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

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: TBD
- Homeworks are due Biweekly
- Grading - 4-5 homeworks (55%), class participation (10%) Final Exam (35%)
- 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 $X_1, \ldots, X_n$. How do we know we have a "good" estimation method?
  - Does $\hat{\theta}_n \to \theta$? This brings us to **Stochastic Convergence**.
- How about the rate of convergence?
  - Can we give any guarantees on how quickly our estimate converges?

$$P(|\hat{\theta}_n - \theta| = \text{large}) = \text{small}$$

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
- 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. $X_n \stackrel{d}{\to} X$ if $P(X_n \leq x) \to P(X \leq x)$ $\forall x$ 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 $P(X_n \leq x) \to P(X \leq x)$ $\forall x$ 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 $\forall \epsilon > 0$, $P(d(X_n, X) \geq \epsilon) \to 0$.

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. $X_n \overset{a.s.}{\to} X$ if $P\left(\lim_{n\to\infty} d(X_n, X) = 0\right) = 1$.

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

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

Unwinding a.s. convergence

- $X_n \overset{a.s.}{\to} X$ implies $P(\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)) = 1$
- What does convergence mean for a sequence of real numbers?

Unwinding a.s. convergence

- $X_n \stackrel{a.s.}{\to} X$ implies $P(\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)) = 1$
- What does convergence mean for a sequence of real numbers?
- $\forall \epsilon > 0$, $\exists n$, $\forall m \geq n$, $|X_n(\omega) - X(\omega)| < \epsilon$
    - Consider a sequence of events $A_1, \ldots, A_n$,
      $A_n = \{|X_n(\omega) - X(\omega)| < \epsilon\}$

Unwinding a.s. convergence

- $X_n \overset{a.s.}{\to} X$ implies $P(\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)) = 1$
- What does convergence mean for a sequence of real numbers?
- $\forall \epsilon > 0$, $\exists n$, $\forall m \geq n$, $|X_n(\omega) - X(\omega)| < \epsilon$
  - Consider a sequence of events $A_1, \ldots, A_n$, $A_n = \{|X_n(\omega) - X(\omega)| < \epsilon\}$
  - $\forall \epsilon > 0$, $\exists n$, s.t. $\forall m \geq n$, $|X_n(\omega) - X(\omega)| < \epsilon$, boils down to:

$$\bigcup_{i=1}^{n} \bigcap_{m \geq n} A_m$$

Unwinding a.s. convergence

- $X_n \overset{a.s.}{\to} X$ implies $P(\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)) = 1$
- What does convergence mean for a sequence of real numbers?
- $\forall \epsilon > 0$, $\exists n$, $\forall m \geq n$, $|X_n(\omega) - X(\omega)| < \epsilon$
  - Consider a sequence of events $A_1, \ldots, A_n$, $A_n = \{|X_n(\omega) - X(\omega)| < \epsilon\}$
  - $\forall \epsilon > 0$, $\exists n$, s.t. $\forall m \geq n$, $|X_n(\omega) - X(\omega)| < \epsilon$, boils down to:

$$\bigcup_{i=1}^{n} \bigcap_{m \geq n} A_m$$

  - Another way of saying this is, A_n^c happens finitely often. (f.o.)
  - $X_n \overset{a.s.}{\to} X$ implies $\forall \epsilon > 0$, $P(\{|X_n - X| \geq \epsilon \text{ f.o.}\}) = 1$

Theorem

$$X_n \stackrel{a.s.}{\to} X \ , X_n \stackrel{q.m.}{\to} X \Rightarrow X_n \stackrel{P}{\to} X \Rightarrow X_n \stackrel{d}{\to} X$$

$$X_n \stackrel{d}{\to} c \Rightarrow X_n \stackrel{P}{\to} c$$

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 $X_n \xrightarrow{P} X$?

- $P(|X_n - X| \geq \epsilon) = P(2|X| \geq \epsilon) \not\to 0 \ \forall \epsilon > 0$. 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 $X_n \overset{a.s.}{\not\to} 0$.

