

- First note that  $S_n^2 = \frac{1}{n} \sum_i X_i^2 - \bar{X}_n^2$

## Using all this

If  $X_1, \dots, X_n$  are i.i.d. random variables with mean  $\mu$  and variance  $\sigma^2$ , prove that  $\sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \xrightarrow{d} N(0, 1)$ .

- First note that  $S_n^2 = \frac{1}{n} \sum_i X_i^2 - \bar{X}_n^2$
- Law of large numbers give  $\frac{\sum_i X_i^2}{n} \xrightarrow{P} E[X^2]$  and  $\bar{X}_n \xrightarrow{P} \mu$ .

## Using all this

If  $X_1, \dots, X_n$  are i.i.d. random variables with mean  $\mu$  and variance  $\sigma^2$ , prove that  $\sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \xrightarrow{d} N(0, 1)$ .

- First note that  $S_n^2 = \frac{1}{n} \sum_i X_i^2 - \bar{X}_n^2$
- Law of large numbers give  $\frac{\sum_i X_i^2}{n} \xrightarrow{P} E[X^2]$  and  $\bar{X}_n \xrightarrow{P} \mu$ .
- So  $(\frac{\sum_i X_i^2}{n}, \bar{X}_n) \xrightarrow{P} (E[X^2], \mu)$  and now using the continuous mapping theorem,  $S_n^2 \xrightarrow{P} \sigma^2$ .

## Using all this

If  $X_1, \dots, X_n$  are i.i.d. random variables with mean  $\mu$  and variance  $\sigma^2$ , prove that  $\sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \xrightarrow{d} N(0, 1)$ .

- First note that  $S_n^2 = \frac{1}{n} \sum_i X_i^2 - \bar{X}_n^2$
- Law of large numbers give  $\frac{\sum_i X_i^2}{n} \xrightarrow{P} E[X^2]$  and  $\bar{X}_n \xrightarrow{P} \mu$ .
- So  $(\frac{\sum_i X_i^2}{n}, \bar{X}_n) \xrightarrow{P} (E[X^2], \mu)$  and now using the continuous mapping theorem,  $S_n^2 \xrightarrow{P} \sigma^2$ .
- Finally,  $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$  using CLT.



## Using all this

If  $X_1, \dots, X_n$  are i.i.d. random variables with mean  $\mu$  and variance  $\sigma^2$ , prove that  $\sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \xrightarrow{d} N(0, 1)$ .

- First note that  $S_n^2 = \frac{1}{n} \sum_i X_i^2 - \bar{X}_n^2$
- Law of large numbers give  $\frac{\sum_i X_i^2}{n} \xrightarrow{P} E[X^2]$  and  $\bar{X}_n \xrightarrow{P} \mu$ .
- So  $(\frac{\sum_i X_i^2}{n}, \bar{X}_n) \xrightarrow{P} (E[X^2], \mu)$  and now using the continuous mapping theorem,  $S_n^2 \xrightarrow{P} \sigma^2$ .
- Finally,  $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$  using CLT.
- Now using Slutsky's lemma,  $\sqrt{n}(\bar{X}_n - \mu)/S_n \xrightarrow{d} N(0, 1)$  using CLT.

## Definition

$X$  is defined to be “tight” if  $\forall \epsilon > 0 \exists M$  for which,

$$P(\|X\| > M) < \epsilon$$

$\{X_n\}$  is defined to uniformly tight if  $\forall \epsilon > 0 \exists M$  for which,

$$\sup_n P(\|X_n\| > M) < \epsilon$$

## Theorem

- $X_n \xrightarrow{d} X \Rightarrow \{X_n\}$  is UT.
- $\{X_n\}$  is UT implies that, there exists a subsequence  $\{n_j\}$  such that  $X_{n_j} \xrightarrow{d} X$ .

# Notation for rates, small oh-pee and big oh-pee

## Definition

- The small  $o_P$ :

$$X_n = o_P(1) \Leftrightarrow X_n \xrightarrow{P} 0$$

$$X_n = o_P(R_n) \Leftrightarrow X_n = Y_n R_n \text{ and } Y_n = o_P(1)$$

$X_n$  is vanishing in probability

- The big  $O_P$ :

$$X_n = O_P(1) \Leftrightarrow \{X_n\} \text{ is UT}$$

$$X_n = O_P(R_n) \Leftrightarrow X_n = Y_n R_n \text{ and } Y_n = O_P(1)$$

$X_n$  lies within a ball of finite radius with high probability

## How do they interact

$$o_P(a_n) + o_P(b_n) = o_P(\max(a_n, b_n)).$$

$$o_P(a_n) + O_P(b_n) = O_P(\max(a_n, b_n)).$$

$$O_P(a_n)o_P(b_n) = o_P(a_nb_n).$$

$$1 + O_P(1) = O_P(1).$$

$$(1 + o_P(1))^{-1} = 1 + o_P(1).$$

$$o_P(O_P(1)) = o_P(1).$$

$$X_n = o_P(a_n), |X_n|^r = o_P(a_n^r)$$

Be careful:

$$e^{o_P(1)} \neq o_P(1)$$

$O_P(1) + O_P(1)$  Can actually be  $o_P(1)$  because of cancellation.