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 $X_n \overset{a.s.}{\not\to} 0$.
- For any $\epsilon > 0$, $P(\{|X_n| > \epsilon\}$ i.o.)
- 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 $X_n \overset{a.s.}{\not\to} 0$.
- For any $\epsilon > 0$, $P(\{|X_n| > \epsilon\}$ i.o.)
- X_n are a sequence of bernoulli's with probabilities $p_n = 1/2^k$ where $k = \lfloor \log n \rfloor$.
- So $X_n \overset{P}{\to} 0$ and $X_n \overset{qm}{\to} 0$

Example

Example

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

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

Example

Example

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

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

Example

Example

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

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

Borel Cantelli

- $X_n \overset{a.s.}{\to} X$ implies $\exists \epsilon > 0$, $P(\{|X_n - X| \geq \epsilon \text{ i.o.}\}) = 0$

Borel Cantelli

- $X_n \overset{a.s.}{\to} X$ implies $\exists \epsilon > 0$, $P(\{|X_n - X| \geq \epsilon \text{ i.o.}\}) = 0$
- Consider a sequence of events $A_1, \ldots, A_n$.
- Infinitely often means $\forall n$, $\exists m \geq n$, s.t. A_m occurs.
- More concretely

$$\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$$

Borel Cantelli Lemma (I)

Theorem
If $\sum_i P(A_i) < \infty$, *then* $P(\{A_n \text{ i.o.}\}) = 0$.

Example
Let $X_n \sim \text{Bernoulli}(2^{-n})$. Then $X_n \overset{a.s.}{\to} 0$.

Check if $X_n = 1$ infinitely often.

Borel Cantelli Lemma (I)

Theorem
If $\sum_i P(A_i) < \infty$, *then* $P(\{A_n \text{ i.o.}\}) = 0$.

- Recall that $\{A_n \text{ i.o.}\}$ is equivalent to $\bigcap_{n=1}^{\infty} \underbrace{\bigcup_{m=n}^{\infty} A_m}_{B_n}$

Borel Cantelli Lemma (I)

Theorem
If $\sum_{i} P(A_i) < \infty$*, then* $P(\{A_n \text{ i.o.}\}) = 0$*.*

- Recall that $\{A_n \text{ i.o.}\}$ is equivalent to $\bigcap_{n=1}^{\infty} \underbrace{\bigcup_{m=n}^{\infty} A_m}_{B_n}$
- Note that $B_{n+1} \subseteq B_n$, and so we have $B_n \downarrow B := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$, hence using monotone convergence we have:

$$\lim_{n \to \infty} P(B_n) = P(B)$$

Borel Cantelli Lemma (I)

Theorem
If $\sum_i P(A_i) < \infty$*, then* $P(\{A_n \text{ i.o.}\}) = 0$*.*

$$P(A_i \text{ i.o.}) = \lim_{n\to\infty} P(B_n) \leq \lim_{n\to\infty} \sum_{m\geq n} P(A_m) = 0$$

Borel Cantelli Lemma (II)

Theorem

If $\sum_i P(A_i) = \infty$ *and* $\{A_n\}$ *are independent then* $P(\{A_n \text{ i.o.}\}) = 1$.

Example

Consider $Z \sim U[0,1]$, $A_n := \{Z \leq 1/n\}$, and $X_n = 1(A_n)$. $\sum_i P(A_n) \to \infty$.

But we know that $X_n \stackrel{a.s.}{\to} 0$.

- Does BC II apply?
- If not, how do you prove it?

Borel Cantelli Lemma (II)

- Start with the complement – we will show $P((A_i \text{ i.o.})^c) = 0$.

Borel Cantelli Lemma (II)

- Start with the complement – we will show $P((A_i \text{ i.o.})^c) = 0$.

$$
\begin{aligned}
P((A_i \text{ i.o.})^c) &= P\left(\bigcup_n \bigcap_{m \geq n} A_m^c\right) \\
&= \lim_{n\to\infty} P\left(\bigcap_{m \geq n} A_m^c\right) \\
&= \lim_{n\to\infty} \prod_{m \geq n} P\left(A_m^c\right) \\
&= \lim_{n\to\infty} \prod_{m \geq n} (1 - P(A_m)) \\
&\leq \lim_{n\to\infty} \exp(-\sum_{m \geq n} P(A_m)) = 0
\end{aligned}
$$

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

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

Example

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

- 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 $X_1, \ldots, X_n$ be i.i.d. with mean μ and variance σ^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 $X_1, \ldots, X_n$ be i.i.d. with mean μ and variance σ^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 μ and variance σ^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 $X_1, \ldots, X_n$ be i.i.d. with mean μ and variance σ^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 $X_n \overset{qm}{\to} X$, then is it true that for continuous f (discontinuous only at a measure zero set), $f(X_n) \overset{qm}{\to} f(X)$?

How about convergence in q.m.?

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

How about convergence in q.m.?

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

How about convergence in q.m.?

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

Portmanteau Theorem

Theorem
The following are equivalent.

- $X_n \xrightarrow{d} X$
- $E[f(X_n)] \to E[f(X)]$ *for all continuous* f *that vanish outside a compact set.*
- $E[f(X_n)] \to E[f(X)]$ *for all bounded and continuous* f.
- $E[f(X_n)] \to E[f(X)]$ *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 $E[X_n] \to ?$
- $E[X_n] = 1$. 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

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

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

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

Putting everything together

Theorem

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

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

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

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

Putting everything together

Theorem

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

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

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

- 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

$$X_n \stackrel{d}{\to} X \text{ and } d(X_n, Y_n) \stackrel{P}{\to} 0 \Rightarrow Y_n \stackrel{d}{\to} X \tag{1}$$

$$X_n \stackrel{d}{\to} X \text{ and } Y_n \stackrel{d}{\to} c \Rightarrow (X_n, Y_n) \stackrel{d}{\to} (X, c) \tag{2}$$

$$X_n \stackrel{P}{\to} X \text{ and } Y_n \stackrel{P}{\to} Y \Rightarrow (X_n, Y_n) \stackrel{P}{\to} (X, Y) \tag{3}$$

- 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 $X_n \stackrel{d}{\to} X$ and $Y_n \stackrel{d}{\to} Y$. But $(X_n, Y_n) \stackrel{d}{\to} (X, -X)$, not $(X_n, Y_n) \stackrel{d}{\to} (X, Y)$.

Putting everything together

Theorem (Slutsky's theorem)
$X_n \stackrel{d}{\to} X$ and $Y_n \stackrel{d}{\to} c$ imply that

$$X_n + Y_n \stackrel{d}{\to} X + c$$

$$X_n Y_n \stackrel{d}{\to} cX$$

$$X_n / Y_n \stackrel{d}{\to} X / c$$

Putting everything together

Theorem (Slutsky's theorem)
$X_n \overset{d}{\to} X$ and $Y_n \overset{d}{\to} c$ imply that

$$X_n + Y_n \overset{d}{\to} X + c$$

$$X_n Y_n \overset{d}{\to} cX$$

$$X_n / Y_n \overset{d}{\to} X/c$$

- *Does $X_n \overset{d}{\to} X$ and $Y_n \overset{d}{\to} Y$ imply $X_n + Y_n \overset{d}{\to} X + Y$?*

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 μ and variance σ^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 μ and variance σ^2, prove that $\sqrt{n}\frac{\bar{X}_n - \mu}{S_n} \xrightarrow{d} N(0,1)$.

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

Using all this

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

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

Using all this

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

- First note that $S_n = \frac{1}{n}\sum_i X_i^2 - \bar{X}_n^2$
- Law of large numbers give $\frac{\sum_i X_i^2}{n} \xrightarrow{P} E[X^2]$ and $X_n \xrightarrow{P} \mu$.
- 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 $X_1, \ldots X_n$ are i.i.d. random variables with mean μ and variance σ^2, prove that $\sqrt{n}\frac{\bar{X}_n - \mu}{S_n} \xrightarrow{d} N(0, 1)$.

- First note that $S_n = \frac{1}{n}\sum_i X_i^2 - \bar{X}_n^2$
- Law of large numbers give $\frac{\sum_i X_i^2}{n} \xrightarrow{P} E[X^2]$ and $X_n \xrightarrow{P} \mu$.
- 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 μ and variance σ^2, prove that $\sqrt{n}\frac{\bar{X}_n - \mu}{S_n} \xrightarrow{d} N(0,1)$.

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

Uniformly tight

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,

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

Uniformly tight

- Give an example of a sequence that is **Not** UT
- $X_n = Uniform([-n, n])$
- $P(|X_n| > n(1 - \epsilon/2)) = \epsilon$, so you cannot find an ϵ such that $P(|X_n| > M) \leq \epsilon$ for all n

Prohorov's theorem

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, big O and big O-pea

Definition

- Big O. Let $g(.)$ be a positive function.

$$f(x) = O(g(x)) \text{ as } x \to \infty$$

$$\exists M, x_0, \qquad |f(x)| \leq Mg(x) \qquad \text{For } x \geq x_0$$

  For large x, $f(x)$ is bounded by $g(x)$ up-to a multiplicative constant

- 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 is likely to lie within a ball of finite radius

Notation for rates, small o and small o-pea

Definition

- The small o:

$$f(x) = o(g(x)) \Leftrightarrow f(x)/g(x) \to 0 \quad \text{as } x \to \infty$$

- 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

How do they interact

Lemma
Let $R : \mathbb{R}^k \to \mathbb{R}$ be a function. Let $X_n = o_P(1)$. Then as as $\|h\| \to 0$, $\forall q > 0$

$$R(h) = o(\|h\|^q) \text{ implies } R(X_n) = o_P(\|X_n\|^q)$$
$$R(h) = O(\|h\|^q) \text{ implies } R(X_n) = O_P(\|X_n\|^q)$$

- *Work out the proof at home.*
- *Hint: apply continuous mapping to $R(h)/\|h\|^q$.*

How do they interact

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

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.